

Spring 2015

Can Judges Make Reliable Numeric Judgments? Distorted Damages and Skewed Sentences

Jeffrey J. Rachlinski

Cornell Law School, jjr7@cornell.edu

Andrew J. Wistrich

U.S. District Court, Central District of California

Chris Guthrie

Vanderbilt University Law School, Chris.guthrie@vanderbilt.edu

Follow this and additional works at: <http://scholarship.law.cornell.edu/facpub>

 Part of the [Judges Commons](#), and the [Remedies Commons](#)

Recommended Citation

Jeffrey J. Rachlinski, Andrew J. Wistrich, and Chris Guthrie, "Can Judges Make Reliable Numeric Judgments? Distorted Damages and Skewed Sentences," 90 *Indiana Law Journal* (2015)

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Publications by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

Can Judges Make Reliable Numeric Judgments? Distorted Damages and Skewed Sentences

JEFFREY J. RACHLINSKI, ANDREW J. WISTRICH & CHRIS GUTHRIE*

In a series of studies involving over six hundred trial judges in three countries, we demonstrate that trial judges' civil damage awards and criminal sentences are subject to influences that make them erratic. We found that the presence of misleading numeric reference points (or "anchors") affected judges' decisions in a series of hypothetical cases. Specifically, judges imposed shorter sentences when assigning sentences in months rather than in years; awarded higher amounts of compensatory damages when informed of a cap on damage awards; imposed different sentences depending upon the sequence in which criminal cases were presented to them; and were influenced by a plaintiff's reference to a damage award seen on a "court TV show." Taken together, the results suggest that unless judges take steps to reduce their susceptibility to anchors, their awards and sentences are apt to be highly unreliable. We also suggest how judges can safeguard against these influences and assign more stable awards and sentences.

INTRODUCTION	695
I. EVIDENCE OF ARBITRARINESS: ANCHORING AND SCALING.....	701
II. ANCHORING AND JUDGES.....	706
III. THE CURRENT STUDY	710
A. STUDY 1: SCALING EFFECTS.....	711
B. STUDY 2: DAMAGE CAPS	718
C. STUDY 3: ORDER EFFECTS	724
D. STUDY 4: DEBIASING	731
IV. IMPLICATIONS.....	736
CONCLUSION.....	739

INTRODUCTION

How many words, on average, did law review articles published in 2012 contain? Before you answer, was it more or less than 11,432,839 words? The word count in law reviews is high, but that is a ridiculous estimate. Even though a correct, objectively determinable answer exists, and 11,432,839 is a ridiculously high estimate, contemplating that ridiculously high reference point is apt to influence estimates of the actual word count. Psychologists refer to the reliance on numeric reference points as a means of making a numeric judgment as "anchoring and adjustment," or simply "anchoring."¹ Anchoring exerts a powerful effect on

* Rachlinski is a Professor of Law at Cornell Law School; Wistrich is a Magistrate Judge in the United States District Court for the Central District of California; Guthrie is Dean and John Wade-Kent Syverud Professor of Law at Vanderbilt Law School.

1. See Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 Sci. 1124, 1128 (1974) (describing anchoring).

judgment.² Even though legal scholars are well aware of the length of articles, in particular the cap of 25,000 words that many law reviews have adopted,³ first contemplating a ridiculously high anchor is apt to influence their estimates.

Estimating the average length of law review articles is just a parlor game for legal academics, but many numeric estimates have important consequences. Numbers abound in civil and criminal cases. Losing defendants pay money or serve time in prison. Sometimes expressing the outcome of a case in dollars or months is a straightforward undertaking, such as when damage awards or sentences are highly constrained.⁴ But judgment is often required to measure damages and impose sentences, and cognitive illusions can systematically skew numeric outcomes.⁵ In particular, such judgments are susceptible to distortions from anchoring and related phenomena.⁶ Our experiments reveal that anchoring exerts a powerful influence on judicial decisions in hypothetical cases across a wide variety of legal contexts. We also find that anchors can arise in subtle and unexpected ways. Our results suggest that anchoring is so persistent, unexpected, and widespread that the process of

2. See DAN ARIELY, PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS 23–38 (rev. & expanded ed., 2009) (describing the influence of anchoring).

3. Many law reviews now state a preference for articles under 25,000 words. See, e.g., *Submissions: How To Submit*, HARV. L. REV., <http://www.harvardlawreview.org/submissions.php>.

4. Criminal sentences imposed using sentencing guidelines are often tightly constrained, as is the case with the Federal Sentencing Guidelines. U.S. SENTENCING COMM’N, GUIDELINES MANUAL (2014). Awards for medical expenses, lost wages, or the cost of covering undelivered consideration in contract can also be relatively uncomplicated relative to nonpecuniary losses. See DAVID BALL, DAVID BALL ON DAMAGES: A PLAINTIFF’S ATTORNEY’S GUIDE FOR PERSONAL INJURY AND WRONGFUL DEATH CASES 27 (1st ed. 2001) (“Intangible damages are hard to get partly because . . . jurors do not know how to figure out how much to give. . . . Jurors have less trouble calculating and giving tangible damages . . . [such as] medical expenses and lost wages, and the amounts can be easily determined.”); Adam G. Winters, Comment, *Where There’s Smoke, Is There Fire? An Empirical Analysis of the Tort “Crisis” in Illinois*, 56 DEPAUL L. REV. 1347, 1349–50 (2007) (“[E]conomic’ damages . . . can be loosely described as any loss that comes with a ‘price tag’ or is easily calculable. Losses considered ‘economic’ include past lost wages due to injury and medical bills, as well as future lost wages and medical expenses. . . . The complimentary piece of the compensatory scheme is collectively labeled ‘noneconomic’ damages. These damages are more difficult to quantify, because there is no underlying bill, record, or pay stub on which to base them. Noneconomic losses include ‘pain and suffering.’” (footnotes omitted)). Some statutes specify a particular amount of damages per violation. See, e.g., 47 U.S.C. § 227(b)(3)(B) (2012) (authorizing “an action to recover for actual monetary loss . . . or to receive \$500 in damages for each such violation, whichever is greater”); 47 U.S.C. § 551(f)(2)(A) (2012) (authorizing “actual damages but not less than liquidated damages computed at the rate of \$100 a day for each day of violation or \$1,000, whichever is higher”).

5. See Valerie P. Hans, Jeffrey J. Rachlinski & Emily G. Owens, *Editors’ Introduction to Judgment by the Numbers: Converting Qualitative to Quantitative Judgments in Law*, 8 J. EMPIRICAL LEGAL STUD. (SPECIAL ISSUE) 1, 1–2 (2011) (“Errors in human judgment arise from the foibles of converting a subjective or qualitative judgment into a linear and quantitative scale.”).

6. See Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 787–94 (2001) (reporting experimental evidence showing that judges are susceptible to the influence of anchoring in assigning damage awards).

assigning awards and sentences might be pervasively unreliable, absent some remedial intervention.

Although converting qualitative assessments into quantitative judgments is challenging,⁷ the rule of law requires that courts make these judgments in a consistent fashion. A criminal defendant's sentence or a civil plaintiff's damage award should arise from the facts and applicable law.⁸ It should not vary with the mood of the judge, the time of day, or the case that came before it. Excessively long criminal sentences constitute an immoral infliction of excess suffering, and overly short sentences fail to reflect society's disapproval and demean the integrity of the victim.⁹ Similarly, erratic damage awards in civil cases are unfair and suggest that the civil justice system is a chaotic game of chance in which some aggrieved victims go uncompensated, while others enjoy a windfall.¹⁰

Deterrence in both civil and criminal cases also depends upon having courts assign sentences and damage awards reliably. A rational, just society should carefully tailor its criminal sentences to create an optimal level of deterrence.¹¹

7. See David M. Studdert, Allen Kachalia, Joshua A. Salomon & Michelle M. Mello, *Rationalizing Noneconomic Damages: A Health-Utilities Approach*, LAW & CONTEMP. PROBS., Summer 2011, at 57, 57 ("Quantifying noneconomic loss is a profound, longstanding, and seemingly intractable problem in the civil justice system.").

8. See 18 U.S.C. § 3553(a)(6) (2012) (requiring the sentencing judge to consider "the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct" when imposing a sentence); Marvin E. Frankel, *The Quest for Equality in Sentencing*, 25 ISR. L. REV. 595, 595 (1991) ("Any system of justice purporting to be civilized must pursue . . . that people similarly circumstanced are to be treated equally under the law . . ."); Paul H. Robinson, Michael T. Cahill & Usman Mohammad, *The Five Worst (and Five Best) American Criminal Codes*, 95 NW. U. L. REV. 1, 1 (2000) (arguing that criminal codes make the understanding and application of laws easier as well as increase the uniformity of applying the law).

9. See ANDREW VON HIRSCH, *DOING JUSTICE: THE CHOICE OF PUNISHMENTS* 71 (1976) ("The offender . . . is being treated as though he deserves the unpleasantness that is being inflicted on him. That being the case, it should be inflicted only to the degree that it is deserved." (emphasis omitted)); Norval Morris, U.S. Dep't of Justice, Address at the University of Denver College of Law: Punishment, Desert and Rehabilitation 22 (Nov. 12, 1976) ("The concept of desert defines relationships between crimes and punishments on a continuum between the unduly lenient and the excessively punitive within which the just sentence may on other grounds be determined.").

10. See MARVIN E. FRANKEL, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* 8 (1973) (asserting that disparities in sentencing violate horizontal equity and undermine public confidence); CASS R. SUNSTEIN, REID HASTIE, JOHN W. PAYNE, DAVID A. SCHKADE & W. KIP VISCUSI, *PUNITIVE DAMAGES: HOW JURIES DECIDE* 3 (2002) ("Our society is deeply committed to employing the force of government with reason and consistency. Discrepant punishments for the same act . . . are inconsistent with that commitment."); Randall R. Bovbjerg, Frank A. Sloan & James F. Blumstein, *Valuing Life and Limb in Tort: Scheduling "Pain and Suffering"*, 83 NW. U. L. REV. 908, 924 (1989) ("[F]undamental fairness requires similarly situated parties to be treated in a similar fashion by the legal system." (emphasis omitted)); Studdert et al., *supra* note 7, at 59 ("Whatever the explanations, the heterogeneity in noneconomic-damages awards among injuries of similar severity is inefficient, inequitable, and has damaging consequences for the legitimacy of personal-injury compensation systems.").

11. See generally Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J.

Sentences that are too long impose unnecessary costs on society and offenders, while sentences that are too short fail to deter crime effectively. Likewise, inadequate civil damage awards fail to deter socially undesirable conduct, while awards that are too high needlessly deter socially useful activities.¹²

Imperfection is, of course, inherent in any legal system. Studies from a wide variety of contexts suggest that litigants' characteristics, such as their race¹³ or gender,¹⁴ influence criminal sentences and damage awards. So does the political orientation of the judge.¹⁵ Moreover, some tasks—notably, assessing some elements of civil damage awards—might be so indeterminate that judges lack the kind of meaningful guidance that would facilitate reliable judgments.¹⁶ These sources of legal indeterminacy, while unfortunate, are at least understood as unwanted parts of an unavoidably imperfect system.¹⁷ Extraneous influences, such as the type of

POL. ECON. 169 (1968).

12. See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 10 (1987).

13. See, e.g., David S. Abrams, Marianne Bertrand & Sendhil Mullainathan, *Do Judges Vary in Their Treatment of Race?*, 41 J. LEGAL STUD. 347, 350 (2012) (describing a study demonstrating racial disparities in sentencing); Ronald Chen & Jon Hanson, *Categorically Biased: The Influence of Knowledge Structures on Law and Legal Theory*, 77 S. CAL. L. REV. 1103, 1159 (2004) (reviewing research indicating that the way individuals view race will cause them to judge the behavior of people differently, even if the behavior being judged is the same).

14. See, e.g., Amy Farrell, Geoff Ward & Danielle Rousseau, *Intersections of Gender and Race in Federal Sentencing: Examining Court Contexts and the Effects of Representative Court Authorities*, 14 J. GENDER RACE & JUST. 85, 91 (2010) (finding that the leniency that women experience may be attributable to criminal justice system authorities viewing women as less dangerous and believing that women “possess a generally greater potential for reform than men”); Ilene H. Nagel & Barry L. Johnson, *The Role of Gender in a Structured Sentencing System: Equal Treatment, Policy Choices, and the Sentencing of Female Offenders Under the United States Sentencing Guidelines*, 85 J. CRIM. L. & CRIMINOLOGY 181, 186 (1994) (“Women are more likely than similarly situated men to receive suspended sentences or probation.”).

15. Research assessing the influence of judges' political preferences is extensive. See generally, e.g., CASS R. SUNSTEIN, DAVID SCHKADE, LISA M. ELLMAN & ANDRES SAWICKI, *ARE JUDGES POLITICAL?: AN EMPIRICAL ANALYSIS OF THE FEDERAL JUDICIARY* (2006); William M. Landes & Richard A. Posner, *Rational Judicial Behavior: A Statistical Study*, 1 J. LEGAL ANALYSIS 775 (2009); Jeffrey A. Segal, Lee Epstein, Charles M. Cameron & Harold J. Spaeth, *Ideological Values and the Votes of U.S. Supreme Court Justices Revisited*, 57 J. POL. 812 (1995).

16. See David Baldus, John C. MacQueen & George Woodworth, *Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittitur Review of Awards for Nonpecuniary Harms and Punitive Damages*, 80 IOWA L. REV. 1109, 1182 (1995) (“[T]here is no ‘correct’ general damages award for any nonpecuniary harm.”); William Zelermyer, *Damages for Pain and Suffering*, 6 SYRACUSE L. REV. 27, 28 (1954) (“[P]ain and suffering have no dimensions, mathematical or financial. Whatever claim or award is based upon these intangibles must, therefore, be conceived at the end of a speculative and uncertain journey.”).

17. See *McCougald v. Garber*, 536 N.E. 2d 372, 374–75 (N.Y. 1989) (“[R]ecovery for noneconomic losses such as pain and suffering and loss of enjoyment of life rests on ‘the legal fiction that money damages can compensate for a victim’s injury.’ We accept this fiction, knowing that although money will neither ease the pain nor restore the victim’s abilities, this device is as close as the law can come in its effort to right the wrong. We have no hope of

defendant or the political attitude of the judge, can be thought of as measurement error in the legal system. Judges and juries use different yardsticks to measure the harm done to women as opposed to men¹⁸ or to assess the appropriate punishment assigned to defendants of different races.¹⁹ Variability among judges is an unfortunate but unavoidable aspect of the nature of justice dispensed by human beings. Some judges, by their nature, will impose more severe sentences than their colleagues.²⁰ Even with intricate sentencing guidelines, detailed sentencing reports prepared by probation officers, and lengthy arguments by attorneys, some judicial discretion will remain to be exercised differently by judges of different predilections.²¹

Invidious biases toward litigants, the influence of political orientation of judges, and case complexity, however, are all such widely acknowledged impediments to consistent adjudication that society attempts to reduce these influences in several ways. Codes of judicial ethics require judges to avoid judgments based on race or gender, and case law requires judges to avoid even the appearance of such improprieties.²² The judicial appointment process also addresses concerns regarding the appointment of judges deemed to be overly partisan.²³ Judges in many jurisdictions also face elections, which are another source of accountability.²⁴ In criminal cases, sentencing guidelines constrain sentences, thereby limiting the degree to which erratic judgment might adversely affect outcomes.²⁵

evaluating what has been lost, but a monetary award may provide a measure of solace for the condition created.” (citations omitted)).

18. See, e.g., Jane Goodman, Elizabeth F. Loftus, Marian Miller & Edith Greene, *Money, Sex, and Death: Gender Bias in Wrongful Death Damage Awards*, 25 LAW & SOC’Y REV. 263, 281 (1991) (reporting experimental evidence that jurors award more in wrongful death suits involving male victims than female victims).

19. See David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts*, 44 J.L. & ECON. 285, 311–12 (2001) (reporting a large racial disparity in sentences imposed under the U.S. Sentencing Guidelines); Jeffrey J. Rachlinski, Sheri Lynn Johnson, Andrew J. Wistrich & Chris Guthrie, *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1221–26 (2009) (concluding that implicit racial biases affect judges’ sentences).

20. See William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505, 527 (2001) (explaining that judges have their own beliefs about particular crimes and how broadly they should be defined, which affects how statutes are interpreted and enforced).

21. See *id.* at 529 (finding that “[i]deological differences, public-interested goals, [and] the reigning institutional culture” can affect or even dominate the behavior of judges).

22. See, e.g., MODEL CODE OF JUDICIAL CONDUCT R. 2.3(B) (2011) (“A judge shall not, in the performance of judicial duties, by words or conduct manifest bias or prejudice, or engage in harassment, including but not limited to bias, prejudice, or harassment based upon race, sex, gender, religion, national origin, ethnicity, disability, age, sexual orientation, marital status, socioeconomic status, or political affiliation . . .”).

