

WHY APPEALS COURTS RARELY REVERSE LOWER COURTS: AN EXPERIMENTAL STUDY TO EXPLORE AFFIRMATION BIAS

Barry C. Edwards*

INTRODUCTION

One of the most striking features of appellate courts in the United States is that they rarely reverse lower court decisions.¹ One could correctly guess the outcome of an appeal about 90% of the time, knowing nothing about the issues involved, by predicting that the appellate court will affirm the lower court decision.² Every other factor thought to explain appellate court decision-making pales in comparison to how the case was previously decided.³

Although many have observed that appellate courts rarely reverse lower courts, there does not appear to be a good explanation for this “puzzling” fact.⁴ Some accounts of high affirmation rates are more troubling than others, but researchers have found no way to adjudicate among various explanations.

Discriminating among potential explanations for high affirmation rates can help us determine whether trial courts rarely make reversible errors or frequently make errors that appellate courts fail to correct.⁵ The fact that appellate courts affirm the overwhelming majority of trial court decisions may make outcomes predictable, but it hard to know what to make of this fact. Indeed, the fact that

* Lecturer, University of Central Florida, Department of Political Science. The author thanks Justin M. Strich and Mogens Jin Pedersen for their assistance with this research, the law professors who graciously allowed us to survey their students, and Tyler Yeargain of the *Emory Law Journal* for his editorial stewardship.

¹ See Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119, 150 (2002); see also CHRISTOPHER P. BANKS & DAVID M. O'BRIEN, THE JUDICIAL PROCESS: LAW, COURTS, AND JUDICIAL POLITICS 265 (2016); ROBERT A. CARP ET AL., JUDICIAL PROCESS IN AMERICA 254 (10th ed. 2017); Kevin M. Clermont, *Litigation Realities Redux*, 84 NOTRE DAME L. REV. 1919, 1968 (2008); Chris Guthrie & Tracey E. George, *The Futility of Appeal: Disciplinary Insights into the “Affirmance Effect” on the United States Courts of Appeals*, 32 FLA. ST. U. L. REV. 357, 358 (2005).

² As discussed below, in Part I, the vast majority of appellate court decisions at the federal and state levels affirm the decision being appealed. One would correctly guess the outcome of 92% of recent federal appeals by always predicting the court to affirm.

³ See, e.g., Christina L. Boyd et al., *Untangling the Causal Effects of Sex on Judging*, 54 AM. J. POL. SCI. 389, 409 tbl.A2 (2010) (independent variable specifying the lower court decision’s direction is statistically significant across all model specifications); Susan Haire et al., *Presidents and Courts of Appeals: The Voting Behavior of Obama’s Appointees*, 97 JUDICATURE 137, 143 tbl.3 (2013) (independent variable identifying conservative lower court rulings has largest coefficient).

⁴ Guthrie & George, *supra* note 1, at 363.

⁵ Appellate courts are supposed to correct errors made in trial courts and ensure that trial courts follow correct legal procedures. See G. ALAN TARR, JUDICIAL PROCESS & JUDICIAL POLICYMAKING 141 (5th ed. 2010).

appellate courts rarely reverse lower court decisions is cited by those who applaud the work of judges,⁶ as well as those critical of the American judicial system.⁷ Given the volume and variety of cases heard by appellate courts, it is imperative that they decide cases correctly. Appellate courts often decide whether someone spends years in prison or goes free. We know the appellate system fails to correct some outrageous errors. Many innocent men and women, including some awaiting execution, have been exonerated by DNA evidence, even after appellate courts reviewed their convictions and affirmed that no harmful errors were made at trial.⁸

This Article proceeds in five Parts. Part I summarizes the best available data on affirmation rates in federal and state appellate courts. Part II identifies several plausible explanations for high affirmation rates in existing literature. Some of these explanations are reassuring; others are disturbing; none are sufficiently tested. Part III describes an experiment designed to estimate the extent to which high affirmation rates can be attributed to cognitive biases and explain why affirmation rates are so high in appellate courts. The results of this experimental research, discussed in Part IV, indicate that an appellate judge's knowledge of the trial judge's original decision increases the probability it is affirmed on appeal—regardless of what the original decision was. The affirmation rate in appellate courts could be inflated by as much as 8% due to a cognitive bias in favor of affirming prior decisions. Part V concludes this Article by identifying some of the limitations of this research and offering some suggestions for future research on appellate court decision-making.

⁶ Some observers use reversal rates to grade the quality of a judge or a court system; low reversal rates suggest a judge or court is performing well. See Stephen J. Choi et al., *What Do Federal District Judges Want? An Analysis of Publications, Citations, and Reversals*, 28 J.L. ECON. & ORG. 518, 521–22 (2012); Maya Sen, *How Judicial Qualification Ratings May Disadvantage Minority and Female Candidates*, 2 J.L. & CTS. 33, 55 (2014); Roy E. Hofer, *Supreme Court Reversal Rates: Evaluating the Federal Courts of Appeals*, LANDSLIDE, Jan./Feb. 2010, at 8, <https://www.ipo.org/wp-content/uploads/2013/03/Supremecourtreversalrates.pdf> (reprinted with permission). The Ninth Circuit, in particular, is often (unfairly) criticized for having its decisions reversed by the Supreme Court. See Michelle Ye Hee Lee, *Are 80 Percent or 0.1 Percent of the 9th Circuit Court's Decisions Overturned?*, WASH. POST (Mar. 21, 2017), <https://www.washingtonpost.com/news/fact-checker/wp/2017/03/21/does-the-9th-circuit-court-overturn-80-percent-or-0-1-percent-of-its-cases>.

⁷ Critics charge that “rubber stamp” appellate courts are complicit in wrongful convictions and other miscarriages of justice. “Lies, cheating, distortions at the lower levels of the system are excused at higher ones.” BARRY SCHECK ET AL., ACTUAL INNOCENCE 175 (2003 ed. 2000); see also Hans Sherrer, *The Complicity of Judges in the Generation of Wrongful Convictions*, 30 N. KY. L. REV. 539, 565–75 (2003).

⁸ See Keith A. Findley, *Innocence Protection in the Appellate Process*, 93 MARQ. L. REV. 591, 593–601 (2009); Brandon L. Garrett, *Judging Innocence*, 108 COLUM. L. REV. 55, 98–106 (2008) (reporting that appellate courts are no more likely to find errors in cases involving exonerated defendants than matched comparison group).

I. AFFIRMATION RATES IN U.S. APPELLATE COURTS

This Part examines the best available data on appellate court outcomes at the federal and state levels. While affirmation rates are almost always very high, they vary depending on subject matter and type of appellate court.

The U.S. Courts of Appeals are intermediate-level federal appellate courts. In the twelve-month period ending June 30, 2018, of all appeals terminated on their merits by a circuit court, only 7.8% resulted in the reversal of the trial court's decision.⁹ However, the reversal rates in federal appeals vary by case type, as shown in Table 1.

