

THE INTERSECTIONALITY OF AGE AND GENDER ON THE BENCH: ARE YOUNGER FEMALE JUDGES HARsher WITH SERIOUS CRIMES?

MORRIS B. HOFFMAN,^a FRANCIS X. SHEN,^b VIJETH IYENGAR^c & FRANK KRUEGER^{d*}

Abstract

We analyzed sentencing data from sixteen years of criminal trials in the State of Colorado, consisting of almost 3,000 individual sentences, and discovered an interaction effect of harm, gender, and age not reported in any of the empirical or experimental literature. Young female judges punished high harm crimes substantially more than their male and older female colleagues. These results, if confirmed, could have significant strategic and tactical implications for practicing lawyers. They may also inform policies surrounding judicial selection, education, training, and retirement.

^a District Judge, Second Judicial District (Denver), State of Colorado; Adjunct Professor of Law, University of Denver; Member, John D. and Catherine T. MacArthur Foundation Research Network on Law and Neuroscience; Research Fellow, Gruter Institute for Law and Behavioral Research.

^b Professor of Law and Faculty Member in the Graduate Program on Neuroscience, University of Minnesota; Instructor in Psychology, Harvard Medical School, MGH Department of Psychiatry; Executive Director, Massachusetts General Hospital Center for Law, Brain, and Behavior; Executive Director of Education and Outreach for the John D. and Catherine T. MacArthur Foundation Research Network on Law and Neuroscience.

^c Brain Health Lead and Technical Advisor to the Deputy Assistant Secretary for Aging, U.S. Administration for Community Living, U.S. Department of Health and Human Services; M.A., Ph.D. (Department of Psychology and Neuroscience), Duke University. The views expressed in this paper are those solely of the author and not necessarily of the U.S. Administration for Community Living, the U.S. Department of Health and Human Services, or any other affiliated organizations of the author.

^d Associate Professor of Systems Social Neuroscience and Faculty Member in the Interdisciplinary Program of Neuroscience, Chief of Social Cognition & Interaction: Functional Imaging (SCI:FI) Laboratory, School of Systems Biology, George Mason University.

* *Acknowledgments:* We thank Warren Cormack, Jennifer Novo, Caitlin Opperman, Mia Kontnik, Tim Leutkemeyer, Abigail Perkins, and Campbell Tow for their research assistance and thoughtful comments. A special thanks to Kristin Wood, District Administrator for the Second Judicial District, for her invaluable assistance in helping us collect the data. The views expressed in this paper are those of the authors and not necessarily of the authors' affiliated organizations.

INTRODUCTION

In 2019, the United States Senate confirmed the nomination of Judge Allison Jones Rushing to the United States Court of Appeals for the Fourth Circuit. Among Rushing's many impressive credentials, including graduating from a top law school and clerking for a Supreme Court Justice, one demographic fact stood out: She was only thirty-six years old. She is the nation's youngest federal judge,¹ joining a group of other young, female appointees such as Holly Teeter (age thirty-nine; District of Kansas), Rebecca Jennings (age forty-one; Western District of Kentucky), Emily Marks (age forty-six; Middle District of Alabama), and Jill Otake (age forty-five; District of Hawaii).

Both major political parties have sought to nominate and appoint younger candidates to life-tenured federal judgeships, no doubt in part to maximize their lingering influence on the federal judiciary when those political parties might no longer be in power.² Despite this trend, almost no academic literature exists examining whether young judges differ from their older peers on the bench. In particular, no one has looked at how age and gender interact in the context of criminal sentencing. This is a significant oversight, as criminal sentencing is a core function of the judiciary. In the words of federal judge Jack Weinstein, criminal sentencing is "perhaps the most difficult task of a trial court judge."³ Will trial judges in their thirties sentence the same thirty years later, when they will likely still be on the bench?

In this Article, we begin to answer this question by presenting the first study to examine longitudinally the interactive effects of judicial age, gender, and crime level in determining sentences in actual criminal cases. After two years of work—including data requests, ensuring data integrity, detailed coding, and extensive statistical analysis—we report here results from a unique database of 2,995 individual state criminal sentences imposed in Colorado, covering 183 different types of crimes and 285 different judges—

¹ Debra C. Weiss, *Newly Confirmed 4th Circuit Nominee Is Now the Country's Youngest Federal Judge*, A.B.A. J. (Mar. 6, 2019), <http://www.abajournal.com/news/article/newly-confirmed-4th-circuit-nominee-is-now-the-countrys-youngest-federal-judge> [<https://perma.cc/2P3K-NB2M>].

² James Joyner, *Packing the Courts with Young Judges*, OUTSIDE BELTWAY (Mar. 6, 2019), <https://www.outsidethebeltway.com/packing-the-courts-with-young-judges/> [<https://perma.cc/FTA4-VU4C>] (discussing the youth of recent Republican judicial nominees); see also Carl Hulse, *Liberals Begin Lining Up Young Judges for a Post-Trump Surge*, N.Y. TIMES (June 8, 2019), <https://www.nytimes.com/2019/06/08/us/politics/liberals-courts-trump.html> [<https://perma.cc/MLS7-GJEL>] (discussing liberal efforts to nominate judges).

³ See generally Jack B. Weinstein, *Does Religion Have a Role in Criminal Sentencing?*, 23 TOURO L. REV. 539 (2007).

180 male and 105 female—over a sixteen-year time span (2001–2016).⁴ Because we observed in the dataset the same judges sentencing similar crimes as they grow older, the data allowed us a window into the complex interaction of age, gender, crime seriousness, and sentencing patterns.

The results of our analysis are striking, as we uncovered a three-way interaction not previously reported in the empirical or experimental literature: For high harm crimes, younger female judges sentenced convicted defendants more harshly than their male and older female colleagues. We controlled for the independent effect of judicial experience, leading to the conclusion that *age*—and not just experience on the bench—is driving the results. On average, young female judges sentenced offenders convicted of high harm crimes to 24% more incarceration (4.9 years more) than did their male colleagues, and to 25% more incarceration (5.1 years more) than did their older female colleagues. This finding has implications for sentencing theory, for practical lawyering, and perhaps even for public policy.

On the theoretical side, although a growing number of studies have examined the impact of a judge's background on sentencing, these studies have typically not accounted for factors that change over time, the most obvious of which is age. The research reported here should pave the way for longitudinal studies examining how different judges' sentencing patterns might change over time. Our results also serve as a reminder for future empirical research of how important it is not just to control for different variables, but also to pay attention to their interactions. In our data, age alone had no impact on sentencing; neither did gender. Even when we considered age and gender together, without distinguishing between the harm levels of the crimes, this two-way interaction had no impact. Only when we considered age, gender, and harm levels together did we see these three factors impact—and impact substantially—the sentences imposed by these judges.

Our results could have practical implications as well. When the stakes are high—as they are with the types of high harm crimes in which our interaction effect emerges—both prosecutors and defense attorneys will look to every possible angle to move the sentence toward their preferred outcome. The complexities of individual sentencing, and the limitations of a single empirical study, caution against the conclusion that all older female judges and all male judges will always be more lenient than young female judges

⁴ We chose Colorado for the simple reason that the judge-author had unique access to the electronic forms of the case data. But Colorado is also in some ways a typical state when it comes to criminal sentencing in that it is, like the vast majority of states, a so-called Model Penal Code state rather than a sentencing-guideline state. See *infra* text accompanying notes 87–88.

with all serious offenders. But the study's real-world data provides both prosecutors and defense attorneys with a stark warning: The interaction of a judge's age and gender, together with the seriousness of a case, likely matters in determining the sentencing outcome. These findings might even impact public policy in arenas including the selection, education, training, and retirement of judges.

In Part I, we survey the existing literature on the impact of judge demographics on criminal sentencing and on "punitiveness" and "third-party punishment" more generally. In Part II, we introduce our new study and discuss its analysis of judicial background variables on sentencing. In Part III, we present the results of our analysis and consider limitations and cautions for interpreting those results. In Part IV, we discuss the implications of these results for future research on judicial sentencing, for practical lawyering, and for judicial selection, education, and retirement policies.

I. Gender, Age, and Sentencing: Conventional Wisdom and Empirical Literature

There is a robust but largely disconnected body of work across multiple disciplines examining the effect that individual background characteristics have on punishment decisions. We briefly review this literature, focusing on "punitiveness" and "third-party punishment," and then review empirical studies examining actual criminal sentences by real judges.

A. Punitiveness

Research on "punitiveness," conducted mostly in sociology, psychology, and criminology, typically asks experimental participants or survey respondents questions about how they would react to certain hypothetical situations—not necessarily criminal situations—and measures their "punitiveness" based on their responses. Investigators then attempt to correlate the subjects' measured punitiveness with various individual characteristics, including race, ethnicity, gender, age, educational background, socio-economic status, and even religious and political affiliations and beliefs.⁵

⁵ See, e.g., Karen K. Dion, *Children's Physical Attractiveness and Sex as a Determinate of Adult Punitiveness*, 10 DEVELOPMENTAL PSYCHOL. 772, 773 (1974); KAREN GELB, PREDICTORS OF PUNITIVENESS: COMMUNITY VIEWS IN VICTORIA 9–10 (2011), <https://researchdirect.westernsydney.edu.au/islandora/object/uws%3A27169/datastream/PDF/view> [<https://perma.cc/RZ35-67Y7>]; Devon Johnson, *Racial Prejudice, Perceived Injustice, and the Black-White Gap in Punitive Attitudes*, 36 J. CRIM. JUST. 198, 200–01 (2008); Brian K. Payne et al., *What Drives Punitive Beliefs?: Demographic Characteristics and Justifications for Sentencing*, 32 J. CRIM. JUST. 195, 199 (2004).

Measures of criminally-relevant punitiveness typically focus on four areas: (1) questions about criminals' rights and the punishment of criminals; (2) questions about the death penalty; (3) questions about support for spending on fighting crime and on the criminal justice system; and (4) questions that relate to confidence and trust in the police and the criminal justice system.⁶

Researchers have explored several individual background variables that affect punitiveness measures. Demographic characteristics, such as gender, age, race, and geographic region, may all influence punitive attitudes to different degrees.⁷ Some studies show that men are, on the whole, more punitive than women.⁸ However, other studies indicate that women are more punitive than men with regard to specific types of crimes.⁹ Age and punitive attitudes appear to have a curvilinear relationship, with the oldest and youngest people being the least punitive.¹⁰ Although members of all races may be punitive, they may have different reasons for their punitiveness¹¹ or be punitive in

⁶ Peter K. Enns, *The Public's Increasing Punitiveness and Its Influence on Mass Incarceration in the United States*, 58 AM. J. POL. SCI. 857, 860–61 (2014).

⁷ See MARC MORJÉ HOWARD, UNUSUALLY CRUEL: PRISONS, PUNISHMENT, AND THE REAL AMERICAN EXCEPTIONALISM 163 (2017) (opining that race, religion, politics, and business can explain increasing American punitiveness); Payne et al., *supra* note 5, at 199–202 (finding that demographic predictors had some impact but that none were particularly strong predictors of punitive attitudes and noting that justifications for sentencing were strongly linked to punitive attitudes). See generally Lynne D. Roberts & David Indermaur, *Predicting Punitive Attitudes in Australia*, 14 PSYCHIATRY PSYCHOL. & L. 56 (2007) (finding that demographic factors are weak to moderate predictors of punitive attitudes).

⁸ Brandon K. Applegate et al., *Public Views Toward Crime and Correctional Policies: Is There a Gender Gap?*, 30 J. CRIM. JUST. 89, 95 (2002); John K. Cochran & Beth A. Sanders, *The Gender Gap in Death Penalty Support: An Exploratory Study*, 37 J. CRIM. JUST. 525, 530 (2009) (finding a robust gender difference for support of the death penalty, with men showing more support than women).

