IF THE JURY ONLY KNEW: THE EFFECT OF OMITTED MITIGATION EVIDENCE ON THE PROBABILITY OF A DEATH SENTENCE

Barry Edwards*

* Barry C. Edwards is the founder and director of Fair Trial Analysis, a non-profit research organization. He received his B.A. from Stanford University, a J.D. from New York University, and a Ph.D. from the University of Georgia. He is co-author of a popular series of political science research methods textbooks and continues to teach part-time for the University of Georgia. He sincerely thanks Indira Rajkumar, Marie Ceske, and the editors at the Virginia Journal of Social Policy & the Law for their thoughtful feedback, diligent cite checking, and helpful suggestions.

IF THE JURY ONLY KNEW: THE EFFECT OF OMITTED MITIGATION EVIDENCE ON THE PROBABILITY OF A DEATH SENTENCE

Barry Edwards

To substantiate a claim for ineffective assistance of counsel, a death row inmate must demonstrate that his counsel was deficient, and that this deficiency caused significant harm. Even when the petitioner can show his counsel failed to meet a professional standard of care, it has been difficult to prove that ineffective counsel caused significant harm. This Article addresses the challenges of proving that ineffective counsel's omission of mitigation evidence caused a reasonable probability of a different outcome in death penalty trials. Lawyers and judges tend to substitute their personal opinions for objective assessments of harm. Researchers' efforts to quantify harm suffer from several shortcomings: insensitivity to litigation context, analysis of non-representative data, and failure to account for the jury deliberation process. Even when an appellate court knows all the facts and identifies all mistakes made at trial, the effect of trial errors and omissions on the probability of a death sentence remains unknown.

INTRODUCTION

In 1976, David Leeroy Washington pleaded guilty to murder in Broward County, Florida. Prosecutors sought the death penalty. Washington's public defender spent a single day preparing for the hearing to decide between life and death sentences. Washington's friends and family had a chance to testify that he was generally a good person who was suffering personal hardships at the time of his crimes.[1] He was sentenced to death. In post-conviction proceedings, Washington argued that his attorney's inadequate investigation and preparation for sentencing violated his right to effective assistance of counsel.[2] After exhausting state processes, Washington petitioned the federal courts for a writ of habeas corpus. Although the Southern District of Florida denied his petition, the U.S. Court of

[1] Washington waived his right to have a jury make a recommendation during the sentencing phase and was sentenced to death by the state trial court judge. The state court's conviction and sentence were upheld by the Florida Supreme Court in 1976. *See* Washington v. State, 362 So. 2d 658 (Fla. 1978), *cert. denied*, 441 U.S. 937 (1979). In 1980, Washington filed an unsuccessful motion for post-conviction relief in state trial court and was denied clemency by the Governor of Florida. The Florida Supreme Court upheld the denial of postconviction relief. *See* Washington v. State, 397 So. 2d 285 (Fla. 1981).

[2] U.S. CONST. amend. VI.

Appeals for the Fifth Circuit vacated and remanded the District Court's decision.[3]

The U.S. Supreme Court eventually agreed to hear the case.[4] *Strickland v. Washington*, decided in 1984, established a key legal standard for evaluating whether a criminal defendant's Sixth Amendment right to effective assistance of counsel has been violated.[5] In *Strickland*, one of the Court's most influential and frequently cited opinions,[6] the Court articulated a two-pronged test for ineffective assistance of counsel claims. First, the petitioner must demonstrate that his trial counsel's performance was deficient, falling below an objective standard of reasonableness. Second, the petitioner must prove that this deficiency prejudiced the defense by demonstrating that there is a "reasonable probability" that, if not for counsel's errors, the outcome of his trial would have been different.[7] The Court clarified that a reasonable probability of a different outcome is more than a mere possibility; it is "a probability sufficient to undermine confidence in the outcome."[8] Despite being able to show his attorney's failure to adequately investigate and present mitigating circumstances was deficient performance under the first prong of the Strickland test, Washington could not prove that the attorney's error was prejudicial. His federal habeas petition was denied, and he was executed later that year.[9]

---

[3] *See* Washington v. Strickland, 673 F.2d 879 (5th Cir. 1982), *aff'd en banc*, 693 F.2d 1243 (5th Cir. 1982).

[4] The U.S. Supreme Court granted certiorari and reversed the Court of Appeals in 1984. *See* Strickland v. Washington, 466 U.S. 668, 668 (1984), *reh'g denied*, 467 U.S. 1267 (1984). The Supreme Court rendered its decision only eight years after Washington was sentenced to death.

[5] *Id.* at 693-94 (stating that defendant/petitioner must show that there is a "reasonable probability" that the trial would have had a different outcome but for the trial error).

[6] One study lists *Strickland v. Washington* as the case most cited by U.S. Circuit Courts of Appeals. *See* Frank B. Cross & James F. Spriggs, *The Most Important (and Best) Supreme Court Opinions and Justices*, 60 EMORY L.J. 407, 434 tbl.2 (2010). According to Google Scholar, the opinion has been cited 159,098 times (as of September 2024), which amounts to approximately eleven citations *per day* since its publication. *See Strickland v. Washington: How Cited*, GOOGLE SCHOLAR, https://scholar.google.com/scholar_case?about=16585781351150334057 (last visited Sept. 30, 2024).

[7] *Strickland*, 466 U.S. at 694.

[8] *Id.* Justice Souter offers insightful commentary on the meaning and application of the "reasonable probability" standard, suggesting that "significant possibility" would be clearer and less prone to misinterpretation. *See* Strickler v. Greene, 527 U.S. 263, 297-301 (1999) (Souter, J., concurring in part and dissenting in part).

[9] Washington filed a second unsuccessful petition for writ of habeas corpus in 1984. *See* Washington v. Wainwright, 587 F. Supp. 525 (S.D. Fla. 1984), *aff'd*, 737 F.2d 922 (11th Cir. 1984). Washington was executed in 1984. *See* Jesus

The *Strickland* prejudice standard appears straightforward: A death row inmate must prove that errors or omissions in his trial created a reasonable probability of different outcome. In post-conviction proceedings, a death sentence is presumptively valid and should stand if the petitioner cannot convince the court that the omitted mitigation evidence would have altered the trial's outcome. If a reasonable probability of a different outcome exists, the defendant did not receive a fair trial and is entitled to relief. Conversely, if there is little to no probability of a different outcome, the defendant's trial was sufficiently fair, and his sentence should stand.

Proving the prejudicial effect of omitted evidence is a challenging task, and *Strickland* claims where attorneys did not present mitigating evidence to the jury usually fail the prejudice test. In direct appeals, courts can evaluate the harmfulness of errors that occurred in the defendant's trial. In post-conviction proceedings, however, the inquiry may shift to errors of omission and the potential impact of evidence that was never presented at trial. *Strickland*'s prejudice requirement necessitates counterfactual reasoning — speculating about the likely outcome of a hypothetical trial in which the jury is presented with evidence that was never introduced during the actual trial. But how can one prove that the trial outcome would have been different if the jury had heard the omitted evidence?[10] This question lies at the intersection of trial strategy, appellate advocacy, and social science.

This Article explores the difficulty of proving the prejudicial effect of omitted mitigation evidence in death penalty cases. Part I examines the role of mitigation evidence in capital sentencing, how it serves to humanize defendants and justify mercy when jurors decide the appropriate sentence for murderers. Part II critiques existing legal frameworks for proving prejudice, underscoring the disconnect between subjective and objective methods of assessing the impact of omitted evidence on trial outcomes. Lastly, Part III explores the challenges of empirically evaluating whether omitted evidence could have altered a trial's outcome. This Part highlights the limitations of social science research in proving the harm caused by omitting evidence in capital cases, including not assessing the effect of omission in a case-specific context, analyzing non-representative data samples, and failing to account for the jury deliberation

---

Rangel, *Confessed Murderer of 3 Executed in Florida*, N.Y. TIMES 24 (July 14, 1984).

[10] This quandary raises the fundamental problem of causal inference: the counterfactual outcome (what would have happened) is never observed. *See* PHILIP H. POLLOCK III & BARRY C. EDWARDS, THE ESSENTIALS OF POLITICAL ANALYSIS 106-07 (6th ed. 2019); KOSUKE IMAI, QUANTITATIVE SOCIAL SCIENCE: AN INTRODUCTION 46-48 (2018). While counterfactual conditions are never directly observed, they may be approximated well enough to make interferences about causes and effects. The best example is a randomized experiment where a control group fairly approximates what would have happened to the treatment group but for the treatment.

process. The Article concludes with a call for more rigorous methodologies to estimate trial fairness.

The purpose of this Article is to identify the challenges petitioners face when attempting to prove that evidence never heard by the jury could have influenced the jury's verdict. This Article is not intended to undermine petitioners' efforts to obtain fair sentences by pointing out flaws in their evidence of harm. Instead, it aims to provide constructive criticism, highlighting both the strengths and weaknesses of their claims to help those who did not have adequate counsel better substantiate their arguments in post-conviction proceedings.

I. MITIGATION EVIDENCE AND THE DECISION TO IMPOSE A DEATH SENTENCE

Mitigating factors are circumstances that can reduce the severity of a defendant's sentence.[11] Evidence of mitigating factors generally decreases the probability of a death sentence.[12] Defendants may introduce a broad range of mitigation evidence, including evidence pertaining to their "background, record, or character," beyond mitigating factors enumerated in statutes.[13]

A significant number of defendants facing capital punishment have documented histories of severe abuse, neglect, or abandonment, which often contribute to their violent tendencies.[14] Severe and prolonged childhood trauma or abuse can derail a child's psychological and emotional development.[15] These experiences often impair a person's ability to regulate emotions, trust others, and form healthy relationships. Repeated exposure to violence or neglect may desensitize a child to aggression, normalize antisocial behavior, and disrupt cognitive development, leading to

---

[11] 18 U.S.C. § 3592(a).

[12] Indeed, all relevant evidence has some bearing on the outcome. *See* Fed. R. Evid. 401.

[13] 18 U.S.C. § 3592(a)(8).

[14] *See* David Lisak & Sara Beszterczey, *The Cycle of Violence: The Life Histories of 43 Death Row Inmates*, 8 PSYCH. MEN & MASCULINITY 118, 122-25 (2007) (reporting that all subjects suffered neglect, nearly all physically abused, more than half sexually abused, and substantial number "subjected to forms of sadism, public humiliation, and the unique degradation of being punished for manifesting the symptoms of an abused child"); David Freedman & David Hemenway, *Precursors of Lethal Violence: A Death Row Sample*, 50 SOC. SCI. & MED. 1757 (2000) (reporting family violence history in all cases, severe physical and sexual abuse in all but two cases).

[15] The negative effects of childhood trauma have been extensively documented. A prominent example is the Adverse Childhood Experiences (ACE) Study, a large, long-term study conducted by the Centers for Disease Control and Prevention and Kaiser Permanente. *See* Vincent J. Felitti et al., *Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study*, 14 AM. J. PREVENTIVE MED. 245 (1998).

impulsive and violent actions in adulthood.[16] Such individuals often develop maladaptive coping mechanisms, including substance abuse, emotional detachment, and heightened aggression, which can culminate in violent criminal behavior, including murder.[17]

Mitigating evidence is crucial in death penalty cases, as it contextualizes the defendant's violent behavior, highlighting the deep-rooted psychological scars that infect their actions and decision-making.[18] Empirical research demonstrates that mitigation evidence significantly influences whether jurors recommend the death penalty.[19] Likewise, testimony from family and friends during sentencing can humanize the defendant. According to Justice Thurgood Marshall, evidence that highlights a defendant's social and familial ties can sway the jury's decision whether to impose a death sentence.[20] If jurors empathize with the defendant — recognizing that he too has been a victim of violence, is still loved by his family, feels remorse, and can lead a purposeful life in prison — they may favor leniency. On the other hand, if the jury finds the defendant cold, calculating, and likely to kill again if given the opportunity, they may favor imposing a death sentence.

An attorney's failure to present mitigation evidence in capital sentencing can be a prejudicial error, as the omission of mitigating evidence generally increases the probability of a death sentence. Ineffective assistance of counsel can inadvertently lead to a death sentence. Many defendants are sentenced to death not solely based on the severity of their crimes, but because their defense counsel failed to investigate the defendant's biography and present mitigating evidence.[21]

---

[16] *See* James Garbarino, *ACEs in the Criminal Justice System*, 17 ACAD. PEDIATRICS S32 (2017); Carrie A. Pettus, *Trauma and Prospects for Reentry*, 6 ANN. REV. CRIMINOL. 423 (2023).

[17] *See supra* note 15.

[18] *See* William M. Bowen, Jr., *A Former Alabama Appellate Judge's Perspective on the Mitigation Function in Capital Cases*, 36 HOFSTRA L. REV. 805, 807–09 (2008); Sean D. O'Brien, *Death Penalty Stories: Lessons in Life-Saving Narratives*, 77 UMKC L. REV. 831 (2009).

[19] *See* David C. Baldus et al., *Arbitrariness and Discrimination in the Administration of the Death Penalty: A Legal and Empirical Analysis of the Nebraska Experience (1973-1999)*, 81 NEB. L. REV. 486, 548-55 (2002); Michelle E. Barnett et al., *When Mitigation Evidence Makes a Difference: Effects of Psychological Mitigating Evidence on Sentencing Decisions in Capital Trials*, 22 BEHAV. SCI. & L. 751 (2004).

[20] Strickland v. Washington, 466 U.S. 668, 718 (1984) (Marshall, J., dissenting). *See also* Gary Goodpaster, *The Trial for Life: Effective Assistance of Counsel in Death Penalty Cases*, 58 N.Y.U. L. REV. 299 (1983).

[21] This occurs most often when public defendants are inexperienced and underfunded. *See* Stephen B. Bright, *Counsel for the Poor: The Death Sentence not for the Worst Crime but for the Worst Lawyer*, 103 YALE L.J. 1835 (1994). Counsel's failure to investigate the defendant's history for evidence of childhood

Public defenders face significant challenges in investigating and presenting mitigating circumstances in death penalty cases, often due to limited resources and time constraints. Unlike private counsel, public defenders frequently handle large caseloads, which leaves them with little time or funding to obtain medical and educational records, interview family and friends, or retain appropriate experts.[22] Additionally, defendants in capital cases often have complex backgrounds involving trauma, mental illness, or substance abuse, so the defendant or their family members may be hesitant to offer cooperative testimony.[23] These hurdles make it difficult for public defenders to develop mitigation evidence that could potentially sway juries against imposing the death penalty.

While jurors may view certain defendants with more sympathy on account of their troubled personal backgrounds, evidence of childhood violence or drug abuse can be a "doubled-edged sword" from which jurors can draw conflicting inferences.[24] This type of evidence may help jurors understand the factors contributing to the defendant's criminal behavior, but it can also lead them to conclude that the defendant bears some responsibility for his own misfortunate, is unlikely to be rehabilitated, and will not lead a useful life in prison.[25]

Numerous petitioners have sought to demonstrate the impact of omitted mitigation evidence in cases reviewed by the U.S. Supreme Court.[26]

---

trauma is especially problematic in light of the documented link between trauma and violence and prevalence of traumatic histories among death row inmates.

[22] *See generally* Benjamin H. Barton & Stephanos Bibas, *Triaging Appointed-Counsel Funding and Pro Se Access to Justice*, 160 U. PA. L. REV (2012); Alex Bunin, *Public Defender Independence*, 27 TEX. J. CIV. LIBERTIES & CIV. RTS. 25 (2021).

[23] *See, e.g.*, Schriro v. Landrigan, 550 U.S. 465 (2007) (defendant directed counsel not to have ex-wife and birth mother testify at sentencing).

[24] *See, e.g.*, Hopkins v. Cockrell, 325 F.3d 579, 585-86 (5th Cir. 2003) (evidence of defendant's brain injury and substance abuse is double-edged); Boyle v. Johnson, 93 F.3d 180, 187-88 (5th Cir. 1996) (evidence of defendant's alcohol abuse and violent childhood is doubled-edged); Mann v. Scott, 41 F.3d 968, 983-84 (5th Cir. 1994) (evidence of defendant's abusive childhood and low intelligence is doubled-edged).

[25] *See* Margaret C. Stevenson et al., *Jurors' Discussions of a Defendant's History of Child Abuse and Alcohol Abuse in Capital Sentencing Deliberations*, 16 PSYCH. PUB. POL'Y & L. 1, 25-30 (2010).

[26] The omission of mitigation evidence of childhood trauma by ineffective counsel is a recurring issue in death penalty appeals heard by the Supreme Court. Of course, the issue is not limited to the U.S. Supreme Court; there are likely a hundred times as many cases in U.S. Circuit Courts of Appeals and state courts of last resort (based on the Supreme Court granting approximately one percent of certiorari petitions). Many ineffective assistance of counsel cases address omitted mitigation evidence of childhood trauma, but petitioners allege other deficiencies. *See, e.g.*, Harrington v. Richter, 562 U.S. 86 (2011) (failure to retain expert testimony on blood evidence); Buck v. Davis, 580 U.S. 100 (2017) (effect of race-laden expert testimony introduced by ineffective counsel); Weaver v.

The Court has consistently heard death row inmates argue that their attorneys failed to properly investigate and present mitigation evidence during sentencing. Table 1 shows, for cases decided since 2000, petitioners, the sentencing court, the type of mitigation evidence that was omitted, and the Supreme Court's decision whether the omission led to prejudice under *Strickland*.

