A SCIENTIFIC FRAMEWORK FOR ANALYZING THE HARMFULNESS OF TRIAL ERRORS

Barry Edwards

Abstract

Judgments about the harmfulness of trial errors have profound consequences for defendants, the criminal justice system, and the public. Judges are expected to assess harm accurately, but they cannot hear directly from jurors and may only speculate about the difference a trial error made. Even experienced judges have a hard time predicting what jurors think and what juries will do. Fortunately, scientific principles and research can assist judges in conducting harmless error analysis. This Article offers a framework for testing claims about the harmfulness of trial errors. It specifies the prosecution's burden to prove a trial error was harmless on direct appeal as well as the defendant's burden to prove a trial error was harmful in post-conviction proceedings. Hypotheses about the harmfulness of errors can be visualized and tested rigorously. Scientific analysis of trial errors can help courts assess the harmfulness of trial errors more accurately, efficiently, and confidently.

About the Author

Barry C. Edwards is an associate lecturer in the Department of Political Science at the University of Central Florida. 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 sincerely thanks the editors at the UCLA Criminal Justice Law Review for their helpful feedback and suggestions.

Table of Contents

Introduction

Appellate courts use harmless error analysis in deciding whether a trial error was so harmful that the defendant is entitled to a new trial, or whether it was harmless enough that the defendant should remain incarcerated.[1] Evaluating the harmfulness of trial errors has profound implications for individual defendants, the criminal justice system, and society at large.[2] The harmless error doctrine is "probably the most cited

1. One should distinguish the plain meaning of "harmless" from the legal definition of "harmless" in this context. Here, harmless does not mean without harm. A trial error may cause harm but be deemed harmless on appeal if the harm was relatively minor. *See* discussion *infra* Sections II.B, IV.D, and V (further discussing the threshold between harmful and legally harmless errors). The term "legally harmless" is not used in harmless error analysis, but social scientists should be alerted to the special meaning of "harmless" in this context.

2. Trial errors, and the inability to accurately distinguish harmful and harmless errors, erode public confidence in the criminal justice system. On one hand, the public loses confidence when appeals courts cannot identify and correct mistakes that lead to wrongful convictions. On the other hand, the public loses confidence when appeals courts allow seemingly endless appeals to those who have been justly convicted. Both problems stem from an inability to accurately evaluate the harmfulness of trial errors.

rule in modern criminal appeals."[3] Harmless error analysis may determine the fate of more appeals than any other legal doctrine.[4]

On the surface, harmless error analysis appears to ask a simple empirical question: Did the trial error cause harm or was it harmless?[5] The seemingly simple question of whether a trial error was harmful raises some complicated issues. Many judges and commentators have argued that the harmless error doctrine is confusing and difficult to apply.[6] Presently, there is no reliable way to assess whether a trial error was harmful. The U.S. Constitution and federal statutes do not contain instructions for conducting harmless error analysis.[7] Judges are expected to measure harm by conducting thought experiments about hypothetical trials.[8] Their analysis of harm is subjective and riddled with problems. Some of the confusion and difficulty comes from the need to make normative judgments about the fairness of trials, but much of the confusion arises from the imprecise definition of concepts, unstated assumptions, and limited concern for practical measurement.[9]

---

3. William M. Landes & Richard A. Posner, *Harmless Error*, 30 J. Legal Stud. 161, 161 (2001).
4. Hon. John M. Walker Jr., *Harmless Error Review in the Second Circuit*, 63 Brook. L. Rev. 395, 395 (1997).
5. *See* Brown v. United States, 411 U.S. 223, 231 (1973) ("[A] defendant is entitled to a fair trial but not a perfect one."); *see also* Fed. R. Crim. Proc. 52(a) ("Any error, defect, irregularity, or variance that does not affect substantial rights must be disregarded."); Jason M. Solomon, *Causing Constitutional Harm: How Tort Law Can Help Determine Harmless Error in Criminal Trials*, 99 Nw. U.L. Rev. 1053, 1063 (2005) ("Fundamentally, the harmless-error inquiry is an empirical, if unanswerable, one: what impact did the error have on the actual jury's verdict?"); D. Alex Winkelman, et al., *An Empirical Method for Harmless Error*, 46 Ariz. St. L.J. 1405, 1414 (2014) ("the question is ultimately one of causation").
6. *See generally* Peck v. U.S., 102 F.3d 1319 (2nd Cir. 1996) (en banc) (Newman, J., concurring) (summarizing the points of conflict and confusion in the Supreme Court's harmless error jurisprudence).
7. All state appeals court follow some version of the harmless error doctrine. *See* Chapman v. California, 386 U.S. 18, 22 (1967); States may, however, apply different standards in civil and criminal cases. Florida, for example, until recently required civil appellants to establish a "miscarriage of justice" to prevail on appeal. *See* J. Robert M. Gross & David R. Maass, *Harmless Error in Civil Appeals*, 89 Fla. Bar J. 10, 14 (2015).
8. Harmless error analysis requires counterfactual reasoning about causal outcomes in hypothetical conditions. *See* David R. Dow & James Rytting, *Can Constitutional Error Be Harmless?*, Utah L. Rev. 483, 500-02 (2000); *see also* discussion *infra* Part III.
9. Judges may agree on the amount of harm caused by a trial error but disagree about the error's harmfulness because they have different levels of error tolerance; this is a normative disagreement. *See* discussion *infra* Part IV. Judges may have the same levels of error tolerance but disagree about the amount of harm caused by a trial error because they focus on different types of harm, consider different counterfactual conditions, apply different burdens of prove, or estimate harm relative to difference reference points; these are not normative disagreements, but rather results of poorly defined measurement strategies. *See also* discussion *infra* Parts II, IV, and V.

This Article provides a precise framework for analyzing the harmfulness of trial errors based on the harmless error doctrine. This framework allows prosecutors and defendants to state their claims about trial errors as testable hypotheses. Moreover, it identifies the quantities to be measured in harmless error analysis and how to interpret results of analysis to a reasonable degree of scientific certainty.[10]

Part I identifies the scope of harmless error analysis. Although harmless error analysis is applied to many types of trial errors, it is not applied to all types of error, and we should be aware of its limitations. Part II addresses several important questions implicated by the harmless error doctrine: Which harms matter? What is the threshold of harm that separates harmful errors from harmless ones? What is the burden of proof and who bears it? Answering these questions will reveal uses for scientific analysis of trial errors and help establish how this analysis should be conducted. Part III discusses some problems judges have in subjectively evaluating the harmfulness of trial errors. It is hard for judges to predict what reasonable jurors would think and do if a trial error did not occur. Scientific analysis of the harmfulness of trial errors can help judges decide whether a trial error was harmful using more objective measurements grounded in research. Part IV outlines a framework for analyzing whether a trial error caused intolerable harm to the defendant. Finally, Part V offers a framework for analyzing the harm an error has caused to a trial.

I. Scope of Harmless Error Analysis

An infinite number of factors potentially affect trial outcomes. Detailing all of the factors thought to influence judges and juries is beyond the scope of this Article, but it is helpful to sort the variables that potentially affect trials and trial outcomes into three categories: (a) variables that affect the structural integrity of trials and are thus not subject to harmless error analysis, (b) variables that affect trial outcomes but are not amenable to judicial remedies, and (c) variables subject to harmless error analysis.

A. *Errors Affecting the Structural Integrity of Trials*

Some errors undermine the integrity of judicial proceedings and warrant reversal even if the defendant is unable to show the error harmed him personally. Fundamental errors that jeopardize parties' substantial rights should be reversed on appeal to protect the integrity of the legal system. "Without these basic protections," the U.S. Supreme Court has stated, "a criminal trial cannot reliably serve its function as a vehicle for determination of guilt or innocence, and no criminal punishment may be

10. This is one of several articles I am writing on the scientific analysis of trial errors. This article presents a theoretical framework for analyzing the harmfulness of trial errors. Other articles show how to measure quantities of interest and demonstrate the validity of the analysis using test cases.

regarded as fundamentally fair."[11] For example, if a judge denies the public access to a trial, the trial lacks integrity and cannot stand. When this type of error occurs, the defendant does not need to show that the error caused him harm;[12] structural errors are presumed to infect everything that happens during a trial from beginning to end.[13] Appellate courts may assess whether an error of this type was harmful on their own initiative, even if neither party raises the issue.[14] Therefore, harmless error analysis is inapplicable to these types of structural errors.

The Supreme Court, through a series of decisions, has identified certain structural errors as inherently harmful. Structural errors that the Court found warrant automatic reversal include:

1. Denial of the right to counsel.[15]
2. Denial of self-representation at trial.[16]
3. Denial of the right to a public trial.[17]
4. Trial conducted by a biased trial judge or prosecutor with personal interest in case.[18]
5. Trial conducted by a magistrate lacking jurisdiction.[19]
6. Racial discrimination in the selection of jurors.[20]

---

11. Rose v. Clark, 478 U.S. 570, 577–78 (1986); *see also* Zachary L. Henderson, *A Comprehensive Consideration of the Structural-Error Doctrine*, 85 Mo. L. Rev. 965 (2020) (discussing structural errors).
12. Even if the defendant did not make a timely objection to the structural error, an appeals court may take correction action. *See* Fed. R. Crim. P. 52(b) ("A plain error that affects substantial rights may be considered even though it was not brought to the court's attention."); *see also* Charles Alan Wright et al., Federal Practice and Procedure 506 (4th ed. 2013) (citing U.S. v. Olano, 507 U.S. 725, 734 (1993)).
13. Neder v. United States, 527 U.S. 1, 8–9 (1999); U.S. v. Gonzalez-Lopez, 548 U.S. 140, 148–149 (2006).
14. *See* Roger J Traynor, The Riddle of Harmless Error 26 (1970) ("Whether or not counsel are helpful, it is still the responsibility of the appellate court, once it concludes there was an error, to determine whether the error affected the judgment.")
15. *See* Gideon v. Wainwright, 372 U.S. 335 (1963); *cf.* Holloway v. Arkansas, 435 U.S. 475 (1978) (trial court requiring defendants to be jointly represented is a structural error); U.S. v. Hamilton, 391 F.3d 1066, 1071 (9th Cir. 2004) (defense counsel's absence is structural error); Geders v. United States, 425 U.S. 80 (1976) (denial of access to counsel during trial is a structural error); Penson v. Ohio, 488 U.S. 75 (1988) (denial of access to counsel on appeal is a structural error); U.S. v. Gonzalez-Lopez, 548 U.S. at 144–148 (denial of access to counsel of choice is a structural error).
16. *See* Faretta v. California, 422 U.S. 806 (1975); McKaskle v. Wiggins, 465 U.S. 168 (1984).
17. *See* Waller v. Georgia, 467 U.S. 39, 49 n.9 (1984); Owens v. US, 483 F.3d 48, 64 (1st Cir. 2007); U.S. v. Thunder, 438 F.3d 866, 868 (8th Cir. 2006).
18. *See* Tumey v. Ohio, 273 U.S. 510 (1927); Young v. U.S. *ex rel.* Vuitton et Fils S.A., 481 U.S. 787, 809–14 (1987).
19. *See* Gomez v. U.S., 490 U.S. 858, 876 (1989) (magistrate presides over jury selection despite defense counsel's objection is a structural error).
20. *See* Batson v. Kentucky, 476 U.S. 79 (1986) (racial discrimination in selection of petit jurors is structural error); Vasquez v. Hillery, 474 U.S. 254 (1986)