23. For a recent discussion of this concern, see generally JOHN R. LOTT, JR., *DUMBING DOWN THE COURTS: HOW POLITICS KEEPS THE SMARTEST JUDGES OFF THE BENCH* (2013).

24. See Joanna Shepherd, *Measuring Maximizing Judges: Empirical Legal Studies, Public Choice Theory, and Judicial Behavior*, 2011 U. ILL. L. REV. 1753, 1760 tbl.1 (documenting the various state judicial selection and retention procedures).

25. *Mistretta v. United States*, 488 U.S. 361, 366 (1989) (noting that the federal sentencing guidelines were intended to remedy two “unjustif[ed]” and “shameful” deficiencies in the previous sentencing regime, the first of which was “the great variation

Efforts to identify and root out systematic deviations in sentences and awards, however, presuppose that if judges all could be induced to measure desert and harm the same way across litigants of different races and genders, then reliability would emerge. This understanding of the criminal and civil justice systems can thus be likened to measurement error in intelligence testing. Test theory suggests that there exists a correct IQ score for any individual,²⁶ but it acknowledges that individual test performance will vary based on mood and mistake and that biases exist in tests themselves that unreasonably influence how people from different races perform.²⁷ But underlying test theory is a faith that there exists a true IQ score and that the results of any one test will measure this true score, along with some error variance.²⁸ This is precisely the model underlying the criminal and civil justice systems. For every offender and every injured plaintiff, there exists a correct sentence and a correct award. Each case produces this correct sentence and award, along with some degree of error variance. As long as the error component is small and not systematic, it can be tolerable. We suspect, however, that the cognitive processes that underlie numeric judgments in court face a more fundamental problem than measurement error. We argue that numeric judgments are constructed in a case-by-case fashion and thus cannot be made consistent simply by inducing judges to ignore invidious factors and harmonizing judges' political and social values.

Our thesis that damage awards and sentences are constructed in an ad hoc fashion arises from a contemporary understanding of human judgment and decision making. Psychologists who study judgment and choice argue that judgments are constructed in context.²⁹ When people estimate what they would be willing to pay for a new home or an iPad, they are not mentally downloading a preexisting, fixed assessment of the value of these commodities; rather, they are creating a set of preferences that are a product of a specific context. This is not to say that the price they are willing to pay is wholly arbitrary. People recognize that a 3500-square-foot home in a nice neighborhood is more valuable than an iPad. The value of the items affects people's mental constructions, but the process is fundamentally one of mental creation rather than measurement. Judgments thus have an arbitrary coherence.³⁰ The amount a customer is willing to pay for a consumer good arises from the customer's mood, other decisions she may have recently made, and a wide range of ephemeral influences.³¹

among sentences imposed by different judges upon similarly situated offenders" (alteration in original) (citation omitted)).

26. See Timo M. Bechger, Gunter Maris, Huub H. F. M. Verstralen & Anton A. Béguin, *Using Classical Test Theory in Combination with Item Response Theory*, 27 APPLIED PSYCHOL. MEASUREMENT 319, 320 (2003).

27. See generally Claude M. Steele, *A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance*, 52 AM. PSYCHOLOGIST 613 (1997) (describing how stereotypes associated with race and gender affect test scores).

28. See Bechger et al., *supra* note 26, at 329.

29. See Dan Simon, Daniel C. Krawczyk, Airom Bleicher & Keith J. Holyoak, *The Transience of Constructed Preferences*, 21 J. BEHAV. DECISION MAKING 1, 1 (2008) ("A substantial body of literature on judgment and decision-making has been devoted to the construction of preferences. . . . It has been suggested that preferences are invented rather than found.")

30. See ARIELY, *supra* note 2, at 26 (describing "arbitrary coherence" in preferences).

31. See Sarah Lichtenstein & Paul Slovic, *The Construction of Preference: An Overview*,

We suspect that judgments made in the legal system are no different than judgments made in other parts of life.³² Judges certainly understand that a murderer deserves a more severe sentence than a shoplifter. But the exact sentence imposed will likely reflect contextual factors and an inherent underlying arbitrariness. Invidious biases, political influences, and simple indeterminacy are thus just the most visible symptoms of a more widespread problem in the legal system. Judges, like most people, lack the cognitive capacity to make reliable quantitative judgments in a complex environment. They can adopt mechanisms to produce a degree of reliability, but their judgments are inherently erratic. Recognizing this endemic characteristic of quantitative judgments provides a more solid footing for efforts to provide regularity to judgments.

I. EVIDENCE OF ARBITRARINESS: ANCHORING AND SCALING

Among the phenomena that demonstrate the arbitrary nature of choice, anchoring stands out as particularly significant. Anchoring refers to the tendency to rely on numeric reference points to make numeric judgments, even when such estimates are arbitrary, irrelevant, or ludicrous.³³ The pervasive influence of anchoring reveals such a strong and systematic tendency to attend to decisional context, regardless of whether it is misleading, that it is prime evidence of an underlying arbitrariness in human numeric judgment.³⁴

The initial studies of anchoring identified the phenomenon as a somewhat innocuous example of the role that simple mental strategies play in judgment.³⁵ Tversky and Kahneman first demonstrated the influence of anchors on numeric judgments by asking undergraduates to estimate the percentage of countries in the United Nations that are African after giving them a transparently arbitrary reference point.³⁶ Subjects were asked to spin a roulette-style wheel that was rigged to land on 65 or 20. When the wheel landed on 65, subjects provided much higher estimates than when it landed on 20. The researchers also demonstrated that when asked to estimate the product of $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$, people provided lower estimates than when asked to estimate the product of $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$.³⁷

Tversky and Kahneman cited anchoring as one of a set of simple mental shortcuts (or heuristics) that people use to make judgments.³⁸ They asserted that these shortcuts

in THE CONSTRUCTION OF PREFERENCE 1, 2 (Sarah Lichtenstein & Paul Slovic eds., 2006) (“The variability in the ways we construct and reconstruct our preferences yields preferences that are labile, inconsistent, subject to factors we are unaware of, and not always in our own best interests.”).

32. See generally Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1 (2007) (presenting evidence that judges use the same kinds of decision-making strategies as most adults).

33. See SCOTT PLOUS, THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 146 (1993).

34. See ARIELY, *supra* note 2, at 33–36 (describing the influence of anchoring on judgment).

35. See Tversky & Kahneman, *supra* note 1 (identifying anchoring as one of a set of mental strategies people use to make decisions).

36. *Id.*

37. *Id.*

38. *Id.* at 1128–32.

are useful but can lead to mistakes in judgment.³⁹ Anchoring is often a sensible strategy for numeric assessment because anchors are commonly reasonable first approximations. For example, consumers know that they are not apt to pay the sticker price listed on a new automobile; rather, they will pay some amount less than this, depending upon their negotiation skills, the dealer's disposition to make the sale, and various other factors. The sticker price, however, is a good initial estimate of the final price. Tversky and Kahneman argued that when the subjects in their studies relied on the random starting point, they were overusing a good strategy for making numeric judgments and failing to put enough cognitive effort into adjusting the initial reference point toward the correct estimate.⁴⁰

This account of anchoring as a useful heuristic that gets overused suggests that people are making fundamentally sound judgments, although these judgments get distorted. But the account does not explain why people fail to recognize how dangerous this heuristic is in situations in which the anchor is obviously irrelevant. In one study, researchers had undergraduate subjects write down the last three digits of their phone number and add 400 to it. The researchers then asked the subjects whether Attila the Hun was defeated in Europe before or after that year.⁴¹ After obtaining that answer, the researchers asked in what year Attila was defeated. The subjects knew perfectly well that their phone numbers had nothing to do with European history, but their final estimates of the year that Attila was defeated nevertheless correlated with their phone numbers. In another study, researchers asked undergraduates to identify the average price of a textbook in the campus bookstore.⁴² For half of the participants, the researchers first asked whether the price was greater or less than \$7163.52. When subjects first had to contemplate that absurd estimate, they provided higher estimates than those who had not been exposed to it.

Anchoring has proven to be a robust phenomenon that affects all manner of judgments.⁴³ In one study, a price provided by an admittedly uninformed person altered subjects' estimates of the value of used automobiles.⁴⁴ In another, the number on a football player's uniform altered estimates of the likelihood that that player

39. *Id.* at 1124 ("In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.").

40. *Id.* at 1128 (noting that adjustments from initial starting points are typically insufficient).

41. J. EDWARD RUSSO & PAUL J.H. SCHOEMAKER, *DECISION TRAPS: TEN BARRIERS TO BRILLIANT DECISION-MAKING AND HOW TO OVERCOME THEM* 90 (1989).

42. PLOUS, *supra* note 33, at 146.

43. See Gretchen B. Chapman & Eric J. Johnson, *Anchoring, Activation, and the Construction of Values*, 79 *ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES* 115, 117 (1999) ("Demonstration of the anchoring effect are [sic] plentiful in studies of judgment."); Nicholas Epley & Thomas Gilovich, *Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors*, 12 *PSYCHOL. SCI.* 391, 391 (2001) ("Countless experiments have shown that people's absolute answers are influenced by the initial comparison with the irrelevant anchor."); Adrian Furnham & Hua Chu Boo, *A Literature Review of the Anchoring Effect*, 40 *J. SOCIO-ECONOMICS* 35, 41 (2011) ("There is now nearly 40 years' worth of research on the anchoring effect which has proved to be extremely robust.").

44. Thomas Mussweiler, Fritz Strack & Tim Pfeiffer, *Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility*, 26 *PERSONALITY & SOC. PSYCHOL. BULL.* 1142, 1144-47 (2000).

would sack the opposing quarterback in an upcoming playoff game.⁴⁵ Researchers found that people asked to predict the likelihood that the United States would deploy military troops to address the conflict in Yugoslavia gave lower estimates when they were first asked whether that likelihood was greater or less than 30% as opposed to 70%.⁴⁶ When people estimated the freezing point of vodka, they tended to start with 32 degrees (the freezing point of water) and adjust insufficiently downward.⁴⁷ When asked to estimate George Washington's age when he died, people fixated on randomly generated numbers that were either too high or too low.⁴⁸ The starting point in negotiations also anchors the final outcome, even when the starting point is not meaningful.⁴⁹ Research on anchoring has also demonstrated that the phenomenon influences estimates of not only what goods might cost but also how much people are actually willing to pay.⁵⁰

But what accounts for the influence of anchors? At its core, anchoring is a demonstration that people construct numeric judgment from the surrounding context—even when that context is misleading and irrelevant.⁵¹ Researchers have identified three cognitive processes that allow irrelevant anchors to influence judgment: cognitive laziness, availability, and scale distortion.⁵²

First, Tversky and Kahneman initially described anchoring as a consequence of cognitive laziness.⁵³ They maintained that cognitive effort is required to mentally adjust away from the initial anchor to the final estimate. This account supposes that in making numeric estimates, people mentally move their estimates to match the evidence that they have available for their understanding of the world. People

45. Clayton R. Critcher & Thomas Gilovich, *Incidental Environmental Anchors*, 21 J. BEHAV. DECISION MAKING 241, 243–45 (2008).

46. Chapman & Johnson, *supra* note 43, at 133–35.

47. Epley & Gilovich, *supra* note 43, at 395.

48. Kevin L. Blankenship, Duane T. Wegener, Richard E. Petty, Brian Detweiler-Bedell & Cheryl L. Macy, *Elaboration and Consequences of Anchored Estimates: An Attitudinal Perspective on Numerical Anchoring*, 44 J. EXPERIMENTAL SOC. PSYCHOL. 1465, 1467–69 (2008).

49. *See generally*, e.g., Adam D. Galinsky & Thomas Mussweiler, *First Offers as Anchors: The Role of Perspective-Taking and Negotiator Focus*, 81 J. PERSONALITY & SOC. PSYCHOL. 657 (2001); Greg Pogarsky & Linda Babcock, *Damage Caps, Motivated Anchoring, and Bargaining Impasse*, 30 J. LEGAL STUD. 143 (2001).

50. *See* ARIELY, *supra* note 2, at 27–31 (describing auctions for actual items in which bids correlated with transparently random starting points such as digits in Social Security numbers).

51. *See* Chapman & Johnson, *supra* note 43, at 116 (“The effect of anchors is one indication that many statements of value and belief are not directly retrieved from memory but rather are constructed online in response to a query.”).

52. *But see* Dan Orr & Chris Guthrie, *Anchoring, Information, Expertise, and Negotiation: New Insights from Meta-Analysis*, 21 OHIO ST. J. ON DISP. RESOL. 597, 601–05 (2006) (describing the “four theoretical accounts of anchoring: (1) the social implications theory; (2) the insufficient adjustment theory; (3) the numeric priming theory; and (4) the information salience theory”). We do not treat the “social implications theory” in our account, as it appears to us either to overlap with other theories or to be the product of motivated reasoning rather than a purely cognitive strategy.

53. Tversky & Kahneman, *supra* note 1, at 1128 (describing anchoring as arising from “insufficient adjustment” from a reference point).

commonly fail to put enough effort into the adjustment process, so the initial anchor exerts too much influence over the final judgment.⁵⁴

A second, competing account of anchoring relies on cognitive availability. This model arose from the concern that adjustment cannot truly account for the range of influence anchors have.⁵⁵ Researchers have proposed that anchors affect the focus of people's attention and consequently affect their judgment.⁵⁶ In rejecting ludicrous anchors, people call to mind the most extreme example that they can imagine.⁵⁷ When asked to estimate the average price of a college textbook, for example, the absurdly high anchor leads people to recall highly expensive books. When they recognize that even the most expensive textbooks are far below the absurd anchor, they can reject the hypothesis that the average book is truly that costly. But when they then estimate the actual average price, they are thinking about expensive books, which produces higher estimates.⁵⁸

The availability model of anchoring, however, fails to explain some of the more bizarre aspects of anchoring. Some of the studies involve anchoring in contexts where availability seems like an implausible explanation.⁵⁹ For example, studies in which anchoring influences estimates involving historic events (such as the year George Washington was elected President) or scientific values (such as the freezing point of vodka) in which people rely on reference points that they know are inaccurate (1776, and 32 degrees Fahrenheit) do not seem to involve cognitive availability.⁶⁰ Furthermore, studies that do not require subjects to reject extreme hypotheses also facilitate anchoring. Shane Frederick and Daniel Mochon demonstrated, for example, that independent serial estimates create anchors.⁶¹ In one of their studies, subjects who first estimated the average weight of an adult male raccoon and then an adult male giraffe gave lower estimates for the giraffe than when they were asked to estimate the weights in the reverse order.⁶² The estimates for the weight of the raccoon were

54. See generally Epley & Gilovich, *supra* note 43 (arguing that anchoring arises from insufficient adjustment).

55. See Fritz Strack & Thomas Mussweiler, *Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility*, 73 J. PERSONALITY & SOC. PSYCHOL. 437, 437–38 (1997) (identifying problems with the adjustment theory of anchoring).

56. See Chapman & Johnson, *supra* note 43, at 119 (“[R]ecent anchoring accounts propose that the anchor makes information that is consistent with the anchor more available.”).

57. See Thomas Mussweiler & Fritz Strack, *The Use of Category and Exemplar Knowledge in the Solution of Anchoring Tasks*, 78 J. PERSONALITY & SOC. PSYCHOL. 1038, 1040 (2000) (“[A] comparative anchoring task that includes an implausible anchor value is likely to be solved on the basis of category knowledge about the target.”); Strack & Mussweiler, *supra* note 55, at 444 (describing anchoring as arising from internal testing of the anchor as an initial hypothesis).

58. See Mussweiler & Strack, *supra* note 57, at 1040 (describing the selective-accessibility model of anchoring).

59. See, e.g., Epley & Gilovich, *supra* note 43, at 391 (Some anchors “are known—from the beginning—to be wrong. There is thus no cause to consider whether the anchor value is correct and thus no engine of heightened accessibility of anchor-consistent information.”).

60. See *id.*

61. Shane W. Frederick & Daniel Mochon, *A Scale Distortion Theory of Anchoring*, 141 J. EXPERIMENTAL PSYCHOL.: GEN. 124 (2012).

62. *Id.* at 127–28.

likewise higher when subjects estimated the weight of the giraffe first.⁶³ In effect, their estimates influenced each other, even though the subjects must have known that the estimates should have been independent because there was no basis for starting with the other animal's weight as a reference point. The results are hard to square with either the cognitive laziness or availability accounts of anchoring.