U.S. SUPREME COURT DECISIONS ABOUT MITIGATION EVIDENCE OMITTED FROM DEATH PENALTY TRIALS, 2000 TO PRESENT

Petitioner Name	*Death Sentence*	*Omitted Evidence**	*Supreme Court Decision*
Terry Williams	Va., 1986	CTA, TBI, MH, ID	Omission prejudicial.[27]
Gary Cone	Tenn., 1982	Mitigation evidence and closing argument	Counsel not deficient.[28]
Kevin Wiggins	Md., 1989	CTA, ID	Omission prejudicial.[29]
Ronald Rompilla	Pa., 1988	CTA, MH, DA	Omission prejudicial.[30]
Jeffrey Landrigan	Ariz., 1986	MH, testimony from witnesses	Omission not prejudicial.[31]
George Porter	Fla., 1988	CTA, MH, TBI, military service record	Omission prejudicial.[32]
Robert Van Hook	Ohio, 1985	MH	Counsel not deficient.[33]
Fernando Belmontes	Cal., 1981	MH, expert witnesses, chance for rehabilitation	Omission not prejudicial.[34]
Demarcus Sears	Ga., 1993	CTA, TBI, DA	Remanded for prejudice analysis.[35]
Scott Pinholster	Cal., 1984	MH, positive family history	Counsel not deficient, omission not prejudicial.[36]
Terence Andrus	Tex., 2008	CTA	Remanded for prejudice analysis.[37]

Massachusetts, 582 U.S. 286, 299-302 (2017) (whether failure to object to trial judge closing courtroom to public is ineffective assistance of counsel).

[27] Williams v. Taylor, 529 U.S. 362 (2000).

[28] Bell v. Cone, 535 U.S. 685 (2002).

[29] Wiggins v. Smith, 539 U.S. 510 (2003).

[30] Rompilla v. Beard, 545 U.S. 374 (2005).

[31] Schriro v. Landrigan, 550 U.S. 465, 475 (2007).

[32] Porter v. McCollum, 558 U.S. 30 (2009).

[33] Bobby v. Van Hook, 558 U.S. 4 (2009).

[34] Wong v. Belmontes, 558 U.S. 15 (2009).

[35] Sears v. Upton, 561 U.S. 945 (2010).

[36] Cullen v. Pinholster, 563 U.S. 170 (2011).

[37] Andrus v. Texas, 590 U.S. 806, 806 (2020) (indicating defense counsel was ineffective for failing to present evidence of traumatic childhood).

Matthew Reeves	Ala., 1996	ID	Counsel not clearly deficient.[38]
Danny Jones	Ariz., 1992	CTA, MH	Omission not prejudicial.[39]

* *Notes:* CTA = childhood trauma and abuse; TBI = traumatic brain injury; MH = mental health issues; DA = drug abuse/addiction (including alcohol); ID = intellectual disability.

Table 1: U.S. Supreme Court Decisions about Mitigation Evidence Omitted from Death Penalty Trials, 2000 to Present

While the cases listed in Table 1 share some recurring themes, each case is a unique story of human tragedy. Supreme Court decisions tell only part of those stories.[40]

The argument that presenting evidence of childhood trauma may lead to a different outcome simply echoes the definition of mitigating factors. General claims about either the humanizing or double-edged effect of mitigation evidence are not enough to evaluate specific cases. The actual effect of omitting mitigation evidence depends on what the evidence is, how it is presented, and other evidence presented to the jury, especially aggravating factors. While omitting mitigation evidence may be a harmful error, the degree of harm is highly case-dependent and is difficult for the petitioners to prove.[41] Indeed, no petitioner in the past fifteen years has convinced the Court that the omission of mitigation evidence caused sufficient harm to overturn his death sentence.

II. SUBSTITUTING PERSONAL OPINIONS FOR COMPARISON OF PROBABILITIES

Deciding whether omitted mitigation evidence caused unfair prejudice in capital sentencing requires comparing two trials. The trial the defendant received is compared to a hypothetical, error-free version of the defendant's trial. This comparison is embodied by the quintessential image of justice: Lady Justice holding a balance scale. If the defendant's actual trial is set on one side of the scale and a hypothetical, error-free version of defendant's trial is set on the other side, would the scale balance?

---

[38] Dunn v. Reeves, 594 U.S. 731(2021).

[39] Thornell v. Jones, 602 U.S. 154 (2024).

[40] Future developments in the cases identified in Table 1 are summarized in the Appendix.

[41] This Article focuses on the recurring issue of assessing the harm caused by ineffective defense counsel, but the measurement problem extends far beyond this particular concern. The same problem arises when courts attempt to assess the effect of ordinary trial errors, constitutional trial errors, and *Brady* violations. In some states, a convict seeking new trial based on the discovery of new evidence must demonstrate that the new evidence could change the outcome. In all these situations, it has been difficult to escape the unhelpful truism that a different trial could have a different outcome.

Justice does not require the scales to balance perfectly because defendants are not entitled to perfect trials. But defendants have the right to fair trials; if the scale reveals significant imbalance, the defendant's right to a fair trial has been violated.

The "reasonable probability of a different outcome" standard under *Strickland* offers a precise conceptual definition of a fair trial. The standard identifies exactly what to measure. Probabilities are well-known mathematical quantities.[42] Trials are observable events with distinct outcomes.[43] The standard also tells judges how balanced the scales of justice must be to say the defendant's trial was fair. The standard works well as an abstract proposition, but practical problems beset any attempt to carry out this analysis.[44]

To begin with, what is a "reasonable" probability? The reference point for the required comparison has not been quantified.[45] It is necessary to define this threshold to judge when it is crossed. That is a hard decision, but it is also a simple one. All probabilities must be a number between zero and one.[46] Defining reasonable probability is as simple as choosing a number between zero and one. I have proposed that .10 is a reasonable

---

[42] *See* ALAN AGRESTI & BARBARA FINLAY, STATISTICAL METHODS FOR THE SOCIAL SCIENCES 73-74 (4th ed. 2009) (defining probability); IMAI, *supra* note 10, at 242-45 (same).

[43] The meaning of "trial" in the context of probabilities, counterfactuals, and causal inference deserves some comment. In this context, a trial is a single instance or occurrence of an event or experiment, which produces an outcome that is observed and analyzed to estimate probabilities or assess hypothetical scenarios. The idea that a trial is repeatable is helpful as it suggests the observed outcome of the defendant's actual trial was the result of a probabilistic process. A trial is fair if the outcome probabilities are approximately the same as the outcome probabilities of an error-free trial; a trial is unfair if there is a reasonable probability of a different outcome compared to the error-free trial.

[44] *See generally* Adam Lamparello, *Establishing Guidelines for Attorney Representation of Criminal Defendants at the Sentencing Phase of Capital Trials*, 62 ME. L. REV. 97, 112-15 (2010) (arguing that prejudice prong calls for distorted, post hoc analysis and it virtually impossible to meet in practice); Richard L. Gabriel, *The Strickland Standard for Claims of Ineffective Assistance of Counsel: Emasculating the Sixth Amendment in the Guise of Due Process*, 134 U. PA. L. REV. 1259, 1279-81 (1986) (discussing practical impossibility of proving prejudice); United States v. Decoster, 624 F.2d 196, 292-93 (D.C. Cir. 1976) (Bazelton, J., dissenting) (noting it is generally impossible to know how proceedings were affected by ineffective assistance of counsel and prejudice cannot be measured); Holloway v. Arkansas, 435 U.S. 475, 490-91 (1978) (noting it is "virtually impossible" to assess "the prejudice resulting from an attorney's failure to undertake certain tasks and doing so requires "unguided speculation").

[45] The Supreme Court defines reasonable probability as "a probability sufficient to undermine confidence in the outcome." *Strickland*, 466 U.S. at 694. This definition is helpful, but it does not quantify the probability that is sufficient to undermine confidence in jury trials.

[46] *See supra* note 40.

probability because empirical research suggests this probability is sufficient to undermine public confidence in the determination that a defendant is guilty.[47] Even if a clear line were drawn to distinguish fair from unfair, courts would still struggle to determine whether that line has been crossed because the probability of a different outcome cannot be quantified.[48]

The "reasonable probability of a different outcome" standard directs courts reviewing death sentences to evaluate how the omission of mitigation evidence affected the trial outcome. The probability of a different outcome should be "*quantitatively assessed* in the context of other evidence presented."[49] Despite this seemingly clear instruction, the probability of a

---

[47] *See* Barry C. Edwards, *A Scientific Framework for Analyzing the Harmfulness of Trial Errors*, 8 UCLA CRIM. JUST. L. REV. 1, 22-23, 43-45 (2024). Research suggests that reasonable doubt corresponds to approximately .10 probability in the minds of jurors. *See* Mandeep K. Dhami et al., *Instructions on Reasonable Doubt: Defining the Standard of Proof and the Juror's Task*, 21 PSYCH. PUB. POL'Y & L. 169 (2015); Francis C. Dane, *In Search of Reasonable Doubt*, 9 LAW & HUM. BEHAV. 141 (1985); Rita James Simon, *"Beyond a Reasonable Doubt": An Experimental Attempt at Quantification*, 6 J. APPLIED BEHAV. SCI. 203 (1970). Reasonable doubt is a possibility sufficient to undermine confidence about a guilty verdict; thus, .10 can be viewed as "a probability sufficient to undermine confidence in the outcome" of a criminal trial.

[48] Chief Justice John Roberts famously stated in his Senate confirmation hearing that "[j]udges are like umpires. Umpires don't make the rules, they apply them." *Confirmation Hearing on the Nomination of John G. Roberts, Jr. to be Chief Justice of the United States: Hearing Before the S. Comm. on the Judiciary*, 109th Cong. 55 (2005). Working with the Roberts analogy, current efforts to judge whether trials were fair are akin to trying to umpire a baseball game without either a defined strike zone or the ability to locate where pitches cross home plate. The strike zone is a topic of some debate within baseball, but knowing its boundary (or quantifying reasonable probability) is not enough to call balls and strikes (or distinguish fair from unfair) if pitches cannot be located (or the probability of a different outcome cannot be estimated). Baseball has, in fact, embraced new technologies that track pitch location, both for the fan experience and to improve umpiring. Some leagues now use automated ball-strike systems during games to call balls and strikes quickly and accurately. *See* Jayson Stark, *Triple-A Games to Start Fully Using Automated Ball-Strike Challenge System*, N.Y. TIMES (Jun. 18, 2024), https://www.nytimes.com/athletic/5573707/2024/06/18/automated-ball-strike-challenge-system-triple-a/.

It should be noted that legal scholars have both praised and criticized the judge-umpire analogy. *See, e.g.*, Michael P. Allen, *A Limited Defense of (at Least Some of) the Umpire Analogy*, 32 SEATTLE U. L. REV. 525 (2009); Aaron Zelinsky, *The Justice as Commissioner: Benching the Judge-Umpire Analogy*, 119 YALE L.J. ONLINE 113 (2010); Joseph Z. Fleming, *Just Like Umpires: Why Chief Justice Roberts Correctly Relied on Baseball to Describe the Supreme Court of the United States*, 5 ALB. GOV'T L. REV. 286 (2012).

[49] Arizona v. Fulminante, 499 U.S. 279, 308 (1991) (emphasis added).

different outcome is not quantified in cases supposedly following this standard.[50]

Rather than consider what ordinary people serving on a jury would have thought about the mitigation evidence, the question often becomes what parties and judges think of the evidence now.[51] When faced with a complex or challenging question, individuals unconsciously substitute it with a simpler, more accessible question that they can answer more easily. Daniel Kahneman describes this cognitive shortcut: "If a satisfactory answer to a hard question is not found quickly, System 1 [fast thinking] will find a related question that is easier and will answer it. I call the operation of answering one question in place of another *substitution.*"[52] The target question is the intended assessment; the heuristic question is the simpler question answered in place of the target question. This substitution happens without conscious awareness, leading to potentially incorrect or biased answers.[53]

In the context of criminal appeals, the target question is, "How would this evidence affect the probability of a different outcome?" The target question requires one to consider complex factors — the hypothetical trial condition, the jurisdiction's jury pool, and the deliberation process — and compare verdict probabilities. The simpler, heuristic question is, "Would this evidence make me more likely to vote for a different outcome?" Notice how quickly the heuristic question replaces the target question in recent oral arguments concerning the effect of evidence suppressed by prosecutors.

> CHIEF JUSTICE ROBERTS: Mr. Waxman, the counsel appointed by the Court argues that a central element of your case is the jury — that the jury would have regarded the matter differently if they knew that the lithium had

---

[50] Probabilities are, by definition, mathematical quantities. I am not aware of any court opinion that attempts to quantify the probability of a different outcome. According to Professor Solomon's study of harmless error analysis in 263 published federal opinions, "judicial discretion in determining harmlessness is largely unguided," and most opinions do not employ "any test at all." *See* Jason M. Solomon, *Causing Constitutional Harm: How Tort Law Can Help Determine Harmless Error in Criminal Trials*, 99 NW. U. L. REV. 1053, 1064, 1068 (2005).

[51] The arguments made in post-conviction proceedings feed into this problem. Petitioners argue that their mitigation evidence would be compelling to a jury because it is compelling to them and should be compelling to the court.

[52] DANIEL KAHNEMAN, THINKING, FAST AND SLOW 97 (2011).

[53] *See id.*, at 97-98. This process is often linked to heuristics, mental shortcuts the brain uses to solve problems quickly. While heuristics can be efficient, they sometimes lead to cognitive biases. Kahneman offers an example of question substitution that closely resembles question substitution in criminal appeals: The target question, "This woman is running for the primary. How far will she go in politics?" is replaced by heuristic question, "Does this woman look like a political winner?" *Id.* at 99.

> been prescribed by a psychiatrist as opposed to someone else because the jury knew about the lithium and what they didn't know is that it was prescribed by a psychiatrist. Do you — do you really think it would make that much of a difference to the jury?
>
> MR. WAXMAN: Well, I think that's not the only material difference here, that the — the — the fact was not only that he was — was that he lied and was allowed to lie when he said that he never saw a psychiatrist, which the defense — which — you know, it is one thing for a witness to stand up in court and testify on the basis of a promise of leniency by the prosecution. It's one thing for a witness to speculate or be inaccurate about what actually happened.[54]

The petitioner's attorney, Seth Waxman, answers a difficult question about the probability of a different jury trial outcome by discussing what "you know" from evaluating the evidence. Rather than discuss the probability of a different outcome, the Court considers whether the witness seems credible. Substitution is not limited to the party attempting to prove prejudice, it is also employed to disprove prejudice. The opposing counsel, Christopher Michel, makes the same substitution when asked how omitted evidence affected the jury's decision.

> JUSTICE KAVANAUGH: Yeah. I think you had said earlier and I want to explore, if you get past all the procedural bars and you get to the point where the prosecutors didn't comply with their obligations, that it still wouldn't have made a difference to the jury had they known that Sneed was bipolar and that he lied on the stand. And I'm having some trouble on that last piece of the argument, if we get there, understanding that, when the whole case depended on his credibility. Can you explain that some more?
>
> MR. MICHEL: Yes. And — and one of the critical arguments in the case — if you read the closing arguments, for example, there's extensive discussion about whether Petitioner was manipulating Sneed. That's probably the issue that comes up the most in the closing arguments, which are not evidence but are a reflection of what was at issue in the trial.[55]

---

[54] Transcript of Oral Argument at 8-9, Glossip v. Oklahoma, (Oct. 9, 2024) (No. 22-7466), https://www.supremecourt.gov/oral_arguments/argument_transcripts/2024/22-7466_h3ci.pdf.

[55] *Id.* at 104-05. The Supreme Court granted Glossip's request for a new trial in 2025. *See* Glossip v. Oklahoma, No. 22-7466, slip op. (Feb. 25, 2025). The

Rather than address the target question, whether the evidence would make a difference to a jury, the discussion turns to what "you know" when "you read" the evidence. The substitution, which occurs in both appellate advocacy and judicial analysis, happens so naturally no one seems to notice. When it comes to evaluating the harmfulness of a trial error or omission, personal opinions substitute for careful analysis of juror preferences and jury behavior.

Judges evaluating the evidence as a proxy for juries would work if judges could reliably estimate juror preferences, but they cannot. Mistakes occur because the substituted question does not directly address the target question, leading to cognitive bias and errors in judgment.[56] Confronted with new evidence, some judges believe the petitioner should still be sentenced to death, while other judges believe his life should be spared. Subjective opinions do not answer an objective question about an error's effect; personal opinions are not facts capable of being true or false.[57] Furthermore, the evidence indicates that trial judges cannot reliably predict jury verdicts and may be biased toward punishment compared to juries.[58] The problem is likely even worse on appeal if the lower court has already decided to affirm the sentence because higher courts have a bias toward affirmation.[59]

---

justices expressed varied opinions about the effect of suppressed evidence. Writing for the Court, Justice Sotomayor opines that the credibility of the key witness "plainly would have suffered." *Id.*, at 19. In dissent, Justice Thomas argues there is "no reasonable likelihood" that the evidence would change the jury's verdict. *Id.* at 26 (Thomas, J., dissenting).

[56] *See generally* Kenneth Williams, *Does Strickland Prejudice Defendants on Death Row*, 43 U. RICH. L. REV. 1459, 1478-81 (2009) (indicating that prevailing prejudice analysis is easily manipulated).

[57] Consider, as an example, the balance of judicial opinions in the case of James McKinney. In federal post-conviction proceedings, McKinney argued he should not be sentenced to death because mitigation evidence of post-traumatic stress disorder was not considered. (The Arizona trial court could not consider this evidence under a then-existing state law limiting mitigation evidence.). In a 6-5 vote, an *en banc* panel of the Ninth Circuit Court of Appeals sided with McKinney. *See* McKinney v. Ryan, 813 F.3d 798 (9th Cir. 2015). The Ninth Circuit's decision was reversed by a 5-4 vote of the U.S. Supreme Court. *See* McKinney v. Arizona, 589 U.S. 139 (2020). Between the Ninth Circuit and Supreme Court, judges were evenly divided 10-10.