7. Improper exclusion of jurors because of their views about capital punishment.[21]
8. Defective reasonable doubt instructions.[22]
9. Subjecting the defendant to double jeopardy.[23]
10. Prejudicial pretrial publicity that denies defendant a fair trial.[24]
11. Denial of the right to a speedy trial.[25]

This list identifies the structural errors the Supreme Court has identified to date, but it is not necessarily exhaustive as new categories may arise in the future. Distinguishing structural errors from other trial errors will likely require case-by-case determinations as novel situations arise.[26]

B. *Some Significant Trial Variables Are Not Amenable to Judicial Remedies*

Some variables that affect trials and trial outcomes cannot be managed or remedied by judges and, therefore, are not subject to harmless error analysis. Many of these variables have been studied by social scientists, and a growing body of research indicates that courtroom decisions are influenced by environmental factors that affect our emotions and capacity to engage in deliberate, logical reasoning.[27] Courtroom

(discrimination in selection of grand jurors is structural error). *See also* Eric L. Muller, *Solving the Batson Paradox: Harmless Error, Jury Representation, and the Sixth Amendment*, 106 Yale L.J. 93 (1996).

21. *See* Gray v. Mississippi, 481 U.S. 648 (1987) (finding potential juror with reservations about death penalty but otherwise qualified to serve should not have been removed for cause).
22. *See* Sullivan v. Louisiana, 508 U.S. 275, 281 (1993) (finding defective jury instructions regrading reasonable doubt are structural errors because "a misdescription of the burden of proof . . . which vitiates *all* the jury's findings"). Most defective jury instructions affect only some aspects of a trial and are subject to harmless error analysis.
23. *See* Price v. Georgia, 398 U.S. 323, 331 (1970).
24. *See* Sheppard v. Maxwell, 384 U.S. 333 (1966) (reasoning harm can be presumed where judge failed to protect trial from media circus).
25. *See* Zedner v. U.S., 547 U.S. 489, 507–09 (2006) (finding violations of Speedy Trial Act not subject to harmless error analysis); *see also* Strunk v. United States, 412 U.S. 434, 439–40 (1973) (finding harms caused by violating right to speedy trial cannot be remedied by granting defendant a new trial).
26. For an interesting example of an unusual error that blurs the line between structural errors and trial errors, *see* Rice v. Wood, 77 F.3d 1138 (9th Cir. 1996). Defendant Rice was not present when jury returned from deliberation and sentencing him to death; court deems trial error and not structural error. *Id.* at 11141. *See also* Susan Yorke, *Jury Nullification Instructions as Structural Error*, 95 Wash. L. Rev. 1441 (2020).
27. A significant body of social science research indicates that voting decisions are influenced by seemingly random and irrelevant events like the weather, the outcome of football games, and shark attacks. *See e.g.* Andrew J. Healy, et al., *Irrelevant Events Affect Voters' Evaluations of Government Performance*, 107 Proc. Nat'l Acad. Scis. 12804 (2010) (football games); Ethan C. Busby et al., *The Political Relevance of Irrelevant Events*, 79 J. Pol. 346 (2017) (football games); Anna Bassi, *Weather, Risk, and Voting: An Experimental Analysis of the Effect of Weather on Vote Choice*, 6 J. Experimental Pol. Sci. 17 (2019) (weather

decisions are likely influenced by the temperature of the courtroom;[28] courtroom lighting and seating arrangements;[29] and whether the decision maker ate recently or is hungry,[30] is pressed for time,[31] or drank a warm beverage while listening to a witness testify.[32] These types of contextual and environmental factors are similar to structural errors because they may affect trials from start to finish; however, they are different from structural errors because reversing a conviction and ordering a new trial will not change the weather outside, traffic conditions, the economy, or other extralegal factors. These trial variables can be studied in scientific experiments, but they are not subject to harmless error analysis because judges cannot offer remedies.

Consider, for example, the effect of a defendant's physical appearance on the outcome of his trial. A significant body of social science research indicates that physical attractiveness is correlated with leniency from juries. Attractive defendants are less likely to be convicted and tend to receive shorter sentences compared to less attractive defendants.[33]

---

events); Christopher H. Achen & Larry M. Bartels, Democracy for Realists: Why Elections Do Not Produce Responsive Government 116-45 (2017) (shark attacks).

28. *See* Xun Irene Huang et al., *Warmth and Conformity: The Effects of Ambient Temperature on Product Preferences and Financial Decisions*, 24 J. Consumer Psychol. 241 (2014); Amar Cheema & Vanessa M. Patrick, *Influence of Warm Versus Cool Temperatures on Consumer Choice: A Resource Depletion Account*, 49 J. Mktg. Rsch. 984 (2012); Yonat Zwebner et al., *The Temperature Premium: Warm Temperatures Increase Product Valuation*, 24 J. Consumer Psychol. 251 (2014).
29. *See* Steven Shepard, Comment, *Should the Criminal Defendant be Assigned a Seat in Court?*, 115 Yale L.J. 2203, 2208 (2006) ("[I]f a defendant is permitted to choose the clothes that he will be seen wearing, then he should be allowed to choose where, and at what distance, he will be seen, because his proximity to the jury will have enormous influence on how jurors perceive him."); *See also* Linda Mulcahy, *Architects of Justice: The Politics of Courtroom Design*, 16 Social & Legal Studies 383 (2007). The related issues of witnesses testifying by live video streaming also affects courtroom decisions and the fairness of criminal trials. *See e.g.* Gail S Goodman et al., *Face-to-Face Confrontation: Effects of Closed-Circuit Technology on Children's Eyewitness Testimony and Jurors' Decisions*, 22 L. & Hum. Behav. 165 (1998).
30. *See* Shai Danziger et al., *Extraneous Factors in Judicial Decisions*, 108 Procs. Nat'l Acad. Scis. 6889 (2011) (noting that decisions of Israeli parole board judges significantly affected by time since last meal).
31. *See* John M. Darley & C. Daniel Batson, *"From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior*, 27 J. Personality & Soc. Psychol. 100 (1973) (noting that willingness to help someone in distress depends on whether someone thinks they are in a hurry).
32. *See* Lawrence E. Williams & John A. Bargh, *Experiencing Physical Warmth Promotes Interpersonal Warmth*, 322 Science 606 (2008) (noting that interviewers who touched hot beverage rated subject more favorably); Yoona Kang et al., *Physical Temperature Effects on Trust Behavior: The Role of Insula*, 6 Soc. Cognitive & Affective Neuroscience 507 (2010).
33. *See* Angela S Ahola et al., *Justice Needs a Blindfold: Effects of Gender and Attractiveness on Prison Sentences and Attributions of Personal Characteristics in*

Given the influence of personal appearance on trial outcomes, how far must trial courts go to allow defendants to look their best for trial? The case law suggests the state cannot make the defendant look guilty at trial, but the state cannot be compelled to make the defendant look attractive.[34] However, an unattractive defendant is not entitled to spa services and teeth whitening before trial.[35] The defendant's appearance can help or harm his chances at trial, but it is an extralegal variable not subject to harmless error analysis because a court cannot change a defendant's basic appearance.

These observations about the effect of the defendant's appearance on trials and trial outcomes can also be made about the attractiveness of the defendant's attorney, the victim, the witnesses, and so on. The personal characteristics of all trial participants, including their race or ethnicity,[36]

a Judicial Process, 16 Psychiatry, Psychol. & L. S90 (2009); Michael E. Antonio, *Arbitrariness and the Death Penalty: How the Defendant's Appearance During Trial Influences Capital Jurors' Punishment Decision*, 24 Behav. Sci. & Law 215 (2006) (discussing the effect of whether the defendant appears engaged and/or remorseful). This "halo effect" is exemplified by the prosecution of serial killer Ted Bundy. *See* Jon Nordheimer, *All-American Boy on Trial*, N.Y Times, Dec. 10, 1978, at SM24 ("those who saw him for the first time agreed . . . [t]here must have been some terrible mistake. Here was a young man who represented the best in America, not its worst. Here was this terrific looking man with light brown hair and blue eyes").

34. For example, the defendant should not be compelled to appear in prison attire because the prisoner's garb suggests guilt. However, the defendant must object to being compelled to wear prison attire and, in some cases, defendants may be compelled to appear at trial in prison garb to protect public safety. *See* Estelle v. Williams, 425 U.S. 501, 507–12 (1976).

35. Criminal defense attorneys make concerted efforts to have their clients dress appropriately for court. In some jurisdictions, public defenders solicit clothing donations for indigent defendants because the state does not buy courtroom attire for them. *See* Dennis McCarthy, *Poor Defendants Deserve Good Clothes, Too*, L.A. Daily News (March 27, 2014, 7:39 AM), https://perma.cc/R63R-XXMB.

36. The scholarly literature on the effect of the defendant's race and ethnicity on criminal trial outcomes is extensive. *See e.g.* Jill K. Doerner & Stephen Demuth, *The Independent and Joint Effects of Race/Ethnicity, Gender, and Age on Sentencing Outcomes in US Federal Courts*, 27 Justice Q. 1 (2010); Ronald S. Everett & Roger A. Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. Quantitative Criminology 189 (2002); Laura T. Sweeney & Craig Haney, *The Influence of Race on Sentencing: A Meta-Analytic Review of Experimental Studies*, 10 Behav. Sci. & Law 179 (1992); Tara L. Mitchell et al., *Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment*, 29 L. and Hum. Behav. 621 (2005).

gender,[37] age,[38] health,[39] and sexual orientation,[40] may significantly affect trial outcomes. Generally, jurors are more lenient on defendants who are similar to them and treat defendants who are different than them more harshly.[41] The personal characteristics of trial participants matter, but there is not much a judge can do about them. A court cannot assign someone else to play the role of the defendant in a new trial to avoid biases jurors may harbor against the actual defendant.