⁹ Jane B. Sprott, *Are Members of the Public Tough on Crime? The Dimensions of Public "Punitiveness"*, 27 J. CRIM. JUST. 467, 468–72 (1999) (finding that women were more punitive in response to the belief that crime was increasing and that women were less punitive than men in cases with teenage defendants). See generally Peter H. Rossi et al., *Just Punishments: Guideline Sentences and Normative Consensus*, 13 J. QUANTITATIVE CRIMINOLOGY 267 (1997).

¹⁰ PETER H. ROSSI & RICHARD A. BERK, JUST PUNISHMENTS: FEDERAL GUIDELINES AND PUBLIC VIEWS COMPARED 180 (1997).

¹¹ Devon Johnson, *Crime Salience, Perceived Racial Bias, and Blacks' Punitive Attitudes*, 4 J. ETHNICITY CRIM. JUST. 1, 11–14 (2007) (finding, based on data from the 2001 Race, Crime, and Public Opinion Study, that fear of crime and individualistic attributions for criminal behavior significantly increase Blacks' punitive

different ways.¹² Finally, while some scholarship suggests the idea of a “southern vigilantism,”¹³ other research has found no significant direct effect of geography on punitive attitudes.¹⁴

Political beliefs, religiosity, and economic perspectives may also play a role. Research has consistently found that people who have conservative political views are more likely to support punitive policies,¹⁵ though of course those policies themselves may

attitudes, but that perceived racial bias and vicarious exposure to incarceration decrease them); Devon Johnson, *Punitive Attitudes on Crime: Economic Insecurity, Racial Prejudice, or Both?*, 34 SOC. FOCUS 33, 41–43 (2001) (finding that whites’ punitive attitudes are not associated with personal economic insecurity, but that Jim Crow and laissez-faire racism are significantly correlated with higher punitiveness by white people); Johnson, *supra* note 5, at 198, 203–04 (finding that racial prejudice and perceived racial bias explain the gap); see also Ted Chiricos et al., *Racial Typification of Crime and Support for Punitive Measures*, 42 CRIMINOLOGY 359, 374–76 (2004) (finding that racial typification of crime is a significant, independent predictor of punitiveness, particularly in white people who are less prejudiced, not southern, not conservative, and have low crime salience); Kelly Welch et al., *The Typification of Hispanics as Criminals and Support for Punitive Crime Control Policies*, 40 SOC. SCI. RES. 822, 830–32 (2011) (finding the same effects as Chiricos et al., *supra*, with respect to typification of Hispanics as criminals as opposed to African Americans). See generally Steven F. Cohn et al., *Punitive Attitudes Toward Criminals: Racial Consensus or Racial Conflict?*, 38 SOC. PROBS. 287 (1991) (finding that punitive attitudes of white people stemmed partly from racial prejudice while punitive attitudes of Black people were associated with fear of crime); Mark Peffley et al., *Racial Attributions in the Justice System and Support for Punitive Crime Policies*, 45 AM. POL. RES. 1032, 1042 (2017) (surveying attitudes about policing held by Blacks, Latinos, and whites).

¹² See, e.g., Payne et al., *supra* note 5 (finding interaction effect between punitive attitudes and justifications for punishment varied across race and gender).

¹³ James D. Unnever et al., *Turning the Other Cheek: Reassessing the Impact of Religion on Punitive Ideology*, 22 JUST. Q. 304, 326–29 (2005).

¹⁴ Alex R. Piquero & Laurence Steinberg, *Public Preferences for Rehabilitation Versus Incarceration of Juvenile Offenders*, 38 J. CRIM. JUST. 1, 1–3 (2010) (finding that the public in four states in different geographic regions was overall willing to pay more in taxes for rehabilitation than incarceration). See generally Marian J. Borg, *The Southern Subculture of Punitiveness? Regional Variation in Support for Capital Punishment*, 34 J. RES. CRIME & DELINQUENCY 25 (1997) (finding little variation between southerners and non-southerners in their support of the death penalty but noting that geographic region “conditions the effects of racial prejudice, religious fundamentalism, and political conservatism” on support for the death penalty).

¹⁵ Unnever et al., *supra* note 13, at 320–21 (citing Melissa Moon et al., *Putting Kids to Death: Specifying Public Support for Juvenile Capital Punishment*, 17 JUST. Q. 663 (2000)). But see Barbara A. Gault & John Sabini, *The Roles of Empathy, Anger, and Gender in Predicting Attitudes Toward Punitive, Reparative, and Preventative Public Policies*, 14 COGNITION & EMOTION 495, 516 (2000) (stating that Democrats and Republicans showed almost no differences in their preferences for punitiveness). See generally Brenda L.

be definitionally part of being conservative. Research on religious beliefs and punitiveness has produced less consistent findings, with some studies finding that conservative theological beliefs are correlated with increased punitiveness and other research showing no correlation.¹⁶ Studies have also found that support for welfare and support for punitive sanctions are inversely related,¹⁷ while a belief in economic individualism is positively related to punitiveness.¹⁸

Not surprisingly, punitive attitudes may vary depending on the type of offender or offense, with people tending to be harsher toward crimes they perceive to be more violent¹⁹ and offenders they believe to be less capable of rehabilitation. A person's belief that others can change results in less punitiveness,²⁰ but evidence that people continue to

Vogel & Ronald E. Vogel, *The Age of Death: Appraising Public Opinion of Juvenile Capital Punishment*, 31 J. CRIM. JUST. 169 (2003)).

¹⁶ MARC MORJÉ HOWARD, *supra* note 7, at 163 (stating that in non-death penalty situations, evangelicals are more punitive than people who are religious but non-evangelical because they tend to blame the individual perpetrator); Unnever et al., *supra* note 13, at 329 (finding that people who have a "rigid and moralistic" approach to religion and view God as a powerful distributor of justice are more likely to have punitive attitudes about offenders, but that people who have a loving image of God are more compassionate toward others and less punitive); Kevin H. Wozniak & Andrew R. Lewis, *Reexamining the Effect of Christian Denominational Affiliation on Death Penalty Support*, 38 J. CRIM. JUST. 1082, 1085–86 (2010) (finding, based on data from the General Social Survey, that affiliation with any Christian denomination increases the likelihood that an individual will support the death penalty compared to nonreligious individuals).

¹⁷ Katherine Beckett & Bruce Western, *Governing Social Marginality: Welfare, Incarceration, and the Transformation of State Policy*, 3 PUNISHMENT & SOC'Y 43, 50–51 (2001); Michael J. Hogan et al., *Economic Insecurity, Blame, and Punitive Attitudes*, 22 JUST. Q. 392, 402 (2005); Ashley T. Rubin, *Punitive Penal Preferences and Support for Welfare: Applying the "Governance of Social Marginality" Thesis to the Individual Level*, 13 PUNISHMENT & SOC'Y 198, 207 (2011).

¹⁸ Ryan Kornhauser, *Economic Individualism and Punitive Attitudes: A Cross-National Analysis*, 17 PUNISHMENT & SOC'Y 27, 34 (2015). This relationship may well be mediated by beliefs in free will. *See infra* text accompanying notes 37–38.

¹⁹ Kristy Holtfreter et al., *Public Perceptions of White-Collar Crime and Punishment*, 36 J. CRIM. JUST. 50, 53 (2008) (finding that the majority of respondents supported punishing violent criminals more severely than white-collar criminals but that over one-third of respondents expressed the opposite opinion); *see also* Russil Durrant et al., *Understanding Punishment Responses to Drug Offenders: The Role of Social Threat, Individual Harm, Moral Wrongfulness, and Emotional Warmth*, 38 CONTEMP. DRUG PROBS. 147, 163–64 (2011) (finding perceptions of moral wrongfulness to be the best predictor of punishment responses across offense types).

²⁰ *See* Shadd Maruna & Anna King, *Public Opinion and Community Penalties*, in ALTERNATIVES TO PRISON: OPTIONS FOR AN INSECURE SOCIETY 83–112 (Anthony Bottoms et al. eds., 2004).

commit crimes increases punitiveness.²¹ Relatedly, studies have generally found that individuals have less punitive attitudes when juvenile offenders are involved compared to adults.²² People are also generally more punitive with sex offenders compared to other kinds of offenders.²³

Although informative to a certain extent, the punitiveness literature has important limitations when it comes to shedding light on actual judicial sentencing biases. For instance, “punitiveness” is often defined in a rather narrow, explicitly pejorative way. A widely-used instrument to measure it is unabashedly called “the vengeance scale,” and its authors describe it as measuring the kind of “revenge . . . implicated in a broad range of criminal and anti-social acts.”²⁴ By asking subjects to put themselves in the shoes of crime victims, rather than in the shoes of judges or juries, this instrument is probably measuring “second-party punishment” rather than “third-party punishment,” to use the terms we will introduce in Part I.B. This is hardly a reliable gauge, at least directly, of the retributive components in a criminal sentence.

B. Third-Party Punishment

Evolutionary theorists coined the term “third-party punishment” to describe the situation where one person punishes another for a norm violation even though the

²¹ VALERIE J. CALLANAN, FEEDING THE FEAR OF CRIME: CRIME-RELATED MEDIA AND SUPPORT FOR THREE STRIKES 12 (2005).

²² See Piquero & Steinberg, *supra* note 14, at 1–3. *Contra* Justin T. Pickett & Ted Chiricos, *Controlling Other People’s Children: Racialized Views of Delinquency and Whites’ Punitive Attitudes Toward Juvenile Offenders*, 50 CRIMINOLOGY 673, 690–92 (2012) (finding that white people who hold racial typifications about delinquency and victimization as well as racial resentment favor more punitive policies for juveniles); Christi Metcalfe et al., *Using Path Analysis to Explain Racialized Support for Punitive Delinquency Policies*, 31 J. QUANTITATIVE CRIMINOLOGY 699, 709–11 (2015) (reaching the same conclusion).

²³ See generally Darren L. Rogers & Christopher J. Ferguson, *Punishment and Rehabilitation Attitudes Toward Sex Offenders Versus Nonsexual Offenders*, 20 J. AGGRESSION MALTREATMENT & TRAUMA 395 (2011).

²⁴ See generally Noreen Stuckless & Richard Goranson, *The Vengeance Scale: Development of a Measure of Attitudes Toward Revenge*, 7 J. SOC. BEHAV. & PERSONALITY 25 (1991) (contextualizing revenge to be measured by the Vengeance Scale); Gizem Uzun, *Vengeance Scale: Reliability and Validity Study with Gender Differences*, 52 QUALITY & QUANTITY 1455 (2018) (describing “punitiveness” as a “rudimental, destructive and illogical personality, accompanied by some personality disorders”).

punisher was not the victim of the violation.²⁵ Third-party punishment is to be distinguished from “second-party punishment”—or what we might also call retaliation or revenge—where the person or animal doing the punishing was the victim of the wrongdoer’s transgression.²⁶

Second-party punishment is a form of self-defense and is widespread throughout the animal kingdom.²⁷ By contrast, third-party punishment, though a human universal,²⁸ seems unique to humans.²⁹ Many evolutionary theorists believe that third-party

²⁵ See generally Robert Boyd et al., *The Evolution of Altruistic Punishment*, 100 PROC. NAT’L ACAD. SCI. U.S. 3531 (2003); Ernst Fehr & Simon Gächter, *Altruistic Punishment in Humans*, 415 NATURE 137 (2002); Ernst Fehr & Urs Fischbacher, *Third-Party Punishment and Social Norms*, 25 EVOLUTION HUM. BEHAV. 63 (2004); Bettina Rockenbach & Manfred Milinski, *The Efficient Interaction of Indirect Reciprocity and Costly Punishment*, 444 NATURE 718 (2006) (discussing that biologists and economists often call third-party punishment “altruistic” or “costly” punishment).