[58] *See* Theodore Eisenberg et al., *Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Zeisel's The American Jury*, 2 J. EMPIRICAL LEGAL STUD. 171 (2005); HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY (1966).

[59] *See* Chris Guthrie & Tracey E. George, *Futility of Appeal: Disciplinary Insights into the Affirmance Effect on the United States Courts of Appeals,* 32 FLA. ST. U. L. REV. 357 (2004); Barry C. Edwards, *Why Appeals Courts Rarely Reverse Lower Courts: An Experimental Study to Explore Affirmation Bias*, 68 EMORY L.J. ONLINE 1035 (2018).

Current approaches to evaluating the fairness of capital sentencing do not inspire confidence that the court system will uphold fair trials and overturn unfair trials with much speed or accuracy.[60] The prosecutor cannot prove that a different trial would have the same outcome any more than the petitioner can prove that the outcome would be different. The prosecutor cannot prove the trial was fair; the petitioner cannot prove the trial was unfair. The analysis of harm remains qualitative, stuck in the realm of educated guesswork, and fairness becomes merely a matter of opinion.[61] The substitution of personal opinion for quantitative assessment is a major obstacle to objectively assessing the harmfulness of omitted mitigation evidence. Until appellate advocates and judges appreciate that personal assessments do not answer the target question, research-based arguments are easily ignored.

III. CHALLENGES OF PROVING WHAT WOULD HAVE HAPPENED IN A HYPOTHETICAL TRIAL

While it is true that mitigation evidence has a general tendency to reduce the probability of a death sentence, one may not deduce that mitigation evidence reduces the probability of a death sentence in all cases. Therefore, petitioners have attempted to substantiate the effect of mitigation evidence in specific cases using scientific studies. For example, in the *Jones* case, multiple organizations filed amicus briefs to document the effect of evidence of childhood trauma, brain injuries, and mental illness.[62]

---

[60] The loss of confidence in the appeals process is particularly acute when it comes to capital punishment. The public is frustrated by seemingly endless litigation. *See* BARRY LATZER & JAMES N. GREEAR CAUTHEN, JUSTICE DELAYED?: TIME CONSUMPTION IN CAPITAL APPEALS: A MULTISTATE STUDY 14-15 (2007). Appeals are procedurally exhausted more than they are substantively decided. The public is losing confidence in the criminal justice system's ability to administer capital punishment. "Assuming that the job of the death penalty is to identify offenders for whom the law prescribed death as a punishment and to carry out the sanction with the swiftness and sureness needed to deter and express revulsion for those offenses, one arrives at the same conclusion: The current system is broken." James S. Liebman, *Opting for Real Death Penalty Reform*, 63 OHIO ST. L.J. 315, 317-18 (2002).

[61] Prevailing interpretations of the harmless error doctrine have been extensively criticized as inconsistent, contradictory, and confusing. *See, e.g.*, Jeffrey O. Cooper, *Searching for Harmlessness: Method and Madness in the Supreme Court's Harmless Constitutional Error Doctrine*, 50 U. KAN. L. REV. 309, 310 (2002) (proliferation of standards has caused "considerable confusion."); Justin Murray, *A Contextual Approach to Harmless Error Review*, 130 HARV. L. REV. 1791, 1794 (2016) ("worrying signs" that appellate courts are "bungling" harmless error analysis); Gavin R. Tisdale, *A New Look at Constitutional Errors in a Criminal Trial*, 48 CONN. L. REV. 1665, 1702 (2016) (reporting pervasive "confusion and misapplication").

[62] *See, e.g.*, Brief of the National Association of Criminal Defense Lawyers and the American Civil Liberties Union as Amici Curiae in Support of the

In studies like those cited by amici in *Jones*, study subjects evaluate a summary of the evidence and arguments from a trial as if they were jurors. For example, the Barnett et al. (2004) finding on the effect of evidence of untreated mental illness is based on the following trial summary.

> The defendant, John Smith, robbed a convenience store at gunpoint on a Saturday night. When he entered the store he tied up the cashier, took the money from the register, and proceeded to take several things off the shelves. Meanwhile, the cashier freed himself and ran out the front door. Smith yelled for him to stop, but when he didn't, Smith shot and killed him. During testimony, it is revealed that Smith has been diagnosed as schizophrenic and sought treatment on and off for years. During the weeks preceding the botched robbery, Smith has not been taking his medication and experienced severe delusions and hallucinations.[63]

After reading the summary, subjects are asked to sentence the defendant to "life in prison" or "the death penalty."[64] 92.7% of study subjects choose life imprisonment.[65] In control conditions, cases with comparable crimes but without mitigating factors, 26.3% of study subjects recommended life sentences.[66] Evidence of a defendant's untreated mental illness increased the subjects' preferences for a life sentence by 66.4 percentage points.

Table 2 presents the estimated effect of mitigation evidence on the proportion of adults who support a death sentence according to the published results of randomized experiments.[67] The table identifies the type of mitigation studied as well as other factors specified in the studies.

Respondent, Thornell v. Jones, 602 U.S. 154 (2024) (No. 22-982); Amicus Curiae Brief of the Arizona Capital Representation Project in Support of Respondent, Thornell v. Jones, 602 U.S. 154 (2024) (No. 22-982). Similar briefs have been filed in other cases. *See, e.g.*, Amicus Curiae Brief for Children's Defense Fund et al. in Support of Petitioner, Andrus v. Texas, 590 U.S. 806 (2020) (No. 21-6001) (substantiating effects of childhood trauma).

[63] Michelle E. Barnett et al., *When Mitigation Evidence Makes a Difference: Effects of Psychological Mitigating Evidence on Sentencing Decisions in Capital Trials*, 22 BEHAV. SCI. & L. 751, 757 (2004). Barnett et al. asked a panel of undergraduate mock jurors to decide whether defendants should be sentenced to death or life imprisonment in different cases (not all identified) given varying types of mitigation evidence. *See id.* at 756.

[64] *Id.* at 760.

[65] *Id.* at 761 tbl.1.

[66] *Id.* at 761-62.

[67] Amici in *Thornell v. Jones* did not cite studies by Nuñez et al. (2017) or Greene & Cahill (2012). Nuñez et al. use random assignment to estimate the effect of offering strong mitigation evidence, rather than weak mitigation evidence,

ESTIMATED EFFECT OF MITIGATION EVIDENCE ON JUROR SENTENCE PREFERENCES

Mitigation Evidence	*Other Factors*	*Study*	*Effect*
Untreated mental illness	Murder during robbery	Barnett et al.	+.66
Neurological tests and images	High danger defendant	Greene and Cahill	+.57
Neurological tests	High danger defendant	Greene and Cahill	+.51
Borderline mental retardation	Not reported	Barnett et al.	+.45
Severe abuse in childhood	Not reported	Barnett et al.	+.29
Strong mitigation evidence	No victim statement	Nuñez et al.	+.26
Drug addiction	Not reported	Barnett et al.	+.23
Strong mitigation evidence	Angry victim statement	Nuñez et al.	+.18
Strong mitigation evidence	Sad victim statement	Nuñez et al.	+.03
Neurological tests and images	Low danger defendant	Greene and Cahill	+.02
Homelessness as adult	Not reported	Barnett et al.	+.01
Neurological tests	Low danger defendant	Greene and Cahill	-.05
Lifetime of migraines	Not reported	Barnett et al.	-.09

Note: Effect is the increase in proportion of respondents who support a life sentence compared to control condition.

Table 2: Estimated Effect of Mitigation Evidence on Juror Sentence Preferences

Amici supporting Jones made a concerted effort to substantiate the harm caused by omitting mitigation evidence. The studies cited do quantitatively assess the effect of mitigation evidence on the decision to impose the death sentence. Unfortunately, these types of studies cannot specify how much harm was caused by the omission of mitigation evidence in a specific case under court review. These academic studies were not designed to answer the target question in real litigation. There is a fundamental difference between testing academic theories for general knowledge and generating information that practitioners and judges need to make decisions on an individualized, case-by-case basis.[68]

---

in trials with different types of victim impact statements. *See* Narina Nuñez et al., *The Impact of Angry Versus Sad Victim Impact Statements on Mock Jurors' Sentencing Decisions in a Capital Trial*, 44 CRIM. JUST. & BEHAV. 862 (2017). Greene and Cahill use randomized mock jury experiments (student samples) to estimate the effect of offering neuropsychological tests and brain images to support an expert's opinion that defendant suffers from psychosis, in trials where the defendant presents either high or low danger of future crimes. *See* Edith Greene & Brian S. Cahill, *Effects of Neuroimaging Evidence on Mock Juror Decision Making*, 30 BEHAV. SCI. & L. 280 (2012).

[68] General findings would be relevant and useful for assessing statutes. For example, general findings on the effect of evidence of childhood trauma is useful to show that categorical prohibition of such evidence is problematic. *See* Lockett

This Part identifies three consequential differences between academic research on jury decision-making and legal reasoning about potential trial outcomes. First, the trial summaries used in research studies do not allow courts to assess the effect of mitigation evidence in a trial context where other facts are present. Second, these studies analyze the sentencing preferences of research subjects who do not fairly represent the relevant jury pool. Third, even if the specific effect of mitigation evidence on jurors' sentencing preferences is known, its effect on the jury deliberation process and the probability of a different trial outcome remain unknown.[69] For these reasons, the research cited to prove the petitioner suffered prejudice fails to substantiate that claim. This Part also discusses how researchers might better inform court decisions.

A. The Impact of Omitting Evidence Depends on the Litigation Context

Trial summaries are not meant to provide research subjects with authentic courtroom experiences, but rather serve to stimulate authentic responses from the respondents.[70] The summary need not replicate a full trial experience but should capture its core effects, akin to how a flavor extract captures the taste of food.[71] Trial summaries should be clear,

---

v. Ohio, 438 U.S. 586, 598-608 (1978) (state law restricting evidence of mitigating factors is unconstitutional); Eddings v. Oklahoma, 455 U.S. 104 (1982) (trial judge did not consider mitigation evidence of childhood trauma; individualized sentencing requires consideration of mitigating factors that without causal nexus to crime).

[69] Even if it can be shown that an omission increased the proportion of jurors who support a death sentence from, say, sixty percent to seventy percent, the error's effect on the probability of a death sentence still cannot be determined. A modest change in individual-level preferences can have a substantial effect in a close case. A substantial shift in preferences may have only a modest effect on the probability of a death sentence. The fairness of defendant's sentence cannot be addressed objectively, even with perfect information about juror verdict preferences.

[70] It is important to distinguish experimental realism from ecological validity. The ecological validity of experiment — which is the extent to which experimental findings can be generalized to the real world — is not determined by the degree to which an experiment replicates a real-world experience. An experiment that mimics the actual experience of jury duty — driving to an unfamiliar part of town, waiting patiently for your name to be called, being guided from place to place — may be high in realism but this kind of mundane realism does not assure us that the treatment effect observed in an experiment would occur in the real world. The real issue with respect to ecological realism is to make sure that strength of the treatment in the experiment is comparable to the strength of the treatment real jurors would experience.

[71] This does not imply that summaries should provide "equal time" to the evidence presented by the prosecution and the defense or style the facts in a way that gives both sides an "equal opportunity" to prevail. Trial summaries should reflect the quantity and quality of evidence presented as trial. Technologies like ChatGPT may be used to efficiently summarize trial conditions based on

simple, and concise.[72] Respondents should be able to complete the survey in one sitting without taking a break.[73] At the same time, summaries need to fairly represent the mitigation evidence at issue as well as other evidence relevant to a jury's sentencing decision.

1. Case-Specific Facts

By the time death sentences are reviewed in post-conviction proceedings, case-specific facts will be known. To prove prejudice, petitioners will identify the omitted mitigation evidence. The court hearing the petition, therefore, knows what the jury heard and what it missed and, if there is any uncertainty, the court can conduct an evidentiary hearing.[74] However, knowing what evidence was omitted is not enough to gauge its effect on the trial outcome. If anything, the record developed during trial and post-conviction proceedings reveals the differences between study conditions and the case before the court.

---

automated analysis of documents, including pleadings, trial transcripts, court opinions, and briefs submitted on appeal. *See* Mahak Gambhir & Vishal Gupta, *Recent Automatic Text Summarization Techniques: A Survey*, 47 A.I. REV. 1 (2017); Aji Prasetya Wibawa & Fachrul Kurniawan, *A Survey of Text Summarization: Techniques, Evaluation and Challenges*, 7 NATURAL LANG. PROCESSING J. 100070 (2024); Deepali Jain et al., *Summarization of Legal Documents: Where Are We Now and the Way Forward*, 40 COMPUT. SCI. REV. 100388 (2021); Joel Hake et al., *Quality, Accuracy, and Bias in ChatGPT-Based Summarization of Medical Abstracts*, 22 ANNALS FAM. MED. 113 (2024) (finding ChatGPT summaries of medical abstracts to be seventy percent shorter, high quality, high accuracy, and low bias). The author did not use this technology to generate trial summaries, but it is a promising aid for summarizing trial conditions.

[72] As more people now access the internet using mobile devices than computers, surveys must be mobile-friendly.

[73] *See* Mirta Galesic & Michael Bosnjak, *Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey*, 73 PUB. OPINION Q. 349 (2009); Melanie Revilla & Carlos Ochoa, *Ideal and Maximum Length for a Web Survey*, 59 INT'L J. MARKET RSCH. 557 (2017) (suggesting ideal duration of ten minutes and maximum duration of twenty minutes); Roberta Sammut et al., *Strategies to Improve Response Rates to Web Surveys: A Literature Review*, 123 INT'L J. NURSING STUD. 104058 (2021) (recommending ten-minute surveys to improve response rates).

[74] Pursuant to 28 U.S.C. § 2255(b), a federal court should hold a hearing "[u]nless the motion and the files of the case conclusively show that the petitioner is entitled to no relief." Petitioner Danny Jones, for example, detailed the nature and extent of the abuse he suffered as a child and what psychiatric experts would say at a new trial. *See* Jones v. Schriro, 450 F. Supp. 2d 1023, 1046 (D. Ariz. 2006) (new mitigation evidence developed in post-conviction evidentiary hearing indicated Jones was physically and emotionally abused by his first stepfather, and may have been sexually abused by his grandfather for a period of five years). *See also* Andrus v. Texas, 590 U.S. 806, 810 (eight-day evidentiary hearing in state court produced "tidal wave" of mitigation evidence not heard by the jury).

The studies cited in Table 1 do not show whether the evidence the *Jones* jury never heard would create a reasonable probability of a different outcome in a new trial. To begin with, the mitigation evidence omitted from a specific trial is different than the mitigation evidence that the study subjects considered.[75] Even if one thinks evidence of severe abuse in childhood increases support for a life sentence by twenty-six percentage points, on average, its effect in specific cases would vary depending on the nature and extent of abuse. In *Jones*, there is clear evidence of physical and mental abuse by petitioner's first stepfather, but the evidence of sexual abuse by petitioner's grandfather is less clear.[76] Jones is a victim of abuse, but he is also a victim of his own self-destructive decisions.[77]

Also, notice that the effect of mitigation evidence reported in Table 1 depends on whether the defendant appears to be dangerous, the emotional content of victim impact statements, and how the crime occurred.[78] Based on these results, the omission of mitigation evidence can have effects ranging from a .66 increase in the proportion of jurors who would sentence the defendant to death to a .09 *decrease* in that proportion.[79] Whether a deficient counsel's failure to introduce mitigation evidence is a tolerable mistake, a tragic mistake, or a blessing in disguise depends on case-specific circumstances.[80] These studies estimate the effect of mitigation

---

[75] Barnett et al. state that subjects were informed that the defendant "was severely physically and verbally abused by his parents during childhood," but does not include the exact content of the severe childhood abuse study conditions in their article. *See* Barnett et al., *supra* note 63, at 757.

[76] Thornell v. Jones, 602 U.S. 154, 168-69 (2024).

[77] *Id.* at 161 (noting long-term substance abuse).

[78] This discussion focuses on studies estimating the effect of evidence of childhood trauma on death sentencing decisions, but similar observations could be made about the effect of other types of evidence or categories of error claims. *See generally* Robert J. MacCoun, *Experimental Research on Jury Decision-Making*, 244 SCI. 1046 (1989). Multiple studies, for example, have attempted to estimate the effect of confessions on guilt decisions. *See, e.g.*, Saul M. Kassin & Katherine Neumann, *On the Power of Confession Evidence: An Experimental Test of the Fundamental Difference Hypothesis*, 21 LAW & HUM. BEHAV. 469 (1997); Saul M. Kassin & Holly Sukel, *Coerced Confessions and the Jury: An Experimental Test of the "Harmless Error" Rule*, 21 LAW & HUM. BEHAV. 27 (1997); Stacy A. Wetmore et al., *On the Power of Secondary Confession Evidence*, 20 PSYCH. CRIME & L. 339 (2014).

[79] The effects reported in Table 1 are based on published results which, in some cases, must be derived from the authors' tables. For example, Nuñez et al. report the effect of victim impact statements, controlling for mitigation evidence, which allows me to estimate the effect of mitigation evidence, controlling for victim impact statements.