Concerned that neither cognitive laziness nor availability can fully account for anchoring, Frederick and Mochon offer a third account. They propose that anchoring consists of a temporary distortion in the process of translating a quantitative sense of some physical property into a quantitative assessment.⁶⁴ When people struggle to decide whether a raccoon weighs 20 pounds, 1700 pounds (the right answer for a giraffe) seems like an extraordinarily great weight for an animal. Likewise, to a subject who concludes that 1700 pounds is right for the giraffe, 20 pounds seems like a paltry weight for an animal.

Frederick and Mochon's account of anchoring reveals not only how potent the effect is but also where the source of arbitrariness leaks into judgment. They demonstrate that anchors do not distort one's sense of the underlying object in other ways (as the availability account suggests).⁶⁵ When they ask questions about animals other than their weight that nevertheless relate to the animal's size, they find no distortion. For example, people's estimates of the average height of a giraffe do not change depending upon whether they are asked to make their estimates before or after they were asked about the weight of a raccoon or how many lions a giraffe carcass would feed. Even their sense of weight on a different scale does not change. Subjects asked to estimate how many grand pianos would balance with an average adult male giraffe are not influenced by previous estimates of the weight of raccoons in pounds either. Therefore, a low anchor generated by contemplating the weight of a small animal does not distort people's sense of how big a giraffe is, but it does influence how people translate that sense into pounds.

Whatever the explanation for anchoring, it is a powerful phenomenon. Real estate agents' estimates of the sale price of homes are heavily influenced by meaningless variations in the initial listing price.⁶⁶ The auction prices that University of Chicago business students are willing to submit for real commodities are influenced by first asking them whether they would be willing to pay more or less than part of their phone number.⁶⁷ The distortion suggests that when people identify the value they would be willing to pay for an item, they might have a good sense of what the item is but no stable way of converting that sense into a price. It is no wonder that retailers use ridiculously high initial prices for retail clothing and then offer items on sale at 50% or 70% off that meaningless initial price. People understand what the item is but do not have coherent ways of converting this understanding into a dollar amount.

63. *Id.* at 124–26 (presenting a “scale distortion” theory of anchoring).

64. *Id.* at 124 (“[T]he perceived magnitude of a number is affected by other numeric values on that scale with which it is compared.”).

65. *Id.* at 126–29 (describing experimental results supporting the scale distortion theory of anchoring).

66. *See, e.g.*, Gregory B. Northcraft & Margaret A. Neale, *Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions*, 39 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 84, 89–91 (1987).

67. ARIELY, *supra* note 2, at 27–31.

II. ANCHORING AND JUDGES

Anchoring can influence a wide variety of judgments in legal contexts,⁶⁸ especially civil damage awards and criminal sentences. Researchers have conducted numerous studies in mock-jury settings documenting a marked effect of anchors on damage-award estimates provided by lay adults.⁶⁹ In one such study, researchers gave ordinary adults a description of a personal injury case and asked them for an appropriate damage award.⁷⁰ The researchers varied the plaintiff's attorney's request for a specific damage award. The plaintiff either asked for no specific amount, for \$20,000, for \$5 million, or for \$1 billion. Even though the range of requests was ridiculous, "[a]s the anchor amount increased, compensation increased."⁷¹ As the title of that article suggests, "the more you ask for, the more you get." Although other research suggests that an excessive request might backfire,⁷² numerous mock-jury studies document a persistent influence of damage award requests on damage awards.⁷³

Judges might or might not be as susceptible to anchoring as nonjudges. Anchoring distorts judgment in so many contexts that it seems likely to influence the kinds of numeric judgments judges must make. In a recent article, Judge Mark Bennett identified the influence of anchoring on judges as a major concern.⁷⁴ Two studies of actual sentencing decisions indicate that judges have difficulty translating their qualitative sense of desert into numerical sentences; in particular, they fixate on certain numbers.⁷⁵ Even the Supreme Court has cited the power of authoritative

68. See generally Yuval Feldman, Amos Schurr & Doron Teichman, *Anchoring Legal Standards* (Mar. 12, 2014) (unpublished manuscript), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2408193.

69. See generally, e.g., Gretchen B. Chapman & Brian H. Bornstein, *The More You Ask for, the More You Get: Anchoring in Personal Injury Verdicts*, 10 APPLIED COGNITIVE PSYCHOL. 519 (1996); Reid Hastie, David A. Schkade & John W. Payne, *Juror Judgments in Civil Cases: Effects of Plaintiff's Requests and Plaintiff's Identity on Punitive Damage Awards*, 23 LAW & HUM. BEHAV. 445 (1999); Verlin B. Hinsz & Kristin E. Indahl, *Assimilation to Anchors for Damage Awards in a Mock Civil Trial*, 25 J. APPLIED SOC. PSYCHOL. 991 (1995); John Malouff & Nicola S. Schutte, *Shaping Juror Attitudes: Effects of Requesting Different Damage Amounts in Personal Injury Trials*, 129 J. SOC. PSYCHOL. 491 (2001); Barry Markovsky, *Anchoring Justice*, 51 SOC. PSYCHOL. Q. 213 (1988) (presenting experimental evidence on the influence of anchoring in a variety of legal contexts, including fairness judgments and criminal sentences); Mollie W. Marti & Roselle L. Wissler, *Be Careful What You Ask For: The Effect of Anchors on Personal Injury Damages Awards*, 6 J. EXPERIMENTAL PSYCHOL.: APPLIED 91 (2000); Jennifer K. Robbennolt & Christina A. Studebaker, *Anchoring in the Courtroom: The Effects of Caps on Punitive Damages*, 23 LAW & HUM. BEHAV. 353 (1999).

70. Chapman & Bornstein, *supra* note 69.

71. *Id.* at 526.

72. Marti & Wissler, *supra* note 69, at 92.

73. See *supra* note 69.

74. Mark Bennett, *Confronting Cognitive "Anchor Effect" and "Blind Spot" Biases in Federal Sentencing: A Modest Solution for Reforming a Fundamental Flaw*, 104 J. CRIM. L. & CRIMINOLOGY 489, 503–13 (2014).

75. Craig Jones & Micah B. Rankin, *Justice as a Rounding Error? Evidence of Subconscious Bias in Second-Degree Murder Sentences in Canada*, 52 OSGOODE HALL L.J.

anchors in holding that criminal defendants must be sentenced under extant sentencing guidelines—even though such guidelines are not binding.⁷⁶

On the other hand, judges' training or experience might give them an advantage. Experience reduces (but does not eliminate) the influence of anchoring.⁷⁷ Forcing decision makers to explain their choices—as judges frequently must do—also reduces the influence of anchors.⁷⁸ One study that directly compares punitive damage awards of judges and juries demonstrates that jury awards are more erratic than judicial awards.⁷⁹ Selecting damage awards and criminal sentences is deadly serious business for judges, and they doubtless put a great deal of effort into the task. But imposing awards and sentences requires the same basic cognitive task seen in the studies of anchoring. Therefore, judges might exhibit the same kinds of difficulties setting awards and sentences reliably as subjects in the research on anchoring.

A series of previous studies suggests that anchoring has a powerful influence on judicial decision making. In one experiment, we asked federal magistrate judges to identify an appropriate damage award in a hypothetical personal injury lawsuit.⁸⁰ The materials indicated that the plaintiff was badly injured in an accident with the defendant's delivery truck. The defendant had admitted to liability, and the only remaining issue was damages. After describing the injury briefly, we asked half of the judges to identify an appropriate damage award. We asked the same question of the other half, but only after asking them to first rule on a motion to dismiss the case for failing to satisfy the minimum amount in controversy for jurisdiction in federal court—\$75,000. Virtually all who were asked to decide the motion denied it, but those who ruled on the motion also awarded the plaintiff less money than those in the control group. The judges who had not ruled on the motion awarded an average of \$1,249,000, whereas judges who first denied the motion to dismiss awarded an average of \$882,000.⁸¹ Even though the motion was frivolous, it injected the low

(forthcoming 2015); Andrew Wiseman, Daniel Fischer & Michael Connelly, *Sentencing and Conventional Number: Preferences: A Research Note*, 8 JUST. RES. & POL'Y 67 (2006).

76. *Peugh v. United States*, 133 S. Ct. 2072, 2084 (2013) (“The federal system adopts procedural measures intended to make the Guidelines the lodestone of sentencing. A retrospective increase in the Guidelines range applicable to a defendant creates a sufficient risk of a higher sentence to constitute an *ex post facto* violation.” (emphasis in original)).

77. See generally Thomas Mussweiler & Fritz Strack, *Numeric Judgments Under Certainty: The Role of Knowledge in Anchoring*, 36 J. EXPERIMENTAL SOC. PSYCHOL. 495 (2000); Andrew R. Smith, Paul D. Windschitl & Kathryn Bruchmann, *Knowledge Matters: Anchoring Effects Are Moderated by Knowledge Level*, 43 EUR. J. SOC. PSYCHOL. 97 (2013).

78. See Philip E. Tetlock & Richard Boettger, *Accountability: A Social Magnifier of the Dilution Effect*, 57 J. PERSONALITY & SOC. PSYCHOL. 388, 388 (1989). But see Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *The “Hidden Judiciary”: An Empirical Examination of Executive Branch Justice*, 58 DUKE L.J. 1477, 1505–06 (2009) (finding that asking judges to explain their decisions did not ameliorate the effect of an irrelevant anchor on judges' damage awards).

79. Theodore Eisenberg, Jeffrey J. Rachlinski & Martin T. Wells, *Reconciling Experimental Incoherence with Real-World Coherence in Punitive Damages*, 54 STAN. L. REV. 1239, 1252–56 (2002).

80. Guthrie et al., *supra* note 6, at 790–94.

81. *Id.* at 791.

anchor into the judges' thinking and thereby influenced their assessment of the appropriate award.

In a second experiment with similar materials, we found that judges attend to anchors that they are explicitly supposed to ignore.⁸² We used the same basic description of the severely injured plaintiff we employed in the first study, again asserting that the defendant had admitted to liability, and the only remaining issue was setting an appropriate damage award.⁸³ We also told the judges to imagine that they had presided over a settlement conference between the parties. In one variant of this study, the materials either informed the judges that the plaintiff had been willing to settle for \$175,000 or stated no amount. Judges who learned of the low demand provided lower awards than the judges who did not.⁸⁴ In another version, the materials either informed the judges that the plaintiff would not accept less than \$10 million or stated no amount. Judges who learned of the high demand provided higher awards than judges who did not.⁸⁵ In all cases, the materials reiterated the evidentiary rule prohibiting the use of any information disclosed during settlement conferences in trials. Anchoring seemed to triumph over this injunction.

We have found that anchoring not only leads judges to ignore the rules of evidence but also leads them to ignore the Supreme Court of the United States. In a third study, we gave bankruptcy judges a hypothetical case involving a financial restructuring known as a "cramdown".⁸⁶ The case required each bankruptcy judge to adjust the interest rate of a loan secured by an independent trucker's rig. Governing law required that the court set the interest rate at the prevailing prime rate, adjusted upwards to reflect the risk of nonpayment. We reminded the judges that the Supreme Court had recently insisted that the original interest rate on the loan was not a relevant consideration.⁸⁷ The judges ignored the Court's command, however. Judges who learned that the initial rate was 21% provided a higher rate than judges who were not told of this amount.⁸⁸

In a fourth study of anchors, we found that even bizarre testimony can produce a numeric anchor that influences judgment.⁸⁹ In this study, we asked administrative law judges attending a national conference to assess a damage award for the plaintiff in an employment discrimination suit. The materials identified the plaintiff as a Mexican-American administrative assistant with an excellent employment record. A new supervisor began calling her racial epithets in front of her coworkers—and her daughter, in one instance. When she complained, he fired her. The materials indicated that she brought suit through a city human rights commission, but because she found another position right away, her damages were limited to "mental anguish" for the wrongful termination. For half of the judges, the plaintiff indicated that she

82. Andrew J. Wistrich, Chris Guthrie & Jeffrey J. Rachlinski, *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, 153 U. PA. L. REV. 1251, 1277–93 (2005).

83. *Id.* at 1288.

84. *Id.* at 1289.

85. *Id.* at 1290.

86. Jeffrey J. Rachlinski, Chris Guthrie & Andrew J. Wistrich, *Inside the Bankruptcy Judge's Mind*, 86 B.U. L. REV. 1227, 1233–37 (2006).

87. See *Till v. SCS Credit Corp.*, 541 U.S. 465 (2004).

88. Rachlinski et al., *supra* note 86, at 1235.

89. Guthrie et al., *supra* note 78, at 1501–06.

had recently seen a “court TV show” in which a similar plaintiff was awarded compensatory damages for mental anguish.⁹⁰ The other half of the judges saw the same testimony but were told that the claimant in the TV show had received a compensatory damage award of \$415,300. This anchor had a huge effect on the judges: those who did not see the anchor awarded a median of \$6,250, while those who saw it awarded a median of \$50,000.⁹¹

An experiment conducted in Germany also showed that judges are vulnerable to anchoring when setting criminal sentences.⁹² The researchers asked judges to impose a sentence for a hypothetical defendant convicted of robbery. The judges first rolled two dice and then were asked whether the sentence should be higher or lower than the total. Just as undergraduates’ phone numbers influenced their estimates of the year of Attila the Hun’s defeat, the roll of the dice influenced judges’ estimates of an appropriate criminal sentence.⁹³ A similar study by two of the same researchers also showed that suggestions from both prosecutors and defense attorneys influenced judges’ sentences in a hypothetical case.⁹⁴

Are anchors as powerful in real cases as they are in the laboratory?⁹⁵ Archival data are harder to assess because it is difficult to control for unobserved variables. Attorneys doubtless ask for more money when the underlying damages are more serious. But what research exists suggests that anchoring influences both damage awards and criminal sentences. A study of both judges and jurors in state courts, for example, showed that economic damages and noneconomic damages are highly correlated, suggesting that they anchor each other.⁹⁶ Another paper showed that the recommendations of probation officers appear to anchor judges’ sentences.⁹⁷ A study

90. *Id.* at 1528.

91. *Id.* at 1504.

92. Birte Englich, Thomas Mussweiler & Fritz Strack, *Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts’ Judicial Decision Making*, 32 PERSONALITY & SOC. PSYCHOL. BULL. 188, 194–96 (2006).

93. See RUSSO & SCHOEMAKER, *supra* note 41, at 90 (describing the study involving Attila the Hun).

94. Birte Englich & Thomas Mussweiler, *Sentencing Under Uncertainty: Anchoring Effects in the Courtroom*, 31 J. APPLIED SOC. PSYCHOL. 1535, 1547 (2001).

95. See, e.g., Shari Seidman Diamond, Mary R. Rose, Beth Murphy & John Meixner, *Damage Anchors on Real Juries*, 8 J. EMPIRICAL LEGAL STUD. 148, 173, 178 (2011) (analyzing videotapes of actual jury deliberations suggesting that actual jurors may react less favorably to a plaintiff’s ad damnum than mock jurors typically have and concluding that “the dangers of bias from these potential anchors offered by attorneys appear to be overstated as applied to the real world of deliberating juries”).

96. Valerie P. Hans & Valerie F. Reyna, *To Dollars from Sense: Qualitative to Quantitative Translation in Jury Damage Awards*, 8 J. EMPIRICAL LEGAL STUD. 120, 141–42 (2011) (“In both judge and jury trials, the economic damage award reached by the decision maker is a significant predictor of the noneconomic damage award. . . . [T]he amount of economic damage that a plaintiff experiences in a case serves as a slightly stronger determinant of the noneconomic award for juries as compared to judges.”).

97. Ebbe B. Ebbesen & Vladimir J. Konečni, *The Process of Sentencing Adult Felons: A Causal Analysis of Judicial Decisions*, in THE TRIAL PROCESS 413, 434 (Bruce Dennis Sales ed., 1981) (finding that “[j]udges and probation officers agreed on the sentence in over 80% of the cases”).

of sentences assigned by Maryland trial judges based upon mishandled sentencing forms also showed that erroneous sentencing recommendations influenced judges' sentences.⁹⁸ Recent research on judges in Taiwan who set compensation rates for unlawful use of land demonstrated that these judges rely heavily on awards suggested by the litigants.⁹⁹

These studies indicate that judges' damage awards and criminal sentences are not always well grounded in reasonable estimates of a case's merits. But the evidence is not conclusive because, with one exception, the anchors in these studies arguably provided some relevant information. In the case of the motion, the defendant could have been conveying its belief that the case was truly worth very little. In the settlement cases, the offer by the plaintiff is actually relevant, even though judges are supposed to ignore it. The judges who saw the reference to \$415,300 in the employment discrimination case also could have interpreted this as an indication that the case was much more serious than the facts suggested. Similarly, the high initial interest rate could have signaled to the bankruptcy judges that the debtor presented enormous credit risks that other facts did not reveal. Likewise, all of the research using archival data involved anchors that were apt to be meaningful, or at least appear to be meaningful, to the judges. Only the German experiment involving the roll of the dice as an anchor provided direct evidence that numeric reference points that are transparently irrelevant affect judges. The study is helpful in this regard, but using dice as a reference point is obviously strange and might have undermined the judges' willingness to take the study seriously. It is hard to imagine what the judges thought that the point of the dice roll was in this experiment. Nevertheless, determining whether arbitrary anchors distort judges' decisions in more realistic settings requires further investigation.