²⁶ MORRIS B. HOFFMAN, THE PUNISHER’S BRAIN: THE EVOLUTION OF JUDGE AND JURY 121–49 (2014). “First-party punishment,” to complete our list of types of evolved punishments, is a way to describe our evolved moral intuitions. *Id.* at 92–120. One of the reasons most of us do not violate certain core norms, perhaps the biggest reason, is not that we fear retaliation from the victim or punishment by a third party but because we know it is wrong. We “punish” ourselves ahead of time with conscience or afterwards with guilt. Seneca put it aptly: “The first and greatest punishment of the sinner is the conscience of sin.” 1 THE EPISTLES OF LUCIUS ANNAEUS SENECA 146 (T. Morrell trans., Palala Press 2015) (1786).

²⁷ Morris B. Hoffman & Frank Krueger, *The Neuroscience of Blame and Punishment*, in SELF, CULTURE AND CONSCIOUSNESS: INTERDISCIPLINARY CONVERGENCES ON KNOWING AND BEING 207, 208 (Sangeetha Menon et al. eds., 2017) (discussing that second-party punishment is likely widespread because it may be rooted in the immunological response). See generally Richard Gordon, *A Retaliatory Role for Algal Projectiles*, 126 J. THEORETICAL BIOLOGY 419 (1987) (describing how algae engages in second-party retaliative punishment by firing projectiles against would-be microscopic predators).

²⁸ DONALD BROWN, HUMAN UNIVERSALS 138 (1991). See generally Frank W. Marlowe et al., *More “Altruistic” Punishment in Larger Societies*, 275 PROC. ROYAL SOC’Y B: BIOLOGICAL SCI. 587 (2008).

²⁹ Katrin Reidl et al., *No Third-Party Punishment in Chimpanzees*, 109 PROC. NAT’L ACAD. SCI. U.S. 14824, 14824–26 (2012) (finding that chimpanzees, one of humans’ closest living relatives, may retaliate against theft but do not engage in third-party punishment). But there are also tantalizing signs that non-human primates, and even some species of fish, have certain kinds of precursors to third-party punishment. See generally, e.g., Toshisada Nishida, *A Within-Group Gang Attack on a Young Adult Male Chimpanzee: Ostracism of an Ill-Mannered Member?*, 36 PRIMATES 207 (1995); Nichola Raihani et al., *Punishers Benefit from Third-Party Punishment in Fish*, 327 SCI. 171 (2010); Claudia Rudolf von Rohr et al., *Impartial Third-Party Interventions in Captive Chimpanzees: A Reflection of Community Concern*, 7 PLoS ONE 1 (2012).

punishment is an evolved trait that enabled our ancestors, despite their genetic heterogeneity, to live in intensely social small groups by deterring norm violators.³⁰

In the course of performing experiments related to third-party punishment, researchers have developed sophisticated tools to unpack the factors that drive criminal punishment. These tools include criminal hypotheticals that vary the amount of harm and the mental state of the hypothetical criminal. Subjects read the hypotheticals and are then asked how much they would punish the hypothetical offender, typically on a scale of zero to nine—zero being defined as no punishment and nine being defined as the most serious punishment the subject can imagine. Here is one example of such a hypothetical, drawn from earlier work by several of the co-authors of the present article, presenting a situation involving somewhat serious harm and the mental state of “knowingly”:

John is doing carpentry work on his house, which abuts a public mountain bike trail. While carrying wood planks, John drops some onto the trail and doesn't pick them up because he wants to start the carpentry, even though he is aware that there is a substantial risk that bikers will hit the planks and be injured. Two bikers passing by at that moment hit the planks, crash as a result, and are seriously injured.³¹

This research has yielded some important results for the law, including that the amount subjects blame hypothetical purposeful crimes that differ in harm is remarkably uniform across subject demographics.³² This research has also demonstrated that while ordinary people are good at distinguishing purposeful, knowing, and negligent crimes, they are terrible at distinguishing knowing crimes from reckless ones.³³

³⁰ See, e.g., Samuel Bowles & Herbert Gintis, *The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations*, 65 THEORETICAL POPULATION BIOLOGY 17, 18–23 (2004) (modeling group effects of punishment and reciprocity). See generally Fehr & Fischbacher, *supra* note 25; Jillian J. Jordan & David G. Rand, *Third-Party Punishment as a Signal of High Continuation Probabilities in Repeated Games*, 421 J. THEORETICAL BIOLOGY 189 (2017); Morris B. Hoffman & Timothy H. Goldsmith, *The Biological Roots of Punishment*, 1 OHIO ST. J. CRIM. L. 627 (2004).

³¹ Francis X. Shen et al., *Sorting Guilty Minds*, 86 N.Y.U. L. REV. 1306, 1328 (2011).

³² See generally, e.g., Paul H. Robinson & John M. Darley, *Intuitions of Justice: Implications for Criminal Law and Justice Policy*, 81 S. CAL. L. REV. 1 (2007); Paul H. Robinson & Robert Kurzban, *Concordance and Conflict in Intuitions of Justice*, 91 MINN. L. REV. 1829 (2007).

³³ Shen et al., *supra* note 31, at 1337–38.

Only a handful of these third-party punishment experiments have examined the effects of the subject punisher's demographics, and those have all been limited to race. In one of the most important studies using actual judges as experimental subjects, researchers in 2009 found that when the race of the defendant was not made explicit, but rather subliminally primed, neither Black nor white judges showed any sentencing biases for or against Black defendants as correlated to the judges' measured implicit biases. But when the defendant's race was explicitly stated, Black judges but not white judges showed sentencing biases in favor of Black defendants that correlated to the Black judges' measured implicit biases.³⁴ The authors' speculation: At least when the race of the defendant was made explicit, white judges compensated for their implicit bias, but Black judges did not.³⁵ There is no comparable experimental literature, using either judges or non-judges, examining the effects of the punisher's gender or age.

Akin to the literature correlating various political beliefs to punitiveness,³⁶ third-party punishment studies have shown that the amount of hypothetical punishment imposed by non-judge subjects is greater when those subjects have a stronger measured belief in free will, although that correlation appeared only for low-harm crimes.³⁷ When the hypothetical crimes were very serious, subjects with weak beliefs in free will punished just as harshly as those with strong beliefs in free will.³⁸ These results were among the first to suggest that the seriousness of the crimes at issue may be mediating other two-way correlations in a three-way interaction, a phenomenon we saw in a different form here.

Of course, both the punitiveness and third-party punishment studies suffer from the same central ecological challenge: It is one thing to test how ordinary people—or even judges—might rank hypothetical crimes or impose hypothetical punishments on

³⁴ Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 8 J. EMPIRICAL L. STUD. 72 (2009); see also Francis X. Shen, *Minority Mens Rea: Racial Bias and Criminal Mental States*, 68 HASTINGS L.J. 1007 (2017) (discussing a between-subjects experiment examining the effect of first names on third-party punishment, in which non-judge subjects showed no punishment differences between vignettes featuring protagonists named John, Jamal, Emily, and Lakisha).

³⁵ Rachlinski et al., *supra* note 34, at 1224.

³⁶ See *supra* text accompanying notes 14–23.

³⁷ See generally Frank Krueger et al., *An fMRI Investigation of the Effects of Belief in Free Will on Third-Party Punishment*, 9 SOC. COGNITIVE & AFFECTIVE NEUROSCIENCE 1143 (2014).

³⁸ *Id.*

hypothetical offenders, but quite another to impose real sentences on real offenders in actual criminal cases.

C. Studying Actual Criminal Sentences

Studying actual criminal sentences has proved to be more challenging than doing punitiveness surveys or third-party punishment experiments, largely because of the difficulty of gathering the sentencing data itself, not to mention data on judge characteristics. Researchers typically face a trade-off between ecological validity on the one hand and experimental control on the other. An important value of doing laboratory experiments on punishment is that the researcher can control all relevant aspects of the fictional vignette. A researcher analyzing real case decisions must account for the fact that no two cases are exactly the same.³⁹ But the sword cuts both ways, as studies utilizing experimental paradigms outside the court system sacrifice much in terms of ecological validity. As Gregory Sisk and colleagues have observed, “[e]ither it is not the same or it is not real—lack of comparability among cases or the absence of authenticity are the Scylla and Charybdis of empirical study of judicial decisionmaking.”⁴⁰ Research to date on the effect of individual demographics on judicial sentencing has had to find ways to address this challenge. We review this literature by examining, in order, research on gender, race, age, judicial ideology, and the interaction of these and other background variables.⁴¹

1. Gender

Research has revealed a number of differences between female and male judges, though the magnitude of those differences appears small.⁴² For example, one study

³⁹ Gregory C. Sisk et al., *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 N.Y.U. L. REV. 1377, 1380 (1998).

⁴⁰ *Id.*

⁴¹ We do not include in this Section any discussion of the vast literature examining the effects of the *defendant's* age, gender, and race on sentencing outcomes. See, e.g., Barbara A. Koons-Witt et al., *Gender and Sentencing Outcomes in South Carolina: Examining the Interactions with Race, Age, and Offense Type*, 25 CRIM. JUST. POL'Y. REV. 299 (2014) (finding females sentenced less harshly than similarly-situated males, but this difference subject to significant interaction by severity of offense; whites sentenced less harshly than Blacks, but Black females sentenced just as harshly as white males; no interactions with age).

⁴² See generally Darrell Steffensmeier & Chris Herbert, *Women and Men Policymakers: Does the Judge's Gender Affect Sentencing of Criminal Defendants?*, 77 SOC. FORCES 1163 (1998); Herbert M. Kritzer &

looking at sentencing outcomes in Pennsylvania state courts from 1991 to 1993 found that female judges were somewhat (11%) more likely to incarcerate offenders than their male counterparts, and that their sentences were on average 1.5 months longer.⁴³ On the civil side, the gender of the judge appears to matter in some, but not all, contexts, notably in cases that deal with sexual harassment or gender-based discrimination.⁴⁴ Likewise, some studies have shown that the presence of a woman on appellate panels affects the decision in cases that involve gender-related issues.⁴⁵

2. Race

A growing literature on the effects of judicial race on decision-making, particularly comparing Black and white judges, has reported a variety of findings but no overarching

Thomas Uhlman, *Sisterhood in the Courtroom: Sex of Judge and Defendant in Criminal Case Disposition*, 14 SOC. SCI. J. 77 (1977) (finding no differences in incarceration rates; lengths not studied); John Gruhl et al., *Women as Policymakers: The Case of Trial Judges*, 25 AM. J. POL. SCI. 308 (1981) (finding that female judges slightly more likely to incarcerate; no difference in lengths); MARTHA A. MYERS & SUSETTE M. TALARICO, *THE SOCIAL CONTEXTS OF CRIMINAL SENTENCING* (1987) (finding no differences in rates or lengths).

⁴³ Steffensmeier & Herbert, *supra* note 42, at 1174–76.