[80] The Greene et al. study assesses the effect of neuropsychologic tests and images in situations based on real criminal trials. Greene & Cahill, *supra* note 67, at 286-87. Their summary of facts was based on *United States v. Saban*, 555 F. Supp. 2d (D. Colo. 2006). The details of the defense expert testimony were based a study patterned on *United States v. Barnette*, 211 F.3d 803 (4th Cir. 2000). *See*

evidence in circumstances unlike those under review. Unless the study conditions match the facts of the *Jones* case, the study results do not accurately estimate the effect of omitted evidence in his specific case.

2. Prosecution's Likely Response to Mitigation Evidence

Summarizing a hypothetical trial presents some challenges that merit discussion. Should the hypothetical trial condition be the defendant's version of an error-free trial, or should it include the prosecutor's likely response to new mitigation evidence? For example, when a post-conviction petitioner argues trial counsel failed to call character witnesses during sentencing, should the analysis consider how the prosecutor would have responded to that testimony?[81]

The hypothetical trial used for comparison typically incorporates the prosecution's foreseeable response to the defendant's new evidence, which may undermine the effect of defense counsel's mitigation evidence.[82] In *Strickland*, the Supreme Court points out that Washington's attorney was able to keep other evidence, like defendant's prior criminal history and the prosecution's psychological evidence, out of the trial by relying on the defendant's plea colloquy and not calling character witnesses.[83] According to the Court, the hypothetical trial would not be as favorable as Washington would like because "his 'rap sheet' would probably have been admitted into evidence, and the psychological reports would have directly contradicted respondent's claim[.]"[84] To assess the harm caused by ineffective counsel, the Court expressly considers the prosecution's likely rebuttal in the hypothetical trial condition.

The hypothetical trial summary should reflect reasonable, foreseeable, and obvious adjustments in the parties' arguments and evidence.[85] A

John F. Edens et al., *The Impact of Mental Health Evidence on Support for Capital Punishment: Are Defendants Labeled Psychopathic Considered More Deserving of Death?* 23 BEHAV. SCI. & L. 603, 610 (2005).

[81] Character evidence is generally inadmissible, but when the defendant introduces character evidence, the door is open for the prosecution to introduce evidence to rebut it. Fed. R. Evid. 404(a).

[82] *See* Wong v. Belmontes, 558 U.S. 15, 20 (2009) (to evaluate harm caused by error, "it is necessary to consider *all* the relevant evidence that the jury would have had before it" if omitted mitigation evidence were offered, such as evidence of prior murder) (original emphasis).

[83] *See* Strickland v. Washington, 466 U.S. 668, 699 (1984).

[84] *Id.* at 701. Similarly, in *Porter*, the Court does not just consider evidence of heroism during war and service-related injuries the defendant would have presented at trial if he were effectively represented, it also considers the effect of prosecution's likely rebuttal that Porter's service record showed he went AWOL multiple times. *See* Porter v. McCollum, 558 U.S. 30, 43-44 (2009).

[85] The best guide to the hypothetical trial summary may be the evidence presented in habeas corpus proceedings when the defendant argues he is entitled to a new trial because of constitutional errors. If the petitioner presents new evidence he contends entitle him to relief, the state may offer new evidence at a hearing to

summary of Washington's hypothetical trial should include his rap sheet and psychological reports because those items were known and are specifically identified in the Court's opinion. That rebuttal evidence requires no speculation. The prosecution should not produce previously unknown evidence and psychological reports.[86] The goal is to determine whether there is a reasonable probability of a different outcome in a new, error-free trial.

A habeas petitioner's argument that an error was harmful is more compelling if the analysis demonstrates harm even if the prosecution attempts to minimize the damage. Similarly, a prosecutor's claim that an error was harmless is more persuasive if analysis shows the error was harmless even if all the alleged errors were addressed in a hypothetical, error-free trial.[87] To proactively address an argument that trial summaries mislead respondents into favorable responses, the party that intends to use survey-based research may want to share proposed summaries with opposing parties, solicit their feedback, and address legitimate concerns before the survey is conducted.[88]

3. Measuring Verdict Preferences

After reading trial summaries, study subjects are typically asked whether they favor life imprisonment or a death sentence. The instructions for completing the survey should be kept as simple as possible; they do

---

counter petitioner's claims. The state is not limited to relying on the evidence it presented at trial; the hearing anticipates future developments in the case. An evidentiary proceeding is akin to the hypothetical trial condition in assessing prejudicial effects. The judge previews the evidence that would be presented in a new trial and asks whether there is a reasonable probability of a different trial outcome.

[86] At times, the Court seems to invent the state's rebuttal to mitigation evidence that was omitted from trial. For example, the Court finds that the evidence omitted from Pinholster's trial did not have a prejudicial effect because it would have "opened the door to rebuttal by a state expert" and new evidence of the defendant's dangerousness. Cullen v. Pinholster, 563 U.S. 170, 201 (2011). But, as Justice Sotomayor pointed out in dissent, the record gave "no reason to know what a state expert might have said." *Id.* at 239 (Sotomayor, J., dissenting).

[87] These claims follow the logic of the standard applied to motions for summary judgment where "all justifiable inferences are to be drawn" in favor of the party opposing the motion. *See* Anderson v. Liberty Lobby, Inc., 477 U.S. 242, 255 (1986).

[88] Additionally, the "court can require the parties to agree on methodology and form before conducting surveys or polls." MANUAL FOR COMPLEX LITIGATION § 523-24 (4th ed. 2004). At the same time, parties should have the opportunity to conduct exploratory analysis to evaluate their arguments without their work product being used against them. *See* Fed. R. Civ. P. 23(b)(3); Hickman v. Taylor, 329 U.S. 495, 510-11 (1947) (purpose of protecting attorney work product from discovery).

not need to be as elaborate as those read to a courtroom jury.[89] Questions that measure verdict preferences should be simple and neutral, ensuring they do not lead the respondent to a particular answer.[90] After weighing the aggravating and mitigating factors, respondents should be asked whether they would recommend a sentence of death or life imprisonment.[91] Subjects should not be told that that they are merely making a sentencing recommendation that a judge will review and correct, if necessary.[92]

---

[89] Respondents know, without being told, that the defendant is entitled to the presumption of innocence and conviction requires proof beyond a reasonable doubt, but it is prudent to remind them of this. Lengthy instruction on the meaning of these standards is not advisable. The content of reasonable doubt instruction has relatively little effect on the threshold of certainty that jurors use to convict. *See* Jason A. Aimone et al., *An Experimental Exploration of Reasonable Doubt*, 212 J. ECON. BEHAV. & ORG. 873 (2023). Moreover, there is no requirement that a "particular form of words be used in advising the jury of the government's burden of proof." Victor v. Nebraska, 511 U.S. 1, 5 (1994). Survey respondents do not need the usual warnings about talking about the case with others or seeking outside information about the trial.

[90] *See* JEAN M. CONVERSE & STANLEY PRESSER, SURVEY QUESTIONS: HANDCRAFTING THE STANDARIZED QUESTIONNAIRE 9-13 (1986); Jon A. Krosnick & Stanley Presser, *Question and Questionnaire Design*, *in* PALGRAVE HANDBOOK OF SURVEY RESEARCH 263, 264 (David L. Vannette & Jon A. Krosnick eds., 2018). The rules of evidence reflect the dangers of asking leading questions. Fed. R. Evid. 611(c) prohibits leading questions in direct examination.

[91] In capital punishment cases, the aggravating circumstances that justify the imposition of the death penalty must be found by a jury. *See* Ring v. Arizona, 536 U.S. 584 (2002); *but see* Spaziano v. Florida, 468 U.S. 447 (1984) (constitution does not require jury to decide whether defendant should be sentenced to death or life imprisonment) *and* McKinney v. Arizona, 589 U.S. 139, 144-45 (2020) (the Constitution does not require jury to determine the balance between aggravating and mitigating factors).

[92] According to the U.S. Supreme Court, jurors should not be misled about the grave consequences of recommending a death sentence. *See* Caldwell v. Mississippi, 472 U.S. 320, 328-29 (1985). Judges cannot override a jury's life imprisonment recommendation and sentence a defendant to death in the sentencing phase of trial. It is error for prosecutors to tell jurors that the judge can overrule a death sentence recommendation because that impermissibly diminishes jurors' responsibility for their sentence recommendation. *See* Driscoll v. Delo, 71 F.3d 701 (8th Cir. 1995).

Jurors may be told that the defendant will be sentenced to life imprisonment without possibility of parole if not sentenced to death. *See* Simmons v. South Carolina, 512 U.S. 154 (1994). This seems unnecessary when respondents are explicitly asked to consider a sentence of life imprisonment. In Florida, state law mandates the punishment of death or life imprisonment without parole for murder. Parole has been effectively abolished in Florida (as of 1983) so Williams and Porter would be facing life sentences without the possibility of parole. *See* FLA. STAT. § 775.082(1)(a).

It is essential to measure respondents' preferred sentence.[93] Some studies do not measure verdict preferences and therefore do not address the target question and relevant legal standard.[94] Some studies cited by amici in *Jones* and *Andrus* measure respondent preferences other than the choice between life imprisonment and a death sentence. For example, Garvey (1998) focused on whether different types of mitigation evidence would make jurors much more, slightly more, just as, slightly less, or much less likely to vote for a death sentence.[95] Montgomery et al. (2005) measured the effects of mitigation evidence on respondents' impressions of the defendant's dangerousness, craziness, and instability on 1 to 4 scales.[96] Along these lines, one may ask respondents how strongly they prefer one outcome over the other and how confident they are about their

---

[93] There is some debate over forcing survey respondents to choose answers without offering a "Don't know" (DK) response, but the option is not advisable, when respondents can make a choice. One review of the literature concludes that

> DKs often result not from genuine lack of opinions but rather from ambivalence, question ambiguity, satisficing, intimidation, and self-protection. In each of these cases, there is something meaningful to be learned from pressing respondents to report their opinions, but DK options discourage people from doing so. As a result, data quality does not improve when such options are explicitly included in questions.

Krosnick, *supra* note 90, at 284-85; *see also* Patrick Sturgis et al., *Middle Alternatives Revisited: How the Neither/nor Response Acts as a Way of Saying "I Don't Know"*, 43 SOCIO. METHODS & RSCH. 15 (2014); Daniela Wetzelhütter, *Scale-Sensitive Response Behavior!? Consequences of Offering Versus Omitting a "Don't Know" Option and/or a Middle Category*, SURV. PRAC., https://doi.org/10.29115/SP-2020-0012; Daniel Laurison, *The Willingness to State an Opinion: Inequality, Don't Know Responses, and Political Participation*, 30 SOCIO. F. 925 (2015).

[94] It is possible to obtain other measures of respondents' verdict preferences, such as their opinions about witness credibility and the defendant's dangerousness. Some surveys have asked respondents to express their verdict preference on a 0 to 10 or 0 to 100 scale. More detail is usually better than less detail, but it is not clear how to translate a numeric verdict preference to a binary decision. The midpoint of the scale is not necessarily the dividing line between life and death sentences given the proof beyond a reasonable doubt standard.

[95] Stephen P. Garvey*, Aggravation and Mitigation in Capital Cases: What Do Jurors Think?*, 98 COLUM. L. REV. 1538, 1551-52 (1998) (recording whether jurors would be more or less likely to vote for death penalty given different mitigating and aggravating factors).

[96] See John H. Montgomery et al., *Expert Testimony in Capital Sentencing: Juror Responses*, 33 J. AM. ACAD. PSYCHIATRY & L. 509, 511 (2005) (asking respondents whether terms like "dangerous" "unstable" and "crazy" describe the defendant very well, fairly well, not well, or not at all).

decision.[97] Respondents can also be asked unstructured questions to gain insight into why they think the defendant should be sentenced to death or life imprisonment.[98] The survey instrument can ask respondents whether specific facts influenced their decision and provide them opportunities to write comments. It is interesting to hear what respondents have to say; hearing respondents explain their answers can satisfy an innate human desire to understand causal processes.[99]

These alternatives to the binary verdict choice question do not inform the relevant inquiry in litigation. Someone being much more or slightly more likely to vote for a life sentence is related to, but different than, someone voting for a life sentence.[100] The effect of a one point change in impressions of dangerousness, craziness, and instability for the life-or-

---

[97] The additional nuance of how strongly a respondent favors one sentence or the other does not improve the analysis. Strength of belief may be a characteristic of respondents rather than a reflection of the case under consideration; the distribution of very strong, fairly strong, and not strong beliefs is approximately the same on both sides in all cases. The analyst can also ask respondents to express their sentence preference using an interval-level measure, such as a 0-100 scale, but it is unclear how to interpret those responses as there is no corollary in the courtroom. Early studies tended to use interval-level measures to obtain more precise measures of individual preferences, but these measures may not speak to the quantities of interest in harmless error analysis and it is not clear how one would use them to evaluate prejudicial effects.

[98] Identifying the causal mechanisms can help us understand why a trial error was harmful (or harmless). Did it affect the prosecution, defense, or both sides? How did it affect the weight subjects assigned to different items of evidence? Understanding these internal linkages can illuminate the jurors' thought process and help explain results, but we should focus on what the effect of an error was rather than why an error had an effect. A trial error may harm a defendant even if we do not fully understand why it caused jurors to vote for a guilty verdict.

[99] At the same time, human decision-makers are not always conscious of factors that influence them. For example, having the defendant's relatives testify on his behalf during sentencing may make a significant statement about his connections with family even if his relatives do not say anything particularly memorable. Jurors may say the family testimony was not important, because the effect is unspoken. Prejudicial evidence presents similar issues; prejudicial evidence is harmful, but respondents may be unwilling to admit to having socially unacceptable prejudicial thoughts.

[100] Garvey (1998) asked respondents whether mitigating factors made them more or less likely to vote for a death sentence. *See* Garvey, *supra* note 95 at 1556 tbl.3. This scaled response is better suited to measuring the effect of omitted mitigation evidence on the jury which is an alternative approach to evaluating the harmfulness of trial errors.

death sentence decision is unclear.[101] Alternate measures do not directly address the outcome of interest and are difficult to interpret.[102]

4. Realism Issue

Skeptics have raised concerns about scientific studies of jurors and juries, primarily arguing that the results of mock jury experiments may not generalize to real-world trials.[103] Indeed, the experience of reading a trial summary and taking a survey is nothing like the experience of being summoned to court to serve on a jury. While mock jurors' trial experiences differ from actual courtroom proceedings, their choice between guilty and not guilty verdicts is realistic. At this point, the external validity of scientific studies of jurors and juries is "widely accepted."[104]

One concern is that the mock juror's decision has no consequences. "Mock jurors," write Brian Bornstein and Sean McCabe, "reach a verdict concerning a paper defendant; but real jurors make decisions concerning a flesh-and-blood defendant, which could entail a prison sentence or hefty damage award."[105] Skeptics argue that mock jurors may acquit without concern for public safety or punish with impunity.[106] The inconsequence

---

[101] Montgomery (2005) measures these defendant characteristics. *See* Montgomery, *supra* note 96.

[102] Another study cited by amici in *Thornell v. Jones*, 602 U.S. 154 (2024), *supra* note 76, is Stetler et al. (2022), which documents cases with strong aggravating circumstances and varying mitigating circumstances where juries opted for life sentences. *See* Russell Stetler et al., *Mitigation Works: Empirical Evidence of Highly Aggravated Cases Where the Death Penalty Was Rejected at Sentencing*, 51 HOFSTRA L. REV. 89 (2022). This study cannot report the probability of a life sentence based on the relative weight of aggravating and mitigating factors because only life imprisonment outcomes were studied.

[103] In scientific literature, this issue is identified as external validity or ecological validity. *See* Derrick M. Anderson & Barry C. Edwards, *Unfulfilled Promise: Laboratory Experiments in Public Management Research*, 17 PUB. MGMT. REV. 1518, 1527-29 (2015).

[104] David L. Schwartz & Christopher B. Seaman, *Standards of Proof in Civil Litigation: An Experiment from Patent Law*, 26 HARV. J.L. & TECH. 429, 470 (2013); *see also* Michael J. Saks, *What Do Jury Experiments Tell Us About How Juries (Should) Make Decisions?*, 6 S. CAL. INTERDISC. L.J. 1, 7-8 (1997) (arguing that common criticisms of mock juries are overstated and misdirected); MacCoun, *supra* note 78, at 1046 ("[M]ock jurors do not appear to reach decisions by a fundamentally different process than actual jurors").

[105] Brian H. Bornstein & Sean G. McCabe, *Jurors of the Absurd? The Role of Consequentiality in Jury Simulation Research*, 32 FLA. ST. U. L. REV. 443, 445 (2005).

[106] *See, e.g.*, Wayne Weiten & Shari Seidman Diamond, *A Critical Review of the Jury Simulation Paradigm: The Case of Defendant Characteristics*, 3 LAW & HUM. BEHAV. 71, 81-83 (1979). In mock jury research, the researcher should attempt to impress upon the subjects the importance of the research. Even if the subject's decision does not directly affect a defendant, the subject's decision will

of mock juror decisions does not appear to systematically influence decisions about guilt and punishment.[107] Research indicates that experimental subjects take scientific research seriously and follow the experimenter's instructions.[108] Additionally, the researcher can randomly vary the order of answer choices and ask questions that check attention to neutralize the effect of careless respondents.[109]

Another concern is summarizing trials with written vignettes, which do not resemble the presentation style of real courtroom trials. Summarizing trials in a few pages is more efficient than staging live reenactments. Additionally, a survey with written trial summaries can be administered online, reaching a larger and more diverse audience than a live production.[110] Researchers have presented trials in various ways, but written vignettes have become the most common method for summarizing trials for mock jurors.[111] While convenient, is something important lost when a trial

---

contribute to research on punishment and jury decision making, potentially having far reaching consequences.