In sum, the weight of evidence suggests that anchors influence judges excessively. Even if some of the anchors in the studies appear to convey meaningful information, the influence of anchoring cuts across so many domains that it appears to be a robust phenomenon.

III. THE CURRENT STUDY

To determine whether judges are subject to arbitrary influences that can produce erratic awards and sentences, we conducted four new studies in which we exposed active trial judges to various forms of anchors. In these studies, we used the same methodology that we have used in past research on judges.¹⁰⁰ That is, judges attending judicial education conferences reviewed the facts of a hypothetical case that called for a numeric judgment. All of the judges reviewed the same case, except we exposed some of the judges to a variation in the fact pattern that was designed to

98. Shawn D. Bushway, Emily G. Owens & Anne Morrison Piehl, *Sentencing Guidelines and Judicial Discretion: Quasi-Experimental Evidence from Human Calculation Errors*, 9 J. EMPIRICAL LEGAL STUD. 291, 298 (2012). Only 10% of the forms were filled out correctly, thereby producing inaccurate sentence recommendations. *Id.*

99. See Yun-chien Chang, Kong-pin Chen, Chang-ching Lin & Yu-sheng Liu, *Anchoring Effects in Real Litigation: An Empirical Study* (2013) (unpublished manuscript) (on file with author).

100. See, e.g., Guthrie et al., *supra* note 78.

test for the anchoring effect. In the first of these studies, we tested whether the metric used to sentence defendants (months versus years) influenced the judges' sentences. In the second study, we explored whether informing judges of a damage cap would alter how they assessed damage awards. In the third, we simply varied the order in which judges sentenced two criminal defendants on the theory that the first sentence might anchor the second. In the final study, we used a plaintiff's offhand reference to an award on a "court TV show" as an anchor, and we also tested an inoculant designed to eliminate or reduce the influence of anchoring.

In all instances, the hypothetical cases were embedded within a group of three to five other cases designed to address other issues. On the last page of the survey, we asked the judges to identify their gender, years of experience as a judge, and political orientation.¹⁰¹ Finally, we gave the judges the opportunity to respond to the survey and participate in the educational program but to withdraw their survey for purposes of data analysis. A small number of judges did this and were removed from the analysis.

A. Study 1: Scaling Effects

– "Five hundred twenty-five thousand six hundred minutes. How do you measure, measure a year?"¹⁰²

The widespread influence of anchoring on numeric judgments suggests that damage awards and sentences do not reflect stable assessments of the appropriate amount of compensation or punishment. There are many scales on which quantities may be expressed. If numeric assessments are as malleable as the research on anchoring suggests, judges might be vulnerable to distortions of the scale they use to render judgments. Even though years and dollars are units with clear meaning that judges should understand well, the research on anchoring suggests that people's perception of such units can vary.

Scale distorts judgments and preferences. A wide range of studies have documented such effects. Consumers prefer rewards programs that are denominated in thousands of points rather than hundreds, even though the units are set arbitrarily.¹⁰³ Stock splits increase share value, even though they have no effect on a firm's true worth.¹⁰⁴ Travelers who have to translate their currency of higher value

101. We used the following question to elicit political orientation: "Which of the two major political parties in the United States most closely matches your own political beliefs?" We then listed the two major parties with a blank next to each for the judges to check. Most judges answered this question. Those who did not or who wrote "other" were retained in the primary analysis of the effect of anchor but were excluded from analysis of their political orientation. We did not ask this question of Canadian judges, military judges, or Dutch judges.

102. JONATHAN LARSON, *Seasons of Love, on RENT: ORIGINAL BROADWAY CAST RECORDING* (Dreamworks Records 1996).

103. See Rajesh Bagchi & Xingbo Li, *Illusionary Progress in Loyalty Programs: Magnitudes, Reward Distances, and Step-Size Ambiguity*, 37 J. CONSUMER RES. 888, 892–95 (2010) (reporting research in which subjects expressed a greater affinity for a consumer reward program in which points and prizes were denominated in thousands).

104. See Subhankar Nayak & Nagpurmanand R. Prabhala, *Disentangling the Dividend Information in Splits: A Decomposition Using Conditional Event-Study Methods*, 14 REV. FIN. STUD. 1083, 1112 (2001) (analyzing data on stock splits and concluding that "46% of split

into one of lower value (such as when Americans use pesos) tend to overspend.¹⁰⁵ Similarly, American consumers express less willingness to pay for a silk tie as the nominal denomination of the currency in which the item is priced increases.¹⁰⁶ Workers are happier with a 5% raise in a time of 12% inflation than with a 7% salary cut in the face of 0% inflation, even though the two are economically identical.¹⁰⁷ Gamblers treat a 5–9 in 100 chance of winning as higher than a 1 in 10 chance.¹⁰⁸ People planning for a wilderness trek state that a battery for a GPS device needed to guide them will be more reliable when the battery life is described as 120 minutes rather than 2 hours.¹⁰⁹ People say they are willing to pay 126 euros for a cellphone with an advertised battery life of 6 days, but only 102 euros for one with a battery life of 144 hours.¹¹⁰ Movie watchers state that they would prefer a movie rental plan that offered 7 movies per week for \$10 per week over one that offered 9 movies per week for \$12 per week; but their preferences reverse when the plans were multiplied out into a whole year; that is, they preferred a plan that offered 468 (9 x 52) movies per year for \$12 per week over one that offered 364 (7 x 52) movies per year for \$10 per week.¹¹¹

As this research suggests, changes in time scales commonly have a large effect on judgment and choice. People find cancer risks to be more frightening when expressed as a risk per day rather than as a risk per year.¹¹² People perceive an

valuation effects are not caused by dividend information implied by splits”).

105. See Klaus Wertenbroch, Dilip Soman & Amitava Chattopadhyay, *On the Perceived Value of Money: The Reference Dependence of Currency Numerosity Effects*, 34 J. CONSUMER RES. 1, 7 (2007) (presenting evidence showing that “consumers . . . anchor[] on the numerosity of the nominal difference between that reference standard and the price of the transaction in order to estimate the real value of that transaction”).

106. See Priya Raghurib & Joydeep Srivastava, *Effect of Face Value on Product Valuation in Foreign Currencies*, 29 J. CONSUMER RES. 335, 338–39 (2002) (“[S]ubjects’ willingness to pay increased with the face value of the foreign currency as the exchange rate became a larger multiple of the home currency.”).

107. See Daniel Kahneman, Jack L. Knetsch & Richard Thaler, *Fairness as a Constraint on Profit Seeking: Entitlements in the Market*, 76 AM. ECON. REV. 728, 731 (1986) (reporting that 22% of subjects stated that a 7% raise in the face of 12% inflation was unfair, but 62% stated that a 7% reduction in salary in the absence of inflation was unfair); see also Eldar Shafir, Peter Diamond & Amos Tversky, *Money Illusion*, 112 Q.J. ECON. 341, 366–72 (1997) (documenting numerous examples of similar effects).

108. See Veronika Denes-Raj & Seymour Epstein, *Conflict Between Intuitive and Rational Processing: When People Behave Against Their Better Judgment*, 66 J. PERSONALITY & SOC. PSYCHOL. 819, 822 (1994).

109. See Y. Charles Zhang & Norbert Schwarz, *How and Why 1 Year Differs from 365 Days: A Conversational Logic Analysis of Inferences from the Granularity of Quantitative Expressions*, 39 J. CONSUMER RES. 248, 255 (2012) (“Participants estimated that a battery life claim of ‘up to 2 hours’ would translate into actual service of 1.49 hours (equal to 89 minutes), whereas a claim of ‘up to 120 minutes’ would translate into actual service of 106 minutes.”).

110. See Christophe Lembregts & Mario Pandelaere, *Are All Units Created Equal? The Effect of Default Units on Product Evaluations*, 39 J. CONSUMER RES. 1275, 1282 (2013).

111. See Katherine A. Burson, Richard P. Larrick & John G. Lynch, Jr., *Six of One, Half Dozen of the Other: Expanding and Contracting Numerical Dimensions Produces Preference Reversals*, 20 PSYCHOL. SCI. 1074, 1076–77 (2009).

112. See Sucharita Chandran & Geeta Menon, *When a Day Means More than a Year: Effects of Temporal Framing on Judgments of Health Risk*, 31 J. CONSUMER RES. 375, 384 (2004).

increase in the length of a product warranty from 84 months to 108 months as a more valuable increase than an increase from 7 years to 9 years.¹¹³ People asked how much they will spend in 12 months provide lower numbers than when they are asked how much they will spend in 1 year.¹¹⁴ When offered a chance to donate to a worthy cause, 52% of MBA students said they would be willing to donate 85 cents per day, even though only 30% were willing to donate \$300 per year.¹¹⁵ Public radio campaigns commonly ask for a donation that amounts to “pennies a day” rather than dollars per year.¹¹⁶ One year, it seems, does not mean the same thing to most people as 12 months or 365 days.

These results can be bewildering, but they begin to make sense in light of how psychologists have found that people make choices. People do not make purchasing decisions by tapping into an underlying sense of value, but instead construct decisions using simple cognitive processes.¹¹⁷ As many of the studies suggest, people prefer goods that come in larger numbers and hence prefer goods with attributes denominated in smaller units—a phenomenon known as “numerosity.”¹¹⁸ Larger numbers often mean more of an item is available and hence make the amount of that item seem more valuable.¹¹⁹ Even chickens will work harder to earn a kernel of corn that has been divided into four parts than one that is whole.¹²⁰ People also assume that the unit is chosen in a meaningful fashion¹²¹ and therefore infer that smaller units connote a more precise estimate.¹²² Furthermore, people like units that are familiar

113. See Mario Pandelaere, Barbara Briens & Christophe Lembregts, *How To Make a 29% Increase Look Bigger: The Unit Effect in Option Comparisons*, 38 J. CONSUMER RES. 308, 310–11 (2011).

114. Gülden Ülkümen, Manoj Thomas & Vicki G. Morwitz, *Will I Spend More in 12 Months or a Year? The Effect of Ease of Estimation and Confidence on Budget Estimates*, 35 J. CONSUMER RES. 245, 253 (2008).

115. John T. Gourville, *Pennies-a-Day: The Effect of Temporal Reframing on Transaction Evaluation*, 24 J. CONSUMER RES. 395, 396 (1998) (reporting that “the percentage of subjects agreeing to donate was significantly higher under the [donation-per-day condition] than under the aggregate framing (52 vs. 30 percent . . .”).

116. See, e.g., *id.* at 395 (“Chicago Public Radio has an ongoing membership drive in which individuals are asked to join their ‘Dollar-a-Day Club.’”).

117. See Lichtenstein & Slovic, *supra* note 31, at 2 (“The variability in the ways we construct and reconstruct our preferences yields preferences that are labile, inconsistent, subject to factors we are unaware of, and not always in our own best interests. Indeed, so pervasive is this liability that the very notion of a ‘true’ preference must, in many situations, be rejected.”).

118. See Brett W. Pelham, Tin Tin Sumarta & Laura Myaskovsky, *The Easy Path from Many to Much: The Numerosity Heuristic*, 26 COGNITIVE PSYCHOL. 103, 104 (1994) (“[P]eople . . . sometimes overinfer quantity from numerosity.”).

119. *Id.* at 105 (“[M]ore pieces of something usually turns out to be more of that something.” (emphasis in original)).

120. John B. Wolf & Martin D. Kaplon, *Effect of Amount of Reward and Consummative Activity on Learning in Chickens*, 31 J. COMP. PSYCHOL. 353, 358–59 (1941).

121. See Zhang & Schwarz, *supra* note 109, at 249 (arguing that norms of communication suggest that people will convey information in relevant units).

122. *Id.* at 251 (“[C]onsumers infer that the real value is closer to the communicated value when it is conveyed in fine-grained rather than coarse units . . .”).

and convey some sense of fluency and meaning.¹²³ Finally, because people tend to prefer commodities that are easier to assess,¹²⁴ expressing the quantity of something in a meaningful metric makes it more attractive.¹²⁵

A study conducted by McAuliff and Bornstein demonstrates the potential power of scale in a legal setting.¹²⁶ These researchers had jury-eligible adults read a description of a personal injury case in which the defendant was clearly liable. The researchers asked their subjects to identify an appropriate award for the 2 years of pain and suffering that the plaintiff had experienced. The plaintiff's attorney varied the request for damages in his closing argument by doing one of the following: not stating an amount (control condition); asking for \$10 per hour for 2 years (hours scale); asking for \$7300 per month for 24 months (months scale); or asking for \$175,000 (anchor condition). The requests for \$175,000 and \$10 per hour produced the highest awards. The \$175,000 likely acted as an anchor, which explains why it produced high awards. The ability of the hours scale to produce a high award relative to the monthly scale shows the importance of scaling. For many of the subjects, \$7300 per month and \$10 per hour can easily be compared to a wage or salary. For most, \$10 per hour might not seem like much, so it seemed like a reasonable basis for an award. But \$7300 per month seems like a lot of money each month (even though it is almost identical to \$10 per hour times 24 hours times 30 days). The study thus shows both the power of anchors and the effect that variation in scale can have on judgment. Different scales convey different meanings and thus produce variations in how much an injury seems worth.

Do variations in scale affect judges when they make quantitative judgments? Unlike the nonjudge adults in McAuliff and Bornstein's study, judges are experts. They might not be affected by specious differences in amounts or durations. On the other hand, variations in scale might convey different impressions and trigger different associations in judges, just as they do in others. One year might not mean the same to a judge as 12 months, and 360 months might not trigger the same impression as 30 years.

1. Methods and Materials

To test whether judges are vulnerable to such scaling effects, we gave judges a one-page description of a criminal case and asked them to impose a criminal

123. See Lembregts & Pandelaere, *supra* note 110, at 1278 ("Given that people are more familiar with default units, consumers should be able to process attribute information more easily when it appears in default units rather than in nondefault units.").

124. See Christopher K. Hsee, *The Evaluability Hypothesis: An Explanation for Preference Reversals Between Joint and Separate Evaluations of Alternatives*, 67 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 247, 255–56 (1996) (describing a series of studies showing that people prefer items that are easier to evaluate).

125. See Lembregts & Pandelaere, *supra* note 110, at 1283–84 (presenting evidence supporting the researchers' prediction that "the processing fluency associated with the unit mediates the effects of default units on product evaluations").

126. Bradley D. McAuliff & Brian H. Bornstein, *All Anchors Are Not Created Equal: The Effects of Per Diem Versus Lump Sum Requests on Pain and Suffering Awards*, 34 LAW & HUM. BEHAV. 164 (2010).

sentence. The materials described a defendant named “Smith” as having “pled guilty to simple voluntary manslaughter.” The materials indicated that Smith admitted stabbing a fellow customer to death in a bar after “the victim informed him that he had had sexual intercourse with Smith’s fiancé.” Smith had not been drinking. The materials also described Smith’s background as follows:

Smith is 23 years old and has no significant criminal history, though he was once arrested, but not charged, following a bar fight while he was intoxicated. He is in good health and does not appear to have a substance abuse problem, although he admits that he occasionally drinks to excess. He has been employed as a custodian for the past four years at a large office complex. He remains unmarried and has no children.

We asked all of the judges to sentence Smith. We posed the following question to half of the judges: “Without regard to the sentence maximum in your own jurisdiction, how many years would you sentence Smith to serve in prison, based on the facts above?” This was followed by a blank labeled “___/years.” For the other half, we replaced the word “years” with “months” and followed this with a blank labeled “___/months.” In effect, we asked half of the judges to sentence the defendant in years and the other half in months.

We presented these materials to 135 judges attending the American Judges’ Association annual conference in Denver, Colorado, in 2010. These judges attended a plenary session entitled “How Do Judges Judge” on the first day of their conference. These judges came from all over the United States, although most were state court judges.¹²⁷ The judges had an average of 13 years of experience as judges (a median of 12 years), ranging from 6 months to 40 years. Of the 128 who reported their gender, 37 (29%) were females. Of the 112 judges who reported their political affiliation as Republican or Democratic (we omitted those who identified as other, independent, or simply refused to answer), 38 (34%) were Republicans and 62% were Democrats.¹²⁸

Substantively, the appropriate sentence for voluntary manslaughter varies by jurisdiction. Procedurally, some of the states rely primarily on sentences assigned in years and some on sentences assigned in months (as the federal system does).¹²⁹ Furthermore, some states follow sentencing guidelines that restrict or guide sentencing, while others do not.¹³⁰ Our materials asked judges to sentence without

127. All but 3 of the judges were trial judges.

128. A total of 23 judges identified themselves as independent, unaffiliated, or did not respond to our question about political affiliation: “Which of the two major political parties in the United States most closely matches your own political beliefs?” We excluded these judges from any analysis of political party so that the judges are simply identified as Republican or Democratic. We treated political orientation the same way in all of our studies. Thus, the percentage of Democratic and Republican judges always adds up to 100%.