Table 1. Reversal Rates in U.S. Courts of Appeals, July 2017–June 2018¹⁰

Type of Appeal	Percentage Reversed	Number of Appeals
Criminal	6.6%	7,073
U.S. Prisoner Petitions	4.8%	3,252
Private Prisoner Petitions	5.2%	5,592
Other U.S. Civil	14.1%	1,599
Other Private Civil	11.7%	6,293
Bankruptcy	7.8%	410
Admin. Agency Appeals	7.5%	3,042
Total	7.8%	27,261

As one might expect, reversal rates in criminal appeals and cases filed by prisoners (which include habeas corpus petitions filed in federal court) are lower than the reversal rate in civil appeals. While private parties typically bear the cost of their appeals, most criminal defendants are provided appellate counsel

⁹ The Administrative Office of the U.S. Courts includes decisions that affirm in part and reverse in part as affirmations. See ADMIN. OFFICE OF THE U.S. COURTS, TABLE B-5—U.S. COURTS OF APPEALS STATISTICAL TABLES FOR THE FEDERAL JUDICIARY (JUNE 30, 2018): U.S. COURTS OF APPEALS—DECISIONS IN CASES TERMINATED ON THE MERITS, BY NATURE OF PROCEEDING—DURING THE 12-MONTH PERIOD ENDING JUNE 30, 2018, at n.1, <http://www.uscourts.gov/statistics/table/b-5/statistical-tables-federal-judiciary/2018/06/30>. The Administrative Office of the U.S. Courts does not report the number of appeals terminated by decisions that affirm in part and reverse in part. However, that number can be estimated using other data. Based on a large data set of U.S. Courts of Appeals decisions, approximately 12% of published decisions of the circuit courts affirm in part and reverse in part. See Barry C. Edwards, *Codebook for Circuit Court of Appeals Database, Phase Three Update: 2003–2010*, <http://www.circuitcourtsdata.com/wp-content/uploads/2017/05/phaseThreeCodebook.pdf> (last visited Feb. 12, 2019); Barry C. Edwards, *Explore Circuit Courts Data*, POL. SCI. DATA, <http://www.poliscidata.com/pages/circuitCourtsCrossTabs.php?rowVariable=TREAT&colVariable=year> (last visited Feb. 12, 2019) (treatment categories 5 and 6 are affirmed in part and reversed in part, and affirmed in part, reverse in part and remanded, respectively).

¹⁰ ADMIN. OFFICE OF THE U.S. COURTS, *supra* note 9. The Table is restricted to appeals terminated on their merits (as opposed to dismissal by parties); miscellaneous applications are also excluded. *Id.*

by the state and have nothing to lose from filing an appeal.¹¹ Indeed, a public defender's refusal to file an appeal may lead to an ineffective assistance of counsel claim,¹² creating a particularly powerful incentive to appeal, even if there is little chance of success.¹³ Prisoner petitions filed in federal courts are generally post-conviction proceedings, which come in the form of motions to vacate, set aside, or correct a sentence by federal prisoners, or habeas corpus filings by state prisoners.¹⁴ Prisoners do not have the right to state-paid counsel in post-conviction proceedings, but the prisoner has little to lose from filing a petition *in forma pauperis* with little chance of success.¹⁵

The affirmation or reversal rates of circuit courts generally have held steady over time.¹⁶ The 7.8% reversal rate in the twelve-month period ending June 30, 2018 is comparable to reversal rates in earlier time periods. The comparable rate was 9.6% in the twelve-month period ending June 30, 2017, and in the five prior twelve-month periods, 8.5%, 7.8%, 7.2%, 6.9%, and 7.2%.¹⁷

Some additional features of federal appellate outcomes are noteworthy. When the U.S. government is a party, it usually wins.¹⁸ In civil cases, defendants are more successful on appeal than plaintiffs.¹⁹ The affirmation rate is lower in published opinions than it is in unpublished opinions.²⁰

¹¹ See *Douglas v. California*, 372 U.S. 353, 357 (1963) (holding that states must provide indigent defendants appellate counsel); *Griffin v. Illinois*, 351 U.S. 12, 18–19 (1956) (holding that indigent criminal defendants must be provided trial record for appeal).

¹² See, e.g., *Roe v. Flores-Ortega*, 528 U.S. 470, 473 (2000) (discussing ineffective assistance of counsel claim based on defense counsel's failure to file notice of appeal without client's consent).

¹³ Eve Brensike Primus, *Structural Reform in Criminal Defense: Relocating Ineffective Assistance of Counsel Claims*, 92 CORNELL L. REV. 679, 681–82 (2007).

¹⁴ See 28 U.S.C. § 2255(a) (establishing a federal prisoner's right to move to vacate, set aside, or correct the sentence). Technically, habeas corpus petitions filed pursuant to 28 U.S.C. § 2254 are civil complaints filed after a state prisoner has exhausted opportunities to appeal his conviction.

¹⁵ See *Pennsylvania v. Finley*, 481 U.S. 551, 555 (1987) ("[T]he right to appointed counsel extends to the first appeal of right, and no further.").

¹⁶ Clermont & Eisenberg, *supra* note 1, at 151; Guthrie & George, *supra* note 1.

¹⁷ These rates were reported by the Administrative Office of the U.S. Courts. For access to these reports, see <https://www.uscourts.gov/data-table-numbers/b-5?pt>All&pn>All<=All&m%5Bvalue%5D%5Bmonth%5D=6&y%5Bvalue%5D%5Byear%5D=>.

¹⁸ See generally Donald R. Songer et al., *Do the "Haves" Comes Out Ahead over Time? Applying Galanter's Framework to Decisions of the U.S. Courts of Appeals, 1925–1988*, 33 LAW & SOC'Y REV. 811 (1999); Donald R. Songer & Reginald S. Sheehan, *Who Wins on Appeal? Upperdogs and Underdogs in the United States Courts of Appeals*, 36 AM. J. POL. SCI. 235 (1992).

¹⁹ Kevin M. Clermont & Theodore Eisenberg, *Appeal from Jury or Judge Trial: Defendants' Advantage*, 3 AM. L. & ECON. REV. 125, 127 (2001).

²⁰ See Michael Hannon, *A Closer Look at Unpublished Opinions in the United States Courts of Appeals*, 3 J. APP. PRAC. & PROCESS 199, 213–15 (2001). The affirmation rate in published opinions may be declining as federal appellate courts make greater use of unpublished opinions.

On the other hand, the U.S. Supreme Court, the country's highest profile appellate court, reverses the majority of appeals it hears.²¹ One should not infer from this fact, however, that the Court frequently reverses decisions. Because its docket is discretionary, the Supreme Court does not need to affirm decisions; it can simply decline to review them. Of the nearly 8,000 appeals made to the Court each year,²² only about 50 or so result in a favorable disposition for the petitioning party.²³ Considering all cases appealed to the Supreme Court (rather than the select few it agrees to hear), the Court reverses less than 1% of the decisions appealed to it.

State appellate courts reverse lower court decisions at similarly low rates. State-level statistics are not as complete or current as federal appellate data, but the data clearly show that state appellate courts uphold most lower court decisions.²⁴ According to the Bureau of Justice Statistics, 11.9% of an estimated 69,348 criminal appeals to state appeals courts in 2010 resulted in the reversal, remand, or modification of a trial court decision.²⁵ The reversal rate is lower in state courts of last resort in the same time frame, with only 6.7% of 18,832 cases reversed, remanded, or modified by state courts of last resort, compared to a 13.9% rate in intermediate state appellate courts.²⁶ State appellate courts also rarely reverse trial courts in civil cases. According to a 2001 civil justice survey

²¹ LEE EPSTEIN ET AL., THE SUPREME COURT COMPENDIUM: DATA, DECISIONS, AND DEVELOPMENTS 274–75 tbls.3, 4, 5, 6 & 7 (6th ed. 2015) (petitioning party has won majority of cases decided by Court from 1954 to 2013).

²² *Id.* at 82–83 tbl.2, 3, 4, 5 & 6.

²³ For the number of Supreme Court decisions granting favorable dispositions to the petitioning party (by reversing a prior decision) from 1946 to 2017, see *Analysis Specifications - Modern Data (1946-2017)*, SUP. CT. DATABASE, <http://scdb.wustl.edu/analysis.php> (last visited Feb. 12, 2019) (under “Case Outcome” and “Winning Party,” select “petitioning party received a favorable disposition” and click “analyze”).