[107] Bornstein & McCabe review five studies that varied the consequentiality of jury decisions; according to them, one found mock jurors were more lenient, another found they were more punitive, and three others found no differences. Bornstein & McCabe, *supra* note 105, at 452-57; *see also* David L. Breu & Brian Brook, *"Mock" Mock Juries: A Field Experiment on the Ecological Validity of Jury Simulations*, 31 LAW & PSYCH. REV. 77, 80 (2007) (comparing four deliberations with subjects told their verdicts would have real or limited consequences).

[108] There is even concern that research subjects take experiments too seriously and try to help the researcher; this is the reason for double-blind random assignment. *See generally* Daniel J. Zizzo, *Experimenter Demand Effects in Economic Experiments*, 13 EXPERIMENTAL ECON. 75 (2010); ROBERT ROSENTHAL, EXPERIMENTER EFFECTS IN BEHAVIORAL RESEARCH (1966).

[109] If mock jurors' supposed carelessness is simply random noise, the results of the experiment remain unbiased, although the experiment might not detect certain effects that could emerge in more attentive samples. *See* KROSNICK & PRESSER, *supra* note 90, at 281 (discussing varying choice order to mitigate response order effects).

[110] It is, of course, possible to incorporate images, audio, and even video into summaries of trial conditions in online surveys. Multimedia should not be used simply to create a more realistic trial experience; that is not the purpose of a trial summary. However, it may be useful to incorporate multimedia to represent trial errors as they would appear to jurors — such as the effect of improper demonstrative evidence, prejudicial photographs, or having the defendant appear in shackles. Incorporating images into vignettes would be particularly useful to estimating the effect of the defendant wearing shackles because this fact would not be explicitly explained to jurors and the prejudicial effect would come from their implicit association of shackles and dangerousness.

[111] *See, e.g.*, D. Alex Winkelman et al., *An Empirical Method for Harmless Error*, 46 ARIZ. ST. L.J. 1405, 1437-38 (2014) (written summaries of trials averaging about two pages); Emily C. Hodell et al., *Factors Impacting Juror Perceptions of Battered Women Who Kill Their Abusers: Delay and Sleeping Status*, 18 PSYCH. PUB. POL'Y & L. 338, 344 (2012) (three-page written

is summarized in writing rather than presented using multimedia or live reenactments? This question has been debated and studied.[112] Research indicates that written vignettes produce responses comparable to reenactments or videotaped trials.[113] Research also indicates that writings may be more comprehensible than live performance because readers can re-read complex or pivotal sections carefully.[114] Presenting a more realistic trial experience would increase costs without appreciable benefit in terms of response accuracy. Vignette experiments may be used to estimate verdict preferences, but trial summaries need to reflect cases under review.

---

summary); Harmon M. Hosch et al., *Effects of an Alibi Witness's Relationship to the Defendant on Mock Jurors' Judgments*, 35 LAW & HUM. BEHAV. 127, 135 (2011) (two page case summaries); Dan M. Kahan, *Culture, Cognition, and Consent: Who Perceives What, and Why, in Acquaintance-Rape Cases*, 158 U. PENN. L. REV. 729, 765 (2010); Tisha R. A. Wiley & Bette L. Bottoms, *Effects of Defendant Sexual Orientation on Jurors' Perceptions of Child Sexual Assault*, 33 LAW & HUM. BEHAV. 46, 49 (2009) (two-page written scenario).

[112] *See* Steffen Bieneck, *How Adequate is the Vignette Technique as a Research Tool for Psycho-Legal Research*, *in* SOCIAL PSYCHOLOGY OF PUNISHMENT OF CRIME 255, 261-63 (Margit E. Oswald et al. eds., 2009) (discussing construction of vigenettes for mock jury research). Experienced litigators may be skeptical of the idea that an elaborate trial, lasting weeks or even months, can be effectively summarized in just a few minutes. However, this brevity is only possible due to the extensive work put into the trial.

[113] According to Professor Kahan, "Comparative evaluations of different testing formats suggest that mock jurors' reactions to detailed trial vignettes is strongly predictive of how they respond to more vivid forms of proof, including the testimony of live witnesses." Kahan, *supra* note 111, at 754 n.103.

The validity of vignette experiments has been demonstrated across multiple disciplines. *See* Karen Skilling & Gabriel J. Stylianides, *Using Vignettes in Educational Research: A Framework for Vignette Construction*, 43 INT'L J. RSCH. & METHOD EDUC. 541 (2020); Rhidian Hughes & Meg Huby, *The Application of Vignettes in Social and Nursing Research*, 37 J. ADVANCED NURSING 382 (2002); Kelly D. Wason et al., *Designing Vignette Studies in Marketing*, 10 AUSTRALASIAN MKTG. J. 41 (2002); Spencer C. Evans et al., *Vignette Methodologies for Studying Clinicians' Decision-Making: Validity, Utility, and Application in ICD-11 Field Studies*, 15 INT'L J. CLINICAL & HEALTH PSYCH. 160, 160-61 (2015); Thom Baguley et al., *Statistical Modelling of Vignette Data in Psychology*, 113 BRIT. J. PSYCH. 1143, 1144 (2022); Koen Migchelbrink & Steven Van de Walle, *When Will Public Officials Listen? A Vignette Experiment on the Effects of Input Legitimacy on Public Officials' Willingness to Use Public Participation*, 80 PUB. ADMIN. REV. 271, 272 (2020).

[114] *See* Lynne Forster-Lee et al., *The Bottom Line: The Effect of Written Expert Witness Statements on Juror Verdicts and Information Processing*, 24 LAW & HUM. BEHAV. 259, 268-29 (2000) (reporting that written witness statements help jurors understand complex testimony).

B. Representativeness of the Subject Pool

The researcher should obtain unbiased estimates of a jury pool's sentence preferences with and without the mitigation evidence at issue. To analyze how mitigation evidence changes jurors' sentencing decisions, one must define the population of interest carefully.[115]

1. Study Sample Should Represent Relevant Jurisdiction

Nationally representative samples, the basis for most public opinion research, should not be used to estimate sentencing preferences because no jury pool is nationally representative. Depending on the location and nature of the offense, the real population of interest may be an entire state, a judicial district within the state, or one of the state's counties.[116]

The sample should accurately reflect the basic demographic features of the relevant jurisdiction.[117] The sample should fairly represent the relevant jurisdiction, just as a jury venire should fairly represent a cross-section of the jurisdiction.[118] Some may think that respondents must be selected at random, but the objective is not randomness per se. No set of survey respondents constitutes a truly random sample, as participation is

---

[115] *See generally* Shari Seidman Diamond, *Reference Guide on Survey Research*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 359, 376-87, (3d ed. 2011).

[116] As a challenging case, consider trials conducted by court martials. Only members of the armed forces on active duty may serve as jurors. *See* 10 U.S.C. § 825(a)-(c). No civilians may serve on the jury, even when the defendant is accused of non-military offense. The unique composition of courts-martial juries presents some challenges, but a researcher could measure sentencing preferences among service members only. As with all cases, the researcher should carefully identify the relevant jury pool and estimate sentencing preferences in that population.

[117] All three studies reported in Table 2, above, use student subjects to some degree. *See* Barnett et al., *supra* note 19, at 756 (study participants are undergraduate students in introduction to psychology course and graduate students in social work program); Greene & Cahill, *supra* note 67, at 286 (participants are jury-eligible undergraduate students in psychology courses); Nuñez et al., *supra* note 67, at 870 (participants are combination of death-qualified undergraduate students and MTurk workers).

[118] The survey respondents, who voluntarily take online surveys, will not be representative of the jurisdiction. Online samples will generally overrepresent young adults, women, and those with college degrees. *See* Kevin E. Levay et al., *The Demographic and Political Composition of Mechanical Turk Samples*, 6 SAGE OPEN 1, 4-5 (2016). When the basic demographic characteristics of the jurisdiction are known, it is possible to calculate sampling weights as a technical adjustment. Properly weighted, a non-random sample can yield unbiased estimates. *See* notes 122 and 123, *infra*.

voluntary.[119] Demographic differences should be expected.[120] The real objective is obtaining unbiased estimates of quantities in the population as if the population was randomly sampled.

The researcher should collect information from respondents about their gender, age, ethnicity, race, income, and education. The sample respondents can then be compared to the population they are meant to represent. When such differences are identified, researchers can calculate and apply sample weights to adjust for overrepresentation or underrepresentation in the sample.[121] The researcher applies weights when analyzing sample data to obtain unbiased estimates of those quantities in the population.[122] Weighting observations is not a magic solution that elevates flawed

---

[119] Participants must provide informed consent to engage in human subject research. The researcher cannot summon respondents at random (even if the researcher could randomly select subjects for recruitment, the researcher cannot compel subjects to participate). The willingness to participate in survey research correlates with specific respondent characteristics, thereby introducing self-selection bias into all research samples.

[120] Some segments of the population are more likely to participate due to the mode of the survey (e.g., whether it is in-person, by phone, or online). Among those invited to take the survey, some are more willing to participate than others are. A truly random sample is not possible, but the researcher should nevertheless take appropriate measures to analyze results in an unbiased manner. *See* Dirk M. Elston, *Participation Bias, Self-Selection Bias, and Response Bias*, J. AM. ACAD. DERMATOLOGY (June 17, 2021), https://doi.org/10.1016/j.jaad.2021.06.025.

[121] A familiar example of analysis that weights observations is calculating a grade point average. Classes with more credit hours count more for GPA purposes than classes with fewer credit hours do and weighting classes by credit hour yields a better overall estimate of class performance compared to weighting all classes equally.

[122] A popular method of calculating sampling weights is iterative proportional fitting, more commonly called raking. It is the most prevalent method for calculating survey weights. *See* Andrew Mercer et al., *For Weighting Online Opt-In Samples, What Matters Most?*, PEW RSCH. CTR. (Jan. 26, 2018), https://www.pewresearch.org/methods/wp-content/uploads/sites/10/2018/01/Weighting-Online-Opt-In-Samples.pdf. One uses the raking method to balance survey samples along several dimensions. The method is called "raking" because it resembles how one uses a rake to level a patch of ground: You rake the ground smooth one direction, then rake sideways to fill in low spots, and keeping alternating until the patch is level in both directions. This iterative process is necessary because the dimensions being balanced may be correlated. The raking method identifies weights that balance the sample along one dimension, then adjusts weights to achieve balance along another dimension, and iterates among the weighting variables to find sampling weights that maximize balance across all desired dimensions simultaneously. For accessible introductions to the raking method of calculating weights, *see* Michael P. Battaglia et al., *Practical Considerations in Raking Survey Data*, 2 SURV. PRAC. 1, 1 (2009); Lew Anderson & Ronald D. Fricker, Jr., *Raking: An Important and Often Overlooked Survey Analysis Tool*, 48 PHALANX 36, 36 (2015); Christoph Waldhauser, *Survey: Computing Your Own Post-Stratification Weights in R*, R-

research to a gold standard, but it is a widely used and accepted statistical practice used to address the fact that some people are more likely to complete a survey than other people are.[123]

The researcher should not estimate jury pool sentence preferences from student samples. Compared to the adult population in general, typical undergraduate research subjects will have higher education, greater belief that people can change, more idealism, all of which suggest that an undergraduate jury pool would be less likely to support death sentences compared to real jury pools.[124] For many research applications, the distinctiveness of students presents no problems.[125] While much juror and

---

BLOGGERS (2014), https://www.r-bloggers.com/survey-computing-your-own-post-stratification-weights-in-r/.

[123] *See* Pierre Lavallée & Jean-François Beaumont, *Why We Should Put Some Weight on Weights,* SURV. METHODS: INSIGHTS FROM THE FIELD (Feb. 20, 2015), https://surveyinsights.org/?p=6255; Graham Kalton & Ismael Flores-Cervantes, *Weighting Methods*, 19 J. OFF. STAT. 81 (2003); Mercer et al., *supra* note 122. While it is standard practice to weight observations when analyzing survey data to estimate quantities in a population, they should be viewed as a technical adjustment and not a cure-all. The sample being weighted must represent the dimensions to be balanced. For example, it is not possible to weight observations to represent all age cohorts if survey respondents are all undergraduate students ages eighteen to twenty-two. To summarize, the sample of respondents does not need to be fully representative like a pure random sample because it is possible to weight observations and obtain unbiased estimates, but the more representative the sample is, the better sampling weights work. The analyst should examine the distribution of calculated sampling weights and should consider trimming or truncating extreme weights. *See generally* Stephen R. Cole & Miguel A. Hernán, *Constructing Inverse Probability Weights for Marginal Structural Models*, 168 AM. J. EPIDEMIOLOGY 656, 660-61 (2008); Frank Potter & Yuhong Zheng, *Methods and Issues in Trimming Extreme Weights in Sample Surveys*, PROC. AM. STAT. ASSOC. SURV. RSCH. METHODS SECTION (Aug. 8, 2015), http://www.asasrms.org/Proceedings/y2015/files/234115.pdf.

[124] *See, e.g.*, Brian H. Bornstein et al., *Mock Juror Sampling Issues in Jury Simulation Research: A Meta-Analysis*, 41 LAW & HUM. BEHAV. 13, 14-15, 22 (2017) (finding that students generally more liberal and less likely to support death sentences); Stacie R. Keller & Richard L. Wiener, *What Are We Studying? Student Jurors, Community Jurors, and Construct Validity*, 29 BEHAV. SCI. & L. 376, 377-78 (2011) (positing that student mock jurors will be more lenient than non-students).

[125] Typically, the researcher is focused on identifying a statistically significant positive or negative effect measured as the difference between two experimental conditions. The researcher is not interested in the average outcome in each group and is instead focused on the existence of a statistically significant difference between groups. The researcher can estimate a treatment effect as the difference observed between treatment and control groups; even if groups of students are not comparable to the general population, they are comparable to each other. In these situations, the issue is whether the treatment effect observed among students is a good estimate of the treatment effect in the adult population and

jury research can be conducted successfully with student subjects, a more diverse sample of adults is preferrable for analyzing a particular case.[126] A research finding based on students at the researcher's university may not represent the population that serves on juries.

2. Qualifications for Jury Service

Jury qualifications and exclusions further modify the relevant population for analysis. To serve on a federal jury, one must meet some basic requirements. All jurors must:[127] be a U.S. citizen, be at least eighteen years of age, have resided in the jurisdiction of the case for at least one year,[128] be proficient in English, have no disqualifying mental or physical condition, have no felony charges, and have no felony convictions.[129] In addition to these basic requirements, federal district courts can impose additional requirements for serving on juries.[130] Similarly, states may define different grounds for disqualifying individuals from service on state court juries.[131] Certain individuals may be excused from jury service: active duty members of armed services;[132] firefighters (volunteer and full-time), police officers, and public officers actively engaged in full-time public duties; individuals over the age of seventy; those who served on a jury

---

research demonstrates that treatment effects observed among students can be generalized to other populations.

[126] As discussed in Section III.C, it is not enough to know the difference between trial conditions. The effect of a change in juror preferences on verdict probabilities depends on what the juror preferences were originally.

[127] 28 U.S.C. § 1865(b)(1)-(5).

[128] Few states require that prospective jurors have lived in the court's jurisdiction for at least one year. The year-residency requirement appears to be confined to federal courts. Federal district courts have larger geographic jurisdictions than state courts; some federal district courts have jurisdiction over entire states.

[129] In most states and in federal courts, felons cannot serve on juries. *See* Brian C. Kalt, *The Exclusion of Felons from Jury Service*, 53 AM. U. L. REV. 65, 150 (2003). California now allows felons to serve on juries after they finish their sentences. *See* Debra Cassens Weiss, *New California Law Allows Felons Who Served Their Time to Serve on Juries*, ABA J. (Oct. 11. 2019). Other states have considered allowing felons to serve on juries. *See* Jacob Rosenberg, *Jury Duty is the Next Big Step for Felons' Rights*, MOTHER JONES (May 21. 2019).

[130] 28 U.S.C. § 1863(a).

[131] *See, e.g.*, FLA. STAT. § 40.013 (enumerating ten exclusions).

[132] 28 U.S.C. § 1863(b)(6). Although members of the armed forces on active duty are exempt from jury duty, the military encourages active-duty service members to serve on juries provided it does not interfere with their military responsibilities. *See* Ryan Guina, *Active Duty Military and Jury Duty Service,* THE MILITARY WALLET (Feb. 17, 2023), https://themilitarywallet.com/active-duty-military-jury-duty-service/; Heidi E. Loredo, *Jury Duty is Civil Duty,* U.S. MARINE CORPS (July 28, 2005), https://www.hqmc.marines.mil/News/Article/Article/551818/jury-duty-is-civil-duty/.

within the last two years;[133] and members of rescue squads and ambulance crews. Further disqualification occurs during *voir dire* questioning.[134]

In death penalty cases, the analyst should "death qualify" survey respondents because potential jurors may be excluded if they hold strong views on capital punishment.[135] The purpose of death qualification is to exclude jurors who have such strong beliefs about capital punishment that it would "prevent or substantially impair" their ability to follow legal instructions and evaluate the evidence.[136] Thus, the relevant population in capital trials excludes those with extreme views on capital punishment.[137]

Standard survey research practices should be adapted to evaluate the effect of omitted mitigation evidence. Standard analysis of a representative sample of adults in a jurisdiction will produce biased estimates because that sample does not represent those eligible to serve on juries. Instead, preferences should be estimated from jury-qualified subset of a representative sample.[138] The basic, supplemental, and death-related jury qualifications enable the analyst to identify respondents who represent a potential jury pool.[139]

---

[133] 28 U.S.C. § 1866(e).