⁴⁴ See Christina L. Boyd et al., *Untangling the Causal Effects of Sex on Judging*, 54 AM. J. POL. SCI. 389, 400–02 (2010) (finding that judicial gender only affects decisions in sex discrimination cases, with female judges more likely to decide in favor of the party alleging discrimination); Sean Farhang & Gregory Wawro, *Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making*, 20 J.L. ECON. & ORG. 299, 319–20 (2004) (finding that female judges are more likely to vote in favor of the plaintiff in employment discrimination cases and that the presence of at least one woman on a three-judge panel increases the likelihood that the panel will vote for the plaintiff); Jennifer L. Peresie, *Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts*, 114 YALE L.J. 1759, 1761, 1776 (2005) (finding that female judges are more likely to vote in favor of plaintiffs in Title VII sex discrimination and sex harassment cases); Donald R. Songer et al., *A Reappraisal of Diversification in the Federal Courts: Gender Effects in the Courts of Appeals*, 56 J. POL. 425, 432–34 (1994) (finding that in an analysis of votes of appeals court decisions, the only difference in decisions between female and male judges was in employment discrimination cases); see generally Thomas G. Walker & Deborah J. Barrow, *The Diversification of the Federal Bench: Policy and Process Ramifications*, 47 J. POL. 596 (1985) (finding that female judges “displayed a distinct pattern of deferring to positions taken by the government”).

⁴⁵ See Boyd et al., *supra* note 44, at 389, 402, 406 (finding that the presence of a woman judge on a panel with men increases the likelihood of the panel ruling in favor of the rights litigant); accord Farhang & Wawro, *supra* note 44, at 324 (finding that “male judges vote more liberally when one woman serves on a panel . . . as compared to all-male panels”); Peresie, *supra* note 44, at 1778 (finding that “[m]ale judges were more likely to find for plaintiffs when at least one female judge was on the panel”).

pattern.⁴⁶ For instance, research suggests that Black judges are more favorable to plaintiffs in affirmative action⁴⁷ and voting rights cases,⁴⁸ and even more favorable if the plaintiffs in those kinds of cases are non-white.⁴⁹ But in other civil contexts, the results remain mixed.⁵⁰

With regard to criminal sentencing, the results are also quite mixed. Some studies find that Black judges are more lenient with Black defendants,⁵¹ some that Black judges

⁴⁶ Maya Sen, *Is Justice Really Blind? Race and Reversal in U.S. Courts*, 44 J. LEGAL STUD. S187, S190 (2015) (noting that on the question of “whether women and minority judges decide cases differently than their white male counterparts . . . [for] the most part, the answer to this question has been yes, but they do so in the context of substantively salient issues”); Allison P. Harris & Maya Sen, *Bias and Judging*, 22 ANN. REV. POL. SCI. 241, 248–49 (2019).

⁴⁷ Jonathan P. Kastellec, *Racial Diversity and Judicial Influence on Appellate Courts*, 57 AM. J. POL. SCI. 167, 179–81 (2013).

⁴⁸ Adam B. Cox & Thomas J. Miles, *Judging the Voting Rights Act*, 108 COLUM. L. REV. 1, 30 (2008).

⁴⁹ See Pat K. Chew & Robert E. Kelley, *Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases*, 86 WASH. U. L. REV. 1117, 1141 (2008) (finding that Black judges are more likely to vote in favor of plaintiffs in racial harassment cases); Cox & Miles, *supra* note 48, at 30 (finding that Black appeals judges are more likely to vote in favor of non-white plaintiffs in Voting Rights Act cases); Jason L. Morin, *The Voting Behavior of Minority Judges in the U.S. Courts of Appeals: Does the Race of the Claimant Matter?*, 42 AM. POL. RES. 34, 47–48 (2014) (finding that Black judges are more likely to favor Black plaintiffs than white plaintiffs in employment discrimination cases); Jill D. Weinberg & Laura Beth Nielsen, *Examining Empathy: Discrimination, Experience, and Judicial Decisionmaking*, 85 S. CAL. L. REV. 313, 345 (2011) (finding that in employment civil rights disputes, “white judges were more likely to dismiss cases involving minority plaintiffs while minority judges were less likely to dismiss cases involving white plaintiffs”).

⁵⁰ Jennifer A. Segal, *Representative Decision Making on the Federal Bench: Clinton’s District Court Appointees*, 53 POL. RES. Q. 137, 147 (2000) (finding that “black and female district court appointees are no more likely to serve the policy interests of their own communities than are [President Clinton’s] white and male appointees”).

⁵¹ Susan Welch et al., *Do Black Judges Make a Difference?*, 32 AM. J. POL. SCI. 126, 132–33 (1988) (looking at 3,418 cases of men convicted of a felony between 1968 and 1979 in an anonymous northeastern city and finding that white trial judges treated Black defendants worse than white defendants while Black trial judges treated Black and white defendants equally); Brian D. Johnson, *The Multilevel Context of Criminal Sentencing: Integrating Judge- and County-Level Influences*, 44 CRIMINOLOGY 259, 272 (2006).

are harsher than white judges,⁵² and others find no differences at all.⁵³ There are also conflicting results on whether Black judges decide differently in death penalty cases.⁵⁴

These mixed findings are also seen in studies that look exclusively at the federal district courts. For example, one study using data collected by the United States Sentencing Commission on offenders sentenced under federal sentencing guidelines between 1992 and 2001 found that there were no significant disparities between sentences imposed by Black judges and those imposed by white judges.⁵⁵ However, in a 2019 study looking at that same federal data source for the years between 1999 and 2015 researchers found that Black judges issued shorter sentences than non-Black judges.⁵⁶ Far less research has been done on races and ethnicities other than Black and white,⁵⁷ partially because there are comparatively fewer Hispanic, Asian and Pacific Islander, and Native American judges serving in the judiciary.

⁵² Darrell Steffensmeier & Chester L. Britt, *Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently?*, 82 SOC. SCI. Q. 749, 755 (2001) (examining Pennsylvania sentencing data from 1991 to 1994 and finding that “black and white judges weighted case and offender information in similar ways when making punishment decisions” with Black judges “more likely to sentence both black and white offenders to prison”).

⁵³ Cassia Spohn, *The Sentencing Decisions of Black and White Judges: Expected and Unexpected Similarities*, 24 L. SOC'Y REV. 1197, 1201 (1990) (examining cases in Detroit where defendants were charged with at least one of eleven felonies between 1976 and 1978 and finding no differences in sentencing between white and Black judges—both groups of judges sentenced Black defendants more harshly).

⁵⁴ Jonathan P. Kastlelec, *Race, Context, and Judging on the Courts of Appeals: Race-Based Panel Effects in Death Penalty Cases* 27 (Oct. 28, 2019) (working paper), https://scholar.princeton.edu/sites/default/files/jkastlelec/files/kastlelec_race_dp_paper_jsj_submission.pdf [<https://4HEP-AZAM>] (examining federal and state death penalty appeals decisions between 1973 and 1995 and finding that “the addition of a single black judge to an otherwise all-nonblack panel significantly increases the probability that a defendant gets relief from a death sentence”). *But see* Farhang & Wawro, *supra* note 44, at 320 (finding that the presence of a Black judge does not impact judicial decisions).

⁵⁵ Max Schanzenbach, *Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics*, 34 J. LEGAL STUD. 57, 80–81 (2005).

⁵⁶ Alma Cohen & Crystal S. Yang, *Judicial Politics and Sentencing Decisions*, 11 AM. ECON. J.: ECON. POL'Y 160, 175 (2019).

⁵⁷ *But see* Morin, *supra* note 49, at 47 (finding that Hispanic judges are less likely to rule in favor of employment discrimination plaintiffs than non-Hispanic colleagues); Malcolm D. Holmes et al., *Judges' Ethnicity and Minority Sentencing: Evidence Concerning Hispanics*, 74 SOC. SCI. Q. 496 (1993) (finding that Hispanic judges are not as affected by defendant ethnicity as white judges).

3. Age

Judicial age has not featured as prominently in the research literature. Indeed, one review of the empirical literature suggested that “age is of minimal value in predicting how judges will vote, particularly once other variables are considered.”⁵⁸ The few studies that have looked at the relationship between judicial age and criminal sentencing have produced no clear pattern of results.⁵⁹ The picture is no clearer in civil cases. One study examined 540 bias rulings and 1,600 decisions in racial and gender discrimination cases tried in the federal district courts between 1984 and 1995 and found that older judges are more likely to favor older plaintiffs in age bias cases.⁶⁰ But a later study using the same data and analyzing it a different way found that there was no relationship between judicial age and decisions in age discrimination cases.⁶¹

4. Ideology

A wide body of research in political science and empirical legal studies has examined the role of judicial ideology in case outcomes, including sentencing. From the political science perspective, judicial background matters, but it is political ideology that most substantially affects judicial decision-making.⁶² A 2019 review of the political science literature found that “differently situated judges might decide cases differently, but that any differences associated with demographics are actually fairly issue-specific and much less pronounced than differences rooted in ideology or partisanship.”⁶³ While the facts of the case and the law remain constraints, and while the relative impact of ideology on

⁵⁸ Tracey E. George, *Court Fixing*, 43 ARIZ. L. REV. 9, 17 (2001).

⁵⁹ Johnson, *supra* note 51, at 272 (finding that older judges are “less likely to incarcerate convicted offenders and sentenc[e] them to shorter periods of confinement”).

⁶⁰ See generally Kenneth L. Manning et al., *Does Age Matter? Judicial Decision Making in Age Discrimination Cases*, 85 SOC. SCI. Q. 1 (2004).

⁶¹ Lee Epstein & Andrew D. Martin, *Does Age (Really) Matter? A Response to Manning, Carroll, and Carp*, 85 SOC. SCI. Q. 19, 19–20 (2004) (finding that Manning, Carroll, and Carp, in Manning, *supra* note 60, could have reached a different conclusion based upon how they grouped their data).

⁶² Harris & Sen, *supra* note 46, at 248–49 (“[T]he literature suggests that differently situated judges might decide cases differently, but that any differences associated with demographics are actually fairly issue-specific and much less pronounced than differences rooted in ideology or partisanship.”).

⁶³ *Id.*

outcomes remains debated, the literature is clear that the ideology of the individual judge is a factor in judicial decision-making, including criminal sentencing.⁶⁴

5. Interaction Effects

In addition to examining gender, age, race, and ideology individually, a handful of studies have examined these variables in combination.⁶⁵ These studies have often produced nuanced findings. For instance, a study of Georgia judges found that older judges were more likely to be punitive than younger judges, but not when offenders were “disadvantaged.”⁶⁶ Similarly, a study of Pennsylvania state sentences imposed between 1999 and 2000 found that “minority judges are somewhat less punitive,” judicial gender does not likely affect sentencing, older judges are “less likely to incarcerate convicted offenders and sentenc[e] them to shorter periods of confinement,” a judge’s prior military experience increased the likelihood of incarceration, and the “tenure of judge was only marginally associated with increased sentence severity.”⁶⁷ A later study conducted by the same researcher using the same data found that older, female, and minority judges are “substantially less likely to sentence offenders to jail or prison terms.”⁶⁸ But a study looking at 440,000 Texas felony cases between 2004 and 2013 concluded that judicial ethnicity, gender, and political orientation had little effect on sentencing.⁶⁹

⁶⁴ See generally Cohen & Yang, *supra* note 56 (finding large discrepancies in federal criminal sentences imposed by Republican and Democratic district judges, with Republican judges being more likely to mete out longer sentences).

⁶⁵ Elaine Martin & Barry Pyle, *Gender, Race, and Partisanship on the Michigan Supreme Court*, 63 ALB. L. REV. 1205, 1222–25 (1999); Carol T. Kulik et al., *Here Comes the Judge: The Influence of Judge Personal Characteristics on Federal Sexual Harassment Case Outcomes*, 27 L. HUM. BEHAV. 69, 75 (2003); SUSAN B. HAIRE & LAURA P. MOYER, *DIVERSITY MATTERS: JUDICIAL POLICY MAKING IN THE U.S. COURTS OF APPEALS* (2015).