Furthermore, courts are skeptical of using evidence that a general practice is harmful to call into question the outcome of a specific case.[42] It is unlikely an individual litigant will be able to point to a general practice as the determining factor in the outcome of his or her trial, but policymakers should consider general practices because minor effects in many trials cause significant impacts when aggregated. While appellate court judges may not be positioned to manage general trial practices that influence whether a defendant received a fair trial, like courtroom design, lighting, and seating, other policymakers may be able to use this type of

37. *See* Jill K. Doerner & Stephen Demuth, *Gender and Sentencing in the Federal Courts: Are Women Treated More Leniently?*, 25 Crim. Just. Pol'y Rev. 242 (2014); Candace Kruttschnitt, *Respectable Women and the Law*, 23 Socio. Q. 221 (1982).

38. *See* Doerner & Demuth, *supra* note 36, at 13–14 (finding that "the youngest defendants are the most likely to receive prison sentences" and that defendants under age 30 receive longer prisoner sentences than do their counterparts, controlling for other factors).

39. *See* Timothy R. Robicheaux & Brian H. Bornstein, *Punished, Dead or Alive: Empirical Perspectives on Awarding Punitive Damages Against Deceased Defendants*, 16 Psych., Pub. Pol'y, & L. 393 (2010) (experiment on the effect of defendant's health on willingness of mock jurors to impose punitive damages).

40. *See* Tisha R. A. Wiley & Bette L. Bottoms, *Effects of Defendant Sexual Orientation on Jurors' Perceptions of Child Sexual Assault*, 33 L. & Hum. Behav. 46 (2009) (mock juror study on willingness to convict gay or straight defendant in child sexual assault case).

41. *See, e.g.*, David R. Shaffer et al., *Hath He Suffered Enough? Effects of Jury Dogmatism, Defendant Similarity, and Defendant's Pretrial Suffering on Juridic Decisions*, 50 J. Personality & Soc. Psychol. 1059 (1986); David A Abwender & Kenyatta Hough, *Interactive Effects of Characteristics of Defendant and Mock Juror on U.S. Participants' Judgment and Sentencing Recommendations*, 141 J. Social Psych. 603 (2001); Dolores A. Perez et al., *Ethnicity of Defendants and Jurors as Influences on Jury Decisions*, 23 J. Applied Soc. Psych. 1249 (1993). The similarity-leniency effect may be mediated by strength of evidence against the defendant. If the evidence against the defendant is very strong, jurors similar to defendant may treat him in an especially harsh manner to distance themselves from him. *See* José M. Marques & Dario Paez, *The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability*, 5 Eur. Rev. Soc. Psych. 37, 47 (1994).

42. Experimental simulation evidence "is often criticized as being too abstract and impersonal to bear on particular case facts, despite its relevance in a technical sense." Brian H. Borstein & Sean G. McCabe, *Jurors of the Absurd? The Role of Consequentiality in Jury Simulation Research*, 32 Fla. St. U. L. Rev. 443, 449 n. 24 (2005).

research to improve trial processes generally, even though a court may not be able to apply it to a specific case.[43]

Therefore, while these variables are important, they are not subject to harmless error analysis because they are not amenable to judicial remedies. Instead, the scientific analysis of trial errors focuses on case-specific variables, like the introduction of a coerced confession during a trial, rather than these extralegal factors that potentially affect many cases.

C. *Harmless Error Analysis for Justiciable and Potentially Harmful Errors*

The preceding Parts identified two categories of trial variables that are not subject to harmless error analysis: structural errors that are grounds for automatic reversal and extralegal variables that courts cannot manage or remedy. What remains are trial errors that potentially, but not always, cause harm and can be managed and remedied by appeals courts using harmless error analysis.

The domain of harmless error analysis is broad, encompassing a constellation of error claims. Most trial errors, even those that affect a defendant's constitutional rights, call for harmless error analysis.[44] In *Chapman v. California*, the U.S. Supreme Court held that constitutional errors should not be reversible *per se* without regard to the impact on the jury's deliberation and verdict.[45] According to *Chapman*, constitutional trial errors do not necessitate overturning a conviction if the error was "harmless beyond a reasonable doubt."[46] Trial errors are now generally subject to harmless error analysis, as the Court explained in *Arizona v. Fulminante*:[47]

> Since this Court's landmark decision in *Chapman v. California*, [. . .] the Court has applied harmless-error analysis to a wide range of errors and has recognized that most constitutional errors can be harmless. [. . .] The common thread connecting these cases is that each involved "trial error" – error which occurred during the presentation of the case to the jury, and which may therefore be quantitatively assessed in the context of other evidence presented

43. I am not suggesting that the effect of contextual and environmental variables on trial outcomes is impervious to any interventions. Courts and policymakers should work to eliminate racism, sexism, and other forms of prejudice from the legal system. I am suggesting, however, that these distortions should be addressed at a macro-level to address widespread, systemic problems, rather than on a case-by-case basis by applying harmless error analysis to extra-legal variables.
44. *See, e.g.*, Glebe v. Frost, 574 U.S. 21 (2014) (improperly restricting the defense's closing argument is not a structural error).
45. Chapman v. California, 386 U.S. 18, 21-22 (1967).
46. *Id.* at 24. Many federal cases address how courts should assess constitutional errors because federal courts do not directly review convictions secured in state courts. Federal court involvement is limited to matters raising constitutional issues (often in collateral, post-conviction proceedings).
47. Arizona v. Fulminante, 499 U.S. 279 (1991).

> in order to determine whether its admission was harmless beyond a reasonable doubt.[48]

Common trial errors include mistakes in the admission and omission of evidence, deficient jury-related trial practices, and erroneous jury instructions. According to Professor Jason Solomon's study of opinions addressing the harmfulness of errors, 72 percent of errors happened during trial, 6 percent were pretrial errors, and 24 percent involved sentencing errors.[49] Solomon also reported that 52 percent of the errors involved the improper admission of evidence and 18 percent of the errors involved improperly omitted evidence.[50] Most of the reported errors (80 percent) occurred in murder trials.[51]

Of all the myriad factors that may potentially influence a jury's decision, the evidence presented at trial constitutes the most important factor. According to Dennis Devine's excellent review of research on juries:

> Although doubted by some legal critics and commentators, the strongest determinant of the jury's ultimate decision in most cases is undoubtedly the strength of evidence (SOE) against the defendant. In dozens of experimental studies with mock jurors, SOE has been systematically varied by providing, withholding, or altering the evidence presented to study participants. These manipulations typically have a substantial impact on resulting decisions, often producing differences in conviction rates of 20–60 percent.[52]

Furthermore, the effects of inadmissible evidence and instructions to disregard inadmissible evidence have been extensively studied.[53] Many controlled scientific experiments with mock jurors confirm the central role of case facts to the jurors' decisions of whether a defendant is guilty or not guilty.[54] Researchers have conducted hundreds of such studies,

48. *Id.* at 306–07.
49. *See* Jason M Solomon, *Causing Constitutional Harm: How Tort Law Can Help Determine Harmless Error in Criminal Trials*, 99 Nw. U.L. Rev. 1053, 1066 (2005).
50. *Id.* Improper admission or omission of evidence can occur during trial or sentencing. These percentages do not sum to 100 percent because some trial errors, such as improper jury instructions, do not involve evidence.
51. *Id* . Solomon reports 80 percent of errors he studied came from cases where the crime was murder. He distinguishes trial and sentencing errors to say 54 percent of the errors were murder trial errors, which means 26 percent were murder sentencing errors.
52. Dennis J. Devine, Jury Decision Making: The State of the Science 122 (2012) (citation omitted).
53. *See* Nancy Steblay, et al., *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, 30 L. & Hum. Behav. 469 (2006) (a meta-analysis of 48 studies on limiting instructions).
54. *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) (case facts affect decision to involuntarily commit sex offender); Stephen P. Garvey, et al., *Juror First Votes in Criminal Trials*, 1 J. Empirical Legal Stud. 371 (2004) (jurors form opinions based on the evidence presented during trial); 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 (2012) (effect of case facts on probability of conviction); Harmon M.

documenting the potential effects of trial errors, including those affecting jury instructions,[55] attorney arguments,[56] and different types of evidence in various trial contexts,[57] on the likelihood of conviction and severity of punishment. This vast literature bears out the potential importance of two types of evidence that exemplify errors of admission and omission: the admission of coerced confessions and the omission of mitigating evidence.

Empirical and experimental work on the effect of coerced confessions shows that these confessions generally have a significant impact on jury trials.[58] Confessions are thought to be "uniquely potent" evidence of the defendant's guilt.[59] "Some jurors," Devine observes, "might even wonder why a trial is necessary in the wake of a confession by the defendant."[60] Confessions are damning even when circumstances suggest they are unreliable, coerced, or possibly fabricated. Confessions significantly impact the probability of a jury rendering a conviction even when reported by a questionable witness, like a jailhouse informant or an accomplice.[61] "[F]alse confessions," Saul Kassin writes, "often trump factual innocence."[62] Even

Hosch et al., *Effects of an Alibi Witness's Relationship to the Defendant on Mock Jurors' Judgments*, 35 L. & Hum. Behav. 127 (2011); Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 Ohio St. L.J. 1391 (2006) (impact of hindsight evaluations in patent litigation); Brian L. Cutler et al., *Expert Testimony and Jury Decision Making: An Empirical Analysis*, 7 Behav. Sci. & Law 215 (1989).

55. *See, e.g.*, Cheryl A. Terrance et al., *Effects of Judicial Instructions and Case Characteristics in a Mock Jury Trial of Battered Women Who Kill*, 24 L. & Hum. Behav. 207 (2000); Neil Brewer et al., *Improving Comprehension of Jury Instructions with Audio-Visual Presentation*, 18 Applied Cognitive Psych. 765 (2004).

56. *See, e.g.*, Judy Platania & Gary Moran, *Due Process and the Death Penalty: The Role of Prosecutorial Misconduct in Closing Argument in Capital Trials*, 23 L. & Hum. Behav. 471 (1999); Thomas A. Pyszczynski & Lawrence S. Wrightsman, *The Effects of Opening Statements on Mock Jurors' Verdicts in a Simulated Criminal Trial*, 11 J. Applied Soc. Psych. 301 (1981).

57. For example, researchers have conducted controlled experiments to assess the effect of expert witness testimony on matters such as DNA evidence and statistics. *See, e.g.*, Julie A. Buck & Amye R. Warren, *Expert Testimony in Recovered Memory Trials: Effects on Mock Jurors' Opinions, Deliberations and Verdicts*, 24 Applied Cognitive Psych. 495 (2010).

58. *See* Saul M. Kassin & Holly Sukel, *Coerced Confessions and the Jury: An Experimental Test of the "Harmless Error" Rule*, 21 L. & Hum. Behav. 27, 42–44 (1997) (higher percentage of mock jurors vote guilty presented with coerced, inadmissible confession); Devine, *supra* note 52, at 145–47.