[134] For-cause strikes can also be applied if the reasons to exclude respondents can be clearly identified. So long as juror qualifications, exemptions, and the basis for strikes can be articulated, they can be implemented.

[135] Some scholars criticized the practice of death-qualifying prospective jurors because the death-qualification process tends to prime jurors toward conviction and excluding jurors who would not impose a death sentence in the penalty phase unfairly stacks the jury in the prosecution's favor in the guilt phase (because jurors opposed to the death penalty are also less likely to find the defendant guilty). According to Dennis Devine, "the biasing effect of death qualification in the guilt phase of capital trials is now well established." DENNIS J. DEVINE, JURY DECISION MAKING: THE STATE OF THE SCIENCE 51 (2012). On the priming effect of death-qualification process, *see* Mike Allen et al., *Impact of Juror Attitudes about the Death Penalty on Juror Evaluations of Guilt and Punishment: A Meta-Analysis*, 22 LAW & HUM. BEHAV. 715, 716 (1998); Michael T. Nietzel et al., *Effects of Voir Dire Variations in Capital Trials: A Replication and Extension*, 5 BEHAV. SCI. & L. 467 (1987). On the increased likelihood of death-qualified jurors finding defendants guilty, *see* Joseph W. Filkins et al., *An Evaluation of the Biasing Effects of Death Qualification*, *in* THEORY AND RESEARCH ON SMALL GROUPS (R.S. Tindale et al. ed. 1998).

[136] Wainwright v. Witt, 469 U.S. 412, 424 (1985).

[137] Potential jurors who believe that anyone who takes a life should always lose his may also be excluded. *See* Morgan v. Illinois, 504 U.S. 719 (1992).

[138] The researcher should recruit surplus respondents knowing that some will not meet jury qualifications. Based on my own research, one should recruit twenty percent more respondents than the target sample size to estimate jury pool preferences in non-capital trial and sixty percent more for a capital trial.

[139] Although some respondents may not qualify for jury service, the analyst should allow them to complete the survey and exclude their responses from later analysis. It is better to err on the side of collecting too much data; it is easy to

3. Respondents Can be Recruited Online

To estimate sentencing preferences in a jury pool, a diverse sample of American adults recruited through online platforms is preferrable to a student sample.[140] Respondents recruited using an online platform like Amazon's Mechnical Turk (MTurk) will be more representative of the general public than a student sample.[141] This is not to say that online respondents are nationally representative, they are not, but they are more diverse and representative of jury pools than student samples, enabling adjustments with sampling weights.[142]

When researchers started conducting online surveys, there was some concern that the data would be low quality, but this concern has not been realized in practice. Survey research is routinely conducted online now, often with respondents recruited from MTurk, not only because it saves time and money, but also because it produces quality data.[143] An

---

exclude some observations afterwards, but it is impossible to collect data from dismissed subjects.

[140] The subject recruiting platform, MTurk, allows us to limit participation to workers located in the United States. The physical location of survey respondents can also be verified using IP addresses. *See* David G. Rand, *The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments*, 299 J. THEORETICAL BIOLOGY 172, 176 (2012) (reporting ninety-seven percent match between MTurkers' self-reported place of residence and IP-located addresses).

[141] *See* Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411, 414 (2010); Levay et al., *supra* note 118, at 2. The representativeness of subjects can be tested by comparing the demographic characteristics of the sample to the population characteristics reported by the Census Bureau.

[142] An interesting question is whether respondents must reside in a specific jurisdiction the same way that jurors must live in the jurisdiction. For my studies, I assume that Americans from any state can represent a particular state. It is possible to weight sample responses to represent major demographic characteristics of the local jurisdiction, but one wonders if there are other relevant and distinctive features of a local jurisdiction that are not represented on demographic dimensions. With the MTurk platform, it is possible to target respondents from a single state, but it is more expensive and reduces the number of potential respondents. Weighting observations to represent a population is a sound practice which improves facial validity, but it does not make a big difference with respect to verdict probabilities. Demographic controls should be sufficient, but one should maintain an open mind and focus on obtained unbiased estimated of verdict preferences in the relevant population.

[143] Keeping costs low is important. The criminal justice system is plagued by economic inequities. *See, e.g.*, Bright, *supra* note 21, at 1836-37; Barton & Bibas, *supra* note 22, at 972-77; Deborah L. Rhode, *Access to Justice*, 69 FORDHAM L. REV. 1785 (2001). If the cost of measuring the harm of omitted evidence is low enough, individuals who now suffer from injustice would be empowered to seek justice without having to rely on others. This includes not only inmates, but also crime victims who want fair sentences upheld. Crime victims are also

impressive body of research speaks to the validity of survey research conducted with respondents recruited from online platforms like Amazon MTurk.[144] MTurk has been used to conduct experimental research on jury decision making.[145] As for litigation, according to Kenneth Plevan, a partner at Skadden, surveys used to show confusion, or secondary meaning, in Lantham Act cases are now commonly conducted online.[146] Overall, studies show that MTurk workers are more attentive and diligent than respondents recruited by polling companies or students taking surveys under direct supervision.[147] The quality of surveys completed by MTurk workers is no doubt a reflection of that system's accountability mechanisms and

---

disadvantaged in the criminal justice system. *See generally* Robert C. Davis & Carrie Mulford, *Victim Rights and New Remedies: Finally Getting Victims Their Due*, 24 J. CONTEMP. CRIM. JUST. 198 (2008); Douglas E. Beloof, *The Third Wave of Crime Victims' Rights: Standing, Remedy, and Review*, BYU L. REV. 255 (2005); Paul G. Cassell, *Protecting Crime Victims in Federal Appellate Courts: The Need to Broadly Construe the Crime Victims' Rights Act's Mandamus Provision*, 87 DENV. U. L. REV. 599 (2009).

[144] *See* Alexander Coppock, *Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach*, 7 POL. SCI. RSCH. & METHODS 613 (2019); John J. Horton et al., *The Online Laboratory: Conducting Experiments in a Real Labor Market*, 14 EXPERIMENTAL ECON. 399 (2011); John Bohannon, *Social Science for Pennies*, 334 SCI. 307 (2011); Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351 (2012); Michael Buhrmester et al., *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 PERSPS. PSYCH. SCI. 3 (2011); Michael D. Buhrmester et al., *An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use*, 13 PERSPS. PSYCH. SCI. 149 (2018); Christoph Bartneck et al., *Comparing the Similarity of Responses Received from Studies in Amazon's Mechanical Turk to Studies Conducted Online and with Direct Recruitment*, 10 PLOS ONE 1 (2015).

[145] *See, e.g.*, Daniel Krauss & Nicholas Scurich, *The Impact of Case Factors on Jurors' Decisions in a Sexual Violent Predator Hearing*, 20 PSYCH. PUB. POL'Y & L. 135 (2014); Schwartz & Seaman, *supra* note 104, at 456-57 (experiment on the effect of varying standards of proof on juror decision in patent infringement lawsuit; short vignette patterned on real case).

[146] "The courts seem to have accepted online surveys as an appropriate approach with no reported discussion about the validity of this new research method. In this respect, the courts appear to have simply followed the lead of the marketing research profession." Kenneth A. Plevan, *Daubert's Impact on Survey Experts in Lanham Act Litigation*, 95 TRADEMARK REP. 596, 599 n.11 (2005).

[147] *See* Eyal Peer et al., *Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk*, 46 BEHAV. RSCH. METHODS 1023 (2014) (high-reputation MTurk workers rarely fail attention checks); Jeremy Kees et al., *An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk*, 46 J. ADVERT. 141 (2017) (respondents recruited via MTurk outperform survey respondents from professional polling organizations); Bingbing Zhang & Sherice Gearhart, *Collecting Online Survey Data: A Comparison of Data Quality Among a Commercial Panel & MTurk*, 13 SURV. PRAC. 1 (2020).

incentives to complete tasks carefully.[148] Researchers are justifiably concerned about the quality of online surveys and the risks posed by bad actors, but the technology provides some mechanisms to address potential problems.[149]

When it comes to issues like effort and inattention, one wonders whether online respondents are any different than courtroom jurors. Judges and litigators have frequently criticized jurors for not paying close attention to the evidence and instructions.[150] Real trials have consequences, but not for jurors because no one with a stake in the outcome can serve on the jury. Real trials are not as exciting as they appear on television; real jurors get bored, distracted, and have a hard time paying attention. In other words, real jurors, like people who take surveys online, are susceptible to the same distractions and critiques, but the research shows

---

[148] Respondents are not automatically paid per task completed; their work must first be approved. A good record of approved work gives MTurkers opportunities to complete more rewarding tasks. Currently, the MTurk system appears to support quality survey research, but that may change. One should not lock into using a particular vendor, but rather should use vendors that provide quality service.

[149] *See, e.g.*, Ryan Kennedy et al., *The Shape of and Solutions to the MTurk Quality Crisis*, 8 POL. SCI. RSCH. & METHODS 614 (2020) (discussing the threat posed by virtual private networks used to disguise location and solutions to problem); Justin M. Stritch et al., *Crowdsourced Data in Public Administration Research: A Review and Look to the Future*, 85 PUB. ADMIN. REV. 581 (2025).

Some surveys include questions designed to identify inattentive respondents and exclude them from analysis. A typical screener question offers a set of answer choices and, within the prompt, instructs the respondent to ignore the rest of the prompt and select a specific answer (proving that they read the entire prompt carefully). While screener questions, or other manipulation checks, serve some benefit, it is not clear that they consistently identify inattentive respondents. It does not appear to be a widespread problem among respondents recruited from MTurk and there may be better ways to neutralize the problem (e.g., by randomizing the order of answers) rather than excluding respondents. Screening out inattentive respondents may even introduce bias because there are inattentive people who do not need instructions thoroughly in the population.

[150] Widespread criticism of juries motivated Kalven and Zeisel's pioneering research. *See* KALVEN & ZEISEL, *supra* note 58, at 3-11. Although Kalven and Zeisel offered a favorable report on American juries, distrust and suspicion has persisted. *See* Kenneth S. Klein, *Unpacking the Jury Box*, 47 HASTINGS L.J. 1325, 1325-26 (1995) (noting the increasing belief that the typical jury is incapable of making the difficult decisions called for in many trials"); Lisa Blomgren Bingham, *When We Hold No Truths to be Self-Evident: Truth, Belief, Trust, and the Decline in Trials*, J. DISP. RESOL. 131 (2006) (discussing loss of confidence in juries, and other institutions, to determine truth); Christopher Robertson & Michael Shammas, *The Jury Trial Reinvented*, 9 TEX. A & M L. REV. 109, 110-15 (2021) (discussing distrust of juries in criminal and civil litigation); James A. Shapiro & Karl T. Muth, *Beyond a Reasonable Doubt: Juries Don't Get It*, 52 LOY. U. CHI. L.J. 1029 (2020) (discussing juror misinterpretations of reasonable doubt standard and proposing reforms).

that ordinary people who occasionally do these jobs take them seriously and do them reasonably well.

C. Accounting for Deliberation Process

Studies of individual verdict preferences, such as the studies summarized in Table 2, do not speak directly to the probability of a different outcome. The relationship between verdict preferences and jury trial outcomes is mediated by a deliberation process. The deliberation process can either magnify or dampen the effect of changed juror sentence preferences. The Court has appropriately recognized that the

> scales of justice may be delicately poised between guilt and innocence . . . [t]hen error, which under some circumstances would not be ground for reversal, cannot be brushed aside as immaterial, since there is a real chance that it might have provided the slight impetus which swung the scales toward guilt.[151]

There is a gap between the outcome measured in juror preference studies and the relevant outcome in "reasonable probability of different outcome" analysis. This Section discusses what is known about the deliberation process and how deliberation modifies the effect of mitigation evidence omitted from death penalty trials.

1. How Juries Deliberate

The American criminal justice system places an extraordinary amount of trust in juries, small groups of ordinary people, randomly selected from the community. At the end of a trial, jurors are instructed to select a foreperson, follow the law, deliberate until they reach a verdict, but they are not told how to deliberate. The jury deliberates behind closed doors, without supervision or oversight.

---

[151] Glasser v. United States, 315 U.S. 60, 67 (1942).

What happens behind closed doors in the jury room?[152] After a bit of small talk, jurors turn to their first task: selecting a foreperson.[153] Jurors typically select a foreperson within the first fifteen minutes of deliberation.[154] Next, most juries take a preliminary vote to discover where they all stand. They are not instructed to take a vote; they simply vote spontaneously by show of hands, written ballots, or by announcing their votes one at a time. Juries typically take their first poll twenty to forty-five minutes into deliberations, before any substantial discussion of the trial.[155]

If the jury must deliberate, its initial poll defines two opposing verdict factions, each attempting to win over the other, like two sides competing in a game of tug-of-war.[156] Deliberation is largely the process of persuading jurors in the minority faction to conform to the majority position.[157] Whichever side wins the jury's initial poll typically prevails in the

---

[152] Our understanding of jury deliberation comes from post-trial interviews with actual jurors, the collective wisdom of judges and litigators, and mock jury research. Mock jury research is particularly informative because researchers can study mock jurors before, during, and after their deliberations. Researchers can manipulate jury size, instructions, composition, and trial evidence. They can even record mock jury deliberations.

Simon's excellent study of the insanity defense was the first to record and transcribe jury deliberations. The experimental setting was very realistic; with the cooperation of local judges, Simon recruited actual jurors summoned to jury duty. *See* RITA JAMES SIMON, THE JURY AND THE DEFENSE OF INSANITY 35-37 (1967). More recently, researchers have conducted large-scale studies with videotaping of jury deliberations. *See* Shari Seidman Diamond et al., *Juror Discussions During Civil Trials: Studying an Arizona Innovation*, 45 ARIZ. L. REV. 1 (2003).

[153] The first juror to bring up the need to select a foreperson often hears another juror say: "So why don't you do it?" *See* Franklin J. Boster et al., *An Information-Processing Model of Jury Decision Making*, 18 COMMC'N RSCH. 524, 538 (1991); Dennis J. Devine et al., *Explaining Jury Verdicts: Is Leniency Bias for Real?*, 34 J. APPLIED SOC. PSYCH. 2069, 2080 (2004).

If no one volunteers for the job, intentionally or by accident, jurors typically select someone who has served on a jury before, has a high-status occupation, is the oldest person in the room, or happens to be sitting at the head of the table. *See* Phoebe C. Ellsworth, *Are Twelve Heads Better Than One?*, 52 LAW & CONTEMP. PROBS. 205, 213-14 (1989); Nancy S. Marder, *Gender Dynamics and Jury Deliberations*, 96 YALE L.J. 593, 594-95 (1986).

[154] *See* DEVINE, *supra* note 135, at 154-55.

[155] *See* Valerie P. Hans et al., *The Hung Jury: The American Jury's Insights and Contemporary Understanding*, 39 CRIM. L. BULL. 33 (2003); Diamond et al., *supra* note 152, at 61.

[156] If the jury's initial poll is unanimous, they do not need to deliberate in any meaningful sense. The jurors would still need to elect a foreperson and may discuss the trial, but they do not need to deliberate to create agreement because they already agree on the verdict.

[157] *See* Solomon E. Asch, *Opinions and Social Pressure*, 193 SCI. AM. 31 (1955); Solomon E. Asch, *Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority*, 70 PSYCH. MONOGRAPHS: GEN. & APPLIED 1, 56-58 (1956).

deliberation. As Professor Harry Kalven and Hens Zeisel observed in their seminal book *The American Jury*, the deliberation process "does not so much decide the case as bring about the consensus, the outcome of which has been made highly likely by the distribution of first ballot votes. The deliberation process might well be likened to what the developer does for an exposed film: it brings out the picture, but the outcome has been predetermined."[158]

The influence of a verdict faction during deliberation is directly related to its size.[159] Social influence is a powerful force. Most jurors respond to social cues, seek common ground with others, and avoid being perceived as disagreeable and obstinate.[160] The probability that a capital jury returns a life sentence increases as the number of jurors in favor of a life sentence increases.

The deliberation process is also shaped by the proof beyond a reasonable doubt standard. This feature of criminal law is believed to create a leniency effect, or bias, in deliberation.[161] If jurors are evenly split between life and death sentences, the defendant should receive a life sentence because the division of opinion is social proof of reasonable doubt.[162] These general observations inform an empirical analysis of jury deliberation.

2. Empirical Analysis of Deliberation

Previous studies of jury deliberation provide a large sample of deliberations where real or mock jurors observe a trial, form preferences about the verdict, and then deliberate in groups to reach a verdict. The data analyzed in this part come from more than half a century of research on jury deliberation in criminal trials.

---

[158] KALVEN & ZEISEL, *supra* note 58, at 489.

[159] *See* Nicole L. Waters & Valerie P. Hans, *A Jury of One: Opinion Formation, Conformity, and Dissent on Juries*, 6 J. EMPIRICAL LEGAL STUD. 513, 536-40 (2009) (discussing that larger factions attract members for a variety of reasons: they can recall more facts and evidence from the trial than smaller factions can, dissenters trust the wisdom of the crowd, and peer pressure encourages conformity); DEVINE, *supra* note 153, at 152-53; REID HASTIE ET AL., INSIDE THE JURY 106-08 (1983).

[160] There is, however, some evidence that the tendency to conform has declined over time. *See* Rod Bond & Peter B. Smith, *Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task*, 119 PSYCH. BULL. 111, 124 (1996).

[161] *See* Robert J. MacCoun & Norbert L. Kerr, *Asymmetric Influence in Mock Jury Deliberation: Jurors' Bias for Leniency*, 54 J. PERSONALITY & SOC. PSYCH. 21, 21 (1988); Waters & Hans, *supra* note 159, at 516.