129. U.S. SENTENCING COMM’N, *supra* note 4, § 5A (depicting a sentencing table denoted in months); *see also, e.g.*, CAL. PENAL CODE § 1170 (West 2014) (years); KAN. STAT. ANN. § 21-4704 (2007) (months); TEX. CODE CRIM. PROC. ANN. art. 42 (West 2014) (years); UTAH CODE ANN. § 77-18-1 (LexisNexis Supp. 2014) (months); MINN. SENTENCING GUIDELINES COMM’N, MINN. SENTENCING GUIDELINES & COMMENTARY § 4A (2014) (months).

130. NAT’L CTR. FOR STATE COURTS, STATE SENTENCING GUIDELINES: PROFILES AND

regard to the sentencing maximum in their own jurisdiction and did not provide any guidelines for judges to follow in setting the sentence.

2. Results

Sentencing format had an enormous effect on the judges. The 71 judges who sentenced in years provided an average sentence of 9.7 years, or 115 months.¹³¹ In contrast, the 60 judges who sentenced in months provided an average sentence of only 66.4 months, or 5.5 years. This difference was statistically significant.¹³² As Table 1 (below) shows, the distribution of the sentences varied markedly between the two formats. Only 22% (13 out of 60) of the judges sentencing in months imposed sentences greater than the median sentence of judges sentencing in years (9 years), and only 1.7% (1 out of 60) of the judges sentencing in months assigned a sentence greater than the 75th percentile of judges sentencing in years (12.5 years).

Table 1. Sentence length by condition: average, percentiles and standard deviation (in years)

Condition (and <i>n</i>)	Average	25th percentile	50th percentile	75th percentile	Standard deviation
Months (60)	5.5	3.0	5.0	8.0	3.5
Years (71)	9.7	5.0	9.0	12.5	6.8

The judges' demographic backgrounds had little effect on these results. Male judges assigned slightly longer sentences on average than female judges (8.1 versus 7.0 years), but this difference was not statistically significant.¹³³ Judges who identified as Democrats assigned essentially identical sentences to those who identified as Republicans (7.8 versus 7.7 years, respectively, which was not a statistically significant difference).¹³⁴ Experience did not correlate with sentence length.¹³⁵ None of these variables interacted statistically significantly with the variation in scale.¹³⁶

The judges sentencing in years expressed greater variability than judges sentencing in months. The standard deviation of the judges' sentences was nearly twice as large among judges sentencing in years as compared to months. This

CONTINUUM 3 (2008) (reporting that 21 states have sentencing guidelines).

131. Three judges did not respond to this question.

132. $t(129) = 4.51, p < 0.001$. Months were divided by 12 to facilitate the comparison. Because the standard errors varied, the *t*-test was performed with no assumption of equal variance in each condition.

133. $t(126) = 1.00, p = 0.32$.

134. $t(110) = 0.09, p = 0.92$.

135. $r = .05, p = 0.56$.

136. This analysis was accomplished with a linear regression of the sentence on the scale (years versus months), experience, gender, party affiliation, and an interaction of these three terms with the scale. The *t* statistics for the three interaction terms were 1.06, 1.05, and 0.14 respectively, with *p* values of 0.29, 0.30, and 0.88.

difference was statistically significant.¹³⁷ That said, in both conditions, the standard deviation was roughly proportional to the average length of the sentence in that condition. Hence, it does not seem that the judges in the years condition sentenced more erratically than the judges in the months condition.

3. Discussion

Judges sentencing in months imposed shorter sentences than judges sentencing in years. To the judges sentencing in years, roughly 9 years was about the right sentence. But fewer than 25% of the judges sentencing in months assigned sentences of that length or greater. Although 9 years seemed like an appropriate sentence to the judges, 108 months seemed far too long.

The result is consistent with the previous findings concerning anchoring in judges. Anchors distort the sense of scale that facilitates a reliable translation of a qualitative sense of the appropriate sentence into a numeric sentence. So too does the actual scale. Apparently, 12 months seems different to judges than 1 year. In this scenario, the scale provided no information to the judges about the case, and yet it distorted their sentences.

This result does not necessarily suggest that jurisdictions that assign sentences in months are more lenient than those that sentence in years. We suspect that judges become accustomed to the scale that they are using over time. Our result suggests only that in an unfamiliar, impressionistic setting, judges reacted to the scale.

At least one jurisdiction has attempted to take advantage of this scaling effect in order to influence the duration of sentences. In 1991, Finnish “[c]ourts were . . . encouraged to use short sentences of imprisonment by instructing them to mete out the shorter sentences in days instead of months.”¹³⁸ Sentence length declined after the change. Because other reforms accompanied the change in metric, it is impossible to determine the exact cause of the decline with certainty, but our research implicates the metric.¹³⁹

The possible difference between judges accustomed to sentencing in months versus years suggests that one potential explanation for the effect is weak mathematics among those judges who normally sentence in years. We believe that judges are perfectly capable of multiplying 9×12 to reach the appropriate sentence. The judges might not have done the math explicitly, however, and may have simply made a rough estimate of the appropriate sentence. The result thus resembles an early study of anchoring conducted by Tversky and Kahneman in which they asked subjects either to determine the product of $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ or the product of $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$.¹⁴⁰ Any undergraduate is capable of solving either of these mathematical functions precisely, but most of the subjects made intuitive estimates instead. Those who started with the lower numbers produced lower estimates than those who started with the higher numbers. Tversky and Kahneman suggested that the subjects multiplied the first few numbers, used that partial product

137. $F(70, 29) = 3.69, p < 0.001$.

138. See Tapio Lappi-Seppälä, *Sentencing and Punishment in Finland: The Decline of the Repressive Ideal*, in SENTENCING AND SANCTIONS IN WESTERN COUNTRIES 92, 113 (Michael Tonry & Richard S. Frase eds., 2001).

139. *Id.* at 114.

140. Tversky & Kahneman, *supra* note 1, at 1128.

as an anchor, and made inadequate upward adjustments. Thus, when the initial numbers used were higher, the subjects produced higher final estimates.

We also suspect that the effect of scale might vary with the severity of the crime. Severe crimes that demand lengthy sentences of 20 years or more might produce the opposite effect. As the number of months grows into the hundreds, judges might start to become insensitive to the length of the sentence. The difference between 300 and 360 months might not seem as notable as the difference between 25 and 30 years. Likewise, sentences at the low end of the scale also reverse the result we identified. A sentence of 24 months might seem longer than a sentence of 2 years.

The variation among sentencing systems might provide one alternative account of our result that suggests that the scale inadvertently provided some information to the judges. Some jurisdictions assign sentences in months (or even days) for less severe crimes and in years for more serious crimes.¹⁴¹ Judges who are familiar with such systems might have assumed that the “years” prompt suggested that the crime was more serious and hence should draw a more severe sentence. We doubt that this is the explanation, though, for two reasons. First, the magnitude of the effect is too large to be explained by the behavior of a small number of judges. The scale seemed to shift the range of sentences, suggesting that the scale influenced most of the judges, not just the small number who reside in a jurisdiction with such a system. Second, the crime used in the survey is a familiar one to judges. Voluntary manslaughter is a serious crime in every jurisdiction and merits a lengthy sentence. Experienced judges could not conceivably have confused this crime with a low-level barroom brawl that might have produced a short sentence for assault and battery. This was a serious crime, and we are confident judges knew it.

B. Study 2: Damage Caps

To test whether anchoring would influence judges even when the anchor cannot plausibly provide them with relevant information, we considered highlighting numeric reference points of which judges were already aware. Our initial study of the effect of filing a motion to dismiss for failing to meet the minimum amount in controversy for federal diversity of citizenship jurisdiction itself was an effort of this sort. We suspect that federal judges are well aware that the jurisdictional threshold for diversity claims is \$75,000. Thus, mentioning the amount in the motion provided no new information but, as we noted above, filing the motion might have suggested to judges that the case was worth far less than they had thought. The legal system, however, provides numerous dollar amounts of various sorts that judges understand well and yet might not relate to a particular case.

Damage caps are one source of such numbers. Numerous jurisdictions have implemented damage caps in some classes of cases over the past two decades, usually as part of an effort at tort reform.¹⁴² Legislatures intend such caps as a response to

141. For example, New York expresses sentences for misdemeanors in months, *see* N.Y. PENAL LAW § 70.15 (McKinney 2009 & Supp. 2014), and felonies in years, *see* N.Y. PENAL LAW § 70.00 (McKinney 2009 & Supp. 2014).

142. *See* Michael I. Krauss, *A Medical Liability Toolkit, Including ADR*, 2 J.L. (1 J. LEGAL METRICS) 349 app. a (2012) (summarizing state-by-state damage caps and other laws applicable to medical liability); Robbennolt & Studebaker, *supra* note 69, at 354 (discussing

the perceived problem of “runaway juries” that award apparently outrageous sums.¹⁴³ Several scholars have studied the effect of such damage caps on both actual awards and settlement amounts.¹⁴⁴ Caps are obviously effective in reducing the prevalence of exceptionally high awards because they prohibit awards in excess of a certain amount. Scholars have also identified two unintended effects, however. First, juries limited by the cap in a particular category of awards tend to compensate by awarding more in another category.¹⁴⁵ For example, juries confronting caps on pain and suffering awards appear to be inclined to award more in punitive damages.¹⁴⁶ Second, caps create anchors that increase awards in cases that would otherwise produce small awards.¹⁴⁷ This effect even extends to parties who are settling a dispute.¹⁴⁸ So influential are damage caps that many have suggested that juries should not be informed of them, lest they have unintended effects on damage awards.¹⁴⁹

The research on anchoring predicts that damage caps will influence damage awards. Each of the theoretical accounts of how anchoring works predicts that damage caps will increase awards in low-value cases.¹⁵⁰ If jurors are simply cognitively lazy, they are apt to use the anchor as a starting point and adjust downward.¹⁵¹ This adjustment is apt to be inadequate for cases that would otherwise produce an award that is a fraction of the cap, so exposure to the cap will produce a

the caps that Alabama, Florida, and Nevada implement by allowing punitive damages only up to a certain multiple of the compensatory damages); Joseph Sanders, *Reforming General Damages: A Good Tort Reform*, 13 ROGER WILLIAMS U. L. REV. 115, 129–43 (2008) (discussing damage caps); Studdert et al., *supra* note 7, at 63–66 (briefly summarizing the genesis, shortcomings, and consequences of damage caps).

143. See Robbennolt & Studebaker, *supra* note 69, at 366 (“[C]ourts and legislatures alike have pursued tort reform efforts largely aimed at reining in what are perceived to be juries out of control. . . . One proposed method for reigning in damage award amounts is to limit, or cap, the amount that can be awarded in punitive damages.”)

144. See generally, e.g., Sandra N. Hurd & Frances E. Zollers, *State Punitive Damages Statutes: A Proposed Alternative*, 20 J. LEGIS. 191 (1994) (arguing that, in some cases, limiting damage awards can inhibit the purpose of punitive damages, such as deterrence); Jacqueline Perczek, *On Efficiency, Punishment, Deterrence, and Fairness: A Survey of Punitive Damages Law and a Proposed Jury Instruction*, 27 SUFFOLK U. L. REV. 825 (1993).

145. See Robbennolt & Studebaker, *supra* note 69, at 356. Although damage caps typically limit awards to selected categories of damages—usually noneconomic damages or punitive damages—at least one state has a comprehensive cap that includes all categories of damage awards in medical cases. See Krauss, *supra* note 142, at 374 (citing VA. CODE ANN. § 8.01-581.15 (2011)).

146. See Robbennolt & Studebaker, *supra* note 69, at 366.

147. *Id.* at 361; see also Ronen Avraham & Álvaro Bustos, *The Unexpected Effects of Caps on Non-Economic Damages*, 30 INT’L REV. L. & ECON. 291, 291 (2010) (“[W]hile caps always reduce the recoveries of plaintiffs with large claims, caps may increase recoveries of plaintiffs with low claims compared to their recoveries in states with no caps.” (emphasis in original)); Hinsz & Indahl, *supra* note 69, at 1006.

148. See Orr & Guthrie, *supra* note 52, at 611.

149. See Michael S. Kang, Comment, *Don’t Tell Juries About Statutory Damage Caps: The Merits of Nondisclosure*, 66 U. CHI. L. REV. 469, 480–86 (1999) (explaining the various effects of disclosing the damage cap to jurors).

150. See Orr & Guthrie, *supra* note 52, at 605–11.

151. *Id.*

higher award. Similarly, if jurors initially attempt to determine whether the case is worth more or less than the cap, they will consider the features of the case that suggest it would merit a high award.¹⁵² Even if jurors reject the notion that the case is worth more than the cap, their attention has been drawn to those aspects of the case that suggest it would merit a high award. Finally, and we believe most plausibly, the cap provides jurors a sense of scale. Jurors contemplating an award of \$10,000 will learn that cases of this type could be worth fifty times as much, which might make the \$10,000 award seem inadequate. In short, damage caps are exactly the kind of anchor apt to have a large effect on judgment.

It comes as little surprise to researchers that damage caps have a large effect on awards provided by juries,¹⁵³ but what about judges? Although no previous research on the influence of damage caps on judges exists, the concerns that scholars have raised about the influence of damage caps assumes that judges are immune from any such effects. The reform efforts all involve blinding juries to the existence of anchors or giving judges greater control over damage awards.¹⁵⁴ Unlike jurors, however, judges cannot remain ignorant of the damage caps. If damage caps have a powerful effect on jurors' damage awards, they seem likely to affect judges as well.

1. Methods and Materials

To determine whether damage caps influence judges, we conducted an experiment involving 115 Canadian trial judges and 65 newly elected trial judges in New York State. The Canadian judges were attending an annual conference organized by the National Judicial Institute in Canada in 2010, 2011, and 2013. The conference is held each year for judges who have served between five and ten years (which they refer to as their "sophomore" years). Judges are not obliged to attend, but each year the week-long conference draws roughly 40 judges. The judges at the conference attend each of the sessions. Our session was entitled "How Judges Decide" in 2010 and "The Working of the Judicial Mind" in 2011 and 2013. We did not ask these judges their political affiliation. Of the 111 judges who informed us of their gender, 51 (or 46%) were female. For the 25 (22%) of the judges from Quebec, the materials were translated into French.

Canada is an ideal jurisdiction in which to test the influence of damage caps on trial judges. In 1978, the Canadian Supreme Court, in a trilogy of cases, imposed a cap of \$100,000 on noneconomic damage awards in civil cases.¹⁵⁵ The court explicitly indicated that the cap must be adjusted for inflation as measured by the Bank of Canada inflation index on a monthly basis.¹⁵⁶ The existence of the cap is

152. *Id.*

153. *Id.* at 605–08.

154. See Kang, *supra* note 149, at 481–86.

155. *Andrews v. Grand & Toy Alberta Ltd.*, [1978] 2 S.C.R. 229 (Can.); *Arnold v. Teno*, [1978] 2 S.C.R. 287 (Can.); *Thornton v. Prince George Sch. Dist. No. 57*, [1978] 2 S.C.R. 267 (Can.).

156. D. BRUCE GARROW & KATHERINE L. AYRE, CONFERENCE ON INT'L AVIATION LIABILITY & INS., THE RECOVERY OF NON-PECUNIARY DAMAGES IN CANADA: THE CAP ON RECOVERY, JURY TRIALS, AND OTHER UNIQUE CONSIDERATIONS FOR GENERAL DAMAGE AWARDS 4 (2009) ("In Canada, general damage awards for pain and suffering are currently capped at \$100,000 and indexed to inflation.").

well known to Canadian trial judges. Because of the monthly adjustment, however, the judges have a rough idea of the amount of the cap, but likely do not know the precise amount.

Because we could not realistically manipulate the existence of the cap, or even its amount, we instead varied whether we informed the judges of the precise amount or not. We presented judges with a case that we expected would produce a nonpecuniary award between \$30,000 and \$50,000—well below the cap. The materials described the case as arising from an automobile accident. The defendant was a large package-delivery company. The plaintiff was “a 31-year-old male software programmer” who was injured by a truck driven “erratically” by one of the defendant’s drivers. The defendant admitted that it was liable. The judges were told that the parties had settled all claims for medical expenses and economic losses, including lost wages. This meant that the only issue for the judges in this bench trial was to determine the appropriate compensatory damage award for pain and suffering.