⁶⁶ See generally Martha A. Myers, *Social Background and the Sentencing Behavior of Judges*, 26 CRIMINOLOGY 649 (1988).

⁶⁷ Johnson, *supra* note 51, at 283.

⁶⁸ See generally Brian D. Johnson, *Judges on Trial: A Reexamination of Judicial Race and Gender Effects Across Modes of Conviction*, 25 CRIM. JUST. POL’Y REV. 159 (2014) [hereinafter Johnson, *Judges on Trial*].

⁶⁹ See generally Claire S.H. Lim et al., *Do Judges’ Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts*, 18 AM. L. & ECON. REV. 302 (2016).

D. Assessing the Literature

Taken together, the literature just reviewed makes clear that although we know much more than we did two decades ago, the relationship between judicial background and sentencing decisions remains unclear. And we know almost nothing about the interaction of demographic variables.

The literature is also unsatisfying in how it resolves the apparent differences between the punitiveness studies, which show some fairly strong patterns based on race, gender, and age, and the studies of real judges, in which these influences seem largely to disappear. One prominent explanation is that sitting judges have been intensely socialized—not just by their experiences on the bench, but also by all the legal experiences that led them to the bench.⁷⁰ Those experiences, according to this explanation, have powerfully taught judges to follow the law regardless of the proclivities they may have had before their intensive training. And, of course, the rule of law itself is grounded on the notion that we expect its operators to be neutral and fair, expectations that undoubtedly go a long way toward eliminating any general punishment biases based on judicial demographics.

Nevertheless, there continues to be a vigorous debate about whether, and why, some demographic biases seen in ordinary subjects seem to disappear—and in some cases reverse—in sitting judges. For example, in a 2014 paper, University of Maryland criminologist Brian Johnson took a second, somewhat more sophisticated statistical look at the 1991–1994 Pennsylvania state judges data, limiting himself to cases that went to trial, and he found that the judge’s race and gender had a larger impact than previously believed.⁷¹ As we do in this study, Johnson recognized that plea-bargained cases—which represent the vast majority of criminal cases—can also include sentence bargaining, which can reduce or even eliminate a judge’s discretion.

Johnson was also one of the few researchers to consider judicial age. Researchers’ more general failure to consider a judge’s age in analyzing sentencing patterns is surprising.⁷² Unlike some other individual characteristics that can be difficult to assess—including race and ethnicity, and perhaps increasingly even gender—age is readily

⁷⁰ See, e.g., Johnson, *Judges on Trial*, *supra* note 68, at 160.

⁷¹ *Id.*

⁷² There has been a similar reluctance to consider the effects of the *offender’s* (as opposed to the *judge’s*) age on sentencing.

determined.⁷³ Moreover, despite the recent trend of appointing younger and younger judges, America's judiciary is old and getting older. The average age of sitting federal judges is sixty-nine years old, older than at any other time in the country's history.⁷⁴ State judges are a little younger, averaging about sixty.⁷⁵ Our nation's judges are getting old, and we know almost nothing about the effects their aging might have on their job performance in general and their sentencing practices in particular.

II. A Novel Approach: Studying the Interaction of Age, Gender, and Crime Seriousness in Real Criminal Cases

To address the gaps identified in the literature, we pursued a study that focused on real world criminal sentencing in the state courts of Colorado and that allowed us to examine the interrelated effects of harm levels and the sentencing judge's gender and age.⁷⁶ Our unique access to the large number of criminal sentences and judge demographics in our data base (almost 3,000 individual sentences) allowed us not only to analyze the individual impacts of these three variables, but also their interactive effects. Before we outline our methodology, we begin with an overview of the Colorado criminal system, the source of our rich dataset.

A. The Colorado Criminal System

As in most states, there are three primary categories of crimes in Colorado: petty offenses, misdemeanors, and felonies. Petty offenses include minor violations such as

⁷³ Although, as discussed below, we did have difficulty determining the ages of some of the judges in our dataset who had retired or died. *See infra* note 107.

⁷⁴ Francis X. Shen, *Aging Judges*, 81 OHIO ST. L.J. (forthcoming 2020). *Cf.* Albert Yoon, *Federal Judicial Tenure*, in THE OXFORD HANDBOOK OF U.S. JUDICIAL BEHAVIOR (Lee Epstein & Stefanie A. Lindquist eds., 2017) (noting that, though the average judicial age is rising, "[t]he average age at commission has declined, albeit modestly, from the Truman to Obama administration").

⁷⁵ The best available data on the age of state judges comes from law professors Stacey George and Albert Yoon. George and Yoon lead a project called "The Gavel Gap," in which they investigate whether the demographics of state court judges reflect the demographics of citizens in that state. Although not the focus of their analysis, they observed birth year data for 5,378 state judges (out of 10,295 in their total dataset). Based on this birth year data, they calculate average state judge's age to be 59.6, with a median age of 60 (max age 88).

⁷⁶ As we discuss in Part IV, the dataset did not allow us to code for offender characteristics, nor for judicial race.

littering⁷⁷ or thefts of less than \$50.⁷⁸ Misdemeanors include thefts of more than \$50 but less than \$2,000,⁷⁹ false reporting to authorities,⁸⁰ and third-degree assault (knowingly or recklessly causing bodily injury but not serious bodily injury).⁸¹ Felonies are the most serious of crimes, and they run the gamut from simple theft of more than \$2,000,⁸² to securities fraud,⁸³ to first- and second-degree assaults,⁸⁴ and to homicide.⁸⁵ The most minor of petty offenses carry only fines with no jail; all misdemeanors and the most serious petty offenses carry potential jail sentences; all felonies carry potential prison sentences.⁸⁶

The Colorado legislature has further divided almost all petty offenses, misdemeanors, and felonies into classes, which determine the potential length of any jail or prison sentence.⁸⁷ The most serious class within any category is Class 1. There are two classes of petty offenses, three classes of misdemeanors, and six classes of felonies. For example, third-degree assault is a Class 1 misdemeanor, meaning it is the most serious kind of misdemeanor. First-degree murder is a Class 1 felony, the most serious kind of felony.

⁷⁷ COLO. REV. STAT. § 18-4-511(4).

⁷⁸ *Id.* § 18-4-401(2)(b).

⁷⁹ *Id.* §§ 18-4-401(2)(c)–(e).

⁸⁰ *Id.* § 18-8-111(2).

⁸¹ *Id.* §§ 18-3-204(1)(b), (2).

⁸² *Id.* §§ 18-4-401(2)(f)–(j).

⁸³ *Id.* §§ 11-51-603(1)–(2).

⁸⁴ *Id.* § 18-3-202(2) (first-degree assault); *id.* § 18-3-203(2) (second-degree assault).

⁸⁵ *Id.* § 18-3-102(3) (first-degree murder); *id.* § 18-3-103(3) (second-degree murder); *id.* § 18-3-104(2) (manslaughter); *id.* § 18-3-105 (negligent homicide).

⁸⁶ One important and common categorical difference between felonies and non-felonies is that felonies are punishable by incarceration in *prison*, while non-felonies are punishable by incarceration in *jail*. *See id.* § 18-1.3-501(1)(b). In Colorado, prisons (except private ones) are state-run and state-funded, while jails are typically county-run and county-funded. Despite this sharp distinction between prison for felonies and jail for misdemeanors, Colorado trial judges can add a short jail sentence (up to ninety days) as a condition of felony probation. *Id.* § 18-1.3-202(1)(a).

⁸⁷ *Id.* § 18-1.3-401(1)(a)(III) (felonies); *id.* § 18-1.3-501(1)(a) (misdemeanors); *id.* § 18-1.3-503 (petty offenses).

Colorado is a so-called Model Penal Code sentencing state, as opposed to a sentencing-guideline state. This means that the presumptive ranges of potential jail or prison sentences are set by the legislature, typically by class of crime, and that Colorado judges generally have discretion to impose a specific point sentence (in days, months, or years) within those presumptive ranges. Likewise, Colorado trial judges generally have discretion whether to impose incarceration or non-custodial sentences such as fines, probation, or, for felonies only, halfway-house sentences.⁸⁸ There are some special circumstances related to the crime or the criminal that can eliminate some or all of this discretion, and which can make incarceration mandatory and even a certain amount of incarceration mandatory.⁸⁹

The only exceptions to the rule that when Colorado judges impose custodial sentences they must impose a point sentence within the legislature's established ranges are certain kinds of sex offense felonies, in which any prison sentence must be indeterminate.⁹⁰ In those cases, if the judge imposes prison, they must impose a minimum within the statutory range. However, the sentence itself is indeterminate, meaning it is entirely up to the parole board to decide if the defendant will ever be released, and thus it is potentially a life sentence.⁹¹

Finally, there are two kinds of Colorado state trial judges—county judges and district judges—both of which were included in our dataset. County judges sit in every one of Colorado's sixty-four counties⁹² and have statutorily limited civil and criminal jurisdiction. On the criminal side, they hear only petty offense and misdemeanor cases.⁹³

⁸⁸ See *infra* text accompanying notes 103–05 (discussing that halfway-house sentences present unique challenges in terms of how to count such a sentence).

⁸⁹ For example, certain crimes are designated “crimes of violence,” which must be punished with prison in a range higher than the presumptive range. COLO. REV. STAT. § 18-1.3-401(11). See *infra* text accompanying note 117 (discussing whether these mandatory sentences infect the reliability of our findings).

⁹⁰ COLO. REV. STAT. § 18-1.3-1004.

⁹¹ See *infra* note 105 (discussing how these kinds of indeterminate sentences also presented unique challenges in terms of how to count them).

⁹² COLO. CONST. art. VI, § 16 (creating county courts); COLO. REV. STAT. § 13-6-101 (establishing them in every county).

⁹³ COLO. REV. STAT. § 13-6-106(1)(a).

District courts are courts of general jurisdiction, and district judges sit in every one of Colorado's twenty-two judicial districts.⁹⁴ On the criminal side, they hear only felonies.⁹⁵

B. Methodology

Our dataset initially included all criminal sentences imposed by Colorado state judges in the sixteen-year period from January 1, 2001, through December 31, 2016.⁹⁶ Because our study is aimed at the sentencing discretion different judges employ, and because many Colorado state judges regularly engage in sentence bargaining—where the sentence itself is also the product of the parties' plea bargaining and thus the judge's sentencing discretion is limited or even eliminated—we excluded all plea-bargained cases, and with the use of a filter we downloaded only cases that resulted in guilty verdicts after trial.⁹⁷ For similar reasons, we also excluded cases in which the judge did not have meaningful sentencing discretion, such as first-degree murder convictions for which the mandatory sentence is life without parole. We excluded all convictions for drug possession, escapes, and attempts of all kinds, because of the difficulty, and controversy, of assessing the "harm" these kinds of crimes cause.

After applying all of these filters, we were left with 2,955 separate sentences representing 183 different crimes imposed by 285 different judges, 180 male and 105 female.

⁹⁴ COLO. CONST. art. VI, § 9 (creating district courts); COLO. REV. STAT. § 13-5-101 (establishing them in judicial districts).

⁹⁵ District courts actually have concurrent jurisdiction with the county courts to hear misdemeanors and petty offenses. COLO. REV. STAT. § 13-6-106(1)(a). But because the county courts have no jurisdiction over felonies, as a practical matter most district courts have become felony-only criminal courts. District judges in our dataset may nevertheless have imposed misdemeanor (or even petty offense) sentences if their defendant was charged with a felony and a misdemeanor (or petty offense) and found guilty of the misdemeanor (or petty offense), or if the jury returned a verdict of guilt on an uncharged misdemeanor (or petty offense) on which it was instructed as a lesser offense.