59. *See* Saul M. Kassin & Katherine Neumann, *On the Power of Confession Evidence: An Experimental Test of the Fundamental Difference Hypothesis*, 21 L. & Hum. Behav. 469, 481–83 (1997).

60. Devine, *supra* note 52, at 145.

61. *See* Jeffrey S. Neuschatz et al., *Secondary Confessions, Expert Testimony, and Unreliable Testimony*, 27 J. Police & Crim. Psych. 179 (2012); Stacy Ann Wetmore et al., *On the Power of Secondary Confession Evidence*, 20 Psych., Crim. & L. 339, 354–55 (2014); Jeffrey S. Neuschatz, et al., *The Effects of Accomplice Witnesses and Jailhouse Informants on Jury Decision Making*, 32 L. & Hum. Behav. 137 (2008).

62. Saul M. Kassin, *Why Confessions Trump Innocence*, 67 Am. Psych. 431, 431

if a confession is repudiated and contradicted by physical evidence, it may lead the jury to believe the defendant is guilty.[63]

Despite the potency of improper confession evidence, the majority in *Fulminante* held that the improper admission of a coerced confession is not an inherently harmful error, but instead, its introduction is only potentially harmful and should, therefore, be subject to harmless error analysis.[64] The impact of a confession depends on the strength of other evidence in the case. Confessions clearly can be damaging, but whether a confession, or any other trial evidence, significantly affects a particular trial depends on the weight of other evidence. If the prosecution presents plenty of other evidence that shows the defendant is guilty, a questionable confession may not change the outcome. If, on the other hand, other evidence is unavailable, a confession, even a questionable one, can serve as the lynchpin of the prosecution's case.

Research also shows that mitigating evidence can influence a jury's decision to recommend a death sentence.[65] It is important to put a defendant's crime in context,[66] and humanize the defendant by having friends and family testify on his behalf.[67] According to Justice Thurgood Marshall, "[e]xperienced members of the death-penalty bar have long recognized the crucial importance of adducing evidence at a sentencing proceeding that establishes the defendant's social and familial connections."[68] Some defendants are sentenced to death not because the facts are so bad, but because their defense counsel do not investigate the defendant's biography and present evidence of mitigating circumstances.[69] If the jury empathizes with the defendant, and considers the

(2012). According to the Innocence Project, twenty-three percent (23 percent) of convictions later overturned by DNA evidence were initially secured using false confessions. *See* Jim Dwyer et al., Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted 92 (2000).

63. Paul G. Cassell, *The Guilty and the "Innocent": An Examination of Alleged Cases of Wrongful Conviction from False Confessions*, 22 Harv. J.L. & Pub. Pol'y 523 (1998).

64. Arizona v. Fulminante, 499 U.S. 279, 306–12 (1991).

65. Mitigating factors are not limited to mitigating factors identified in statutes. Jurors may find anything a mitigating factor.

66. *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).

67. *See* 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). Some jurors, however, view a defendant's troubled history as evidence of future dangerousness and inability to rehabilitate. *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).

68. 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).

69. *See* Stephen B. Bright, *Counsel for the Poor: The Death Sentence Not for the*

defendant remorseful and unlikely to recidivate, they may favor leniency. If, on the other hand, the jury finds the defendant cold, calculating, and likely to commit more crimes if given the opportunity, they may favor a death sentence. Therefore, an attorney's failure to admit this sort of evidence is potentially a harmful error because the omission of mitigating evidence may substantially increase the probability that the defendant is sentenced to death.

However, while jurors may view some defendants more sympathetically on account of their difficult personal histories, evidence of a defendant's abusive childhood and alcohol abuse may be a "doubled-edged sword" because jurors can draw conflicting inferences from it.[70] Such evidence may help the jury understand what led the defendant to commit crimes, but it may also show that the defendant is at least partly to blame for his own misfortunate and is not a good prospect for rehabilitation. The effect of mitigation evidence, then, may vary from one case to the next. Wrongfully omitting evidence is potentially a harmful error, but the amount of harm depends on the defendant's specific circumstances. This variability in turn requires a judge to apply harmless error analysis to determine whether in fact the omission of mitigating evidence harmed a defendant's case to an intolerable degree so as to warrant a new trial.

It should be noted that some trial errors, while not considered inherently harmful, structural errors are judged by standards other than harmless error analysis. Confrontation Clause violations, for example, do not result in automatic reversal (like a structural error) nor are they evaluated using harmless error analysis.[71] Instead of analyzing whether a violation affected the outcome of the trial, the court should consider whether the violation affected the jury's evaluation of the witnesses the defendant was not allowed to confront.[72] Similarly, when the State withholds exculpatory evidence in violation of *Brady v. Maryland*,[73] the defendant must show there is a "reasonable probability" that the result of a proceeding would have been different had the evidence not been withheld, so those claims are not also subject to harmless error analysis.[74]

Worst Crime But for the Worst Lawyer, 103 YALE L.J. 1835 (1994).

70. *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).
71. *See* Delaware v. Van Arsdall, 475 U.S. 673 (1986).
72. Similar methods could be used to scientifically evaluate jurors' perceptions of witnesses. The key modification would be measuring jurors' perceptions of witnesses rather than their verdict preferences or impression of the trial.
73. Brady v. Maryland, 373 U.S. 83 (1963).
74. *See* Kyles v. Whitley, 514 U.S. 419, 435–36 (1995). The "reasonable probability" standard for judging *Brady* violations suggests evaluating the effect of the violation in manner akin to harmless error analysis but imposing a more

Additionally, it is important to note that a trial event is considered a trial error only if it violates either legal rules or the standard of care owed by attorneys. In any trial, the judge and the attorneys must make quick decisions based on their professional experiences, intuitions, and instincts. An appellate court may then determine the wrongfulness of decisions that gave rise to an error based on rules of professional responsibility, evidence, and criminal procedure. Some of these decisions made during the trial may harm the defendant but are not considered errors because they were reasonable at the time they were made. A defense attorney's strategic decision of whether to introduce potentially mitigating evidence should not be considered an error by the appellate court if that decision turns out to harm the defendant's probability of acquittal when analyzed with the benefit of hindsight.[75] In the case of defense attorneys, if a decision was reasonable when made, courts often determine there was no error.[76] The same should apply to trial judges, where if a decision was reasonable when made, it should not be deemed an error subject to harmless error analysis.[77]

Lastly, to clarify the role appellate judges play in determining the errors subject to analysis, it is important to distinguish the question of error from the question of harm. In many cases, the harmlessness of error is a secondary argument, such as when a prosecutor argues that there was no error, and alternatively, if there was, the error was harmless. This Article is concerned with the effect of a trial error, not whether the defendant's attorney violated a standard of care, or whether the presiding judge acted contrary to applicable law. In some cases, evidence that a decision harmed the defendant speaks of the wrongfulness of a decision, but it is important to recognize that the violation of the standard of care is conceptually distinct from the harm the decision caused. So, while the source of error may be the violation of a standard of care, a question of law for the appellate judge, the focus of the harmless error inquiry is on

stringent burden on the defendant.

75. To prevail on an ineffective assistance of counsel claim, a petitioner must show that his defense counsel violated a duty of care; additionally, a petitioner must show that the defense counsel's breach of duty caused sufficient harm to warrant a new trial. *See* Strickland v. Washington, 466 U.S. 668, 687 (1984); *see also* Elizabeth Gable & Tyler Green, Wiggins v. Smith*: The Ineffective Assistance of Counsel Standard Applied Twenty Years After* Strickland, 17 Geo. J. Legal Ethics 755, 756-59 (2004).

76. Reasonable strategic decisions that defense attorneys make for their clients are not deficient if they turn out to be harmful when viewed retrospectively. *See, e.g.*, Harrington v. Richter, 562 U.S. 86 (2011) (decision to forgo expert forensic analysis was reasonable strategic decision, not deficient representation); Yarborough v. Gentry, 540 U.S. 1 (2003) (defense counsel's limited closing argument was reasonable strategic choice).

77. Trial judges' decisions are reviewed in a similar manner. To successfully challenge a discretionary decision by the trial court, the appellant must show the trial judge abused his or her discretion; the appellant must also show that discretionary decision was harmful. *See* J. Richard M Markus, *A Better Standard for Reviewing Discretion*, 2004 Utah L. Rev. 1279, 1296–97 (2004).

the level of harm caused by the particular breach, which is a question of fact more difficult for judges to independently assess.

Scientific analysis of trial errors can be used to assess the level of harm caused by evidentiary errors. Parts IV and V will explore how research of the type outlined in this section can be applied to a scientific framework to help judges make these fact-based determinations.

II. Measurable Outcomes, Tolerable Harms, and Burdens of Proof

Before turning to this Article's proposed scientific framework, it is important to note that legal scholars have identified two distinct approaches to assessing the harm caused by a trial error: the error-based approach, and the overwhelming evidence approach.[78] The error-based approach is a process-oriented inquiry that asks whether the improperly admitted evidence was the type of information reasonable jurors would rely on to make their decisions.[79] The overwhelming evidence approach, in contrast, is results-oriented and asks whether the error likely affected the probability of acquittal.[80]

Although these approaches can be defined with reasonable clarity, attempts to implement either approach are frustrated by confusing and conflicting directives. Most federal judges seem to have stopped trying to identify which approach they are employing to evaluate the harmfulness of an error.[81] The Supreme Court has even questioned whether competing approaches to harmless error analysis differ in name only.[82] Much of the confusion surrounding harmless error analysis results from failing to identify who or what is harmed, how much harm is tolerable, and who bears the burden of proof.

78. *See* Solomon, supra note 5, at 1062 n.44 (citing Jeffery O. Cooper, *Searching for Harmlessness: Method and Madness in the Supreme Court's Harmless Constitutional Error Doctrine*, 50 U. Kan. L. Rev. 309, 311 (2002)).

79. Jeffrey O. Cooper, *Searching for Harmlessness: Method and Madness in the Supreme Court's Harmless Constitutional Error Doctrine*, 50 U. Kan. L. Rev. 309, 328-29 (2002).

80. *See Id.* at 311.

81. In Solomon's sample of cases, only 14 percent of decisions employed any harmless error test. Solomon, *supra* note 5, at 1068.