The materials described the plaintiff’s symptoms in some detail:

The evidence concerning the nature and extent of the plaintiff’s residual symptoms was reasonably clear. Shortly after the accident (it is now approximately four years later), the plaintiff was diagnosed with cervical and thoracic strain and a severe concussion. The effects of the concussion persisted for about a year, during which time the plaintiff experienced memory problems. The plaintiff suffered no other cognitive problems and the memory problems have not recurred.

Although the plaintiff’s doctors originally concluded that he would recover fully within about a year, the symptoms resulting from the cervical and thoracic strain gradually plateaued. The plaintiff still has stiffness in his neck, which is annoying and causes him to experience severe headaches that are sometimes debilitating. The plaintiff also experiences numbness in his left hand if he works at the computer for more than a few hours at a stretch.

The plaintiff’s job requires that he use a computer for several hours a day. Only the more severe headaches cause him to lose any notable time at work. He manages to work through the less severe headaches with the assistance of over-the-counter pain medication and the numbness in his left hand dissipates if he takes a short break. He testified that his pain makes it hard for him to function as a parent for his two children, making him short-tempered and tired. He complained that he is often unable to help his children with their homework or play sports with them.

Although his doctors have not identified any specific physical injury as the cause of the plaintiff’s symptoms, he might have suffered some slight brachial nerve damage during the accident. Doctors expect that his stiffness, headaches, and numbness will persist.

For half of the judges, the materials indicated that the defendant contended “that the injuries are not serious and do not warrant a significant damage award.” The materials then asked the judges, “How much would you award the plaintiff for pain and suffering?” For the other half of the judges, we included an additional sentence. In 2010, the additional sentence stated: “The defendant also reminded you that an award of noneconomic damages cannot exceed the cap imposed by the Canadian Supreme Court, which is currently \$332,236.” In 2011, we included the same sentence for half of the judges, except that we indicated that the *plaintiff* (rather than

the defendant) provided the reminder, and we also updated the amount to \$342,535. In 2013, the materials reminded the judges of the cap (then updated to \$367,160) but did not attribute the reminder about the cap to either party.

We also gave a similar set of materials to 65 newly elected trial judges in New York State. New York does not have a damage cap akin to the Canadian limitation, but for half of the judges, the materials added the following: "Further suppose that the parties have stipulated that a statutory damage cap for common-carrier defendants of \$750,000 applies in this instance. Thus, you may not award more than \$750,000." These judges had no previous judicial experience and were attending an annual conference for new judges in New York. Among the 63 judges who identified their gender, 35% were female. Among the 62 who identified their political orientation, 26% were Republicans and 74% were Democrats.

2. Results

Overall, the reference to the damage cap influenced the Canadian judges. The median award among the Canadian judges who learned about the cap was \$85,000, as opposed to \$57,500 among those who did not.¹⁵⁷ This difference was statistically significant.¹⁵⁸ This effect was driven almost entirely by the variation in which the defendant mentioned the cap.¹⁵⁹

Table 2. Median award by condition (in thousands of dollars) and sample size

Party who identifies damage cap		Condition	
		No reference to cap	Cap referenced
Canada	Defendant	45.0 (19)	100.0 (26)
	No attribution	55.0 (15)	60.0 (16)
	Plaintiff	87.5 (14)	97.5 (22)
	Total	57.5 (48)	85.0 (64)
New York	No attribution	100.0 (28)	250.0 (30)

157. Three judges did not respond; 2 of these judges were in the condition in which the plaintiff referred to the damage cap, and 1 was in the condition in which the reference to the damage cap was unattributed.

158. Using the Mann-Whitney test, $p = 0.01$. The studies involving damage awards generally produced positively skewed distributions of awards with some high outliers, making untransformed parametric statistical analysis unreliable. We therefore report the more stable median awards, rather than the means. When we compare only two samples, we use the Mann-Whitney rank-sum test as our test statistic. When multiple conditions were compared or when we tested for the influence of demographic factors, however, we transformed the data, analyzing the square root of each award rather than the raw award.

159. Analysis of variance (ANOVA) of the square root of the awards on the anchor condition and each of the three versions reveals a significant main effect of version ($F(2, 106) = 5.52, p = 0.005$) and a significant interaction ($F(2, 106) = 4.84, p < 0.01$).

The 64 male and 43 female judges did not provide significantly different awards, nor did the anchor affect them differently.¹⁶⁰ The 25 French-speaking and the 90 English-speaking judges likewise did not provide different awards, and neither did they react differently to the anchor.¹⁶¹

The effect of the anchor was even more pronounced among the new New York judges. Those who learned of the damage cap awarded a median of \$250,000, as compared to \$100,000 among those who did not learn about it, which was a significant difference.¹⁶² None of the judges awarded more than \$750,000, and hence the damage cap did not meaningfully constrain the judges—instead, it acted as an anchor that increased awards.

Among the judges in New York, gender and political orientation had some effect on awards. The damage cap influenced the female judges more than male judges. The 39 male judges awarded a median of \$100,000, as compared to \$137,500 among the 18 female judges. The anchor also seemed to have a greater effect on the female judges: male judges awarded a median of \$50,000 in the no-anchor condition and \$200,000 in the anchor condition, while female judges awarded a median of \$100,000 in the no-anchor condition and \$400,000 in the anchor condition. Both gender and the interaction of gender and anchoring were marginally significant.¹⁶³ The 42 judges who identified themselves as Democrats awarded a median of \$150,000, as compared to \$100,000 among the 13 judges who identified themselves as Republicans. The anchor had a greater effect on the Republican judges: Democratic judges awarded a median of \$100,000 in the no-anchor condition and \$150,000 in the anchor condition, while Republican judges awarded a median of \$25,000 in the no-anchor condition and \$450,000 in the anchor condition. The main effect of political orientation was not significant,¹⁶⁴ but an interaction between the anchoring condition and party affiliation was statistically significant.¹⁶⁵

3. Discussion

References to damage caps affected the judges in our study. Telling judges not to award more than the cap led them to award more than when no reference was made to that cap. Because the Canadian judges were already aware of the existence of the cap, the effect must have arisen from increasing its salience. Because the case is worth so much less than the cap, the judges who were not reminded of it were likely not thinking about it. But when the materials highlighted the cap, judges became aware of the scale on which they were making their assessment. The cap made the small award that they were considering seem paltry relative to the available range.

160. F 's < 1.00, p 's > 0.5.

161. F 's < 1.00, p 's > 0.5.

162. Using the Mann-Whitney test, $p < 0.005$. 7 judges did not respond; 4 of these judges were in the no-anchor condition, and 3 were in the anchor condition.

163. $F(1, 53) = 3.60, p = 0.06$ and $F(1, 53) = 3.60, p = 0.06$. This analysis was conducted on a square root of the awards, as the distribution of the awards was positively skewed.

164. $F(1, 52) = 0.00, p > 0.5$. This analysis was also conducted on the square root of the awards.

165. $F(1, 52) = 6.04, p < 0.02$.

The mechanism by which we identified the anchor, however, could have inadvertently signaled to the judges that the case might be worth more than they otherwise would have thought. The anchoring effect occurred largely in the version in which the defendant referred to the cap. The cap had less influence when the plaintiff mentioned it and when the reference was unattributed. The defendant's statement that the award could not exceed the cap might have implied that the defendant feared that the case was worth much more than the cap. Although we provided judges with most of the information about the nature of the injury that they would have had in a real case, the materials are naturally somewhat less detailed. The judges were also apt to have been less motivated to measure the plaintiff's damages accurately than they would have been in a real case and might have been looking for a simple cue to approximate its value. Therefore, the defendant's reminder might have led the judges to assume that the injuries were more serious than they otherwise would have thought. The plaintiff's reminder, on the other hand, might have seemed self-interested and thus less credible than the defendant's, so it understandably had less influence on the judges.

Nevertheless, we believe that, as a normative matter, the judges should not have been influenced by any reference to the damage cap, regardless of its source. The facts provided the judges with enough information to determine an appropriate award without relying on an offhand remark. We believe that the results reflect an unfortunate influence of anchoring.

C. Study 3: Order Effects

The order or sequence in which information is learned or tasks are performed can alter judgment.¹⁶⁶ The sequence in which a judge sentences offenders provides another possible source of misleading anchors. Just as Frederick and Mochon found that the order in which subjects estimated the average weight of animals affected their judgments,¹⁶⁷ so too might the order in which judges assign sentences affect the relative length of their sentences.¹⁶⁸ A short sentence for a minor offense might serve as an anchor that shortens a subsequent sentence imposed for a more serious crime. Similarly, a long sentence imposed for a more serious crime might serve as an anchor that lengthens a subsequent sentence assigned to a less serious crime. Such a result would provide convincing evidence that anchors influence judgment even when they are wholly irrelevant. Obviously, a sentence assigned to an unrelated defendant for

166. See S. E. Asch, *Forming Impressions of Personality*, 41 J. ABNORMAL & SOC. PSYCHOL. 258, 271 (1946) (demonstrating that the sequence in which a person's traits are revealed influences our perception of that person and explaining that "the first terms set up in most subjects a direction which then exerts a continuous effect on the latter terms" (emphasis omitted)); see also Birte English, Thomas Mussweiler & Fritz Strack, *The Last Word in Court—A Hidden Disadvantage for the Defense*, 29 LAW & HUMAN BEHAV. 705 (2005) (presenting evidence that defense attorneys might anchor their own recommendations on prosecutors' recommendations, which are presented first).

167. Frederick & Mochon, *supra* note 61, at 126.

168. See Critcher & Gilovich, *supra* note 45, at 248 (concluding that "incidental numbers present in the environment influenced participants' estimates of uncertain values").

an unrelated crime provides a judge with no meaningful insight into the appropriate sentence for a subsequent case.

Anchoring is not the only possible extraneous consequence of sentence order. Judges might also suffer from a contrast effect. The contrast effect is the tendency to base judgments on comparisons with similar examples even though the judgments were intended to be independent.¹⁶⁹ The contrast effect produces the opposite result from anchoring on the influence of serial judgments; anchoring suggests that the initial judgment will pull the second closer, while the contrast effect suggests that the first judgment will push the second further away.¹⁷⁰

In our previous research on anchoring, we found anchoring effects rather than contrast effects, but other research suggests that contrast effects might occur in some circumstances. If contrast effects influence sentencing in judges, then reviewing a minor crime might make a more serious crime seem even worse in contrast, thereby producing a longer sentence for the more serious crime. Likewise, reviewing a serious crime might make a minor crime seem even less serious, thereby producing a shorter sentence for the less serious crime. Research by Kelman and his colleagues has demonstrated the influence of the contrast effect and the influence of scaling in lay assessments of criminal cases.¹⁷¹ In their study, mock jurors were more likely to sentence a hypothetical defendant to community service rather than jail if an inadequate community service sentence was presented as a third option.¹⁷² Other research, however, has failed to uncover related contrast effects in similar circumstances,¹⁷³ suggesting that the influence of contrast effects on assessments of criminal cases is not a robust phenomenon.

One previous study suggests that sentence order influences judges. Research by Danziger and his colleagues found that decisions made by an Israeli parole court were different during different times of the day.¹⁷⁴ This court granted only 20% of the applications for parole made immediately before lunch, but the success rate soared to 60% for the first case after lunch. The success rate declined again in the lead up to the court's afternoon break, and rose again after that break.¹⁷⁵ The

169. See Itamar Simonson & Amos Tversky, *Choice in Context: Tradeoff Contrast and Extremeness Aversion*, 29 J. MARKETING RES. 281, 281 (1992) ("Contrast effects are ubiquitous in perception and judgment."); see also Jeffrey J. Rachlinski, Andrew J. Wistrich & Chris Guthrie, *Altering Attention in Adjudication*, 60 UCLA L. REV. 1586, 1597–609 (2013) (documenting the influence of contrast effects in judges). See generally Cass R. Sunstein, Daniel Kahneman, David Schkade & Ilana Ritov, *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153 (2002) (describing research showing that people often make comparative judgments about fault in legal settings, even when absolute judgments are required).

170. See Frederick & Mochon, *supra* note 61, at 127–28 ("[P]rior judgments can affect respondents' interpretation of numeric labels even on objective scales.")

171. Mark Kelman, Yuval Rottenstreich & Amos Tversky, *Context-Dependence in Legal Decision Making*, 25 J. LEGAL STUD. 287, 295–97 (1996).

172. *Id.* at 296–97.

173. See, e.g., Jeffrey J. Rachlinski & Forest Jourden, *The Cognitive Components of Punishment*, 88 CORNELL L. REV. 457, 482 (2003) (finding that the data obtained in this study "[did] not show evidence of contrast effects in decision making about the death penalty").

174. See generally Shai Danziger, Jonathan Levav & Liora Avnaim-Pesso, *Extraneous Factors in Judicial Decisions*, 108 PROC. NAT'L ACAD. SCI. 6889 (2011).

175. *Id.*

researchers interpret these results as evidence that the judges grew hungry or tired, encouraging the somewhat easier decision to deny parole (most applicants are denied).¹⁷⁶ The researchers also provide an alternative account, however, suggesting that the judges had an implicit quota for each day;¹⁷⁷ having granted parole in a number of cases, the judges had to become stingy so as not to exceed their quota as the day progressed. Others have argued, however, that the order in which the Israeli parole board hears cases is not truly random.¹⁷⁸

The impact of sentence order thus remains uncertain. Judges might be subject to anchoring or contrast effects or might resist any such influence. A sentence-order effect would be pernicious because each defendant is entitled to be sentenced according to the merits of his or her own case. To ascertain whether judges resist this influence, we undertook a study of case order.

1. Methods and Materials

We studied the effect of case order with three different groups of judges: newly appointed military judges,¹⁷⁹ judges attending the annual Arizona Judicial Conference in 2011,¹⁸⁰ and Dutch judges attending a conference sponsored by the Dutch Judicial Council in the Netherlands.¹⁸¹ Each of these three groups assessed a similar pair of hypothetical cases. One of the cases consisted of an assault or a similar battery. We intended this case to constitute the less serious crime. The second case consisted of a voluntary manslaughter similar to that used in Study 1, which produced an average sentence of 9.7 years. The materials provided the judges with the following information: "Assume that your court calendar this morning includes the sentencing of two defendants: Jones and Smith. Both cases arise from separate incidents, but both have pled guilty." Half of the judges reviewed the less serious crime first, and the other half reviewed the more serious crime first.

176. *Id.*

177. *Id.*

178. See Keren Weinshall-Margel & John Shapard, *Overlooked Factors in the Analysis of Parole Decision*, 108 PROC. NAT'L ACAD. SCI. E833, E833 (2011) (arguing "[t]he phenomenon of favorable decisions peaking after a meal break is likely an artifact of the order of case presentation" and contending that such breaks are not random).

179. These judges were nearing the completion of a training course for newly assigned trial judges, and hence were inexperienced judges. As part of their training, they were required to participate in a session identified as reviewing "judicial decision making." For these judges, completing the survey was compulsory (and indeed, the response rate was 100%), although each judge was afforded the option of withdrawing his or her survey from further analysis (anonymously). The survey was presented at both the 2010 and 2011 sessions of this conference, and the results were aggregated for analysis.

180. The Arizona Judicial Conference is an annual conference attended by most of the judges in Arizona. The data come from a breakout session entitled, "How Judges Think," which was optional for the judges attending the conference.

181. This was a day-long conference sponsored by the Dutch Judicial Council and was optional for the judges. The conference was entitled, "Judgment Day," and the materials for this conference were translated into Dutch.

a. Military Materials

Although we used the same fact pattern for the severe crime in all three groups, the less severe crime varied to suit the jurisdiction. For the military judges, the less severe crime consisted of threatening with an unloaded weapon. The facts were as follows:

Jones pled guilty to simple assault with an unloaded firearm; as part of an agreement, he would not be prosecuted for any other crime related to this incident. The act occurred when Jones got into a loud argument with another service member while on the base at which they were both stationed. The source of the argument is unclear, but may have involved an accusation by Jones that the service member owed him some money. Jones walked away from the argument only to return a few minutes later with a service pistol. He pointed the pistol at the service member and threatened to shoot him if he did not pay. At a moment of inattention, bystanders wrestled Jones to the ground and removed the pistol, which was not loaded. Jones asserts that he checked to ensure that the weapon was not loaded before he pointed it at the service member.

Jones is 19 and has served for 6 months. He has not had any other disciplinary problems. His service record is otherwise unremarkable to date. He has agreed to accept a dishonorable discharge, and the only remaining issue is the length of confinement. Simple assault with an unloaded weapon carries a maximum sentence of 3 years of confinement.