⁹⁶ The sentencing data was derived from the official electronic Colorado state records of these proceedings. Although Colorado state criminal courts did not move entirely from paper files to electronic ones until January 2017, they had always recorded certain information electronically, including the sentence date, the crime of conviction, the sentencing judge's name, and the length of the sentence. It was this electronic information we downloaded to populate our dataset.

⁹⁷ See *supra* text accompanying notes 71–72 (discussing that criminologist Brian Johnson also made the decision to exclude plea-bargained cases when he re-examined the 1990s Pennsylvania sentencing data in his 2014 paper, for the same reason).

For each of these 2,955 sentences we recorded the date of the sentence, the crime for which the defendant was being sentenced,⁹⁸ the statutory classification for that crime (petty offense, misdemeanor, or felony, and the class of crime within those categories),⁹⁹ the level of harm for that crime (low, medium, or high), the judge's gender, the judge's age at the time of the sentence, the judge's length of time on the bench at the time of the sentence, and the length of the sentence.¹⁰⁰ We recorded no other demographic information about the sentencing judge¹⁰¹ and only limited demographic information about the defendant being sentenced.¹⁰² What follows are summaries of how we measured the length of sentence, the harm, and the judge's age, as well as a discussion of the problem that some of our judges imposed substantially more sentences over this time period than others.

⁹⁸ What we recorded was the actual statutory citation to the crime for which the defendant had been convicted.

⁹⁹ See *supra* text accompanying notes 77–86 (discussing these classifications under Colorado law).

¹⁰⁰ See *infra* text accompanying notes 102–06 (discussing the conventions we used to count the length of the custodial sentences).

¹⁰¹ There was simply no such information recorded in the official electronic files we used to populate our data base, and we decided not to make guesses, or to search other extrinsic sources, for judge demographics. See *infra* text accompanying notes 112–15 (discussing this limitation).

¹⁰² The official electronic files we used to download our data had spotty and sometimes ambiguous information about the defendant's race or ethnicity. There is a place in the formatting of those records to record the defendant's "race," but our understanding is that that record is made by data-entry personnel in the clerk's office who base it on their review of the complaint, the affidavit in support of the arrest warrant, and/or the probable cause statement. These documents sometimes refer to a defendant's race or ethnicity, but much more often they do not. As a result, the "race" entries in a large percentage of these electronic files are blank. Even when they have been populated, they reflect the usual tension between race and ethnicity. For example, and especially in the early years, it appears that almost all Hispanic defendants were labeled as Caucasian, with no recording of their ethnicity. Of course, the boundaries between race and ethnicity are complex and controversial, as census officials who have been struggling with this issue for many years have come to realize. See, e.g., Jens Manuel Krogstad & D'Vera Cohen, *U.S. Census Looking at Big Changes in How It Asks About Race and Ethnicity*, PEW RES. CTR. FACTTANK (Mar. 14, 2014), <http://www.pewresearch.org/fact-tank/2014/03/14/u-s-census-looking-at-big-changes-in-how-it-asks-about-race-and-ethnicity/> [https://perma.cc/N55S-WRU7]. In any event, because of this incomplete information about defendants' race or ethnicity, we did not consider those demographics in our analysis. See *infra* text accompanying notes 112–15 (discussing this limitation).

1. Length of Sentence

We looked at initial sentences only, not re-sentencings after probation revocations, halfway-house revocations, reconsiderations of the sentences, or remands from appeal. We quantified each sentence by the length of custodial time the judge imposed in the initial sentence. Thus, if a defendant received probation coupled with a jail sentence, we counted that as the length of the jail sentence. We did not consider any pre-sentence confinement credit, nor any good-time or earned-time credits that jail or prison officials might award. We did not consider that some felony defendants would be released early on parole. We counted jail time the same as prison time and did not consider any mandatory parole period a defendant sentenced to prison was required to serve once released. We made all these definitional choices in an effort to focus on the one phenomenon we were trying to measure: the sentencing discretion of individual trial judges. For example, although the amount of presentence time a defendant must be credited is certainly an important matter for the defendant, that time is a simple reflection of the defendant being unable to post bond pending trial and not a reflection of the judge's sentencing discretion. Similarly, parole is mandatory in Colorado and its length is determined by the level of the felony of which the defendant is convicted; the trial judge has no discretion in this regard.

As has been done in other studies,¹⁰³ we counted all halfway-house sentences as 120 days, roughly reflecting the “custodial” time most of those sentences represent.¹⁰⁴ We counted indeterminate sentences as the minimum of the indeterminate range that was imposed plus ten years.¹⁰⁵

¹⁰³ Morris B. Hoffman et al., *An Empirical Study of Public Defender Effectiveness: Self-Selection by the Marginally Indigent?*, 3 OHIO ST. J. CRIM. L. 223, 235, 235 n.55 (2005).

¹⁰⁴ Typical halfway-house defendants serve around four months in what is known as the “residential” phase of the program, regardless of how many years the halfway sentence is. They are, in effect, under house arrest and may not leave the facility. This residential status eventually transitions into work release and then into non-residential status, where the defendant lives off-site and is at that point, for all functional purposes, simply on probation for the balance of the halfway-house sentence.

¹⁰⁵ In the prior study, see Hoffman et al., *supra* note 103, we counted indeterminate sentences differently than we do here. At the time of that earlier study, indeterminately sentenced sex offenders were simply not being released at all on parole. See, e.g., Phillip Cherner, *Felony Sex Offender Sentencing*, 33 COLO. LAW. 11, 18 (2004) (explaining that as of 2004, eight years after indeterminate sentencing began, not a single inmate had been paroled from an indeterminate sentence). We therefore counted indeterminate sentences as 25% of the difference between 110 years (which is how we arbitrarily quantified a life sentence) and the minimum of the indeterminate range. Hoffman et al., *supra* note 103, at 235, 235 n.56. But since that study, parole officials

Finally, we counted each sentence separately, even if they were imposed in the same case and run concurrent to one another.¹⁰⁶

2. Harm Ratings

For our harm ratings, and following the conventions used in many third-party punishment studies, we divided all the sentenced crimes into one of three levels of harm—low, medium, or high—based on the severity of the offense. We did that by dividing the crimes into categories, not by looking at the individual facts in all 2,955 sentences. We used the statutory classifications of the crimes to put them into these three harm categories:

Type/Class	Harm Level
Class 1, 2, or 3 felonies.....	High
Class 4, 5, or 6 felonies.....	Medium
All petty offenses and misdemeanors.....	Low

3. Age

Our 285 different judges' average age was 54.63 years.¹⁰⁷ We decided to use just two age splits divided by that median, because creating smaller age groups would have significantly reduced our statistical power.¹⁰⁸ Thus, our “older” judges were those who

have been releasing these defendants much more frequently than before. COLO. DEP'T OF CORRECTIONS ET AL., LIFETIME SUPERVISION OF SEX OFFENDERS: ANNUAL REPORT (2018), <https://drive.google.com/file/d/0B21TrpBx507cVnVYRHVuWJKRk5BQzM1amxXS0kyNkFNnVBV/view> [<https://perma.cc/YF9A-FP5W>]. This change in how we counted those sentences was designed roughly to reflect this newer parole reality.

¹⁰⁶ This is also a different convention than we used in the prior study, where we assumed concurrent sentences. Hoffman et al., *supra* note 103, at 234, 234 n.52. But in that study we were interested in the sentences from the defendants' points of view—how much actual time will they have to serve. Although that question is, of course, also pertinent to measuring the judge's exercise of discretion in sentencing, we simply could not pull every one of our 3,000 sentences and manually determine which ones were concurrent and which consecutive.

¹⁰⁷ This is lower than the national average of about sixty. *See supra* note 75 and accompanying text. There were a handful of judges whose ages or dates of appointment we were unable to determine. We excluded their sentences.

¹⁰⁸ *See infra* text accompanying notes 113–16 (discussing this statistical power problem).

were 54.6 and older at the time of the sentence. Our “younger” judges were those who were younger than 54.6 at the time of the sentence.

Our female judges were on average a little younger (mean \pm standard deviation: 51.82 ± 7.99 years) than our male judges (55.72 ± 6.08 years). We decided to use the 54.63 average age without distinguishing between gender, both because that approach gave us more statistical power than using two different age averages and because we intended in any event to examine the interaction of age and gender.

We also looked at judges’ length of service on the bench, in contrast to just their chronological age. We anticipated, as some of the punitiveness literature suggests and as we have discussed,¹⁰⁹ that any sort of age or gender bias might be culturally wrung out of judges over time as they get comfortable with their roles and gain experience. We therefore anticipated that a judge’s time on the bench would be a more important factor than his or her chronological age. We were surprised to discover that length of service was not correlated to any of our sentence outcomes. We therefore used length of service only as a covariate.

4. Collapsing Multiple Sentences by the Same Judge

There was a wide variation in the number of trials conducted, and therefore the number of sentences imposed, by any given judge among our 285 different judges, ranging from a single sentence over this 16-year period to 74 sentences imposed by a single county judge. A judge’s appointment date and duration on the bench may largely account for this difference. But even within a single year there was wide variation in the number of sentences imposed by individual judges. Some judges, mostly in busy urban courts, tried many more cases in a given year than other judges. In addition, there was a wide annual variation even within busy urban courts, probably because the judges in some of those courts sit in different subject-matter divisions and rotate among those divisions. Judges in Denver, for example, might not try a single criminal case for several years during their assignment to non-criminal divisions (divorce and civil).

Because we were concerned that the wide variations in the number of sentences imposed by a given judge might skew our data, we collapsed and averaged all multiple sentences imposed by the same judge in the same calendar year within the same harm level. We then kept track of, and controlled for, these multiple sentences by using the sentence multipliers as a covariate.

¹⁰⁹ See *supra* text accompanying notes 69–73.

This left us with 1,215 separately-counted sentences, 921 of which were imposed by male judges and 294 by female judges; 608 of which were imposed by older judges, 607 by younger judges; 581 of which were low harms, 331 medium, and 303 high.¹¹⁰ See Table 1.

Table 1. Description of Judges Included in Data Set, by Harm Level, Age, and Gender

Harm Level	Gender	Age	Number of Judges
Low	Female	Younger (< 54.6 years)	111
		Older (\geq 54.6 years)	41
	Male	Younger	197
		Older	232
Medium	Female	Younger	52
		Older	27
	Male	Younger	105
		Older	147
High	Female	Younger	38
		Older	25
	Male	Younger	105
		Older	135

III. Analysis and Results

We analyzed these 1,215 individual sentences using sentence length as the dependent variable; the offense harm level, the judge's age, and the judge's gender as the independent variables of interest; and the sentence multipliers and judge's length of service as additional covariates. In this Part, we discuss the results of that analysis and a series of cautions in interpreting the data.