82. The Supreme Court's non-decision in the case of *Vasquez v. US* is interesting on this point. The Court granted cert and heard oral arguments in this case. Vasquez argued that the Seventh Circuit failed to make the proper harmless error inquiry when it asked whether a reasonable jury would still convict Vasquez based on the untainted evidence (the guilt-based approach). Transcript of Oral Arguments at 3-7, Vasquez v. U.S., 565 U.S. 1057 (2012) (No. 11-199); see also U.S. v. Vasquez, 635 F.3d 889 (7th Cir. 2011). Vasquez argued that courts must focus their analysis on how the error affected the actual jury, setting up a "battle of standards" in the high court. The battle, however, was anticlimactic. Several justices questioned whether there was any meaningful difference between the two tests. Transcript of Oral Arguments at 8-9; Tejinder Singh, *Opinion Analysis: Harmless Error Case Dismissed as Improvidently Granted*, SCOTUSblog (Apr. 2, 2012, 10:34 AM), https://www.scotusblog.com/2012/04/opinion-analysis-harmless-error-case-dismissed-as-improvidently-granted [https://perma.cc/RZF3-YJJC].

For instance, the process and results-oriented approaches to harmless error analysis both ask whether the error "made a difference" but what victim(s) of harm do we have in mind? Clearly, a trial error can harm the defendant. The defendant enjoys the right to a fair trial and a trial error jeopardizes his liberty and, possibly, his life. In addition to harming the defendant, a trial error also deprives the public, represented at trial by a jury, the opportunity to discover truths and engage in a democratic decision-making process.[83] These harms are less acute but have a broader social impact. Trials are not simply a means to an end; trials tell stories, reveal truths, and provide society an opportunity to express shared values.[84] There are procedural benefits to jury deliberation that are independent of any effect that deliberation has on the outcome.[85] Trial by jury is an important institution and social ritual. Although people may grumble when they are summoned to jury duty, there is compelling evidence that jurors are satisfied by the process.[86]

A trial infected by error tells a false story. It denies the trial's audience a genuine opportunity to participate in criminal justice. A trial error can harm not only the defendant, but also the trial's audience who have the right to attend a fair, public trial, hear truths, and contribute to the ongoing articulation of community values. Jurors sit closest to the trial, but the audience also includes the defendant's friends and family, crime victims and their friends and families, the media, and the general public. For all these people, trials reveal truths and provide answers. Therefore, along with the defendant, the trial itself is considered a relevant object of harm caused by a trial error.

Defining who, or what, is harmed by a trial error helps to distinguish the two main approaches to harmless error analysis discussed in caselaw and legal scholarship.[87] One approach focuses on harm to the defen-

83. Just as a medical treatment could affect a variety of health outcomes, not just mortality but also quality of life, so too can a trial error affect not just the verdict but the quality of the deliberative process used to reach the result.

84. Public trials are an important social ritual. Systems of resolving disputes are important to many cultures. *See generally* Oscar Chase, Law, Culture, and Ritual: Disputing Systems in Cross-Cultural Context (2005).

85. Jury service is educational and may encourage voting and other civic behaviors. *See* John Gastil et al., *Civic Awakening in the Jury Room: A Test of the Connection between Jury Deliberation and Political Participation*, 64 J. Pol. 585 (2002). Deliberation offers jurors an opportunity to express their views and have opinions heard by other jurors. Even if the side a juror initially took in deliberation is not ultimately reflected in the jury's verdict, it is important that jurors can express their views to other jurors and deliberate.

86. Inferences from this research should be discounted somewhat, however, by self-selection bias. While jury summons may be issues randomly, the subset of those summons who serve on juries are those who respond to the summons and then are picked to serve on a jury. Many people who receive jury summons will not appear, excuse their appearance, or express an inability to serve or unwillingness to render an unbiased verdict.

87. *See, e.g.*, Hon. Harry T. Edwards, *To Err is Human, but not Always Harmless: When Should Legal Error be Tolerated?*, 70 N.Y.U.L. Rev. 1167 (1995); Stuart

dant while the other considers harm to the trial process and its audience. The difference is the object of harm; therefore, these approaches are best understood as the effect-on-defendant approach and the effect-on-trial approach. Although the merits of the effect-on-defendant and effect-on-trial approaches are the subject of ongoing debate, their differences are not irreconcilable. Both approaches will often yield the same conclusion about the harmfulness of error; however, these approaches may support different conclusions, and judges should thus consider which harms matter in their harmless error analysis and use data that speak to their approach. This Article does not take a position on which approach should be used, but instead this Article demonstrates how a scientific framework can be applied to either approach to achieve more accurate outcomes.

A. *Differentiating the Effect-on-Defendant and Effect-on-Trial Approaches*

The effect-on-defendant approach focuses on how the error affected the jury's verdict.[88] It considers whether the trial error made a guilty verdict or death sentence more likely. Following this approach, an appellate judge compares the outcome of the actual trial to the potential outcome of a hypothetical, error-free trial.[89] A harmful error in this view is one that deprives the defendant of his freedom by increasing the probability of his conviction or of receiving a death sentence. If the outcome of the hypothetical trial would be the same for the defendant, the trial error did not harm the defendant. If the trial error made the outcome significantly

P. Green, *The Challenge of Harmless Error*, 59 La. L. Rev. 1101 (1999); Cooper, *supra* note 79; Daniel Epps, *Harmless Errors and Substantial Rights*, 131 Harv. L. Rev. 2117 (2018); John M. Greabe, *Criminal Procedure Rights and Harmless Error: A Response to Professor Epps*, 118 Colum. L. Rev. 118 (2017); Daniel Epps, *The Right Approach to Harmless Error*, 120 Colum. L. Rev. 1 (2020); Gregory Mitchell, *Against "Overwhelming" Appellate Activism: Constraining Harmless Error Review*, 82 Cal. L. Rev. 1335 (1994).

88. This approach is alternatively described as the guilt-based approach, the overwhelming evidence test, or the hypothetical trial approach.

89. Some opinions disavow analysis of a hypothetical error-free trial and purport to evaluate a trial error's effect on the defendant without comparing the defendant's actual trial to a hypothetical trial. Although some may find analysis of a hypothetical, error-free trial uncomfortable or inappropriate, it is simply unavoidable. One cannot estimate causal effects without some counterfactual analysis. Donald Rubin has shown that causal analysis necessitates counterfactual analysis of unobserved, potential outcomes to address the inherent challenge in causal inference: understanding what would have occurred if individuals had experienced a treatment different from the one observed. *See* Donald B. Rubin, *Causal Inference Using Potential Outcomes: Design, Modeling, Decisions*, 100 J. Am. Stat. Assoc. 322 (2005). Rubin's framework recognizes that, in any given situation, individuals can only undergo one treatment condition, making the alternative outcomes unobservable. By comparing observed outcomes to these counterfactuals, researchers aim to isolate the causal impact of a treatment. For further discussion of the Rubin Causal Model with applications to legal studies, see Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 Ann. Rev. L. & Soc. Sci. 17 (2011).

worse for the defendant, it was a harmful error. In Part II subpart B, I consider how much worse the outcome must be to warrant a new trial.

The effect-on-trial approach, in contrast, focuses on how much the trial error influenced the trial process and the jury's understanding of the case.[90] This approach focuses on how the trial error shaped the jury's perception of the case. If the trial error played a significant role in the trial or left something important out of the story, it harmed the jury's understanding of the facts and denied jurors the opportunity to fairly participate in the truth-seeking process.

An error that significantly affects the trial outcome will, almost by definition, significantly affect the trial process. If an error was large enough to significantly change the outcome of the trial, that error must also have had a significant effect on the process that produced that outcome, assuming that the verdict was a rational result of a reasonable jury's deliberations.[91] Therefore, an error deemed harmful following the effect-on-defendant approach should also be considered harmful using the effect-on-trial approach.[92]

Conversely, an error that significantly affects the trial process may or may not significantly affect the outcome. There may be cases where errors affect the trial narrative and what the audience understands to be true, but they do not change the verdict or sentence.[93] A trial error may have a significant effect on jury deliberation without changing the jury's verdict. Character evidence, for example, may show the defendant in a more or less positive light without changing what his actions were or the

90. This approach also identified as the error-based approach, the effect-on-the-verdict test, or the actual trial approach.

91. This assumption is generally valid. Juries focus on evidence and consider each other's opinions. *See generally* Harry Kalven, Jr. & Hanz Zeisel, The American Jury (1966). During deliberations, jurors discuss their interpretations of evidence, often using narratives. *See* James A Holstein, *Jurors' Interpretations and Jury Decision Making*, 9 L. & Hum. Behav. 83 (1985).

92. Trial errors that affect jury verdicts likely affect what jurors thought about a trial because jurors discuss their impressions of a trial during deliberation to reach a verdict, but it is conceivable that a trial error could affect the jury's verdict without affecting jurors' conscious thoughts and deliberations. A trial error could affect jurors' thoughts in subtle ways they may not notice or may be reluctant to admit.

93. The distinction between the process-oriented and outcome-oriented approaches is seen in the initial federal court opinion that granted petitioner some relief in the *Washington* case: "[A]lthough a habeas petitioner seeking relief on the basis of a claim of ineffective assistance of counsel must generally make a showing of prejudice, this prejudice requirement is satisfied by demonstrating that but for his counsel's ineffectiveness *his trial*, but not necessarily *its outcome*, would have been altered in a way helpful to him. While our cases have not articulated a standard for how material the change *in the trial* would have to be in order for a petitioner to meet his burden of showing prejudice, it is clear from reviewing those cases that the change must be something more than insubstantial or de minimus" Washington v. Strickland, 673 F.2d 879, 901–902 (5th Cir. 1982) (emphasis added).

harm he caused. The jury's deliberation may take many different paths that arrive at the same outcome.

A trial error may, therefore, be deemed harmful following the effect-on-trial approach but harmless following the effect-on-defendant approach. The effect-on-trial approach identifies more trial errors as harmful than the effect-on-defendant approach does. According to Jason Solomon's study, when courts follow the effect-on-defendant approach, they found the error to be harmless 93 percent of the time, compared to 47 percent of the time when they followed the effect-on-trial approach.[94]

There are critics and proponents of both approaches to harmless error analysis. Critics of the effect-on-defendant approach argue that it essentially allows appellate judges to enter directed verdicts in violation of the right to a jury trial guaranteed by the Sixth Amendment.[95] In criminal cases, judges cannot enter directed verdicts for the prosecution.[96] A trial court judge is constitutionally prohibited from entering a directed verdict "regardless of how overwhelmingly the evidence may point in that direction."[97] When appellate judges decide a trial error is harmless and upholds a conviction, they are arguably directing a guilty verdict in a hypothetical version of the trial that no jury has ever heard. Deciding that an error did not harm the defendant is thus functionally equivalent to directing a guilty verdict after a new trial. Appellate judges should not usurp the role of jurors on appeal, even if they can render a correct verdict, because doing so deprives the public of its opportunity to meaningfully participate in trials.[98]

Even if one believes that trials have non-instrumental value and prefers the effect-on-trial approach, it is not clear how one would implement this approach to decide if a trial error was harmful. Proponents of the effect-on-trial approach make strong theoretical arguments but offer little practical guidance to someone who would like to follow their approach. A recent article in the *Harvard Law Review*, for example, makes a compelling argument that harmless error analysis should not focus narrowly on trial outcomes, but instead should consider the "constellation of interests" served by criminal procedures and reverse

94. *See* Solomon, *supra* note 5, at 1071.