For the military judges, the more severe crime remained involuntary manslaughter and involved another service member. The materials also indicated that this defendant had pled guilty and agreed to a dishonorable discharge. Involuntary manslaughter carries a maximum sentence of 15 years in the military justice system. The materials provided the following background about the defendant:

Smith is 23 years old and has served for over three years. He has been reprimanded once in the past for an incident arising from a bar fight while he was intoxicated. His service record is otherwise clean, but unremarkable. He served for 18 months in Afghanistan with his unit.

Sentencing in a military court-martial is not conducted entirely by the trial judge.¹⁸² Both the jury and the judge independently recommend a sentence. The judge then recommends the less severe of the two sentences to the general in command of the unit, who may depart from the recommendation either upward or downward. In the case of the military judges, we were thus studying an important sentence recommendation, rather than the sentence actually imposed.

We did not ask the military judges for political affiliation. We did not ask for their experience as judges either because all were new to the position; but all had significant experience in the military.¹⁸³ We asked their gender, but because only 17

182. See *MANUAL FOR COURTS-MARTIAL UNITED STATES*, A21, at 72 (2012) (“Sentencing in courts-martial may be by the military judge or members.”).

183. The judges had an average of 18.7 years of experience in the military (the median was

of the judges (24%) were female, we did not analyze gender differences for these judges. A total of 71 military judges completed the sentencing questions:¹⁸⁴ 35 in 2010 and 40 in 2011.

b. Arizona Materials

The materials for the Arizona judges were similar to those used with the military judges, although the less severe crime differed slightly. Because the analogous crime in Arizona is punished much more severely, we had to eliminate the use of the firearm and substitute conduct that would still draw a similar range of sentences.¹⁸⁵ The facts we used are described below:

Jones pled guilty to aggravated assault under A.R.S. § 13-1204(A)(5) (assault that occurs in the private home of the victim). The incident occurred when Jones got into a loud argument with a coworker while the two were leaving their workplace (an office) at the end of the day. The source of the argument involved an accusation by Jones that his coworker owed him some money. Jones apparently followed the coworker home, waited until the coworker unlocked the door, and pushed him inside. Jones, who is several inches taller than his coworker, shoved the victim around inside his living room and threatened to “beat him badly” if he did not pay. A neighbor overheard the shouting, and called the police, who arrived quickly and arrested Jones. Jones asserts that he did not intend to hurt his victim, and only wanted the money he was owed. Nevertheless, the victim suffered a mild concussion.

We also varied the defendant’s background slightly to avoid references to his military record:

Jones is 19, unmarried, and has no previous arrests. He was employed as a full-time administrative assistant until the incident and was living on his own in an apartment. His parents both testified at his hearing that he had never been in trouble before. His mother stated that she felt that Jones was worried about money because his company had been threatening to close his office and lay everyone off.

In Arizona, the judge assigned to the case imposes the sentence. Arizona, however, has somewhat more elaborate substantive restrictions in the form of limited maximum and minimum sentences and a presumptive sentence. In Arizona, the

18 years).

184. Due to an error in creating the surveys in 2011, 4 judges sentenced only the manslaughter defendant. These judges were kept in the analysis among judges who sentenced the manslaughter defendant first, but they provided no data on the threat.

185. The same crime in Arizona would constitute aggravated assault with a deadly weapon, which is a class 3 felony. ARIZ. REV. STAT. ANN. § 13-1204(A)(2) (2010). It carries a presumptive sentence of 3.5 years. ARIZ. REV. STAT. ANN. § 13-702(D) (2010). In light of this fact, we changed the crime to an assault that occurs in the private home of the victim. § 13-1204(A)(5).

crime at issue is a class 6 felony.¹⁸⁶ As the materials noted, “[f]or first time-offenders with no aggravating or mitigating circumstances, it carries a minimum sentence of 0.5 years, a maximum of 1.5 years, and a presumptive sentence of 1 year.”¹⁸⁷

The circumstances of the voluntary manslaughter were identical to those in Study 1, above. In Arizona, this crime would be a class 2 felony,¹⁸⁸ which, for this defendant’s background, would carry a minimum sentence of 7 years, a maximum of 21 years, and a presumptive sentence of 10.5 years.¹⁸⁹

A total of 39 Arizona judges completed the survey. These judges had an average of 11.5 years of experience as judges (a median of 9.5 years), ranging from 0 to 35 years (only 37 reported their experience). Of the 38 who reported their gender, 15 (39%) were female. Of the 37 judges who reported their political affiliation, 18 (46%) identified themselves as Republicans and 19 (54%) as Democrats.¹⁹⁰ The group included 8 appellate judges (out of 38 who responded to the demographic page—we assumed the individual who did not complete the demographic questions was also a judge), who were included in the analysis.

c. Dutch Materials

The materials for the Dutch judges were identical to those of the Arizona judges in terms of their facts. We changed the names of the defendants to Jasperen and Smulders instead of Jones and Smith, respectively. We also anticipated that many of the Dutch judges would insist upon a suspended sentence for the lesser crime, so we offered them that option. For both the less severe and more severe crimes, the Dutch judges could impose a suspended sentence, a sentence to be served, or both.

A total of 62 Dutch judges completed the questions on sentencing.¹⁹¹ These judges had an average of 13.2 years of experience as judges (a median of 12 years), ranging from just under 1 year to 27 years. Of the 62 who reported their gender, 23 (39%) were female. We did not ask the Dutch judges questions concerning their political affiliation. The sample included 17 appellate judges (including 3 judges from the High Court).

2. Results

a. Military Judges

The order of sentence influenced the military judges. Table 3a, below, presents the average sentence in each condition. Judges who sentenced the manslaughter

186. § 13-1204(D) (“Aggravated assault pursuant to subsection A, paragraph . . . 5 . . . is a class 6 felony.”).

187. See § 13-702 (D).

188. ARIZ. REV. STAT. ANN. § 13-1103(C) (2010).

189. See § 13-702 (D).

190. Twenty-three judges identified themselves as independent, unaffiliated, or simply did not respond to our question about political affiliation: “Which of the two major political parties in the United States most closely matches your own political beliefs?”

191. Sixty-three judges were in attendance at the conference; only 1 did not complete the sentencing questions.

case first provided longer sentences than the judges who sentenced the threat first (1.61 years versus 1.15 years). Likewise, judges who sentenced the threat first provided shorter sentences than judges who sentenced the manslaughter first (6.65 years versus 8.52 years). Both effects were statistically significant.¹⁹² The result is consistent with an anchoring effect. The first sentence effectively acted as an anchor for the judges, even though the cases were clearly labeled as independent crimes.

Table 3a. Mean sentence by condition among military judges

Crime	First crime: mean sentence (and <i>n</i>)		Test of difference	
	Threat	Manslaughter	<i>t</i> -statistic	<i>p</i> -value
Threat	1.15 (31)	1.61 (40)	2.65	0.01
Manslaughter	6.65 (31)	8.52 (44)	2.46	0.02

b. Arizona Judges

Sentencing order influenced the Arizona judges, albeit in a more limited way. Table 3b, below, presents the average sentence in each condition. As was the case with the military judges, Arizona judges who sentenced the manslaughter case first provided longer sentences than the judges who sentenced the threat first (0.82 years versus 0.56 years); this difference was statistically significant. Assigning the lengthy sentence to the manslaughter defendant seemed to anchor the judges, biasing their sentences for the less severe crime. We did not observe a similar effect on the manslaughter sentences, however. Judges provided virtually identical sentences for manslaughter, regardless of which defendant they sentenced first. Owing to the relatively small sample size, we did not assess the influence of the demographic variables.

Table 3b. Mean sentence by condition among Arizona judges

Crime	First crime: mean sentence (and <i>n</i>)		Test of difference	
	Battery	Manslaughter	<i>t</i> -statistic	<i>p</i> -value
Battery	0.56 (18)	0.82 (21)	2.13	0.04
Manslaughter	9.06 (18)	9.12 (21)	0.12	0.91

c. Dutch Judges

Table 3c, below, presents the average sentence in each condition. Despite differences between the two types of judges and the jurisdictions, the results are comparable to those of the Arizona judges. The order influenced the sentences judges assigned to the less serious crime but not to the more serious crime.

192. We report the statistical test results in the Tables for this study.

Table 3c. Mean sentence by condition among Dutch judges

Crime	Sentence type (and <i>n</i>)	Condition		Test of difference	
		Battery (1st)	Manslaughter (1st)	<i>t</i> - statistic	<i>p</i> - value
Battery	Unconditional (62)	0.05	0.24	1.98	0.05*
	Conditional (62)	0.09	0.52	3.20	0.002*
	Total (62)	0.14	0.76	3.71	0.0004*
Manslaughter	Unconditional (62)	5.13	5.43	0.49	0.62
	Conditional (62)	0.66	1.03	0.96	0.34
	Total (62)	5.79	6.46	1.05	0.31

*Statistically significant

3. Discussion

Sentence order influenced the judges. Even though the cases clearly had nothing to do with each other, sentencing one defendant influenced the sentencing of the next. Sentencing a defendant to roughly 1 year in prison made a lengthy sentence seem excessive to the judges, even though the second case was a serious one. Additionally, sentencing a defendant to a lengthy sentence made a short sentence seem too short by comparison. We saw no evidence of the contrast effect. Instead, anchoring prevailed.

The effect was more pronounced on the less serious case than the more serious one. In all three versions, sentencing the more serious crime first raised the sentence of the less serious crime, but the more serious crime was unaffected in two of the three variations. We are not certain why this occurred. Possibly, the judges felt that they had more discretion to raise the somewhat short sentence for the less serious crime than they had to lower the sentence for the more serious crime.¹⁹³

Because the two crimes are independent, it is hard to see how they could have provided relevant information about each other. Perhaps judges assumed that, despite our assertion to the contrary, crimes tend to be batched together in terms of severity. We think not, however. Judges are certainly used to the idea that different unrelated cases appear before them in the same session. The most plausible account of the results is anchoring.

D. Study 4: Debiasing

In experiment after experiment, we find that judges rely on numeric anchors, even when doing so is normatively indefensible. In some studies, one might be able to identify an appropriate reason to rely on the numeric reference point, but the weight of evidence across the preceding studies (as well as the experiments on anchoring in

193. See also Bushway et al., *supra* note 98, at 298 (suggesting that anchoring might be more effective at increasing sentences than decreasing sentences).

judges that we have previously reported elsewhere) suggests that anchors distort judges' judgment.¹⁹⁴ In particular, the pattern of results is most consistent with the account of anchoring developed by Frederick and Mochon.¹⁹⁵ That is, numeric reference points distort judges' sense of scale. As an example, less serious injuries do not necessarily seem more serious when judges learn of (relatively) high damage caps, but their sense of where the injury fits on the scale is affected.

This account of anchoring suggests a mechanism for producing judgments that are less heavily affected by anchoring. If anchoring distorts judges' sense of scale, then reliable judgments depend upon establishing a reliable sense of scale. In civil cases, giving judges a reliable sense of how to convert injuries into dollars might ameliorate the influence of anchors.

1. Methods and Materials

To determine whether providing a reliable scale can thwart the influence of anchors, we asked 242 judges attending a statewide judicial conference in Ohio to assess a case involving a serious personal injury. These judges had an average of 14.2 years of experience, ranging from newly appointed to 42 years of judicial service, with a median of 14 years. Of the 239 judges who informed us of their gender, 20% were female. Of the 224 judges who answered our question concerning political orientation, 62% were Republicans and 38% were Democrats.

We constructed six conditions to induce an anchoring effect and then inoculate the judges against its influence. One-third of the judges evaluated a scenario in which we requested a damage award but provided no anchor; one-third of the judges evaluated the same scenario with a low anchor; and one-third of the judges evaluated the same scenario with a high anchor. We also created a variation in which judges reviewed a discussion of a newspaper article describing three reasonable damage awards in other cases. We wanted to test whether reviewing this article would serve as a kind of inoculant that would reduce the influence of the anchors. Thus, half of the judges also reviewed this inoculant, thereby creating a 3 x 2 between-subjects design (low anchor, no anchor, or high anchor crossed with either no inoculant or inoculant).

The injury was one we used in our original experiment on anchoring with federal magistrate judges. We described the plaintiff as a 31-year-old school teacher and the defendant as a large package-delivery service. The materials indicated that the defendant's employee had been driving one of its trucks erratically and had sideswiped the plaintiff. The defendant had admitted liability, and the parties had settled economic claims and medical expenses, leaving only the noneconomic damages unresolved. The materials described the injury as follows:

As a result of the accident, the plaintiff broke three ribs and severely injured his right arm. He spent a week in the hospital and missed six

194. See, e.g., Guthrie et al., *supra* note 6, at 790–94; Wistrich et al., *supra* note 82, at 1291–93.

195. Frederick & Mochon, *supra* note 61, at 124 (“[O]ur theory of scale distortion suggests that anchoring effects may also occur because of a shift in the use of the response scale itself and not because of any deeper change in the underlying representation of the target being judged.”).

weeks of work. The injuries to his right arm were so severe as to require amputation. (He was right-handed.)

.....
The evidence presented at trial included testimony from the plaintiff, a young father, that he could no longer play recreational softball, or even play catch with his son. Although the plaintiff has continued teaching, he testified that doing his job is somewhat more difficult, and that he is subject to periodic ridicule by the students. Plaintiff also described the severe pain he endured before arriving at the hospital, during the surgery to amputate his arm, and during his post-surgical therapy.

The materials created variations for the anchoring conditions by using the manipulation we used in a previous study involving administrative law judges.¹⁹⁶ That is, we inserted a statement by the plaintiff into the materials. For the low-anchor condition, this statement read as follows: "During his testimony, the plaintiff also mentioned as an aside that he recently saw a case on a 'court television show' where 'a victim like me' received a '\$150,000 award.' You sustained the defendant's objection to this comment as irrelevant." The high-anchor condition was identical, except that it stated the amount as \$10 million. The no-anchor condition included the same statement, but omitted the dollar amount. For all judges, the materials ended by asking: "How much would you award the plaintiff in this case for pain and suffering and loss of his right arm?"

The inoculant consisted of the following additional paragraph inserted after the reference to the plaintiff's testimony and before the call of the question:

The case reminds you of a recent legal newspaper article discussing damage awards upheld on appeal in personal injury cases in Ohio, including: a \$214,000 award for a plaintiff who had lost a thumb in a railroad accident; a \$1,050,000 award for a plaintiff who had suffered crippling back pain and an amputated leg from a car accident; and an \$11,000,000 award for a 24 year-old-plaintiff whose injuries made him a quadriplegic with severe brain damage.

We hypothesized that the inoculant would negate the effect of the anchoring by giving the judges a reasonable sense of how a variety of personal injury cases with varying awards might translate into dollar amounts.

2. Results

Of the 242 judges, 226 provided a damage award.¹⁹⁷ Table 4, below, reports the results.

196. Guthrie et al., *supra* note 78, at 1502-03.

197. The missing data were distributed fairly evenly through all six cells of the 3 x 2 design: 3 in the no-inoculant, low-anchor condition; 4 in the no-inoculant, no-anchor condition; 3 in the no-inoculant, high-anchor condition; 2 in the inoculant, low-anchor condition; 3 in the inoculant, no-anchor condition; and 1 in the inoculant, high-anchor condition.

Table 4. Mean, 25th percentile, median, and 75th percentile of awards by condition

Inoculant?	Anchor (and <i>n</i>)	Mean	25th percentile	Median	75th percentile
No inoculant	Low (35)	741	225	500	750
	No (36)	1278	325	750	1250
	High (34)	1483	500	1000	2000
Inoculant	Low (42)	1079	500	750	1000
	No (37)	1828	750	1000	2500
	High (42)	1841	500	1000	2000

Statistical tests to assess these effects were made challenging by the highly skewed nature of the awards. Although a square root transformation produced a distribution that more closely approximated a normal distribution, it was still somewhat skewed. We thus conducted both an analysis of the medians, which is robust to outliers and skewed data, and an analysis of the square root of the awards. In both cases, we regressed awards against a dummy variable for the low-anchor condition; a dummy variable for the high-anchor condition; a dummy variable for the inoculant conditions; and two interaction terms between the variables for the low anchor, high anchor, and inoculant. We report the regression results in Table 5, below.