²⁴ In California, the country’s largest court system, official state reports indicate that between 2012 and 2013, of cases disposed of by written opinions, California courts of appeal (intermediate appeal courts) affirmed 7,999, reversed 870, and dismissed 286 cases. JUDICIAL COUNCIL OF CAL., 2014 COURT STATISTICS REPORT: STATEWIDE CASELOAD TRENDS, 2003–2004 THROUGH 2012–2013, at xv (2014), <http://www.courts.ca.gov/documents/2014-Court-Statistics-Report.pdf>. The rate at which these courts affirm trial court decisions appears constant; the affirmation rate in criminal appeals by defendants was 94% between 2011 and 2013. *Id.* at 26. These statistics include as affirmation decisions those that affirm in full or with some modification of the trial court opinion. *Id.* at xv and 69.

²⁵ NICOLE L. WATERS ET AL., U.S. DEP’T OF JUSTICE, OFFICE OF JUSTICE PROGRAMS, CRIMINAL APPEALS IN STATE COURTS 1 (2015), <https://www.bjs.gov/content/pub/pdf/casc.pdf>. For further analysis of the variables affecting appellate results in criminal cases in state courts, see generally Michael Heise et al., *State Criminal Appeals Revealed*, 70 VAND. L. REV. 1939 (2017).

²⁶ NICOLE L. WATERS ET AL., *supra* note 25, at 4–5 tbls.1 & 2. In the twelve states without intermediate appellate courts, parties have the right to appeal directly to the state’s court of last resort. *Id.* at 1. The reversal rate in criminal appeals in these states’ courts of last resort was 11.5%. *Id.* at 4 tbl.1. The reversal rate in criminal appeals heard by permission in other states’ courts of last resort was 5.6%. *Id.* This report classified decisions to reverse, remand, or modified any component of a trial court’s decision as reversals. *Id.* at 1.

of state courts on 1,204 general civil trials in which a litigant sought appellate review, the reversal rate was 17%.²⁷

The best available data on federal and state court appeals indicate that the overwhelming majority of lower court decisions are affirmed on appeal. This reality raises two questions which are addressed in the next Part: Why do appellate courts affirm at such a high rate? And, given the likely futility of appealing to a higher court, why do so many litigants file appeals?

II. CONFLICTING ACCOUNTS OF HIGH AFFIRMATION RATES

This Part reviews several plausible explanations for the low reversal rates observed in federal and state appellate courts. These factors are not mutually exclusive; each may explain some percentage of observed outcomes. For example, some theories hold the appellate system is healthy and dispensing justice effectively. Other plausible explanations for low reversal rates are more disturbing—they suggest that appellate courts might fail to fairly weigh the merits of appeals they hear and leave errors uncorrected. It is important to distinguish the plausible explanations for high affirmation rates observed in appellate courts because they have very different normative implications.

The most plausible reason appellate courts rarely reverse lower court decisions is that lower court judges decide cases correctly more often than not.²⁸ Real trial court judges are not flipping coins to decide which party prevails; they have training, practical experience, and take their jobs seriously. As discussed above, many criminal defendants have nothing to lose from filing a low-merit appeal.²⁹ The fact that a lower court judge ruled against the appellant is a relatively reliable signal that its argument is weaker than the appellee's argument is.

Furthermore, if trial court judges think strategically about their decisions, one would expect their decisions to be shaped, in part, by the preferences and likely actions of appellate court judges. While trial court judges may prefer to issue opinions that reflect their personal preferences, they are constrained by the prospect of appellate review. Given the possibility of reversal, lower court

²⁷ Nicole L. Waters, *Caseload Highlights: Civil Trials on Appeal – Part I*, 14 CT. STAT. PROJECT 1, 3 (2007), <http://www.courtstatistics.org/~media/Microsites/Files/CSP/DATA%20PDF/Vol14Num1CivilTrialsOnAppeal1.ashx>. No decision was made in 43% of civil appeals; 40% were affirmed and 17% were reversed. *Id.* The Court Statistics Project report included in the “reverse” category any disposition other than affirmed in full. *Id.*

²⁸ See generally Jon O. Newman, *Study of Appellate Reversals*, 58 BROOK. L. REV. 629, 631 (1992).

²⁹ See *supra* notes 14–16.

judges may temper their personal preferences to avoid getting reversed on appeal.³⁰ Strategic trial court judges weigh the benefit of writing opinions more to their liking against the possibility of reversal.

Advocates of the case selection theory answer this overarching inquiry by asking a hypothetical question about litigants: why do parties appeal cases that were correctly decided in the trial court? This theory posits that most litigants are rational actors, capable of evaluating potential litigation outcomes, and committed to maximizing their expected utility in litigation.³¹ If litigants were fully rational, economic agents, one would expect appellate court decisions to be evenly split between affirmations and reversals. According to Professors Clermont and Eisenberg's review of empirical research on the appellate phase of litigation:

Appeals that clearly favor either the appellant or the appellee would tend to be settled readily, because both sides could save costs by so acting in light of their knowledge of all aspects of the case. Difficult appeals falling close to the applicable decisional criteria would tend not to settle, because the parties would be more likely to disagree substantially with respect to their predicted outcomes. These unsettled, difficult appeals entailing divergent expectations would fall more or less equally on either side of the decisional criterion, regardless of both the position of that criterion and the underlying distribution of cases. Case selection, then, should leave for appellate adjudication a residue of appeals exhibiting some non-extreme affirmation rate. Indeed, under simplifying assumptions, and as a limiting implication, case-selection theorizing would even predict a 50% affirmation rate. That is clearly wrong, as the data prove.³²

Real litigants are not fully rational, economic agents who make litigation decisions using mathematical models of expected utility. Many litigants are insulated from the costs of appeal. Some litigants seek revenge even if it is costly. Sunk costs also cause economically irrational decisions—the appellant may consider the cost of appeal justified by the steep cost already paid to litigate

³⁰ See Clermont & Eisenberg, *supra* note 1, at 151–52; Kirk A. Randazzo, *Strategic Anticipation and the Hierarchy of Justice in U.S. District Courts*, 36 AM. POL. RES. 669, 673–75 (2008); Joseph L. Smith & Emerson H. Tiller, *The Strategy of Judging: Evidence from Administrative Law*, 31 J. LEGAL STUD. 61, 63 (2002). However, only about one-fifth of cases decided by lower federal courts are appealed and only a fraction of these appeals result in reversal, so strategic judges, forecasting the prospects of reversal, may not be strongly inclined to temper their personal preferences when making decisions. *Id.* at 63–64.

³¹ George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 4–5 (1984).

³² Clermont and Eisenberg, *supra* note 1, at 150–51.

a matter unsuccessfully in lower court.³³ Additionally, appellants and their attorneys are likely to systematically overestimate their chances on appeal.³⁴ Given this falsely confident predisposition, one would not expect appellate dockets to be limited to “difficult appeals” with equally matched arguments.

Another plausible explanation for lower reversal rates is that appellate courts are “supposed” to defer to lower courts. The standard of review that applies in a case may require the appellate court to defer to the judgment of the lower court. For example, an appellate court may overlook a minor mistake by a lower court judge who did not commit a clear error, abuse discretion, or act in an arbitrary and capricious manner. “[M]ost appeals courts’ decisions involve routine examinations of lower court outcomes, primarily using highly deferential standards of review, such as abuse of discretion or plain error.”³⁵ In some cases, the lower court’s reasoning may be faulty, but if it made the right decision, the appellate court can affirm the decision by substituting their better legal reasoning for the lower court judge’s reasoning. And even if the lower court made an error, the error may have been harmless. “Appellate judges should and do lean toward affirmation as the usual course.”³⁶

Affording lower courts this level of deference makes sense in many cases and can be defended on policy grounds. The trial court is closer to the evidence. Appellate courts cannot observe whether witnesses testified confidently or evasively; they can only read transcripts of what was said.³⁷ The local trial court judge is, therefore, in a much better position to make decisions about evidence and procedure.³⁸ Some authority must be delegated to lower courts to maintain an efficient division of labor in the judicial system.³⁹ Low reversal rates, to some extent, reflect how the federal and state appeals courts were intended to operate.