95. *See* Cooper, *supra* note 79, at 334; *See also* Green, *supra* note 87, at 1104.

96. Fed. R. Crim. P. 29. Judges may acquit defendants but are authorized to direct a verdict to convict defendants. Defendants have the right to a jury trial.

97. Rose v. Clark, 478 U.S. 570, 578 (1986) (quoting United States v. Martin Linen Supply Co., 430 U.S. 564, 572-93 (1977)). *See also* Kotteakos v. United States, 328 U.S. 750, 763 (1946) ("[I]t is not the appellate court's function to determine guilt or innocence. Nor is it to speculate upon probable reconviction and decide according to how the speculation comes out. Appellate judges cannot escape such impressions. But they may not make them sole criteria for reversal or affirmance. Those judgments are exclusively for the jury.") (citations omitted).

98. "[T]he question is, not were [the jurors] right in their judgment, regardless of the error or its effect on the verdict. It is rather what effect the error had or reasonably may be taken to have had upon the jury's decision." Kotteakos, 328 U.S. at 764.

conviction "if the redressable harm that the error caused to those interests is substantial enough to justify reversal."[99] The article makes strong arguments against verdict-oriented analysis, but an alternative approach is not clearly specified. The author does not identify the constellation of interests, which vary from one case to the next, nor does the author explain how to identify them.[100] A would-be follower is not told how to measure harm to those interests, how to weight or aggregate varied harms to yield a net measure of harm, or at what point the harm is substantial enough to justify reversal; there is no recipe for the suggested approach, so there is no way to know if someone has done it correctly. Trial by jury may not be merely a means to an end, but the jury's verdict is a specific, concrete, and meaningful outcome: the defendant is either guilty or not guilty. The process-oriented, non-instrumental values of criminal trials—like the quality of jury deliberations, the maintenance of democratic traditions and punishment rituals, truth-seeking, and storytelling—are harder to identify and measure.[101]

Although these approaches are generally consistent, there may be cases where a trial error has a substantial effect on the trial, but the effect is in the defendant's favor. For example, an improperly admitted coerced confession may support a defendant's claim that he was framed by police and cause jurors to question the reliability of the entire police investigation.[102] Does a significant effect on the trial in the defendant's favor justify a new trial? A strict effect-on-trial approach may deem this as a harmful error because the improper admission of a confession caused the jury to think more about the police investigation than the defendant's conduct. The improper evidence distorts the narrative and misdirects the jury's attention, harming the process-oriented values of criminal trials, but the distortion and misdirection favor the defendant. Holding a new trial in this situation seems like a misuse of resources because a new jury would think the defendant's conduct was worse than his original jury did, but one could argue the public should hear an error-free account of the defendant's conduct. Unfortunately, this situation is not fully addressed in the literature.

---

99. Justin Murray, *A Contextual Approach to Harmless Error Review*, 130 Harv. L. Rev. 1791, 1810-1811 (2017).
100. *Id.* at 1811.
101. Some may argue that process-oriented values cannot or should not be measured numerically. Although the abstract values served by criminal procedures are more difficult to measure than observed outcomes, abstract concepts like trust, self-efficacy, and representation have been successfully operationalized and measured by social scientists for decades. Quantification does not debase or devalue process-oriented interests. To the contrary, to respect and protect these interests, courts must be able to identify them and know when trial errors harm them. To make good management decisions, it is necessary to measure important values. *See generally* Peter F. Drucker, Management (2011).
102. An error in defendant's favor may also occur when ineffective defense counsel fails to introduce character evidence during sentencing which keeps the prosecution from introducing evidence of defendant's bad character.

B. *Tolerance for Harms*

Imagine, for the sake of argument, that we have decided which harms matter and can measure those harms perfectly on a 100-point scale. How much harm to the defendant or to the trial is too much harm? Just as people will argue whether a thermostat should be set to 78 or 82 degrees in the summer, deciding how much harm is intolerable, even if we knew exactly what to measure and how to measure it, depends on individual and collective determinations.

The question of when harm becomes intolerable has not been answered precisely, but the Court has offered useful guidance. The Court's discussion of the harm threshold in the *Strickland* case is worth quoting at length:

> It is not enough for the defendant to show that the errors had some conceivable effect on the outcome of the proceeding. Virtually every act or omission of counsel would meet that test, and not every error that conceivably could have influenced the outcome undermines the reliability of the result of the proceeding. [. . .] On the other hand, we believe that a defendant need not show that counsel's deficient conduct more likely than not altered the outcome in the case. [. . .] The result of a proceeding can be rendered unreliable, and hence the proceeding itself unfair, even if the errors of counsel cannot be shown by a preponderance of the evidence to have determined the outcome. [. . .]
>
> The defendant must show that there is a reasonable probability that, but for counsel's unprofessional errors, the result of the proceeding would have been different. A reasonable probability is a probability sufficient to undermine confidence in the outcome.[103]

In *Strickland v. Washington*, the Court thus held that the defendant is required to show that the trial error had more than just some harmful effect, but he is not required to show that the error was so severe it changed the likely outcome.[104] In quantitative terms, a trial error is not deemed harmful merely because its effect is greater than 0 percent, but the threshold is also not as high as a 50-percentage point increase in the probability of conviction.

A harmful error occurs when there exist reasonable doubts about the fairness of the defendant's trial. If trial errors are minor imperfections in a generally fair trial, one should accept the jury's verdict even if it seems improbable and one would not have reached the same conclusion. However, if the trial errors are serious enough to suggest the trial was unfair, the errors should be deemed harmful even if a new trial would

103. Strickland v. Washington, 466 U.S. 668, 695–96 (1984).

104. *Id*. The Court's usage of "defendant" is a source of confusion when it comes to properly identifying the burden of proof. Washington was the plaintiff/petitioner in post-conviction proceedings, having already exhausted his direct appeals as defendant. He was not the defendant in this case; he was the plaintiff. Read out of context, the opinion gives the mistaken impression that the criminal defendant has the burden to prove harmfulness on appeal.

likely yield the same result, meaning the jury made the correct decision despite the trial error. Even if his probability of prevailing is low, he is entitled to have his case heard fairly and to take his chances at trial. The defendant is entitled to a fair fight, even if he is likely to lose, and he is entitled to relief if his trial was unfair.

How much does an error have to affect the jury to be considered harmful and not incidental? This determination may depend on whether the error affected how jurors thought about the case and how they reached their verdict. Is the subject of the error something they had in mind when they were deciding which verdict to support? Did they talk about it when they deliberated? If so, at what point do these thoughts or discussions go too far in impacting the jury's understanding of the case? The tolerable effect-on-trial is not clear. Extrapolating from the preceding discussion of the threshold of a harmful effect on a defendant in *Strickland*, it is not enough that the trial error had some effect on the jury's understanding of the trial, but the defendant is not required to show that the trial error tainted most of what the jury understood to be true.

C. *Burdens of Proof*

Which party bears the burden of proof when a court considers the harmfulness of an error? The allocation of the burden of proof has important implications for the scientific analysis of trial errors. In some cases, the trial error is presumed to be harmless, meaning the defendant must prove that the error was harmful to prevail. In other cases, the trial error is presumed to be harmful, meaning the prosecution needs to demonstrate the error was harmless to sustain a conviction. Whether the burden of proof belongs to the defendant or prosecutor depends on the procedural context of the appeal.

1. Appeals

When a defendant appeals his conviction based on trial error, the prosecution carries the burden to prove that the trial error was not harmful.[105] On direct appeal, the reviewing court must "be able to declare a belief that [a federal constitutional error] was harmless beyond a reasonable doubt."[106]

---

105. According to Wright and Henning's treatise on appellate procedure, "harmless error requires the prosecutor to disprove prejudice." Wright, *supra* note 12, at 506.

106. Davis v. Ayala, 576 U.S. 257, 267 (2015) (quoting Chapman v. California, 386 U.S. 18, 24 (1967)). Some courts suggest that the prosecution's burden of proof is relaxed in cases of nonconstitutional trial error. The prosecution may be required to demonstrate a nonconstitutional error was harmless "by a preponderance" or with "fair assurance." *See* U.S. v. Kettles, 970 F.3d 637, 643-645 (6th Cir. 2020) (discussing four different standards used to analyze nonconstitutional evidentiary errors); *see also* U.S. v. Lane, 474 U.S. 438, 471-475 (J. Stevens concurring in part and dissenting in part) (questioning standard applicable to violation of misjoinder rule). If the prosecution must prove a nonconstitutional error was harmless by a preponderance of evidence, the

The burden of proof is not always explicitly stated in court opinions, and many appellate decisions appear to improperly place the burden on the defendant.[107] Since the prosecution has the burden of proof in an appeal, if a court has grave doubts about the impact of an error and is unable to confidently classify the trial error as harmful or harmless, the "tie" between parties should be resolved in the defendant's favor.[108]

2. Post-Conviction Proceedings

The harmfulness of a trial error is not always raised by a criminal defendant on appeal; it can be raised by a petitioner-prisoner through post-conviction habeas corpus proceedings, which are civil lawsuits even though they address the constitutionality of a criminal conviction.[109]

In post-conviction proceedings, the burden of proof shifts to the defendant since he is the moving party challenging a presumptively valid conviction. In *Washington v. Strickland*, the Court states that it is the defendant's responsibility to show that the error was harmful: "[A]ctual ineffectiveness claims alleging a deficiency in attorney performance are subject to a general requirement that the defendant affirmatively prove prejudice. [. . .] Even if a defendant shows that particular errors of counsel were unreasonable, therefore, the defendant must show that they actually had an adverse effect on the defense."[110]

When an appeals court is engaged in the collateral review of a conviction, the trial error must be deemed "substantial and injurious" to be deemed harmful. In collateral proceedings, the petitioner is not entitled to relief unless he can establish the error resulted in "actual prejudice."[111] Therefore, in collateral proceedings, the harmless error test is

standard adopted for statistical significance more lenient than the conventional .05 level which has the effect of narrowing the confidence interval around an estimate of harm. *See* Section IV-C for discussion of standards for statistical significance with illustrative examples.

107. *See* Solomon, *supra* note 5, at 1068. This is not the burden of proof in all criminal appeals. For example, if the defendant challenges the sufficiency of properly admissible evidence to sustain a conviction, he bears the burden of proving the jury's decision was not rational, but if the defendant argues the jury's decision was based on inadmissible evidence, the prosecution bears the burden to prove the trial error was harmless. *See* Cooper, *supra* note 79, at 335.