Table 5. Regression results for median quantile regression and linear regression on the square root of damage awards (in thousands)

Variable	Model 1: Median regression			Model 2: Square root		
	coefficient	<i>t</i> - statistic	<i>p</i> - value	coefficient	<i>t</i> - statistic	<i>p</i> - value
Constant	750	—	—	976	—	—
Low anchor	-250	3.26	0.001	-263	1.97	0.05
High anchor	250	3.33	0.001	111	0.83	0.41
Inoculant	250	3.21	0.002	235	1.79	0.09
Low x inoculant	0	0.00	1.000	-17	0.09	0.93
High x inoculant	-250	2.40	0.020	-129	0.70	0.49

The analysis produced a complex pattern of results. The medians shifted in ways that we expected. That is, exposure to the low anchor lowered the median by \$250,000, and exposure to the high anchor raised the median by \$250,000. The inoculant had no effect on the influence of the low anchor, but it reduced the influence of the high anchor. At the same time, the inoculant increased the median damage award. The transformed data, however, show a somewhat different pattern. These results suggest that the low anchor influenced the judges' awards, but the high anchor did not. A nonparametric analysis of the comparison between the no-anchor and the high-anchor condition did not produce a significant result.¹⁹⁸ The inoculant

198. Using the Mann-Whitney test, $p = 0.37$.

increased awards overall and did not eliminate the influence of the low anchor. Because there was no significant effect of the high anchor, the inoculant could not eliminate the influence of the high anchor.

Female judges were more generous than male judges. They provided an overall median award of \$1,000,000, as compared to a median of \$750,000 among male judges. This was a marginally statistically significant difference.¹⁹⁹ Democratic judges provided a median award of \$1,000,000, as compared to a median of \$750,000 among Republican judges, although this difference was not statistically significant.²⁰⁰ Years of experience did not significantly correlate with award size.²⁰¹ None of the demographic variables interacted significantly with the influence of the anchor.²⁰²

3. Discussion

The results suggest that judges find it challenging to convert injuries into awards without being influenced by anchoring. Our effort to reduce anchoring by providing an inoculant was, at best, only partly successful. The inoculant had no apparent effect on the low anchor, although it produced median awards that were identical in the high-anchor and no-anchor conditions. The analysis of the transformed data and the nonparametric analysis, however, suggest that the high anchor had little or no effect on the overall distribution of awards (even though the median award was higher in the high-anchor condition). Furthermore, the inoculant did not eliminate the effect of the low anchor. The inoculant also had the unintended effect of increasing the size of awards. Therefore, although exposing judges to sensible numeric reference points might aid their judgment and diminish the distortion of anchoring, such an approach would have to be more carefully tailored and might have unintended consequences.

Some have suggested that case-specific anchors may influence judges less than jurors because judges may possess contextual or comparative information about damage awards that jurors lack.²⁰³ Assuming that such a disparity in non-case-specific information exists, it may not be very helpful. First, our experiments as a whole suggest that despite their arguably broader exposure to a range of non-case-specific damage awards, judges are nevertheless powerfully influenced by case-specific anchors. Second, the results of Study 4 indicate that contextual or comparative information can itself distort damage awards. The ability of non-case-specific information to reduce distortions caused by anchoring may depend on the quality of that information and the manner in which it is presented.

199. Using the Mann-Whitney test, $p = 0.06$.

200. Using the Mann-Whitney test, $p = 0.15$.

201. $r = .07$ (using the square root of the awards), $t(223) = 1.09$, $p = 0.27$.

202. This was assessed by regressing the square root of the awards on the dummy variables denoting the low- and high-anchor conditions, the three demographic variables (gender, party, and years of experience), and the interaction terms. All t statistics for the coefficients for the interaction terms were less than 1.0, and all p -values were greater than 0.25.

203. Hans & Reyna, *supra* note 96, at 144 (“[A]nchors presented during trial should have a stronger impact on jury assessments than on judge assessments because judges are apt to have contextual and comparative information about other cases that may counteract or outweigh the impact of anchors presented during trial.”).

IV. IMPLICATIONS

Anchoring is a heuristic that profoundly distorts numerical estimates in a variety of settings. As this study—especially taken together with our previous experiments on the topic—confirms, American, Canadian, and Dutch judges are not immune. Although some of the anchors in these studies arguably had some relevance to the underlying case, even hopelessly arbitrary anchors affected the decisions of the judges we studied. Without remediation, damage awards and criminal sentences might.

Remediating the untoward influences of anchors is not an easy task. We tested a plausible strategy for overcoming anchoring by embedding an anchor among a set of other potential anchors, but the result was not entirely effective. Other studies suggest that merely warning people about anchors also does little to avoid their effects.²⁰⁴ What, then, might be done to minimize or eliminate the impact of irrelevant anchors?

In previous articles, we suggested steps that might be taken to minimize the distortion caused by anchoring. These included attempting to train judges to avoid the impact of anchoring,²⁰⁵ prohibiting litigants from mentioning numbers that might operate as anchors (such as a damage cap or plaintiff's ad damnum),²⁰⁶ separating decision-making functions,²⁰⁷ requiring explanations for the amount of damages awarded or the sentence imposed,²⁰⁸ relying on aggregated data,²⁰⁹ and cabining discretion with sentencing guidelines²¹⁰ and damage schedules.²¹¹

Unless further research uncovers a better inoculant strategy than the one we used in this study, minimizing the effect of anchors requires either (1) avoiding them, (2) facilitating exposure to meaningful anchors, or (3) restricting judges' discretion in imposing sentences or damages.

Prohibiting the mention of figures that might serve as anchors seems a promising strategy. Some jurisdictions have adopted it.²¹² A separation of decision-making functions serves the same purpose indirectly by ensuring that if an anchor is mentioned to a judge acting as a case manager, settlement officer, or evidence gatekeeper, a

204. Guthrie et al., *supra* note 78, at 1505–06.

205. Guthrie et al., *supra* note 32, at 38–40.

206. See Kang, *supra* note 149, at 493; see also 43 U.S.C. § 1981a(c)(2) (2012) (prohibiting mention of damage caps to juries in employment discrimination cases). *But see* Rebecca Hollander-Blumoff & Matthew T. Brodie, *The Effects of Jury Ignorance About Damage Caps: The Case of the 1991 Civil Rights Act*, 90 IOWA L. REV. 1361, 1404 (2005) (arguing that “concealing the cap has the potential to undermine the integrity and legitimacy of the jury system” and advocating as an alternative remedy “provid[ing] juries with more complete knowledge of caps and their context”).

207. Guthrie et al., *supra* note 32 at 42; Wistrich et al., *supra* note 82, at 1325–27.

208. Guthrie et al., *supra* note 32, at 36–38; Guthrie et al., *supra* note 78, at 1501, 1504–06.

209. Guthrie et al., *supra* note 6, at 823.

210. *Id.* at 794.

211. *Id.*

212. See, e.g., *Reese v. Hersey*, 29 A. 907, 908 (Pa. 1894); *Porter v. Zenger Milk Co.*, 7 A.2d 77, 78 (Pa. Super. Ct. 1939); see also *EEOC v. Wal-Mart Stores, Inc.*, 276 F.R.D. 637, 639–40 (E.D. Wash. 2011) (“[I]f Plaintiff intends to suggest a specific amount to the jury for emotional distress damages, yet fails to supplement its Rule 26 disclosures to provide Defendant with a computation of damages, Plaintiff may be foreclosed from suggesting that specific amount for emotional distress damages to the jury at trial.”).

different judge or a jury acting as the fact finder will be shielded from exposure to it. Some of the experiments reported in this Article (particularly Study 3), however, suggest that this approach may be too limited. Anchors can arise from contextual sources perhaps too numerous and varied to be effectively controlled in this way.²¹³

Mechanisms to facilitate exposure to meaningful anchors might be promising. The legal system includes numerous meaningful, sensible anchors. As the research demonstrates, for example, recommendations from prosecutors and defense attorneys influence judges.²¹⁴ If these actors make sensible recommendations, then the judges will find them to be stabilizing influences.²¹⁵ Although parties might abuse the insights in this Article to make outrageous requests,²¹⁶ lawyers' suggestions as to sentences or awards might make for more reliable awards and sentences when both parties suggest them. Competing anchors are apt to stabilize awards and sentences.²¹⁷ Furthermore, knowing that an adversary will suggest an award or sentence might discipline each party into making reasonable suggestions. In civil cases, judges could facilitate this by obtaining competing suggestions in the context of settlement discussions and permitting bifurcation to allow a defense attorney to suggest an anchor without admitting liability.²¹⁸

Controlling discretion with sentencing guidelines and damage schedules also can minimize anchoring effects. The former are common,²¹⁹ and the latter are gradually gaining acceptance.²²⁰ Guidelines and schedules are not likely to eliminate the

213. See Critcher & Gilovich, *supra* note 45, at 248 (reporting that "incidental numbers present in the environment influenced participants' estimates of uncertain values").

214. See English & Mussweiler, *supra* note 94, at 1547–49.

215. Lawyers, however, are not immune to the influence of anchoring either; actors might anchor on each other. One study suggests that defense attorneys who make their recommendations after prosecutors might adjust their recommendations toward those offered by the prosecutor. See generally English et al., *supra* note 166.

216. See Don Rushing, Linda Lane & Erin Bosman, *Anchors Away: Attacking Dollar Suggestions for Non-Economic Damages in Closings*, 70 DEF. COUNSEL J. 378, 379–81 (2003).

217. Studies in which multiple anchors are available suggest that counteranchors might be effective. See generally Glen Whyte & James K. Sebenius, *The Effect of Multiple Anchors on Anchoring in Individual and Group Judgment*, 69 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 75 (1997). One study in a legal setting in which researchers used competing anchors, unfortunately, did not compare versions in which the defendant did not offer an anchor to ones in which the defendant did offer one. Malouff & Schutte, *supra* note 69, at 493–96. A more recent study suggests that defense attorneys can blunt the anchoring effect of a plaintiff's demand by offering a counteranchor, although doing so might also increase the defendant's chances of being found liable. John Campbell, Bernard Chao, Christopher Robertson & David Yokum, *Countering the Plaintiff's Anchor: Jury Simulations To Evaluate Damages Arguments* (Arizona Legal Studies, Discussion Paper No. 14-25, 2014), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2470066.

218. See Randy Wilson, *Do You Suggest a Number?*, ADVOCATE, Winter 2011, at 107, 107 ("The defense attorney is in . . . a quandary. To suggest . . . anything might be perceived as a concession of liability.").

219. See generally NAT'L CTR. FOR STATE COURTS, *supra* note 130.

220. See, e.g., *Arpin v. United States*, 521 F.3d 769, 776 (7th Cir. 2008) (reversing an award of damages for loss of consortium and directing the trial court to determine the average ratio of loss-of-consortium damages because the trial court failed to consider awards in similar cases); *Geressy v. Digital Equip. Corp.*, 980 F. Supp. 640, 635, 654–62 (E.D.N.Y. 1997)

distortion caused by anchoring, but these efforts will reduce the variability of these judgments by constraining their range.²²¹ Such guidelines themselves might also act as anchors, although judges could minimize their influence by first assessing qualitative factors and only consulting the guidelines after identifying a tentative sentence.²²² Many such existing devices, however, may leave too much room for discretion and hence for distortion. There is, for example, a continuum of strictness along which the states employing sentencing guidelines may be arrayed.²²³ Those that are most strict may still permit some individualization and thereby also permit some distortion but may confine the latter within acceptable bounds.²²⁴ States with more relaxed guidelines, on the other hand, might tolerate too much distortion. The post-*Booker*²²⁵ U.S. Sentencing Guidelines may fall nearer the relaxed end of the continuum than those of some states, such as Minnesota.²²⁶ A similar continuum exists with respect to damage schedules.²²⁷

(determining whether a damage award “deviated materially from what would be reasonable compensation” by comparing it to previous awards in similar cases (citation omitted)). See also Oscar G. Chase, *Helping Jurors Determine Pain and Suffering Awards*, 23 HOFSTRA L. REV. 763, 775–76 (1995) (suggesting that jurors be provided with a chart summarizing damage awards in cases involving similarly situated plaintiffs); Shari Seidman Diamond, Michael J. Saks & Stephan Landsman, *Juror Judgments About Liability and Damages: Sources of Variability and Ways To Increase Consistency*, 48 DEPAUL L. REV. 301, 321–22 (1998) (advocating that attorneys be permitted, albeit with judicial supervision, to present to the jury a set of damage awards made by previous juries in similar cases); Roselle L. Wissler, Allen J. Hart & Michael J. Saks, *Decisionmaking About General Damages: A Comparison of Jurors, Judges, and Lawyers*, 98 MICH. L. REV. 751, 817 (1999) (recommending that parties or the court “pool jury awards made for similar injuries, and . . . present these cases and their award distributions to juries for guidance in reaching their general damages awards”).

221. See Rachlinski et al., *supra* note 86, at 1236–37.

222. Judge Mark Bennett has suggested this procedure. See Bennett, *supra* note 74, at 530 (“I suggest that the sentencing judge should review and study the information in a PSR’s non-Guidelines § 3553(a) first.”).

223. See generally NAT’L CTR. FOR STATE COURTS, *supra* note 130.

224. See Sanders, *supra* note 142, at 149 (“Here we face a tradeoff. The proposals that are likely to have the greatest impact in reducing horizontal inequity are the ones that most clearly limit jury discretion, and those that leave the most discretion in the hands of jurors are likely to leave the most residual inequities.”).

225. *United States v. Booker*, 543 U.S. 220, 245 (2005) (concluding that the U.S. Sentencing Guidelines are “effectively advisory”).

226. See generally Mustard, *supra* note 19 (noting that race and gender disparities persist despite sentencing guidelines); Lydia Brashear Tiede, *Disparity in Federal District Court Sentencing*, LAW & CTS. (Am. Political Sci. Ass’n, Columbia, S.C.), Summer 2012, at 29, 32 (“[I]n several of the circuits, judges chose to sentence within the Guideline ranges over 70% of the time, while in other circuits . . . judges stayed within the Guideline ranges for only 35% of the cases. Further, in several other circuits . . . judges stayed within the federally mandated Guidelines less than 50% of the time.”).

227. See Sanders, *supra* note 142, at 145 (distinguishing between “proposals . . . that would restrict juries to a range of general damage amounts and those that only wish to give juries guidance without any real constraint on their ultimate decision”); see also Joseph Sanders, *Why Do Proposals Designed To Control Variability in General Damages (Generally) Fall on Deaf Ears? (And Why This Is Too Bad)*, 55 DEPAUL L. REV. 489, 496–507 (2006) (summarizing various

CONCLUSION

One of the defining characteristics of the justice system is that its outputs are frequently expressed in numbers. The amount of bail to be posted to obtain pretrial release, the length of a sentence to be served, the amount of compensatory or exemplary damages to be paid—all require quantification. Sometimes determining the right number is easy, such as when there is a fixed bail schedule, a precisely prescribed sentence, a liquidated damages clause, or a claim for medical expenses that can be calculated by simply tallying the doctors' bills. In an appreciable percentage of cases, however, quantification is necessary, but not so simple. This is the realm—the amount of damages for pain and suffering, the duration of discretionary sentences, and so on—in which anchors can and, as we have shown, do have a distorting, and sometimes dramatic, impact. Because anchoring is a powerful phenomenon, reducing its influence will not be easy. Although we have experimented with debiasing techniques, we have yet to find one that is unqualifiedly successful.

As we have previously suggested, the only reliable solution might be to confine judicial discretion with relatively tight sentencing guidelines and workers' compensation-like damages schedules.²²⁸ Of course, adopting such solutions entails costs. The solutions would reduce individualization²²⁹ and could prove dispiriting to judges who might feel that they have been relegated to serving as calculating machines. Is defeating anchoring worth the costs of adopting such measures? That is a question policymakers should consider if they want to avoid arbitrariness in judges' criminal sentences and civil damage awards.

scheduling proposals); Studdert et al., *supra* note 7, at 81 tbl.1 (summarizing the strengths and weaknesses of various options for scheduling noneconomic damages).

228. See Wistrich et al., *supra* note 82, at 1328–29.

229. See Richard Abel, *General Damages Are Incoherent, Incalculable, Incommensurable, and Inegalitarian (But Otherwise a Great Idea)*, 55 DEPAUL L. REV. 253, 296 (2006) (“All schedules . . . impose slot-machine justice, sacrificing nuance and individualization to convenience.”); Thomas C. Galligan, Jr., *The Tragedy in Torts*, 5 CORNELL J.L. & PUB. POL’Y 139, 172 (1996) (“It should be noted that recent proposals to provide jurors with guidelines or schedules for general damages are inconsistent with traditional rules concerning general damages. The proposals would abstract away from the particular plaintiff.” (footnote omitted)). *But see* Murray B. Rutherford, Jack L. Knetsch & Thomas C. Brown, *Assessing Environmental Losses: Judgments of Importance and Damage Schedules*, 22 HARV. ENVTL. L. REV. 51, 101 (1998) (“[A] damage schedule would necessarily be somewhat arbitrary. However . . . it might in practice be less arbitrary and more equitable than ad hoc measurement.”); Leandro M. Zanitelli, *Determining Pain-and-Suffering Awards Accurately: General or Case-by-Case Law?*, 28 QUINNIPIAC L. REV. 183, 214 (2009) (concluding that “no greater accuracy is to be expected from case-by-case damage awarding as compared to general rules”).