A. Results

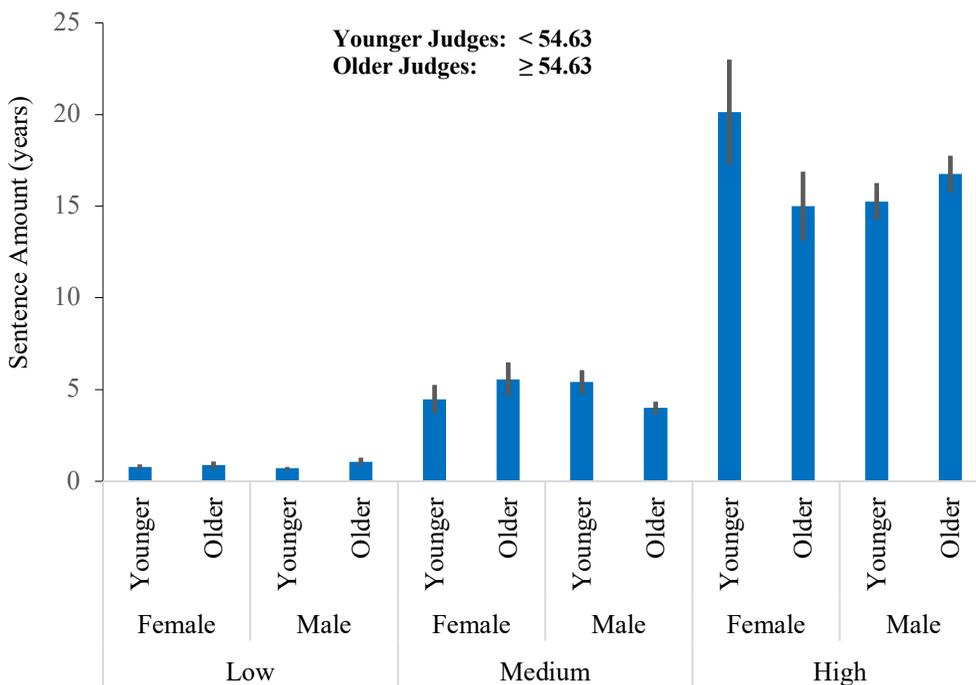
If we ignore age entirely, male judges sentenced their defendants roughly the same as their female colleagues. If we ignore gender entirely, older judges sentenced their

¹¹⁰ Because male judges tended to have more longevity, this manipulation also marginally improved our male-female imbalance. See *infra* text accompanying notes 114–17 (discussing that imbalance).

defendants roughly the same as their younger colleagues. By “roughly the same” we mean there was no statistically significant difference between the two sets of sentence distributions.¹¹¹ Similarly, when we look at all cases without distinguishing harm levels, there was no significant interaction between age and gender. That is, averaged across all types of harm levels, younger female judges do not sentence differently than anyone else.

But when we also considered harm level, we found a significant three-way interaction between harm level, judicial age, and judicial gender, which is reflected in Figures 1 and 2. Younger female judges sentenced defendants in high harm cases to sentences that averaged approximately 24% (4.9 years) longer than sentences imposed by their male colleagues ($p < 0.05$) and approximately 25% (5.1 years) longer than those imposed by their older female colleagues ($p < 0.05$). Figure 1 shows the average sentencing amounts for all the possible combinations of our age, gender, and harm variables.

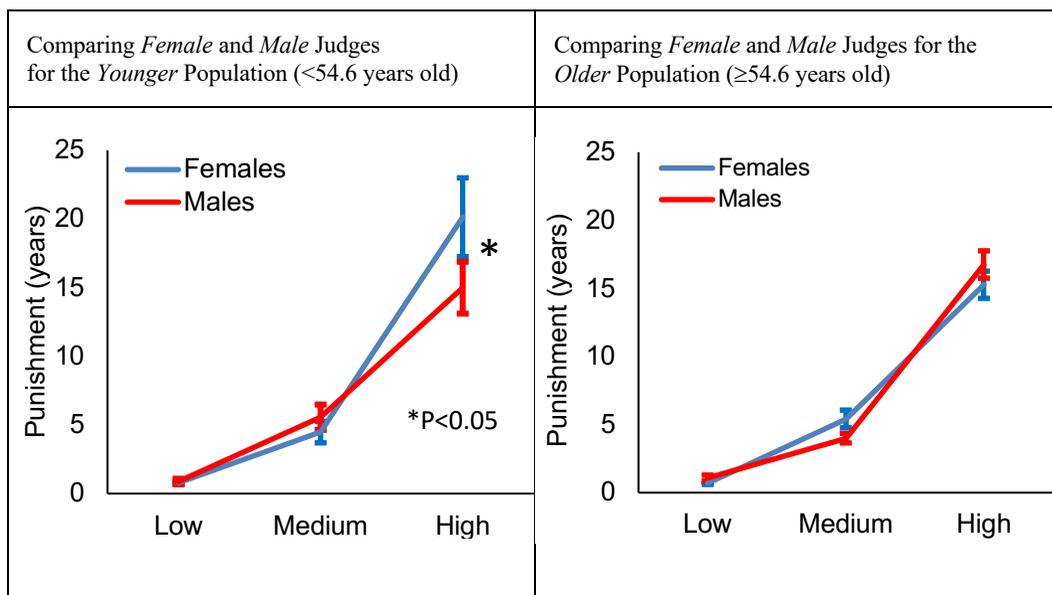
¹¹¹ “Significance,” as measured by p-values, is a statistical measure of the likelihood that a correlation is the product of chance. P-values less than .05 are sometimes colloquially described as representing a “significant” correlation; p-values less than .001 as a “highly significant” correlation. As set forth in the paragraph in the text immediately below this signal, our three-way interactions had p-values of less than .05, meaning they are statistically “significant.”

Figure 1. Average Punishment Results Separated by Age, Gender, and Harm

In the low and medium harm cases, all the judges impose statistically indistinguishable sentences, regardless of age or gender. But in the high harm cases, one group stands out: younger female judges.

This result is depicted more starkly in Figure 2, which compares men and women judges, with the younger judges on the left graph and the older ones on the right.

Figure 2. Comparison of Criminal Sentencing by Female and Male Judges, by Crime Seriousness



This Figure compares average criminal sentences, measured in years on the y-axis, by low, medium, and high crime harm levels indicated on the x-axis. The primary result to note is on the left-hand side of the Figure, where the data show a statistically significant ($p < 0.05$) difference between average sentences by younger female as compared to younger male judges. Younger female judges sentence their defendants, on average, to 4.9 (~24%) more years than their younger male counterparts for high-harm crimes. For older judges, there are no significant differences in sentencing.

B. Cautions in Interpreting the Results

Before we discuss the possible explanations and implications of these results in Part IV, it is important to acknowledge the limitations of our study and raise cautions about interpreting its results.

Perhaps most importantly, we were not able to code for a host of additional variables likely to affect sentencing outcomes. Unobserved variables in this data include race or ethnicity of the judge, ideology and partisanship of the judge, race or ethnicity of the offender, the defendant's criminal history, and the specific nature of the crime—although, of course, the nature of the crime is partially captured by our harm assessments. Because we did not have access to these additional variables, which the literature identifies as

likely affecting criminal sentences,¹¹² we need to recognize that our results here do not paint a complete picture of the factors that drive judicial sentencing.¹¹³

That said, our primary finding would be most threatened if we had reason to believe that offenders with certain criminal histories and case facts were systematically being assigned to or away from younger female judges. We do not believe this to be the case, and indeed in all multi-judge districts in Colorado, cases are generally assigned randomly.

A second limitation comes from the requirements of statistical power. Because we were attempting to measure the interaction of three different independent variables—two with two conditions (gender and age) and one with three conditions (harm levels)—and their effects on the dependent variable of sentence length, the statistical power was not great enough to further investigate other possible interactions. The story of sentencing is likely to be more nuanced; for instance, it could be that the effects we observed are driven by very young (or very old) judges. Power considerations prevented us from slicing our age cohorts any more finely than above and below the mean of 54.6 years old. Likewise, it is entirely conceivable that only a certain kind of case—for example, sex assaults or crimes against children—are driving the three-way interaction we discovered, a fact that would be very important as we speculate about what the interaction means. Again, power considerations prevented us from delving in any more detail into the nature of the crimes beyond sorting them into high, medium, and low harm levels.¹¹⁴

Further, we not only had an unequal number of judges by gender (180 men to 105 women), but on average our male judges were older than our female judges. Our male judges also imposed more sentences per judge than our female judges (a total of 921

¹¹² See, e.g., Rodney L. Engen et al., *Modeling the Effects of Legally Relevant and Extralegal Factors Under Sentencing Guidelines: The Rules Have Changed*, 38 CRIMINOLOGY 1207, 1217–18 (2000).

¹¹³ For instance, a study examined “the effects of emotional shocks associated with unexpected outcomes of football games played by a prominent college team in the state” on judge sentencing and found that a loss by the local college football team produced harsher juvenile sentences. See generally Ozkan Eren & Naci Mocan, *Emotional Judges and Unlucky Juveniles*, 10 AM. ECON. J.: APPLIED ECON. 171 (2018); Adam N. Glynn & Maya Sen, *Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women’s Issues?*, 59 AM. J. POL. SCI. 37 (2015) (finding that judges with daughters “consistently vote in a more feminist fashion on gender issues” than judges with just sons and that these trends are stronger with Republican judges).

¹¹⁴ In the hope that we might be able to delve further into the types of crimes, we coded every one of our 183 crimes for 6 characteristics: violent/non-violent; child/adult victim; sex/non-sex crime; death/non-death of victim; DUI; other driving offense.

sentences for the male judges and 294 for the females), further exaggerating this gender disproportion. Although we are confident the age/gender/harm interaction we discovered was statistically significant, we would be more confident in our results if the gender and age splits were closer. We attempted to take a smaller sample of male judges to equalize that split, but, once again, power limitations prevented us from meaningfully doing so.

It is possible that some or all of the harm piece in our three-way interaction is a statistical artifact springing from the fact that high-harm cases tend to have the widest and of course highest ranges for possible incarceration. The three-way interaction we observed could be just a two-way interaction between age and gender, which expresses itself only in high-harm cases simply because there is more room for it to be expressed in those cases. Of course, far from being a limitation, this observation would broaden the study's potential impacts.

By design, we examined only criminal sentences imposed after a trial, in order to maximize our judges' sentencing discretion. But by doing so, we necessarily ignored the ninety-plus percent of criminal cases that did not go to trial but were instead plea bargained.¹¹⁵ This not only means our criminal sample was a tiny slice of all criminal cases, it was a slice that has come under particular attention by scholars looking at what they call the "trial penalty." There are studies that suggest that judges sentence defendants more harshly after a trial than they sentence similarly-situated defendants who take a plea bargain.¹¹⁶ We are unaware of any literature suggesting this so-called trial penalty might be affected by a judge's age or gender, or by the seriousness of the offense, though that is a possibility we cannot eliminate.

¹¹⁵ In 2010 (the latest date for which data was available from this source), 97.4% of all non-dismissed federal criminal cases plea bargained. SOURCEBOOK OF CRIM. JUST. STAT. ONLINE, TABLE 5.22.2010: CRIMINAL DEFENDANTS DISPOSED OF IN U.S. DISTRICT COURTS (2010), <https://www.albany.edu/sourcebook/pdf/t5222010.pdf> [<https://perma.cc/8XM8-SSLE>]. In 2006 (the latest date for which data was available from this same source), an average of 94% of non-dismissed state criminal cases plea bargained. SOURCEBOOK OF CRIM. JUST. STAT. ONLINE, TABLE 5.46.2006: PERCENT DISTRIBUTION OF FELONY CONVICTIONS IN STATE COURTS (2006), <https://www.albany.edu/sourcebook/pdf/t5462006.pdf> [<https://perma.cc/Q3A5-VSJJ>].

¹¹⁶ See generally, e.g., Anthony Walsh, *Standing Trial Versus Copping a Plea: Is There a Penalty?*, 6 J. CONTEMP. CRIM. JUST. 226 (1990). But see generally Andrew Chongseh Kim, *Underestimating the Trial Penalty: An Empirical Analysis of the Federal Trial Penalty and Critique of the Abrams Study*, 84 MISS. L.J. 1195 (2015) (no trial penalty); Shi Yan & Shawn D. Bushway, *Plea Discounts or Trial Penalties? Making Sense of the Trial-Plea Sentence Disparities*, 35 JUST. Q. 1226 (2018).