³³ See generally Hal R. Arkes & Catherine Blumer, *The Psychology of Sunk Cost*, 35 ORG. BEHAV. & HUM. DECISION PROCESSES 124 (1985).

³⁴ See generally Lyle A. Brenner et al., *Overconfidence in Probability and Frequency Judgments: A Critical Examination*, 65 ORG. BEHAV. & HUM. DECISION PROCESSES 212 (1996).

³⁵ Guthrie and George, *supra* note 1, at 372.

³⁶ Clermont and Eisenberg, *supra* note 1. For an extended analysis of this point, see Thomas Y. Davies, *Affirmed: A Study of Criminal Appeals and Decision-Making Norms in a California Court of Appeal*, 7 AM. B. FOUND. RES. J. 543 (1982).

³⁷ See *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 855 (1982) (explaining that trial court judges have unique opportunity to evaluate witnesses and evidence).

³⁸ See generally Tyler Q. Yeargain, Comment, *Discretion Versus Supersession: Calibrating the Power Balance Between Local Prosecutors and State Officials*, 68 EMORY L.J. 95, 134 (2018).

³⁹ The Supreme Court has made this argument in support of deference to trial courts. “Duplication of the trial judge’s efforts in the court of appeals would very likely contribute only negligibly to the accuracy of fact determination at a huge cost in diversion of judicial resources.” *Anderson v. Bessemer City*, 470 U.S. 564, 574–75 (1985).

Although there are some positive explanations for high affirmation rates, other cognitive factors may contribute to “affirmation bias.”⁴⁰ Affirmation bias in an appellate case is the tendency to affirm a prior decision for reasons unrelated to the relative merits of the parties’ arguments or the applicable standard of review.⁴¹

While judges have extensive knowledge and experience, they face many of the same cognitive limitations encountered by other types of decision makers.⁴² They face uncertainty and are limited in their ability to obtain complete information.⁴³ Decades of psychological research has revealed that humans use two systems to make decisions: System 1 is fast, automatic, and instinctive and System 2 is slow, deliberate, and analytic.⁴⁴ Both systems are useful—indeed vital for survival—but sometimes come into conflict, particularly when System 1 reaches conclusions before System 2 begins analyzing information.⁴⁵ As such, even while pursuing rational decisions in earnest, judges are like other decision makers who may unknowingly take mental shortcuts, such as the subconscious reliance on heuristics, to make complicated decisions.⁴⁶

Affirmation bias is a consequence of subconscious processes, rather than deliberate decision-making based on expressly stated legal principles. While scholars have yet to fully consider the role of a bias towards affirmation—or “affirmation bias”—in appellate decision-making, other types of bias in the

⁴⁰ Affirmation bias is not the same as confirmation bias. Confirmation bias is a well-studied tendency of individuals to search for or interpret information in a manner that supports or is consistent with his or her preexisting beliefs. See generally Eva Jonas et al., *Confirmation Bias in Sequential Information Search After Preliminary Decisions: An Expansion of Dissonance Theoretical Research on Selective Exposure to Information*, 80 J. PERSONALITY & SOC. PSYCHOL. 557, 557–58 (2001); Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175, 175–77 (1998). Thus, an appellate judge may experience confirmation bias in an en banc review after he or she made an initial decision in the case, but not in ordinary appellate review of another judge’s decision.

⁴¹ Affirmation bias is comparable to the “affirmance effect” discussed by Guthrie and George, *supra* note 1, but the term affirmative bias is more descriptive of a causal process that may explain appellate court outcomes.

⁴² A growing body of legal scholarship applies these psychological insights to judging. See generally Chris Guthrie, *Misjudging*, 7 NEV. L.J. 420 (2007); Brian Sheppard & Andrew Moshirnia, *For the Sake of Argument: A Behavioral Analysis of Whether and How Legal Argument Matters in Decisionmaking*, 40 FLA. ST. U. L. REV. 537 (2013); Eyal Zamir & Ilana Ritov, *Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation*, 41 J. LEGAL STUD. 165 (2012).

⁴³ HERBERT A. SIMON, *ADMINISTRATIVE BEHAVIOR: A STUDY OF DECISION-MAKING PROCESSES IN ADMINISTRATIVE ORGANIZATIONS* 93–96 (4th ed. 1997).

⁴⁴ For an accessible summary of this research, see generally DANIEL KAHNEMAN, *THINKING, FAST AND SLOW* (2011).

⁴⁵ *Id.* at 415–16.

⁴⁶ This view of individual decision-making is informed by a wealth of research in cognitive psychology. For a seminal account, see generally Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974).

judicial system have garnered considerable attention. For example, prior research has examined how the race, gender, and other characteristics of a defendant can affect judges' sentencing decisions.⁴⁷ Under certain conditions the outcomes of football games can even affect the decisions judges make.⁴⁸ In addition to legal decisions, researchers have found biases in the assignment of majority opinion-writing in state supreme courts, with the graduates of "elite" educational institutions more likely to receive majority opinion-writing assignments on salient cases, while both female and minority justices are less likely to receive such assignments.⁴⁹ These are but a few examples of where unconscious and implicit biases have been examined in the context of the courts.

The courts and the federal judiciary are characterized by many decisions that must be made under conditions of uncertainty. This uncertainty might arise from conflicting versions of events or opposing views on the implications of specific facts. Uncertainty may lead decision makers, in this case judges, to rely on different heuristics and cognitive shortcuts. This can lead to bias in different decision outcomes.

First, the tendency to affirm others' judgments is related to the natural human tendency to conform. The Asch conformity experiments suggest multiple forms of social pressure to confirm a prior decision.⁵⁰ When participants were compelled to announce their opinions verbally in front of the group, they felt a social pressure to fit in.⁵¹ When participants wrote their answers privately on paper, they could give nonconforming answers without worrying about the group's judgment, but also could take the group's vote into account as relevant information.⁵² Judges make decisions knowing the results will be documented,

⁴⁷ The research literature on racial biases in the judicial system is extensive. See, e.g., Jill K. Doerner & Stephen Demuth, *The Independent and Joint Effects of Race/Ethnicity, Gender, and Age on Sentencing Outcomes in U.S. Federal Courts*, 27 JUST. Q. 1, 23 (2010); Ronald S. Everett & Roger A. Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANT. CRIMINOLOGY 189, 190 (2002); Darrell Steffensmeier & Stephen Demuth, *Does Gender Modify the Effects of Race-Ethnicity on Criminal Sanctioning? Sentences for Male and Female White, Black, and Hispanic Defendants*, 22 J. QUANT. CRIMINOLOGY 241, 252–56 (2006); Laura T. Sweeney & Craig Haney, *The Influence of Race on Sentencing: A Meta-Analytic Review of Experimental Studies*, 10 BEHAV. SCI. & L. 179, 190–92 (1992).

⁴⁸ Ozkan Eren & Naci Mocan, *Emotional Judges and Unlucky Juveniles* 3 (Nat'l Bureau of Econ. Research, Working Paper No. 22611, 2016).

⁴⁹ Robert K. Christensen et al., *Race and Gender Bias in Three Administrative Contexts: Impact on Work Assignments in State Supreme Courts*, 22 J. PUB. ADMIN. RES. & THEORY 625, 637 (2012).

⁵⁰ For the seminal study of conformity, see Solomon E. Asch, *Opinions and Social Pressure*, 193 SCI. AM. 31 (1955). There is, however, increasing evidence that the tendency to conform has declined over time. See Rod Bond & Peter B. Smith, *Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task*, 119 PSYCHOL. BULL. 111, 124 (1996).