108. O'Neal v. McAninch, 513 U.S. 432, 436 (1995).

109. Harmless error analysis applies equally to collateral review and direct appeal. *See* Skilling v. U.S., 561 U.S. 358, 414 n. 46 (2010). The standards used for the analysis differ, however, depending on whether the trial error is considered on direct appeal or in post-conviction proceedings. *See* Davis, 576 U.S. at 267.

110. Strickland v. Washington, 466 U.S. 668, 693 (1984).

111. Davis, 576 U.S. at 267–68 (citing Brecht v. Abrahamson, 507 U.S. 619, 637 (1993) (must show more than "reasonable possibility" that trial error was harmful)). The Court did not seem to follow this rule in O'Neal v. McAninch, where it held that judge with "grave doubts" about the harmlessness of trial error should rule in petitioner's favor, but the Court limits its analysis to the "narrow circumstances" of the case which it distinguishes from the burden of proof. Although the Court does not say as much in O'Neal, "grave doubts" about harmlessness indicate a reasonable degree of certainty about the harmfulness of error. 513 U.S. at 439-442.

more stringent than it is on direct appeal; the reviewing courts must have "grave doubts" about the jury's verdict.[112] This more stringent standard may result from the fact that conducting another trial after a petition for habeas corpus is granted and the defendant's conviction set aside imposes significant social costs and ought to be avoided except in cases where defendants have been clearly and severely wronged.[113]

It also seems logical to impose a different burden of proof for errors of omitted evidence compared to errors of admitted evidence based on which party was in a better position to prevent the error at trial. Take, for instance, errors of omission in ineffective assistance of counsel claims raised in post-conviction proceedings. It is not possible for the defendant to assert errors of omission in direct review because omitted evidence, by definition, is not part of the trial record. If the prosecutor caused error by admitting a coerced confession at trial, the prosecution should have the burden of proving that error was harmless. Conversely, if the defendant and his attorney caused the error by failing to introduce mitigating evidence at trial, the defendant and his appellate counsel should have the burden of proving that the error was harmful. For both types of trial errors, the burden of proof thus falls on the party in a better position to prevent the error at trial.

The Antiterrorism and Effective Death Penalty Act further restricts the grounds for reversing a conviction due to trial error in habeas corpus proceedings. Under this Act, a federal court reviewing a state court proceeding cannot grant the prisoner's petition unless the trial error "was contrary to, or involved an unreasonable application of, clearly established Federal law, as determined by the Supreme Court of the United States."[114] This statute has been interpreted to limit relief to cases where the state courts applied harmless error review in an objectively unreasonable manner.[115] The Act does not change the burden of proof in post-conviction proceedings; the defendant-petitioner still must prove the trial error was harmful.[116] However, the Act's "objectively unreasonable" standard does narrow the scope of justiciable state court trial errors insofar as federal courts are now limited to reviewing egregious errors.[117]

112. Davis, 576 U.S. at 268. At his murder trial, Ayala's attorney objected to the prosecution's seemingly race-based peremptory challenges but were excluded from the judge's inquiry into the prosecutor's reasons for those strikes (so the prosecutor could avoid disclosing trial strategies). Ayala could not prove his attorney's presence in the Batson hearing would have affected jury selection.
113. *See* Brecht, 507 U.S. at 637; *see also* Nicholas Beekhuizen, *Post-AEDPA Compromise: Increased Habeas Corpus Relief for Capital Cases and Tighter Restrictions for Noncapital Cases*, 10 Ind. J.L. & Soc. Equal. 321, 327 (2022); Jordan M. Barry, *Prosecuting the Exonerated: Actual Innocence and the Double Jeopardy Clause*, 64 Stan. L. Rev. 535 (2012) (criticizing re-trial after habeas corpus relief granted to those wrongfully convicted).
114. 28 U.S.C.A. § 2254(d).
115. *See* Mitchell v. Esparza, 540 U.S. 12, 19 (2003).
116. *See* Fry v. Pliler, 551 U.S. 112, 120 (2007); Davis, 576 U.S. 257, 271 (2015).
117. *See generally* Hon. Diane P. Wood, *The Enduring Challenges for Habeas Corpus*,

The appropriate criteria for evaluating the harmfulness of trial errors depends on when the errors are reviewed. To prevail on direct appeal, the defendant must show that an error occurred at his trial, and the prosecution must fail to prove that the error was harmless. In post-conviction proceedings, the defendant must show that an error occurred at his trial, and he must also prove that the error was harmful. If the defendant challenges a state conviction in federal post-conviction proceedings, he must show that an egregious error occurred at his trial, and he must prove that the error was harmful. This article does not address the rules governing admissibility of evidence or the professional responsibilities of defense attorneys, so it does not help courts decide whether an error occurred at defendant's trial. This article does, however, help courts evaluate whether an error is proven harmless or proven harmful. As discussed in the next section, a scientific framework for evaluating harm is urgently needed.

III. Why Courts Need Scientific Analysis of Trial Errors

Understanding the purpose and scope of the harmless error doctrine is essential to judging trial errors, but these abstract principles are not enough. General principles must be applied to specific cases, and in this undertaking is where problems arise. This Part identifies four reasons that judges have difficulty evaluating the harmfulness of trial errors and are unlikely to self-correct mistaken estimates of harmfulness. This inquiry also begins to reveal areas where scientific analysis may offer an improvement.

A. *Judges Cannot Use Direct Evidence of Trial Error's Effect*

The harmless error rule asks appellate judges to assess the effect a trial error had on the jury's deliberations and/or verdict. The best and most direct evidence of a trial error's effect would come directly from the jurors who heard the evidence, but the jurors' insights are off-limits and cannot be used by appellate courts.[118] Jurors deliberate in secret. There is no record

95 Notre Dame L. Rev. 1809 (2020). The difference is evident in post-conviction claims for ineffective assistance of counsel. A petitioner challenging a federal court conviction must show his counsel's representation was unreasonable; a petitioner challenging a state court conviction must go further. "Federal habeas courts must guard against the danger of equating unreasonableness under *Strickland* with unreasonableness under § 2254(d). When § 2254(d) applies, the question is not whether counsel's actions were reasonable. The question is whether there is any reasonable argument that counsel satisfied *Strickland's* deferential standard." Harrington v. Richter, 562 U.S. 86, 105 (2011).

118. Fed. R. Evid. 606(b) expressly bars testimony from jurors on their thought process or their jury deliberations. The Rule's exception for juror testimony regarding "extraneous prejudicial information" applies only to information jurors obtain outside the trial, like news stories, not prejudicial information obtained from the trial. According to the Supreme Court, "evidence about the actual process of decision, if not part of the record of the proceeding under review . . . should not be considered in the prejudice determination." Strickland

of how often improper evidence was mentioned during jury deliberation, by whom, or in what context.[119] Appellate judges can only speculate about what evidence jurors discussed, found compelling, and used to reach their verdict; meanwhile, the jurors' actual thoughts are unknowable.[120]

While there are certainly good reasons to preserve the sanctity and secrecy of jury deliberations, doing so puts appellate judges in a bind. The harmless error rule asks judges to evaluate a trial error's actual effect on jurors' deliberations and/or verdict, but no one may hear from those jurors. On appeal, the judge must disregard the actual jurors and, instead, try to channel the thoughts of an imaginary set of reasonable jurors.[121]

Even if transcripts of jury deliberations were available and jurors could testify about their decision making, the actual effect analysis would still be problematic. Research shows that human decisionmakers are influenced by variables they do not consciously perceive.[122] If they were permitted to testify, jurors would deny making impermissible inferences, having personal biases, or feeling pressure to conform their views.[123] If direct evidence were available, jurors could tell us how they rationalized their decision, but their recollections of deliberation would only offer a limited account of how they made decisions.[124] The next parts of this Section criticize the decisions judges make about the harmfulness of errors, but we should recognize that they do not know what happened during jury deliberation and cannot ask them how they made their decision. They do not have access to information needed to assess the effects

---

v. Washington, 466 U.S. 668, 695 (1984).

119. Even in cases where the thoughts and opinions of the fact finder are knowable, as is the case in judge trial, appellate courts do not admit testimony from the fact finder on his or her thoughts and impressions. Washington v. Strickland, 673 F.2d 879, 903–906 (1982). This case provides an interesting example of the unwillingness to hear why a verdict or sentence was rendered because the conviction resulted from a judge trial, not a jury trial, and the trial judge was available to testify about his decision. However, in the subsequent habeas proceeding, the district court ruled that the state trial court judge's testimony could not be used to explain why he convicted the defendant.

120. "[I]n harmless-error analysis, the jurors are neither deposed nor put on the stand, and their thought processes forever remain a source of pure speculation." Solomon, *supra* note 5, at 1082.

121. In Harrington v. California, 395 U.S. 250, 254 (U.S. 1969), Justice Douglas makes this clear: "We of course do not know the jurors who sat. Our judgment must be based on our own reading of the record and on what seems to us to have been the probable impact of the two confessions on the minds of an average jury."

122. *See generally* Daniel Kahneman, Thinking, Fast and Slow (2011); Richard H. Thaler & Cass R. Sunstein, Nudge: Improving Decisions about Health, Wealth, and Happiness (2009).

123. Consider, for example, how jurors in the famous OJ Simpson murder trial have explained their verdict. *Compare* Deborah Hastings, *Jurors Say Acquittals Were Based on Lack of Evidence, Nothing Else*, A.P. News (Oct. 8, 1995), *to* Tim Molloy, *OJ Simpson Juror: Not-Guilty Verdict Was 'Payback' for Rodney King*, Yahoo! News, (June 15, 2016).

124. This is not a fatal flaw for survey research experiments. Subjects express a verdict preference that may reflect both deliberate and instinctive reasoning.

of trial errors and, therefore, are unable to make accurate and objective decisions about the harmfulness of trial errors.