Another challenge is that the judges imposing these sentences did not have unfettered discretion—the sentences they chose had to be within the statutorily mandated ranges.¹¹⁷ These ranges necessarily distorted the sentencing data. A given judge imposing the mandatory minimum sentence of ten years for an aggravated robbery might have imposed much less had there not been a ten-year floor. A different judge imposing the maximum of thirty-two years for an aggravated robbery might have imposed a longer sentence but for the thirty-two-year ceiling. We have no way of knowing whether these distortive effects were felt equally across age, gender, or even harm levels. Again, though, we have no reason to believe these kinds of cases with mandatory sentences to incarceration were distributed anything but randomly as among our judges.

IV. Discussion

Recognizing all the caveats in Part III.B, if our results can nevertheless be confirmed by future studies, they could have significant theoretical, practical, and policy implications.

Criminal sentencing is a core function of the judiciary, yet the individual judicial background factors that drive those sentencing decisions remain poorly understood. Any insights we can gain about this black box are important in their own right. Moreover, as both a theoretical and methodological matter, our results suggest that future empirical studies of judicial behavior recognize that as judges age, their decision-making may change in ways that are relevant for criminal sentencing.

The age dynamic emerged in our data only when the sentencing judge was female and the harms high. Why? Our data do not answer that tantalizing question, and we do not intend by the discussion that follows to suggest that we have any concrete answers. But because considering *possible* explanations will inform where future research might need to be aimed, we offer a few speculative explanations before we turn to implications.

A. Possible Explanations

The three-way nature of the interaction suggests that no single explanation will be sufficient. As already mentioned, the simple explanation that judges become less harsh as they gain judicial experience does not explain these results, because it was only female judges who became less harsh. And they didn't become less harsh across the board, they only became less harsh in high-harm cases. Moreover, our statistical analysis makes clear

¹¹⁷ See *supra* text accompanying notes 87–88.

that it is the judge's age, not judicial experience, that drives this interaction. Likewise, any suggestion that some kind of broad generational difference might explain these results must somehow account for the fact that women, but not men, seem to express that generational difference, and only for high harms. Similarly, explanations that account for gender differences alone will be inadequate. Theories that emphasize differences in male and female judges' sentencing patterns cannot explain why older female judges sentence the same as male judges.

Finally, any explanation that properly considers both age and gender must explain why this age/gender interaction is seen only for high harms. "For such-and-such reason women, but not men, get harsher as they age" does not explain why our judges, men and women alike, sentenced low and medium harms roughly the same. Neither would the theory, supported by some occupational studies,¹¹⁸ that young women judges might be trying to fit into a traditionally man's world by acting more like men.

It is possible that our three-way interaction is more nuanced than our data allowed us to investigate. For instance, data constraints required us to bifurcate judges into two groups: older and younger. But the interaction may look different if age were examined as a continuous variable. It is likewise possible that it is not driven by all high-harm crimes, but only by a few kinds of crimes that particularly threaten the integrity of the family—burglary, domestic violence, or sex assault on a child, for example. If these two sets of sub-factors are actually driving our findings, then they may be part of a literature showing women tend to be more punitive than men when it comes to certain kinds of antisocial norm violations, namely men-on-women violence and crimes against children.¹¹⁹ As far as we know, none of that literature examined whether this phenomenon is more robust among younger women, but it is not implausible to think so.

Our single study cannot distinguish any of these speculative explanations, but it does make clear the need for further empirical research in this area. Specifically, future research should (1) explicitly consider the effect of judicial age and gender on sentencing outcomes, perhaps by following individual judges longitudinally to see if their sentencing practices change over time; (2) look at more refined age splits than the fifty-five year median we were forced to use in our data; (3) examine the interaction of judicial age with other judicial demographics, including race and ideology; (4) attempt to determine

¹¹⁸ See generally Abdelmagid M. Mazen & Jeanne Parr Lemkau, *Personality Profiles of Women in Traditional and Non-traditional Occupations*, 37 J. OCCUPATIONAL BEHAV. 46 (1990).

¹¹⁹ Michelle K. Ryan et al., *Who Cares? The Effect of Gender and Context on the Self and Moral Reasoning*, 28 PSYCHOL. WOMEN Q. 246, 250 (2004).

whether this interaction is influenced by the judge's own family structure and/or by certain kinds of family-significant crimes; and (5) attempt to include defendants' demographics as a factor influencing sentencing and interacting with the judges' demographics.¹²⁰

If, after these follow-ups, our results generally hold, then they could have practical and perhaps even policy significance. Here, we sketch a few of those potential implications.

B. Implications

No one with an even rudimentary awareness of human nature will be surprised to learn that “judge shopping,” the process by which attorneys attempt to find the judge most likely to be favorable to their client, is a standard part of litigation practice.¹²¹ Indeed, judge shopping has been institutionalized in a handful of states, which allow each side in a case (criminal or civil) a one-time option to remove the assigned judge and have a new one assigned to the case.¹²² Regardless of the folk psychology lawyers bring to bear in making this decision—whether they think, for example, that former criminal defense lawyers will be more lenient or less lenient than, say, former prosecutors—our results suggest they also need to consider the judge's age and gender.

Even in jurisdictions that do not allow lawyers to peremptorily challenge judges, our results may have application in jury selection. In those handful of states that allow jurors

¹²⁰ Future work should also examine whether the phenomenon we uncovered in a criminal sentencing context has any application beyond the criminal law. For example, do trial judges (and jurors) exhibit a similar three-way interaction when deciding quasi-criminal issues, such as whether to award punitive damages?

¹²¹ Theresa Rusnak, *Related Case Rules and Judge-Shopping: A Resolvable Problem?*, 28 GEO. J. LEGAL ETHICS 913, 923 (2015) (“As long as different judges continue to hand down inconsistent sentences for similar crimes, attorneys will have incentives to judge shop to obtain the most favorable outcomes for their clients, whether that be through marking cases as related or refusing to disclose judicially-rejected plea agreements.”).

¹²² ALASKA STAT. § 22.20.022 (2002); ALASKA R. CIV. P. 42(c) (2009); ARIZ. R. CIV. P. 42.1; ARIZ. R. CRIM. P. 10.2; CAL. CIV. PROC. CODE § 170.6 (2004); IND. CODE ANN. § 235-36.5-1; MONT. CODE ANN. § 3-1-805 (2003); MO. R. CIV. P. 51.05; NEV. SUP. CT. R. 48.1; N.M. STAT. ANN. § 38-3-9; N.M. R. CIV. P. 1-088.1; N.M. CRIM. P. 5-106; WIS. STAT. ANN. § 971.20 (1998) (applying to criminal cases only); WYO. R. CIV. P. 40.1(b) (2019). Some states permit peremptory challenges to judges designated by special appointment, such as visiting judges, but do not extend the privilege to the trial courts generally. *See, e.g.*, TEX. GOV'T CODE ANN. § 74.053 (2003). Peremptory challenge procedures have been proposed for federal judges but have never been adopted. *See, e.g.*, Peremptory Challenge Act of 2011, H.R. 3196, 112th Cong. (2011).

to impose criminal sentences,¹²³ prosecutors and defense lawyers in high-harm cases may want to think about their prospective jurors' ages and genders. Even in the vast majority of jurisdictions where judges, not jurors, impose criminal sentences, our results may nevertheless have some application in jury selection. Aspects of the three-way interaction we saw in a sentencing context may spill over into jurors' roles as factfinders. Perhaps lawyers already sense that female jurors may have special feelings about crimes that threaten families, feelings that might impact their verdicts; they may not realize that, if our results hold up in such circumstances and extend beyond sentencing judges to jurors, those feelings may tend not to be shared even by men with children, and may tend to disappear in women as they get older.

Our results might even inform debates about judicial selection and mandatory retirement. Those relatively young people appointed by recent administrations to the federal trial bench, mentioned at the beginning of this Article,¹²⁴ may sit for a long time, but our study suggests that young male appointees will not be as tough on serious crime as their young female colleagues, and that even the women will tend to age out of their sentencing toughness.

Do our results require any kind of policy response? Not necessarily. On the one hand, the interaction we discovered may be viewed simply as being a very small part of the large and changing canvas on which a judge's whole life plays out in the form of sentencing attitudes and practices. Not every demographic influence on judges, even if statistically significant, justifies policy changes. There is no doubt a host of other as yet undiscovered effects that arguably do not need policy corrections. What if future researchers discover, for example, that judges over six feet tall are harsher or easier sentencers than their shorter colleagues? Such a result might be important to lawyers as they contemplate which judges or jurors to peremptorily excuse, but we doubt anyone would suggest that this single factor should have much to do with policies surrounding judicial selection or retirement. Focusing on any particular interaction at the expense of other more salient policy considerations seems short-sighted. Finally, of course, who is to say that the younger female judges in our study were "wrong" for being harsher sentencers, and their male and older female colleagues "right" for being more lenient, as opposed to the other way around? Sentencing, by its very nature, is normative, and thus

¹²³ In noncapital felony cases, only five states—Arkansas, Missouri, Oklahoma, Texas, and Virginia—permit juries to make the sentencing decision. ARK. CODE ANN. § 5-4-103 (2019); MO. ANN. STAT. § 557.036 (2014); OKLA. STAT. ANN. tit. 22, §§ 926.1, 927.1 (2019); TEX. CODE CRIM. PROC. ANN. art. 37.07(2)(b) (2019); VA. CODE ANN. § 19.2-295 (2012). All noncapital sentencing in the federal courts is done by the judge. FED. R. CRIM. P. 32.

¹²⁴ See *supra* text accompanying note 1.

not generally susceptible to these kinds of comparative judgments.

On the other hand, the effects we saw were statistically significant and large. What do we tell a criminal defendant who wants to know why his young female judge sentenced him to a term five years longer than the term a similarly-situated defendant received from a male or older female judge? What do we tell the victim if the judge in her case was male or an older female and imposed a lenient sentence? Age and gender, unlike height, are significant, character-affecting traits whose broad influences should not be surprising to anyone, and which are therefore arguably game for policy considerations. Finally, there are arguments to be made that the younger female judges in this study were in fact “wrong” in their harsh sentences in the sense that they were the age and gender outliers. Their male and older female colleagues sentenced less harshly than they did, and significantly so. Moreover, it appears that, on average, those younger female judges will themselves tend to return to the norm of less harsh sentences as they age.

In the end, we doubt that any policymakers will, or should, use these results to pivot dramatically from any given policy trajectories. But our results probably will matter at the margins of these policies and in case-specific applications, as nominating authorities consider specific judicial candidates and perhaps even as legislatures consider proposed legislation setting or resetting specific mandatory judicial retirement ages.

Judicial education may also be an appropriate response. Just calling this data to the attention of trial judges could go a long way toward closing this age-gender sentencing gap. Young female judges apprised of this trend, especially if further research shows it is driven by a few kinds of crimes, may become less harsh with those crimes just knowing that they are outliers compared to their male and older female colleagues.

Conversely, presenting these results as part of judicial education may well cause older female judges to reassess their sentencing patterns for a few kinds of high-harm crimes and perhaps even recapture the sense of appropriate harshness they once had when they were younger judges. It might even cause male judges to reassess their relative leniency.

CONCLUSION

Our examination of a database of tried criminal cases in Colorado shows a three-way interaction not previously reported in the literature: Younger female judges sentence high-harm cases significantly more harshly than their male and older female colleagues. More work needs to be done to confirm these results and to ferret out details in finer age

cuts and subsets of high-harm crimes. But if they hold, researchers, practicing lawyers, judicial educators, and perhaps even appointing authorities and legislatures should take note.