⁵¹ Asch, *supra* note 50, at 3.

⁵² Solomon E. Asch, *Studies of Independence and Conformity: A Minority of One Against a Unanimous*

publicly available, and subject to review by higher courts. An implicit pressure to affirm prior decisions could exist because deviating from the status quo by overturning the lower court's decision could be more highly scrutinized. If one thinks that others have rendered judgment, one might feel compelled to override his or her own independent judgment and conform to group decisions.⁵³

These implicit pressures might also drive a bias towards the status quo.⁵⁴ "All other things being equal, individuals tend to prefer an option that is consistent with the status quo rather than one that requires change from the status quo."⁵⁵ Although status quo bias is, in some respects, comparable to the doctrine of stare decisis, this cognitive bias differs from a fully rationalized decision to follow a legal doctrine. Judges may cite the doctrine of stare decisis and articulate its rationale for keeping matters settled; in contrast, judges may be unaware that their judgments are compromised by cognitive bias and certainly will not attribute their opinions to bias.

As with the status quo bias, people also have a bias towards omission of action.⁵⁶ Individuals tend to react more strongly and negatively to reversal, an outcome produced by action, reversal, than affirmation, which is an inaction or omission. In this context, an appellate judge may experience more discomfort from an error caused by their decision to reverse a previous decision than from affirming a case with an error the judge did not themselves make—which may, in turn, be attributed to the trial court judge. "The omission bias thus induces people not to act, to just 'leave things as they are.'"⁵⁷ Furthermore, judges may not necessarily want to create more work for other judges. State and federal courts deal with tremendous caseloads and are often constrained by limited resources.⁵⁸ If the appellate court reverses a decision, the trial judge may need to order a new trial or additional hearings.

⁵³ Majority, 70 PSYCHOL. MONOGRAPHS: GEN. & APPLIED 1, 56–58 (1956).

⁵⁴ See Burton M. Atkins, *Judicial Behavior and Tendencies Towards Conformity in a Three Member Small Group: A Case Study of Dissent Behavior on the U.S. Court of Appeals*, 54 SOC. SCI. Q. 41, 43 (1973).

⁵⁵ See generally Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSP. 193 (1991); William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7 (1988).

⁵⁶ Guthrie and George, *supra* note 1, at 377.

⁵⁷ See generally Terry Connolly & Jochen Reb, *Omission Bias in Vaccination Decisions: Where's the "Omission"? Where's the "Bias"?*, 91 ORG. BEHAV. & HUM. DECISION PROCESSES 186 (2003).

⁵⁸ Guthrie and George, *supra* note 1, at 380.

⁵⁹ There is some evidence that affirmation rates in circuit courts are positively correlated with their caseloads. See Guthrie and George, *supra* note 1, at 361. Experimental research indicates that legal decisions are strongly influenced by time pressures. See generally Brian Sheppard, *Judging Under Pressure: A Behavioral Examination of the Relationship Between Legal Decisionmaking and Time*, 39 FLA. ST. U. L. REV. 931 (2012).

Professors Chris Guthrie and Tracey George discuss different explanations for what they call an “affirmance effect” in circuit courts.⁵⁹ Their work is largely exploratory, raising more questions than answers. “[E]ach of these accounts sheds some light on this question, but none of them is capable of providing a definitive answer.”⁶⁰ To fill this gap, the next Part introduces and explains the results of an experiment conducted to identify how affirmation bias affects appellate review of lower court decisions more broadly.

III. AN EXPERIMENT EXPLORING AFFIRMATION BIAS

The fundamental problem with observational analysis is that *we do not know what the affirmation rate should be when no bias exists*. Given a sample of appellate court cases, no researcher could practically determine what the courts got “right” and what they got “wrong.” There is no reliable method of coding how cases “should” have been decided and, thus, no reliable way of assessing whether the affirmation rate is “too high” using observational data. Because an observational strategy is unlikely to shed light on bias in affirmation bias in appellate courts, I conducted an experiment to overcome the limitations of empirical analysis and isolate the effect of affirmation bias. This Part details how the experiment was conducted.

Experimental research designs give the researchers the ability to change variables that cannot be manipulated in observational research.⁶¹ In this case, I randomly manipulated the direction of a hypothetical lower court decision synthesized from a real case, and held all other case factors, such as the weight of evidence and applicable appellate review standard, constant to isolate and measure the extent of affirmation bias in decision-making.

In the experiment, I presented political science majors and law school students at a large, public research university with a vignette that described a motor vehicle stop and asked subjects to decide whether there was reasonable suspicion to stop and search the vehicle. I recruited students to participate in the experiment at the end of regular class sessions.⁶²

⁵⁹ See generally Guthrie and George, *supra* note 1.

⁶⁰ *Id.* at 358.

⁶¹ WILLIAM R. SHADISH ET AL., EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE 7–9 (2002).

⁶² Although participation was voluntary and not part of any class requirement, nearly all students in the classes recruited participated in this experiment. The research was approved by the University’s Institutional Review Board and participants signed consent documents (on file with author).

The vignette is a modified version of *United States v. Jimenez-Medina*, a case decided by the Ninth Circuit.⁶³ The vignette presents the case of a truck driver stopped by a border patrol agent who suspected the driver was smuggling illegal aliens into the United States.⁶⁴ The facts are not disputed but support a range of reasonable inferences about the likelihood of criminal activity. For example, the driver violated no traffic laws, but was driving unusually slowly on an interstate highway at 2:00 a.m., arousing the border patrol agent's suspicion.⁶⁵ The agent pulled the truck over.⁶⁶ The agent searched the vehicle and found that the driver was smuggling drugs, not people, across the border.⁶⁷ The facts of the case presented to participants never varied.

Jimenez-Medina is a good hypothetical case for this research. First, there is not an obvious “right” answer.⁶⁸ In the actual case, two federal judges (the trial court judge and one circuit court judge) concluded that there was reasonable suspicion to make the stop; two other judges (the circuit majority) reached the opposite conclusion and sided with the defendant.⁶⁹ Second, this appeal is a good example of de novo review.⁷⁰ There is no dispute over the facts or laws; the dispute solely centers on the application of the law to the facts. Third, this case is representative of a question frequently decided by appellate courts and one that involves balancing individual rights against public safety. If reasonable suspicion existed, the stop was permissible, and the evidence obtained by searching the vehicle is admissible in a criminal trial. If reasonable suspicion did not exist, the stop was illegal, and evidence subsequently obtained by searching the vehicle is not admissible.

⁶³ 173 F.3d. 752 (9th Cir. 1999).

⁶⁴ A complete copy of the survey instrument is included in the Appendix.

⁶⁵ *Jimenez-Medina*, 173 F.3d. at 753–54. Defendant Jimenez-Medina was stopped around 11:15 p.m. *Id.* Some details were modified in the hypothetical case summary.

⁶⁶ *Id.* at 754.

⁶⁷ *Id.*

⁶⁸ If the case involved a “smoking gun,” subjects would find it easy to disregard the subtle suggestion of a prior decision and apply the law in a manner consistent with facts. Easy decisions are not particularly interesting to analyze. This is not to suggest that all appeals are difficult cases, but that reasonable minds are inclined to disagree about many of them.

⁶⁹ Judge Donald Molloy, a federal district judge from Montana sitting by designation, wrote the opinion, which was joined by Judge John Noonan. *Jimenez-Medina*, 173 F.3d at 753. Judge Stephen Trott dissented. *Id.* at 756. The trial court judge was Judge Paul Rosenblat of the United States District Court for the District of Arizona. *Id.* at 753. The case was not heard by the U.S. Supreme Court. Jimenez-Medina was sentenced to fifty-five months imprisonment after pleading guilty to possessing cocaine with intent to distribute. *Id.*

⁷⁰ See *Ornelas v. United States*, 517 U.S. 690, 699 (1996) (deciding whether reasonable suspicion existed to make warrantless stop should be reviewed de novo). In contrast, a magistrate’s determination that probable cause exists to issue a search warrant is not subject to de novo review and entitled to great deference. See *Illinois v. Gates*, 462 U.S. 213, 236 (1983).