[162] Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 PSYCH. PUB. POL'Y & L. 622, 693 (2001) (cautioning that "the strong leniency effect observed in laboratory studies may be weaker or less reliable in actual juries").

Most of the data points are derived from the Devine et al. compilation of twenty-four jury deliberation studies, published in 2001.[163] I have updated that work with five additional studies: Devine et al. (2007),[164] Devine and Kelly (2015),[165] Devine et al. (2004),[166] Sandys & Dillehey (1995),[167] and Hannaford-Agor et el. (2002).[168] Data from these twenty-nine studies allow for the analysis of the relationship between juror preferences and jury verdicts, based on thousands of data points.[169] The available evidence, summarized in Figure 1, reveals the relationship between a jury's initial poll and the verdict it ultimately returns.[170]

---

[163] *Id.* at 691 tbl.6 (showing an excellent compilation of twenty-four studies by Devine and colleagues that provides the starting point for this analysis).

[164] Dennis J. Devine et al., *Deliberation Quality: A Preliminary Examination in Criminal Juries*, 4 J. EMPIRICAL LEGAL STUD. 273, 293 tbl.4 (2007) (showing the data from the field study of Indiana juries).

[165] Dennis J. Devine & Christopher E. Kelly, *Life or Death: An Examination of Jury Sentencing with the Capital Jury Project Database*, 21 PSYCH. PUB. POL'Y & L. 393, 396-402 (2015). For additional analysis of the Capital Jury Project Database, see Theodore Eisenberg et al., Jury Responsibility in Capital Sentencing: An Empirical Study, 44 BUFF. L. REV. 339 (1996); Scott E. Sundby, *War and Peace in the Jury Room: How Capital Juries Reach Unanimity*, 62 HASTINGS L.J. 103 (2010).

[166] Devine et al., *supra* note 153, at 2082 tbl.2 (showing data by jury size from the field study). I use data from this study as discussed by Kerr and MacCoun. *See* Norbert L. Kerr & Robert J. MacCoun, *Is the Leniency Asymmetry Really Dead? Misinterpreting Asymmetry Effects in Criminal Jury Deliberation*, 15 GRP. PROCESSES & INTERGROUP RELS. 585, 588 tbl.2 (2012) (summarizing the data from Devine's 2004 study).

[167] Marla Sandys & Ronald C. Dillehay, *First-Ballot Votes, Predeliberation Dispositions, and Final Verdicts in Jury Trials*, 19 LAW & HUM. BEHAV. 175 (1995). I use data from this study as adjusted by Kerr & MacCoun. *See* Kerr & MacCoun *supra* note 166, at 592 tbl.4 (summarizing the adjusted data from Sandys and Dillehay's study).

[168] Paula L Hannaford-Agor et al., *Are Hung Juries a Problem?*, NAT'L CTR. FOR STATE CTS. (Sept. 30, 2002), https://www.ncsc-jurystudies.org/__data/assets/pdf_file/0018/6138/hung-jury-final-report.pdf (containing 382 cases from four state courts systems). The data from this study are publicly available. *See* Paula L. Hannaford-Agor et al., *Evaluation of Hung Juries in Bronx County, New York, Los Angeles County, California, Maricopa County, Arizona, and Washington, DC, 2000-2001*, INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (Mar. 30, 2006).

[169] Some observations come from death penalty deliberations, but most do not. Controlled analysis, the results of which are reported in Table 3, allows us to assess whether death and guilt deliberations are different, and shows us that they are not different, once we account for juror preferences.

[170] Hung juries that fail to reach a verdict are excluded from the analysis.

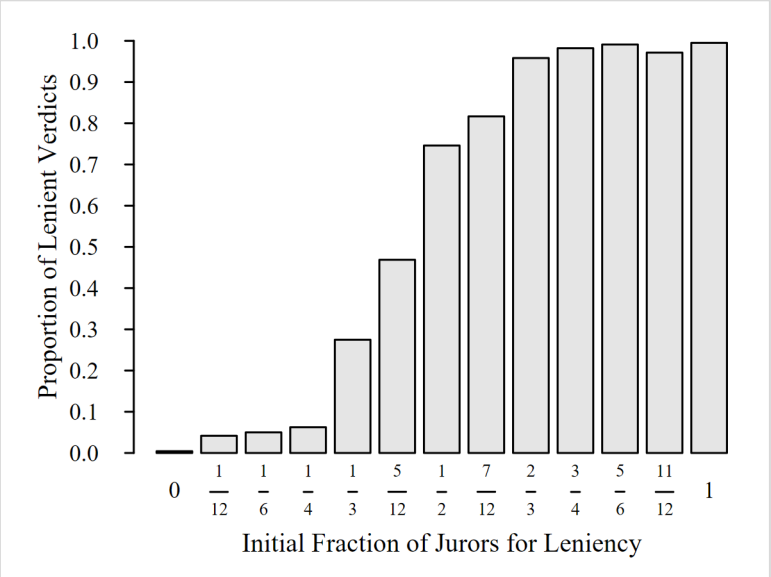

Figure 1: Initial Jury Polls and the Proportion of Lenient Verdicts

Figure 1 describes the relationship between the fraction of jurors who initially favor leniency and the probability of a lenient verdict. The effect of one additional juror in lenient verdict faction is not constant; it depends on how many other jurors currently favor leniency. If there are two or fewer votes for leniency, an additional vote for leniency does not create much probability of a lenient verdict. There is not much difference in the height of the first few bars in Figure 1. If there are eight or more votes for leniency, the probability of a lenient verdict is already so high, another vote for leniency has little effect. The last five bars in Figure 1 are roughly the same height. If three to seven jurors vote for leniency, another vote for leniency has a dramatic effect on the probability of a lenient verdict. In this middle range, Figure 1's bar heights increase dramatically.

We can analyze this relationship more carefully with logistic regression analysis, a special type of regression analysis used to study binary outcomes, like a jury's decision between life and death sentences.[171] Similar to well-known linear regression analysis, logistic regression analysis

[171] In logistic regression analysis, a link function translates a linear equation into predicted probabilities of the outcome of interest. The relationship between jurors' initial preferences and the probability of a life sentence is not linear, but the relationship between jurors' initial preferences and the *logged odds* of a jury returning a life sentence may be a linear relationship. The odds of an event occurring equal the probability of the event occurring divided by the probability of the event not occurring; $odds = p / (1 - p)$ where p is the probability of the outcome of interest. The logged odds of an outcome are a linear function of one or more explanatory variables.

produces the terms of an equation that best fit observed data points.[172] The logistic regression equation is initially estimated using only jurors' initial preferences, measured as the proportion of jurors who initially favor a leniency, to predict jury verdicts.[173] An additional equation which accounts for possible effects of deliberating the death penalty and using six-person juries is also estimated.

The results of logistic regression analysis, reported in Table 3, confirm that initial juror preferences have a profound impact on jury verdicts. Consider the initial equation estimated. The coefficient for jurors' initial preferences in Model 1 is a positive and statistically significant value.[174] This is not a new or surprising finding, but empirical analysis allows us to quantify the terms of this relationship and avoid relying on general, qualitative descriptions.[175]

---

[172] Linear regression and logistic regression utilize different criteria for identifying the terms that "best fit" observed data. The criteria used for linear regression analysis is known as ordinary least squares and for logistic regression it is maximum likelihood.

[173] Measuring initial preferences as the proportion of jurors who favor leniency, rather than the raw number, allows for the analysis of different jury sizes and makes efficient use of the full sample. There is some debate over how to measure initial preferences when jurors abstain from the first poll or say they are undecided. Devine and his colleagues assume these jurors have reasonable doubts and prefer a not guilty verdict. Devine et al., *supra* note 153, at 2078. Kerr and MacCoun suggest a different approach for undecided and abstaining jurors. Kerr and MacCoun divide their votes between guilty and not guilty to measure initial support for a guilty verdict. Kerr & MacCoun, *supra* note 166, at 589-90. Following the suggestion of Kerr and MacCoun, I divide undecided votes and abstentions between guilty and not guilty votes where possible. First vote tallies in the Hannaford-Agor et al. study's cases were calculated from juror surveys. *See* Hannaford-Agor et al., *supra* note 168, at "Description of Variables" (2006).

[174] Table 3's Model 1 coefficients are essentially a simple line equation: $y = -4.91 + 11.78x$ where x is the proportion of jurors who initially favor leniency and y is the logged odds of a lenient verdict. The standard errors reported below the coefficients quantify the uncertainty of the coefficient estimates that is inherent to analyzing a finite sample of data. Here, the coefficient 11.78 is estimated from $n = 2{,}303$ observations. Given this sample size, this coefficient estimate will random vary from its true value by .50, on average. Given this uncertainty, it is reasonable to think the true coefficient for jurors' initial preferences could be within one or two standard errors of the estimated value, but it is not reasonable to think that the true value of coefficient is zero (which would imply that jurors' initial preferences have zero effect on jury verdicts). Because the random error inherent to analyzing finite samples does not account for the estimated value, 11.78, we say this result is statistically significant. It is the signal of a systematic relationship and is not simply noise.

[175] Jurors' initial preferences explain jury verdicts very well. The simple model's logistic regression coefficient for initial juror preferences, 11.78, is positive and statistically significant (P-value < .001). The various measures of model fit are strong. This sole predictor reduces null deviance by 59%, reduces outcome

EMPIRICAL ANALYSIS OF JURY DELIBERATIONS

Variable	*Model 1*	*Model 2*
Juror's initial preferences	11.78***	11.63***
	(0.50)	(0.50)
Death penalty case		-0.21
		(0.26)
Six-person jury		0.37*
		(0.17)
Constant	-4.91***	-5.06***
	(0.22)	(0.26)
Likelihood ratio	1854.67	1864.23
Pseudo *R*-square	0.59	0.59
Reduction in error	73.7%	73.7%
Correctly classified	88.6%	88.6%

Notes: Sample size = 2,303; dependent variable is the logged odds of a lenient verdict (not guilty or life sentence); standard errors in parentheses; * = p < .05, *** = p < .001.

Table 3: Empirical Analysis of Jury Deliberations

As mentioned, we can analyze the effect of jurors' initial preferences while controlling for other variables that may affect jury verdicts, independent of their influence on jurors' initial preferences: capital punishment and jury size.[176] We may assess whether juries deliberating guilt decisions or six-person juries are different than twelve-person juries deliberating death sentences.

prediction errors by 73.7%, and correctly classifies 88.6% of sample outcomes. For further discussion of measures of logistic regression model fit, *see* POLLOCK III & EDWARDS, *supra* note 10, at 295-97.

[176] Variables affecting both initial preferences and deliberation may confound causal analysis and distort regression results. Some variables may influence how jurors deliberate without altering their initial preferences. Such variables may affect how preferences translate to verdicts, but not whether jurors initially favor a lenient verdict. For example, if the jury's vote for a death sentence does not need to be unanimous, more death sentences should occur, even without changes in jurors' initial preferences, because the threshold for that decision is lower.

Similarly, cases involving multiple charges, or lesser included offenses may affect how jurors deliberate. The option of convicting on lesser offenses, or on only some charges can lead to compromise verdicts, altering how juror preferences are aggregated to reach unanimous verdicts. Unanimity rules and verdict forms do not affect juror preferences, and therefore do not confound analysis of the relationship between juror preferences and jury verdicts. Section IV.B discusses how complex verdict forms complicate the relationship between juror preferences and jury verdicts and require further research.

The punishment phase of a capital trial differs from the guilt phase, but it is unclear whether deliberating punishment is different than deliberating guilt. The relationship between initial preferences and verdict probabilities could be stronger, weaker, or no different than it is for guilt decisions.[177] If the death penalty deliberation is significantly different than deliberating guilt, it is important to account for those differences when analyzing the effect of mitigation evidence on the probability of a life or death sentence.

Jury size has been the subject of ongoing debate and scholarly research, but the impact of using a jury with fewer than twelve members remains unclear. Smaller juries deliberate differently from larger ones.[178] Smaller juries take less time to deliberate but are more likely to hear from everyone.[179] It remains unclear whether six-person juries are more or less

---

[177] On the one hand, jurors may be reluctant to impose the ultimate punishment, potentially heightening the leniency effect in death penalty cases. On the other hand, the jury's advisory role and the judge's authority to override a death sentence recommendation in favor of life imprisonment may diminish jurors' sense of responsibility. Generally, a judge cannot override jury recommendations for life imprisonment, as the Sixth Amendment requires a jury to find every fact necessary to impose a death sentence. *See* Hurst v. Florida, 577 U.S. 92 (2016).

Alabama remains the only state that permits trial judges to override a jury's life imprisonment recommendation and impose a death sentence. *See* Patrick Mulvaney & Katherine Chamblee, *Innocence and Override*, 126 YALE L.J. F. 118, 118 (2016). Despite ordering Alabama to reconsider its laws in light of *Hurst*, the Supreme Court has subsequently refused to hear challenges to Alabama's death penalty laws. *See* Richard Wolf, *Supreme Court Lets Alabama Judges Impose Death Penalty*, USA TODAY (Jan. 23, 2017), https://www.usatoday.com/story/news/politics/2017/01/23/supreme-court-alabama-florida-death-penalty-judge-jury/96947280/.

It is also possible that the relationship between initial juror preferences and jury verdicts does not differ in the punishment phase of a capital trial. Prior studies suggest that death sentences are deliberated in the same manner as guilt decisions. One mock jury study reported no heightened leniency effect when jurors deliberate whether to impose death sentence. *See* Mona Lynch & Craig Haney, *Capital Jury Deliberation: Effects on Death Sentencing, Comprehension, and Discrimination*, 33 LAW & HUM. BEHAV. 481, 491 (2009). Another study found that the jury's recommendation of a death sentence is "strikingly similar" to its decision to convict the defendant. *See* Theodore Eisenberg et al., *Forecasting Life and Death: Juror Race, Religion, and Attitude Toward the Death Penalty*, 30 J. LEGAL STUD. 277, 283-84 (2001) (analyzing death penalty juries in South Carolina).

[178] Jury size also affects jurors' initial preferences because it affects the distribution of initial preferences observed on juries drawn from the jury pool. Statistical evidence cited in *Ballew* indicated that the "prejudices of individuals were frequently counterbalanced, and objectively reduced" in larger jury groups. Ballew v. Georgia, 435 U.S. 223, 233 (1978).

[179] Smaller juries spend less time deliberating and are more likely to reach verdicts. *See* Michael J. Saks & Mollie Weighner Marti, *A Meta-Analysis of the Effects of Jury Size*, 21 LAW & HUM. BEHAV. 451, 457-61 (1997). *But see* Barbara

lenient than twelve-person juries.[180] Research on the impact of jury size on jury verdicts is inconclusive.[181] For present purposes, we need not take sides in the jury size debate. No state allows six-person juries to decide between life and death sentences. At the same time, we should account for the different jury sizes employed in studies of deliberation.[182]

The coefficient for death penalty cases is not statistically significant, indicating that the effect of deciding a death penalty case is indistinguishable from zero.[183] The coefficient for six-person juries is positive and statistically significant. Empirical analysis suggests that six-person juries are

---

Luppi & Francesco Parisi, *Jury Size and the Hung-Jury Paradox*, 42 J. LEGAL STUD. 399, 404 (2013) (positing that smaller juries are not more likely to reach verdicts).

[180] It is particularly interesting to compare 5-1 and 10-12 votes for guilty verdicts. Although these initial splits are mathematically equivalent fractions, there may different deliberation dynamics. There are reasons to believe a 5-1 jury will be more punitive than a 10-2 jury. A dissenting opinion is more likely to resist conforming to majority opinion if joined by an ally. Two jurors, allied in dissent, may hold up better against ten jurors than one lone vote against five jurors. The Supreme Court has observed that "a person in the minority will adhere to his position more frequently when he has at least one other person supporting his argument." *Ballew*, 435 U.S. at 236. *See also* Robert H. Miller, *Six of One is not a Dozen of the Other: A Reexamination of Williams v. Florida and the Size of State Criminal Juries*, 146 U. PENN. L. REV., 621, 654 (1998).

However, there are also reasons to believe a 5-1 jury will be *less punitive* than a 10-2 jury. Five people exert less peer pressure than ten people do. People are less intimidated to speak up in small group discussions. *See* Nicolas Fay et al., *Group Discussion as Interactive Dialogue or as Serial Monologue: The Influence of Group Size*, 11 PSYCH. SCI. 481, 481 (2000). *See generally* Rod Bond, *Group Size and Conformity*, 8 GRP. PROCESSES & INTERGROUP RELS. 331, 332 (2005) (questioning the assumption that there is a single function that describes the relationship between group size and conformity). This social dynamic may allow one to resist the influence of five more than two can resist the influence of ten.

[181] *See* Joan B. Kessler, *An Empirical Study of Six-and Twelve-Member Jury Decision-Making Processes*, 6 U. MICH. J.L. REFORM 712, 715-16 (1973); J. Clark Kelso, *Final Report of the Blue Ribbon Commission on Jury System Improvement*, 47 HASTINGS L.J. 1433, 1489-90 (1995); Saks & Marti, *supra* note 179, at 465; Norbert L. Kerr & Robert J. MacCoun, *The Effects of Jury Size and Polling Method on the Process and Product of Jury Deliberation*, 48 J. PERSONALITY & SOC. PSYCH. 349, 354 (1985); Luppi & Parisi, *supra* note 179, at 402-05; Adam M. Chud & Michael L. Berman, *Six-Member Juries: Does Size Really Matter*, 67 TENN. L. REV. 743, 756 (1999).