Consistent with the applicable standard of review in an appeal like this, the experimental subjects were instructed to decide the case as if there were no prior decision, giving no deference to the prior decision in the case. When judges engage in de novo review, they are supposed to decide matters without any deference to a prior decision from a lower court. This is an important point and one that is emphasized in the instructions to test subjects:

The standard of review in a case like this is called de novo review. This means you should view the case from the same position as the trial court judge. You should consider the matter anew, as if there were no prior ruling on the admissibility of this evidence. No deference is given to the District Court Judge's decision.

Although deference to lower court decisions is justified in many cases, this instruction was intended to eliminate the possibility of deliberate deference to the prior decision. After reading about the case, subjects were asked whether there was reasonable suspicion to stop the vehicle. No time limit was imposed.

The only variation in the experimental vignettes is the procedural history. Subjects were randomly assigned one of three possible procedural histories: (1) deciding the case after a trial court judge ruled in the prosecution's favor (finding that reasonable suspicion existed), (2) deciding the case after a trial court judge ruled in the defendant's favor (finding that reasonable suspicion did not exist),⁷¹ or (3) deciding the case without a prior ruling (in the role of a trial court judge). The randomized vignettes and response forms were distributed in identical folders to effect double-blind assignments. For reasons discussed below, the final treatment condition (no prior ruling) is not strictly necessary; no assumption is made with respect to whether reasonable suspicion existed, but it provides a useful reference point. Of the participants asked to decide an appeal after a lower court ruling had been made, there was a 50% chance of being assigned to Group 1 (prosecutor won in lower court) and a 50% chance of being assigned to Group 2 (defendant won in lower court).

Randomly assigning experiment subjects to varying treatment conditions effectively controlled for other variables that may affect how subjects could have evaluated the evidence of reasonable suspicion in this case. One would expect,

⁷¹ A prosecutor may appeal a trial court judge's pretrial decision to exclude evidence in a criminal case. 18 U.S.C. § 3731 (2012). After a trial begins (a jury is sworn or the first witness in a bench trial is sworn), jeopardy attaches, and if the prosecutor loses the case, appealing acquittal violates constitutional protections against double jeopardy. See Scott J. Shapiro, Note, *Reviewing the Unreviewable Judge: Federal Prosecution Appeals of Mid-Trial Evidentiary Rulings*, 99 YALE L.J. 905, 905 n.1 (1990). Defendants cannot sit on motions to exclude evidence until trial, however, to defeat the government's right to appeal adverse evidentiary rulings. See FED. R. CRIM. P. 12(b)(3)(C).

for example, that political conservatism would make some students more inclined to find reasonable suspicion to stop a suspected immigrant smuggler. Because conservative students were randomly assigned to both treatment groups, the group told that the prosecutor won in the lower court would be, on average, just as conservative as the group told the defendant won in the lower court. Therefore, differences between the groups cannot be attributed to political differences or any other confounding variable.

In this experiment, one can derive the expected affirmation rate when decision-making is unbiased. The decision maker affirms the prior decision in two possible situations: when the decision maker in Group 1 agrees with the trial court that there *was* reasonable suspicion and when the decision maker in Group 2 agrees that there *was not* reasonable suspicion. The probability of each of these situations can be calculated. Let RS be the percentage of subjects who believe that reasonable suspicion existed in the hypothetical fact pattern and would vote in favor of the prosecution in a fair hearing of this appeal. The inverse quantity, $100\% - RS$, equals the percentage of respondents who do not think reasonable suspicion existed and would vote in the defendant's favor. In a fair hearing without affirmation bias (H_0),⁷² the expected affirmation rate equals the percentage of respondents who think there was reasonable suspicion (RS) multiplied by the probability of being told the prosecution won in the trial court (50%) plus the percentage of respondents who do not think there was reasonable suspicion ($1 - RS$) multiplied by the probability of being told the defendant won (50%).

$$\begin{aligned} H_0: \text{Affirmation Rate} &= 50\% \times RS + 50\% \times (100\% - RS) \\ &= 50\% \times RS + 50\% - 50\% \times RS \\ &= 50\% \end{aligned}$$

$$H_A: \text{Affirmation Rate} > 50\%$$

Applying some simple algebra to the formula for the expected affirmation rate in a fair hearing without affirmation bias, the RS terms cancel out (no matter what the value of RS is) and 50% is the expected affirmation rate.⁷³

For those less mathematically inclined, it may help to think of grading a true-false test question that was answered by flipping coins. Regardless of whether

⁷² The notation H_0 signifies the null hypothesis and H_A the alternative hypothesis (also called the research hypothesis).

⁷³ The expected 50% affirmation rate in the experiment is a mathematical result of randomized treatments. It is not based on the case selection theory articulated by Clermont and Eisenberg, *supra* note 1, at 150–51, nor does it require the probability of finding reasonable suspicion in the hypothetical fact pattern to be 50%.

the answer to the question is true or false, the coin has a 50% chance of getting the answer right. In this experiment, I essentially flipped coins to answer the question of whether reasonable suspicion existed and asked subjects whether the coins, disguised as judges, gave the right answer. If subjects grade coin-flip decisions without bias, we should expect them to affirm 50% of the answers as correct and mark the other 50% as errors. Random coin flips sometimes get more questions right than expected, but we can quantify how much the percentage scored correct will vary as a result of random error.

The research hypothesis (H_A) is that more than half of respondents will affirm the prior decision. This does not imply that the law and facts are such that votes should be evenly split between prosecution and defense; there is no hypothesized value of RS.⁷⁴ Because this experimental design allows us to calculate the expected affirmation rate in a fair hearing, we can define and calculate affirmative bias as the difference between the affirmation rate in the experiment and 50%:

$$\text{Affirmation Bias} = \text{Affirmation Rate} - 50\%$$

In the next Part, I present the results of the experiment, test the hypothesis, and offer an estimate of affirmation bias.

IV. RESULTS

This Part presents, discusses, and analyzes the results of the experiment, starting by evaluating whether the subjects were effectively randomized between treatment conditions. This is important because effective randomization allows us to rule out alternative explanations for any significant differences observed between the groups. I asked subjects some demographic questions to compare the treatment groups. Descriptive statistics on the sample are reported below.

⁷⁴ As noted above, for experimental purposes, this experiment was designed to avoid extreme RS values close to 100% (everyone thinks there is reasonable suspicion) or 0% (no one thinks there is reasonable suspicion).

Table 2. Description of Subjects in Sample by Treatment Condition

	Winner in Trial Court		Overall
	Prosecution	Defendant	
Female	53.3%	48.2%	50.9%
Law Students	61.7%	60.0%	60.9%
Non-White	18.5%	21.4%	19.9%
Political Ideology*	5.4	5.4	5.4
Religiosity*	5.4	6.1	5.7
Not Employed	67.0%	67.1%	67.1%
Family Income over \$75,000	19.8%	21.3%	20.5%
Father Has B.A. or More	69.6%	72.9%	71.2%
Mother Has B.A. or More	72.8%	71.8%	72.3%
Sample Size	85	93	178

*Political Ideology and Religiosity are measured on 10-point scales with higher values corresponding to a more conservative ideology and more religious, respectively.

The two groups are not identical because the assignment of subjects to one group or the other was random. There are some slight differences between the groups, including the number of subjects randomly assigned to the groups, but none of the differences summarized in Table 2 are statistically significant.⁷⁵ Thus, if Groups 1 and 2 decide the appeal in a significantly different manner, one may infer that the varying treatment conditions (i.e., the difference in prior decision) caused the observed difference because that is the only significant difference between the groups.

Participants told they were reviewing a decision in which the prosecution won in the trial court found that reasonable suspicion existed and decided the appeal in the prosecution's favor 64.5% of the time. In contrast, participants told the defendant won in the lower court decided the appeal in the prosecution's favor 49.4% of the time.⁷⁶ Participants told that the prosecutor won in the lower court found reasonable suspicion at higher rates than those told the defendant won, despite both groups being instructed to disregard the prior ruling. Table 3 summarizes whether experimental subjects thought there was reasonable suspicion to stop the vehicle in the hypothetical scenario according to which party won in the trial court. A test of statistical significance indicates that the

⁷⁵ Differences between group were assessed with two-sample *t*-tests (for Political Ideology and Religiosity variables) and difference of proportions tests (all other variables listed in Table 2).

⁷⁶ Fifty-six percent of students deciding the matter in the "clean slate" condition found reasonable suspicion and decided the matter for the prosecution, which is very close to the overall percentage of students in experimental conditions who opted for the prosecution (57.3%).

effect of the prior decision on the outcome is unlikely to occur by chance.⁷⁷ Applying the logic of null hypothesis testing, one rejects the null hypothesis, H_0 , because the evidence favors the alternative hypothesis, H_A .

Table 3. Summary of Decisions Made in the Experiment*

Winner on Appeal	Winner in Trial Court		Overall
	Prosecution	Defendant	
Prosecution (RS existed)	64.5% (60)**	49.4% (42)	57.3% (102)
	35.5% (33)	50.6% (43)	42.7% (76)
Totals	100% (93)	100% (85)	100% (178)

* Chi-Square Test Statistic = 4.14 (p-value 0.042).

** Count in each cell appears in parentheses.



= Affirmations



= Reversals

Recall that subjects affirm the prior decision in two possible situations: first, when they agree that there was reasonable suspicion (60 of 93 subjects presented with a prior ruling of reasonable suspicion affirmed that ruling) and, second, when they agree that there was not reasonable suspicion (43 of 85 presented with a prior ruling that there was not reasonable suspicion affirmed that decision). Overall, the experimental subjects agreed with the prior decision 57.9% of the time.⁷⁸ The expected affirmation rate when the decision on appeal is unbiased is 50.0%. The difference between the observed and expected affirmation rates represents the best estimate of the affirmation bias: 7.9%.

Because the true value of RS is unknown, one should not infer that subjects in Group 1 exhibited affirmation bias but those in Group 2 did not. If, for

⁷⁷ The Chi-Square test is a venerable statistical test of the independence of two variables with discrete values. See generally PHILIP H. POLLOCK III, THE ESSENTIALS OF POLITICAL ANALYSIS 165–70 (5th ed. 2016). The test statistic corresponds to the deviation between the frequencies observed in a cross tabulation (like Table 3) and the frequencies expected when H_0 is true and there is no relationship between the two variables. *Id.* The higher the test statistic, the greater the deviation of observed frequencies from expected frequencies. *Id.* The p-value quantifies the probability of observing the test statistic if H_0 is true. *Id.* It is conventional in social sciences to reject H_0 when the p-value of a test statistic is less than .05. See generally Michael Cowles and Caroline Davis, *On the Origins of the .05 Level of Statistical Significance*, 37 AM. PSYCHOLOGIST 553 (1982); James K. Skipper, Jr. et al., *The Sacredness of .05: A Note Concerning the Uses of Statistical Levels of Significance in Social Science*, 2 AM. SOCIOLOGIST 16, 16 (1967).

⁷⁸ The overall percentage of subjects who affirmed the prior decision is equal to the number of subjects who agreed with the prior ruling divided by the overall sample size, (60 + 43)/178.

example, reasonable suspicion should be found 65% of the time, Group 2's finding reasonable suspicion only 49.4% of time is too low (their affirmation rate should only be 35%). This is the problem with a strictly observational approach: the true value of RS is unknown and unknowable. As demonstrated mathematically in the preceding Part, it is only when we random experimental conditions that the RS term drops out and we can expect an overall 50% affirmation rate from unbiased decision makers.

The limited use of a “clean slate” version suggests that affirmation bias cuts both ways and is not strictly prosecutorial bias, but this observation is merely a starting point for future research. 56.0% of students randomly assigned to the group with no prior decision ruled in favor of the prosecution whereas 44.0% ruled in favor of the defendant.⁷⁹ The prosecution’s win rate following a prior decision in its favor is 64.5%, 8.5% higher than its win rate in the clean slate condition (56.0%). The defendant’s win rate following a prior decision in its favor is 50.6%, 6.6% higher than his win rate in the clean slate condition (44.0%). These appear to be comparable gains for both the prosecution and defense.

This experimental analysis of legal decision-making suggests that the affirmation rate in appellate courts could be as much as 8% higher than it should be due to a cognitive bias in favor of affirming prior rulings. Of course, the results are suggestive and do not definitely prove that appellate decision-making is infected by affirmation bias. A number of potential limitations to this research are discussed in the next Part. Despite the limitations of this analysis and the need for future research, given the volume and importance of cases decided in federal and state appellate courts, even a small level of affirmation bias seriously compromises the fairness and impartiality of our justice system.

V. DISCUSSION OF LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This Article is intended to define and discuss the possibility of affirmation biases in appellate courts. Drawing on theories from social psychology, specifically those related to the limits of human rationality and the use of decision heuristics, this research shows how knowledge of a prior decision can subconsciously affect a decision process. Specifically, decision makers have a marked tendency to affirm prior decisions, even prior decisions no more

⁷⁹ While the outcomes are not perfectly balanced, the facts of the case appear to lead to a relatively even split on the question of reasonable suspicion. Statistically, 56.0% is not different from 50.0% in a one-sample difference of proportions test given the small sample size ($n = 25$). For the purposes of this research, the hypothetical case is a “toss-up.”

informative than a coin flip that they are instructed to disregard. The results presented here help us understand a prominent feature of the American judicial system, extremely high affirmation rates in appellate courts, that has been, until now, not fully appreciated by legal scholars.

This Article reports the results of an experimental study of law students and political science majors to measure the extent to which knowledge of a prior decision could affect legal judgments when all other facts are held constant. One must recognize a tradeoff between internal and external validity in this type of controlled experiment. To obtain an internally valid measure of affirmation bias, I used a controlled, artificial setting and sacrificed some external validity—the ability to generalize these findings to appellate court judges in the real world.⁸⁰ One should be cautious about generalizing these findings to appellate court judges. Cognitive biases might become attenuated with professional experience among judges.⁸¹ At the same time, law students are only a couple years away from being eligible to serve as law clerks for appellate court judges. Clerks' responsibilities vary, but it would not be unusual for law clerks to write first drafts of opinions or internal memoranda regarding the merits of cases. Prior experimental research on legal decision-making based on student samples has proven surprisingly generalizable, particularly with respect to the influence of biases and heuristics.⁸²

Another potential limitation is the strength of the treatment. The prior judge's decision, which is subtly mentioned at the end of vignette, is relatively mild and easy to overlook. In this experiment, the lower court decision was rendered by a hypothetical judge, not a real federal judge with social connections and a lifetime appointment. Experimental subjects ran no risk of offending a federal judge by calling his or her judgment into question. In practice, the prior ruling is not merely a bit of information that hints at the correct decision, but rather is a decision rendered by a fellow judge. District court judges frequently serve by designation on appellate panels and this practice plays an important role in the socialization of federal judges.⁸³ Many circuit court judges served on

⁸⁰ See generally Derrick M. Anderson & Barry C. Edwards, *Unfulfilled Promise: Laboratory Experiments in Public Management Research*, 17 PUB. MGMT. REV. 1518 (2015).