[182] Observed deliberations in the sample are 54.8% six-person juries and 45.2% twelve-person juries.

[183] This does not necessarily imply that death penalty decisions are no different than guilt decisions. The gravity of capital punishment may affect initial juror preferences but having accounted for jurors' initial verdict preferences in these cases, there does not appear to be a direct effect of death penalty type cases on deliberation.

slightly *more likely* to return lenient verdicts, holding jurors' initial verdict preferences constant.[184]

The estimated relationship between the number of jurors in favor of a life sentence and the probability of a life sentence can be visualized. Figure 2 presents the predicted probabilities of life sentences, based on logistic regression analysis.[185] Figure 2 does not paint a different picture of

[184] The difference between empirical results and the theoretical expectation that small juries are more likely to convict because dissenting juror has an "ally" may be due to the impact of hung juries. According to Devine et al., of 17 twelve-person juries observed with 10-2 initial vote distributions, there were 11 guilty verdicts, 0 not guilty verdicts, and 6 hung juries. Devine et al., *supra* note 162, at 691 tbl.6. Thus, the conviction rate was 68.75% if hung juries are included but 100% if hung juries are excluded. Of the 164 six-person juries observed with initial 5-1 votes, there were 136 guilty verdicts, 5 not guilty verdicts, and 23 hung juries. *Id.* Thus, the conviction rate starting from 5-1 initial votes was 82.93% including hung juries and 96.45% if hung juries are excluded. When hung juries are excluded, the conviction rate is lower starting from a 5-1 vote (96.45%) than it is starting from a 10-2 vote (100%).

[185] The logistic regression coefficients in Table 3 are reported in terms of logged odds (the natural logarithm of the odds of a guilty verdict). While they are necessary for logistic regression analysis, logged odds are not an intuitive metric for communicating the likelihood of a guilty verdict. Fortunately, logged odds can be translated into predicted probabilities to better understand the results of logistic regression analysis. *See* POLLOCK III & EDWARDS, *supra* note 10, at 289-91; ANDREW GELMAN & JENNIFER HILL, DATA ANALYSIS USING REGRESSION AND MULTILEVEL/HIERARCHICAL MODELS 79-82 (2007).

the deliberation process than Figure 1, but does allow us to quantify the effect of deliberation more precisely than descriptive analysis does.

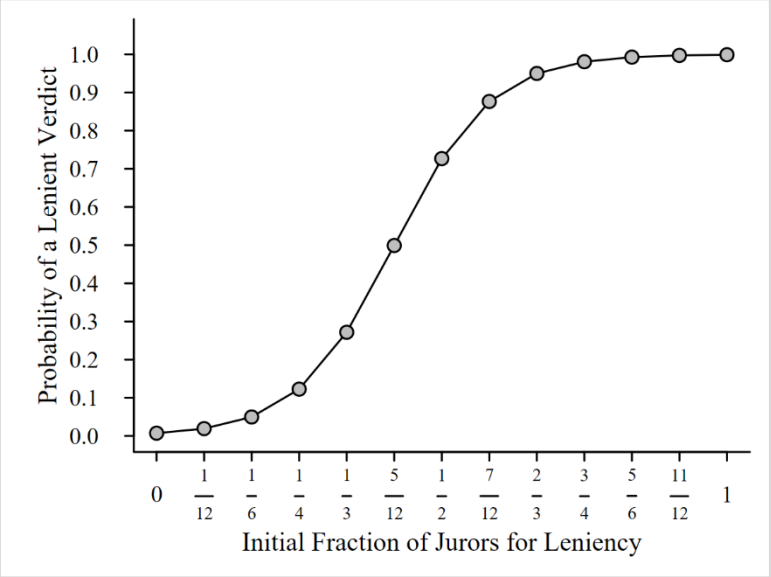

Figure 2: Initial Jury Polls and the Proportion of Lenient Verdicts

As expected, the marginal effect of an additional vote for a life sentence is greatest when the jury is evenly divided and smallest when there is a strong majority for or against a life sentence. The marginal effect of an additional vote for life is most significant when the jury is deeply divided. Based on logistic regression analysis, a twelve-person jury with three votes for life has a .086 probability of returning a life sentence. With one additional vote for life, the probability of a life sentence increased to .198. A fifth vote for life increases the probability of a life sentence to .394 and a sixth vote for life increases the probability of a life sentence to .631.[186] If the life-sentence faction gains a seventh member, the probability of a life sentence increases to .819. If eight jurors favor a life sentence, the probability increases to .922. With nine or more jurors in favor of a life sentence, a life sentence is nearly certain. The effect of one juror changing his or her vote cannot be assessed without knowing how other jurors vote.

3. Implications of the Varying Effects of One Changed Mind

The intervening process of deliberation highlights the inadequacy of approximating the harmfulness of omitting mitigation evidence by one's opinion of that evidence. Even if one's opinion is indicative of what jurors think, the impact of a juror's opinion depends on what other jurors think

[186] Note the asymmetry of Figure 1, which supports the leniency shift of criminal juries.

and cannot be judged in isolation. Even if you think like a typical juror, the effect of your opinion depends on the opinions of other jurors. The evidence may only have a minor effect on juror opinions, but a minor change in juror preferences will have an outsized effect on the deliberation process if opinions are divided. The evidence may have a major effect on juror opinions, but its effect on deliberation may be limited if it is not widely shared.

Accounting for the deliberation process underscores the importance of estimating the verdict preferences of a sample that represents the jury-qualified adults from the relevant jurisdiction.[187] The effect of additional mitigation evidence on the probability of a life sentence depends on the baseline percentage of jurors who would support a life sentence without hearing that mitigation evidence. It is not enough to accurately estimate the change in preferences. One must also accurately estimate the baseline level of support for a life sentence with and without the mitigation evidence at issue.[188]

Understanding deliberation effects challenges the assumption that there is a reasonable probability of a different outcome in a capital trial if there "a reasonable probability that at least one juror would have struck a different balance" between death and life imprisonment.[189] The "one juror" language implies that a single vote for life imprisonment can prevent a death sentence; however, a death sentence is nearly certain if 11 jurors favor it. Again, the impact of one juror's changed verdict preference depends on how many other jurors support a life sentence.

When it comes to assessing the effect of omitted mitigation evidence on the probability of a death sentence in litigated cases, one should not assume, based on the jury returning a death sentence, that the jury started off unanimously in favor of a death sentence and that the "one juror" who strikes a different balance, after hearing mitigation evidence, would be a lone voice for leniency. A jury that returns a death sentence after deliberation may have started with a 7-5 vote for the death penalty. If that were the case, one juror striking a different balance, thus starting deliberation from a 6-6 vote, would have reduced the probability of a death sentence by .237. In this situation, where the jury is deeply divided on the sentence, the change in verdict probability is nearly *three times* the change in verdict

---

[187] *See* discussion Subsection II.B.1-2 above.

[188] As noted in Subsection II.B.1, *supra*, it is normally enough for the researcher to accurately estimate a treatment effect which can be done by comparing treatment and control groups composed of student subjects. To accurately estimate the effect of some treatment (like the addition of mitigation evidence) on the probability of a life sentence, one must accurately estimate the values being compared in the population, not just the difference between them, because the relationship between juror preferences and jury verdicts is not linear.

[189] Wiggins v. Smith, 539 U.S. 510, 537 (2003).

preference.[190] In close cases, those attempting to demonstrate the prejudicial effect of omitted mitigation evidence by showing its effect on verdict preferences have underestimated the probability of a different outcome by failing to account for the jury deliberation process.[191]

CONCLUSION

The challenge of proving the prejudicial effect of omitted mitigation evidence in death penalty cases is both profound and urgent. Mitigation evidence is a critical mechanism for humanizing defendants and enabling jurors to make informed, individualized sentencing decisions. The omission of such evidence has the potential to alter the course of a trial, raising serious questions about the fairness of death sentences imposed without a full consideration of the defendant's background and circumstances.

Despite the centrality of this inquiry, appellate courts lack a systematic approach to evaluate whether a trial error or omission created a "reasonable probability of a different outcome."[192] Courts have largely relied on intuition and abstract standards rather than engaging in rigorous, empirical assessments of how such omissions affect jury deliberations and verdicts.

Even when a trial error or omission is well-defined, existing methods do not allow the effect of a trial error or omission to be "quantitatively assessed in the context of other evidence presented," leaving the fairness of trials in doubt.[193] Empirical tools can estimate the effect that mitigation may have had on jurors' decisions in certain situations. Properly designed research has the potential to lift analysis of the effects of trial error and omissions from the realm of opinion and speculation. However, existing studies do not permit courts to quantitatively assess the extent to which ineffective assistance of counsel increased the probability of a sentence in

---

[190] When a jury is 7-5 in favor of a death sentence and one juror moves to the life sentence faction, this 1/12th change in juror verdict preferences (.083 preference change) corresponds to .237 increase in the probability of a life sentence. The deliberation process increases the effect of preference change by a factor of 2.85.

[191] It is possible to account for the deliberation process when estimating the effect of omitted mitigation evidence on the probability of a different trial outcome. A full account of deliberation process exceeds the scope of this article, but as a general idea, the probability of selecting 0, 1, 2, . . . 12 jurors who prefer a death sentence given the proportion of jurors in the pool who prefer a death sentence can be calculated from a famous statistical formula. Analysis of deliberation, like that reported in Subsection II.D.2, identifies the probability of a death sentence starting with 0, 1, 2, . . . 12 jurors who prefer a death sentence. Thus, one may calculate the probability of a death verdict based on jury pool preferences and the change in verdict probabilities that results from a change in preferences. *See* Edwards, *supra* note 47, at 34-37; Barry Edwards, *SATE: Scientific Analysis of Trial Errors,* THE COMPREHENSIVE R ARCHIVE NETWORK (Sept. 11, 2024), https://cran.r-project.org/package=sate.

[192] Strickland v. Washington, 466 U.S. 668, 694 (1984).

[193] Arizona v. Fulminante, 499 U.S. 279, 308 (1991).

a litigation context. Even if courts avoid the substitute question heuristic and consider research that quantifies the probability of a different outcome, current research designs do not address the target question in specific cases. The disconnect between theory and practice underscores the need for a more precise, evidence-based approach to determining prejudice.

The broader implications of this issue cannot be overstated. When courts fail to properly assess the impact of omitted evidence, they risk perpetuating miscarriages of justice in cases where someone's life is on the line. Even when we know what was presented and omitted from his trial, it is still not possible to objectively say, one way or the other, whether the defendant received a fair trial. How frustrating it must be for the parties to keep arguing about the fairness of a trial for decades until the review process is finally exhausted. In a system that strives to balance justice with fairness, such failures erode public confidence in the administration of justice and the legal process as a whole. Only by embracing a more rigorous approach to evaluating the effects of trial errors and omissions can we hope to achieve fairness and uphold the integrity of our justice system.

***

APPENDIX

SUBSEQUENT HISTORY OF CASES WITH OMITTED MITIGATION EVIDENCE

Petitioner Name	*Subsequent History*
Terry Williams	Williams agreed to serve a life sentence without possibility of parole.[194]
Gary Cone	Cone's post-conviction litigation reached the Supreme Court two more times.[195] In 2013, his petition for relief based on a *Brady* violation was denied.[196] Cone died of natural causes in 2016.[197]
Kevin Wiggins	Wiggins agreed to serve a life sentence with possibility of parole.[198]
Ronald Rompilla	Rompilla agreed to serve a life sentence without possibility of parole.[199]
Jeffrey Landrigan	Landrigan was executed three years after the Court's decision.[200]
George Porter	Porter died of natural causes in prison in 2016.[201]
Robert Van Hook	Van Hook was executed in 2018.[202]
Fernando Belmontes	Belmontes died on death row from unknown causes in 2017.[203]

[194] *See* Brooke Masters, *Deal Gets Inmate Off Death Row*, WASH. POST (Nov. 14, 2000), https://www.washingtonpost.com/archive/local/2000/11/15/deal-gets-inmate-off-death-row/2edd2e13-a860-482f-bf30-86b84e4e0774.

[195] *See* Bell v. Cone, 543 U.S. 447 (2005) (upholding constitutionality of state's "especially heinous, atrocious, or cruel" aggravating circumstance); Cone v. Bell, 556 U.S. 449 (2009) (vacating death sentence because prosecution may have withheld material evidence).

[196] *See* Cone v. Colson, 925 F. Supp. 2d 927, 1020 (W.D. Tenn. 2013).

[197] *See* Amanda Haggard, *Gary Cone Dies on Death Row*, NASHVILLE SCENE (Apr. 20, 2016), https://www.nashvillescene.com/news/gary-cone-dies-on-death-row/article_51f9e342-a3fa-534d-83ee-abb932c68193.html.

[198] *See* Eric Rich, *Md. Inmate Gets Life in '88 Case*, WASH. POST (Oct. 7, 2004), https://www.washingtonpost.com/archive/local/2004/10/08/md-inmate-gets-life-in-88-case/2d2ac0e7-182a-49d9-9dcc-7d3e6b6f7516.

[199] *See Murderer to Spend Life in Prison*, THE MORNING CALL (Aug. 14, 2007), https://www.mcall.com/2007/08/14/murderer-to-spend-life-in-prison-death-sentence-voided-in-torture-slaying-of-allentown-bar-owner.

[200] *See* Edecio Martinez, *Jeffrey Landrigan Execution: "Boomer Sooner" Last Words Before Ariz. Executes Inmate*, CBS NEWS (Oct. 27, 2010), https://www.cbsnews.com/news/jeffrey-landrigan-execution-boomer-sooner-last-words-before-ariz-executes-inmate-27-10-2010.

[201] *See Inmate Release Info. Detail*, FLA. DEPT. CORR. (Nd), http://www.dc.state.fl.us/offenderSearch/detail.aspx?Page=Detail&DCNumber=110825&TypeSearch=IR (last visited Sept. 30, 2024).

[202] *See* Cameron Knight, *"I'm No Good." Ohio Executes "Homosexual Panic" Murderer and Killer of Hyde Park Man*, CINCINNATI ENQUIRER (July 17, 2018), https://www.cincinnati.com/story/news/crime/crime-and-courts/2018/07/17/execution-homosexual-panic-murderer-set-10-m/792535002.

[203] *See Condemned Inmate Fernando Belmontes Jr., Dies of Unknown Causes*, CAL. DEPT. CORR. & REHAB. (Oct. 3, 2017), https://www.cdcr.ca.gov/

Demarcus Sears	Following the Supreme Court's decision, the state habeas court determined that the omission of mitigation evidence did not prejudice Sears.[204] Sears then turned to the federal court; he recently obtained a favorable decision from the Eleventh Circuit.[205]
Scott Pinholster	Pinholster has continued post-conviction litigation since the Court's decision.[206] He remains on death row.[207] Gov. Gavin Newsome has imposed a moratorium on executions in California.[208]
Terence Andrus	Following the Court's decision, the Texas Court of Criminal Appeals determined that ineffective counsel's omission of a "tidal wave" of mitigation evidence did not prejudice Andrus.[209] The Supreme Court declined to review that decision.[210] Andrus committed suicide on death row six months later.[211] Andrus's own perspective on the litigation can be read in *Harvard Law Review*.[212]
Matthew Reeves	Reeves was executed in 2022.[213]
Danny Jones	Jones remains on death row in Arizona.[214]

***

news/2017/10/03/condemned-inmate-fernando-belmontes-jr-dies-of-unknown-causes.

[204] *See* Sears v. Humphrey, 751 S.E.2d 365, 368 (Ga. 2013) (upholding denial of writ of habeas corpus).

[205] *See* Sears v. Warden, 73 F.4th 1269 (11th Cir. 2023) (granting writ of habeas corpus due to prejudicial effect of an inequitable expert discovery rule).

[206] *See* Marisa Gerber, *An L.A. Court Mistakenly Destroyed Evidence A Death Row Inmate Says Would Free Him. Now What?*, L.A. TIMES (Dec. 17, 2017), https://www.latimes.com/local/lanow/la-me-death-row-evidence-destruction-20171217-story.html.

[207] *See Condemned Inmate List*, CAL. DEPT. CORR. & REHAB. (Oct. 8, 2024), https://www.cdcr.ca.gov/capital-punishment/condemned-inmate-list-secure-request (last visited Oct. 15, 2024).

[208] *See* Scott Shafer & Marisa Lagos, *Gov. Gavin Newsom Suspends Death Penalty in California*, NPR (Mar. 12, 2019), https://www.npr.org/2019/03/12/702873258/gov-gavin-newsom-suspends-death-penalty-in-california.

[209] *See* Ex Parte Andrus, 622 S.W.3d 892 (Tex. Crim. App. 2021).

[210] *See* Andrus v. Texas, 142 S. Ct. 1866 (2022).

[211] *See Texas Death Row Prisoner Commits Suicide*, DEATH PENALTY INFO. CTR. (Feb. 1, 2023), https://deathpenaltyinfo.org/texas-death-row-prisoner-commits-suicide.

[212] *See* Terence Andrus, *Reflection on Andrus v. Texas*, 134 HARV. L. REV. F. 78 (2020).

[213] *See* Jodi Lopez, *Lawyer for Executed Matthew Reeves: "He Was My Responsibility and I Loved Him"*, AL.COM (Apr. 11, 2022), https://www.al.com/news/2022/04/lawyer-for-executed-matthew-reeves-he-was-my-responsibility-and-i-loved-him.html.

[214] *See Inmate Data Search*, ARIZ. DEPT. CORR., https://inmatedatasearch.azcorrections.gov/PrintInmate.aspx?ID=092576 (last visited Mar. 12, 2025).