⁸¹ Prior research shows that professional experience might mitigate the effect of decision heuristics leading to biased outcomes. See Mogens Jin Pedersen et al., *Punishment on the Frontlines of Public Service Delivery: Client Ethnicity and Caseworker Sanctioning Decisions in a Scandinavian Welfare State*, 28 J. PUB. ADMIN. RES. & THEORY 339, 350 (2018).

⁸² Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, 23 LAW & HUM. BEHAV. 75, 78–81 (1999); Avishalom Tor, *The Methodology of the Behavioral Analysis of Law*, 4 HAIFA L. REV. 237, 285–86 (2008).

⁸³ See James J. Brudney & Corey Ditslear, *Designated Diffidence: District Court Judges on the Courts*

district courts before being appointed to their circuit court.⁸⁴ The American judiciary can be characterized as a social network among different types of judges and their law clerks.⁸⁵ These professional considerations cannot be replicated effectively in a short vignette, but we know the decision-making context is important.

One possibility, not tested in this research, is that the *de novo* standard of review “backfires” and has the unintended effect of bolstering the prior judge’s decision. In this experiment, subjects were told that a judge has already made a decision, but were then asked to disregard the prior decision and consider the matter anew. Research on limiting instructions given to juries in trials suggests that an instruction to ignore a legally irrelevant fact tends to make that fact more salient in the minds of jurors.⁸⁶ If the instruction to disregard the prior decision is comparable to a limiting instruction to disregard irrelevant evidence, it may have the perverse effect of bolstering, rather than negating, the influence of the prior decision. This possibility could be tested by omitting the instruction to disregard the prior decision from some vignettes.

Additional avenues for future research include examining variables that correlate to affirmation bias and whether it “cuts both ways.” This sample was not large or diverse enough to rigorously examine whether personal characteristics, like gender, family background, or political ideology, make subjects significantly more likely to affirm prior decisions or if the decisions of some judges are more likely to be affirmed than decisions made by other judges.⁸⁷ Future research may consider whether winning in the trial court has the same effect for defendants and prosecutors on appeal. Given prosecutors’ higher winning percentages, one might suspect that a prior decision in the

of Appeals, 35 LAW & SOC’Y REV. 565, 597–600 (2001); Paul M. Collins & Wendy L. Martinek, *The Small Group Context: Designated District Court Judges in the U.S. Courts of Appeals*, 8 J. EMPIRICAL LEGAL STUD. 177, 180–81 (2011) (noting that district judges serve on appellate panels in at least 10% of cases decided on merits). In fact, the *Jimenez-Medina* opinion was written by a district court judge sitting on a Ninth Circuit panel by designation. 173 F.3d. 752, 753 (9th Cir. 1999).

⁸⁴ BARRY J. McMILLION, CONG. RESEARCH SERV., R43538, U.S. CIRCUIT COURT JUDGES: PROFILE OF PROFESSIONAL EXPERIENCES PRIOR TO APPOINTMENT 6–7 (2014).

⁸⁵ See generally Daniel M. Katz & Derek K. Stafford, *Hustle and Flow: A Social Network Analysis of the American Federal Judiciary*, 71 OHIO ST. L.J. 457 (2010).

⁸⁶ See Joel D. Lieberman & Jamie Arndt, *Understanding the Limits of Limiting Instructions: Social Psychological Explanations for the Failures of Instructions to Disregard Pretrial Publicity and Other Inadmissible Evidence*, 6 PSYCHOL. PUB. POL’Y & L. 677, 689–91 (2000); Nancy Steblay, *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, 30 LAW & HUM. BEHAV. 469, 486 (2006).

⁸⁷ For an empirical analysis of the effect of judge characteristics, see Maya Sen, *Is Justice Really Blind? Race and Reversal in U.S. Courts*, 44 J. LEGAL STUD. S187, S220–21 (2015).

prosecutor's favor confirms prior expectations while a prior decision in the defendant's favor is a more ambiguous cue. This research suggests that affirmation bias cuts both ways and is not strictly prosecutorial bias, but this is merely a starting point for future research.

CONCLUSION

This Article reports the results of a randomized experiment used to test a theory of judicial decision-making. Experimental research is useful in this context because it is impossible to distinguish among potential causes for high affirmation rates with observational methods. This experimental research controls for plausible explanations for high affirmation rates to isolate and measure the effect of a cognitive bias. This is not the only application for experimental research on judicial decision-making. While conducting experiments on judges may be impractical, conducting experiments on undergraduate students in prelaw majors, law students, and lawyers, can be a practical method for gaining insight on decision-making in the legal content.

APPENDIX: VIGNETTE USED IN EXPERIMENT

This hypothetical situation asks you to decide whether a Border Patrol Agent had reasonable suspicion to make a stop that led to the discovery of illegal drugs. Please read this information carefully and then mark your answers to the questions below.

You are an appellate court judge. In this case, the Prosecution and Defense disagree about the admissibility of evidence obtained by stopping and searching a vehicle. The parties agree about the facts and about the law, but disagree whether the facts satisfy the legal standard for reasonable suspicion. A District Court Judge has already made a ruling and that ruling is being appealed.

A Border Patrol Agent may not stop and detain a vehicle without reasonable suspicion. This is a right protected by the U.S. Constitution. An Agent has reasonable suspicion when the Agent has particular and objective reasons to suspect the person stopped of criminal activity. The facts should be considered as a whole. Reasonable suspicion cannot be based on broad profiles which cast suspicion on entire categories of people.

The standard of review in a case like this is called de novo review. This means you should view the case from the same position as the trial court judge. You should consider the matter anew, as if there were no prior ruling on the admissibility of this evidence. No deference is given to the District Court Judge's decision.

Here is a summary of what happened:

The defendant was driving his pickup truck on the Interstate at approximately 11:15 p.m. The defendant was observed by a Border Patrol Agent. The Agent first noticed the pickup because it was traveling approximately 45–50 mph in an area posted with a 75-mph speed limit. The truck was not being driven on a spare tire (which would require reduced speed). Based on his eleven years of experience, eight of them patrolling in the area, the Agent knew the Interstate was used as a route for smuggling illegal aliens into the United States (alien smuggling is a crime). It is also the Agent's experience that open-bed pickups are often used to smuggle aliens.

The Agent began to follow the pickup in his marked sedan and performed a registration check. While waiting for this information, he observed the pickup weave within its lane, leading the Agent to believe that the driver was preoccupied with the Agent's presence. The Agent knew from physical observation that there was no one in the back of the pickup, but the pickup had tinted windows, preventing the Agent from seeing how many people were inside the truck. A

registration check showed that the owner of the pickup lived in a city in Mexico where there had been a recent increase in alien smuggling. This information led the Agent to conclude the pickup had “recent border access.” A border patrol shift change takes place near 10:00 p.m., the approximate time the Agent guessed the pickup would have passed through the border check point if it had been continuously traveling on the Interstate at the observed speed. The Agent believed it was not uncommon for alien smugglers to pass through border check points during shift changes in order to avoid detection.

Based upon the facts as interpreted by the Agent, he activated his lights and pulled the pickup over.

The District Court Judge held a pre-trial hearing, found that reasonable suspicion [existed/did not exist], and ruled that the evidence is [admissible/inadmissible]. The [Defense/Prosecution] appeals this ruling to your court. The [Defense/Prosecution] argues that the Agent [did not have/had] reasonable suspicion to stop the vehicle and, as a result, the evidence [is not/is] admissible. The [Prosecution/Defense] argues that the Agent [had/did not have] reasonable suspicion to stop the vehicle and, therefore, the evidence is [admissible/inadmissible].