B. *Judges' Estimates of Harm are Subjective and Speculative*

At present, a judge's assessment of harmfulness is merely a matter of opinion. If one judge thinks an error was harmful and another thinks the error was harmless, there is no way to determine which judge is right. Their opinions may be informed by facts and experience, but their determinations about harmfulness are nevertheless opinions and not facts.[125] Whether a judge deems a trial error harmful or harmless, the judge's opinion about harmfulness cannot be falsified, only overturned by a higher court's opinion.[126]

Although case law articulates two main approaches to harmless error analysis, the effect-on-defendant and effect-on-trial approaches, judges rarely follow any specific approach to assess the harmfulness of trial errors. Solomon's analysis of 263 published federal appeals court opinions found that "[l]ess than 20% of the analyses used a test for determining harm."[127] When cases involving Confrontation Clause violations are excluded, "only 14% employed any test at all."[128] Accordingly, Solomon writes "judicial discretion in determining harmlessness is largely unguided."[129]

Even if judges agreed to follow the effect-on-defendant approach or the effect-on-trial approach, specified which party has the burden of proof, and qualified how much harm was tolerable, their assessments of harmfulness would still be subjective opinions.[130] [131] No verbal

125. "Whether and to what extent an error influenced a given jury verdict is therefore necessarily an exercise in speculation – perhaps principled or reasoned speculation, but nonetheless speculation, about what a jury would or would not have done with or without the offending evidence, instructions, or comment." Hon. Dennis J. Sweeney, *An Analysis of Harmless Error in Washington: A Principled Process*, 31 Gonz. L. Rev. 277, 280 (1995).

126. *See* Brecht v. Abrahamson, 507 U.S. 619 (1993). Judicial analysis of the trial error at issue exemplifies this point. In that case, the prosecutor improperly used defendant's post-Miranda silence for impeachment purposes. Was that error harmful? The Court observed that it was the sixth court to decide whether the trial error was harmful with each court disagreeing with the one before it on the harmfulness of the error. *Id*. at 636. In a 5-4 decision, the Court deemed the error harmless.

127. *See* Solomon, *supra* note 5, at 1067.

128. *Id.* at 1068.

129. *Id.* at 1064.

130. The problem is analogous to calling balls and strikes in baseball. Right now, there are multiple points of subjectivity: players can argue over the definition of the strike zone, how it is affected by the batter's stance, and whether the strike zone depends on game conditions. Even if we resolved all issues of definition, umpires would still have estimate where pitches cross the plate.

131. *See* Charles Alan Wright et al., Federal Practice and Procedure Vol. 11, 614 (2012) ("The difficulty is in gauging when an error is sufficiently serious that it has affected the 'substantial rights' of a party.")

formulation can "avoid the subjectivity that necessarily inheres in determinations of this kind."[132]

Some argue that appellate judges can avoid subjective speculation by focusing on the trial error's actual effect on the trial process rather than the error's effect on the trial outcome. In some opinions, the Supreme Court advises appellate judges to focus only on the effect of an error on the actual jury and avoid speculating on what a hypothetical jury would have done if the trial were different.[133] Judges are not supposed to rely on "mere conjecture"[134] or "unguided speculations"[135] about the effect of a trial error on a jury.

These warnings are well-intended but ultimately ineffective. The "actual reliance" approach does not replace subjective speculation with an objective measurement of harm. First, as discussed in the prior Part, appellate judges cannot hear from the trial's actual jurors and must instead speculate on the jurors' thought processes. Second, the actual reliance approach is unworkable in cases of wrongfully omitted evidence. If the juror never heard mitigating evidence, it is impossible to assess its actual influence on the jury.

More generally, one cannot assess the causal effect of some intervention, like the admission or omission of some trial evidence due to error, without comparing two states of the world, one observed and the other, unobserved. To assess the impact of the trial error on jurors, one compares their mental impressions of the actual trial to the mental impressions they would have had of a hypothetical, error-free trial. The difference between those two outcomes—one observed and one hypothetical—is the effect of the trial error on jurors.

The use of counterfactual analysis is more subtle in the effect-on-trial approach than it is in the effect-on-defendant approach. The outcome of interest in the effect-on-trial approach, the impression created by a trial, is not directly observable and the outcome of interest in the effect-on-defendant approach, the jury's verdict, is announced in court. A trial error's effect on the trial may be less tangible than its effect on the defendant, but counterfactual analysis is necessary in either case. Simply put, we cannot assess how an error affected a jury's impressions of a trial without comparison to a jury's impressions of a hypothetical, error-free trial. To say whether an error made a difference, one must compare two states of the world and determine if they are different. Analyzing causal effects thus requires counterfactual analysis.

The decisions appellate courts make about the harmfulness of trial errors are not unreliable simply because they require judges to speculate about the outcome of a hypothetical, error-free trial; rather, the issue lies with imprecise defining of hypothetical conditions. Therefore,

132. *Id.*
133. *See* Sullivan v. Louisiana, 508 U.S. 275, 279–80 (1993).
134. Stewart v. United States, 366 U.S. 1, 7 (1961).
135. Holloway v. Arkansas, 435 U.S. 475, 491 (1978).

while some speculation is necessary in conducting harmless error analysis, to use counterfactual analysis effectively one must carefully define the hypothetical condition used for comparison and devise methods for objectively evaluating hypothetical trial outcomes.

C. *Judges' Verdict Predictions are Inaccurate and Guilt-Biased*

To assess the harmfulness of a trial error, appellate judges must speculate about jurors' perceptions of trials, but research shows that judges are not very good at predicting jury decisions.[136] When an appellate judge considers the record and concludes a trial error had a significant effect on a jury trial, or concludes the error did not have a significant effect, there is a strong chance the judge is wrong.

The accuracy of judges' predictions about criminal jury verdicts has been tested empirically. Harry Kalven and Hans Zeisel's landmark American Jury Project research asked judges to make their own determinations as to guilt in criminal trials and compared their judgments with those made by juries in 3,567 criminal jury trials. Excluding trials that resulted in hung juries, they found that judges and juries agreed 78 percent of the time.[137] When judges predicted what the jury decision should be based on their firsthand observations of the same trial, their error rate was 22 percent.[138] This seminal finding on the rate of judge-jury agreement is supported by Theodore Eisenberg et al. who reported only 75 percent agreement in a sample of criminal jury trials (excluding hung jury trials).[139]

Even judges who attend the same trials as juries struggle to predict jury verdicts. Why is it so difficult to predict what a jury will do? Suppose we know that twelve jurors begin their deliberation initially split nine-to-three in favor of conviction. According to physicist Pierre-Simon LaPlace, if we have complete information about the mass and direction of objects, "nothing would be uncertain," all future interactions can be calculated, and the future, is entirely predictable.[140] But jury deliberation is not predictable like the path of inanimate objects in a physics problem. We cannot say for certain where the jury will end up, even if we know

136. There is a considerable body of research that compares the prediction accuracy of experts and simple algorithms. See generally Philip E. Tetlock, Expert Political Judgment (2005); Nate Silver, The Signal and the Noise: Why So Many Predictions Fail-but Some Don't (2012).

137. In 14 percent of cases both judge and jury would acquit; in 64 percent of cases both judge and jury would convict. *See* Kalven et al., *supra* note 91.

138. *Id.*

139. *See* Theodore Eisenberg et al., *Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Zeisel's* The American Jury, 2 J. of Empirical Legal Stud. 171, 181 (2005). *See also* "[A]ll available evidence indicates that judges are not very good at determining the impact of various pieces of evidence on jurors." Solomon, *supra* note 5, at 1086.

140. This extreme view of determinism is attributed to French physicist Pierre-Simon Laplace. *See* Mark A. Stone, *Chaos, Prediction and Laplacean Determinism*, 26 Am. Phil. Q. 123 (1989).

where it starts. Empirical analysis suggests that a jury initially split nine-to-three in favor of conviction will return a guilty verdict 80 percent of the time.[141] A guilty verdict is the more likely outcome, but if we predict a guilty verdict, there is a 20 percent chance our predicted verdict will be wrong. Even if we somehow knew jurors' initial verdict preferences following trial, our predictions would not be much better than the judges studied by Kalven and Zeisel (22 percent error rate) or Eisenberg et al. (25 percent error rate).

Rather than accept that juries often return unexpected verdicts and discuss how trial errors change the probability of a guilty verdict, appellate courts try to eliminate uncertainty by assuming the case is heard by a "rational jury."[142] Presumably, a rational jury would view the evidence the same way a judge does; the rational jury's verdicts should be predictable. But the rational jury assumption does not make the problem go away. So long as rational jurors can disagree about the appropriate verdict, the deliberation process can yield either verdict. Rational jurors can disagree because there is not only one rational perspective on the trial.[143] Jurors bring varied life experiences and beliefs into the jury room. According to Kalven and Zeisel, jury verdicts reflect non-legal values, for example, when a jury awards increased leniency when a defendant has already suffered shame or injury because of his crimes or has dependents who would suffer if the defendant cannot go to work.[144] When appellate courts conduct harmless error analysis, they try to estimate how a trial error affected a jury; their assessments are often wrong, but they do not confront the possibility of predictive error. As discussed in Subpart IV.A, jury trial outcomes should be quantified using probabilities that limit the possibility of predictive error by including a range of possible outcomes.

Moreover, the available empirical evidence suggests judges are not merely bad at guessing outcomes (which would produce random but unbiased errors), but also that their expectations about jury verdicts are systematically biased in favor of the prosecution. Actual jury decisions are more lenient than judges predict them to be. Kalven and Zeisel found that in 19 percent of trials the judge would have convicted but the jury acquitted the defendant while in only 3 percent of trials the judge

141. The author extensively discusses the relationship between juror verdict preferences and jury verdicts in a separate article. It is a non-linear relationship tilted in the defendant's favor due to presumption of innocence and guilt beyond a reasonable doubt standard. For discussion and analysis of jury deliberation, see Kalven et al., *supra* note 91; Robert J. MacCoun & Norbert L. Kerr, *Asymmetric Influence in Mock Jury Deliberation: Jurors' Bias for Leniency*, 54 J. Personality & Soc. Psychol. 21 (1988); Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 Psych., Pub. Pol'y, & L. 622 (2001); Dennis J. Devine, Jury Decision Making: The State of the Science (2012).

142. Roger J. Traynor, The Riddle of Harmless Error 33 (1970).

143. *See* Solomon, *supra* note 5, at 1070–71.

144. *See* Kalven et al., *supra* note 91.

would have acquitted but the jury convicted the defendant.[145] Eisenberg et al. (2005) report a similar asymmetry in judge-jury disagreement.[146] The data indicate that trial court judges cannot reliably predict jury decisions and tend to be less lenient than juries.[147]

D. *Hindsight Bias and Overconfidence*

Appellate judges may have an even more difficult time predicting hypothetical trial outcomes than the trial judge does in predicting what the jury will do before it announces its decision. The appellate court judge sees the trial in hindsight, which is problematic because the judge will be influenced by knowledge of what happened. Humans have a natural tendency to perceive events as more rational, inevitable, and predictable than they are.[148] We want our lives to make sense. We identify patterns, create narratives, and recall information selectively to support our narratives. This tendency helps us adapt to complex environments but makes it difficult to evaluate probabilities and hypothetical situations.

Once a verdict is entered, it is natural to see it as the most likely outcome of the evidence presented at trial. The verdict that was reached seems the most plausible one after the fact;[149] however, there was a non-trivial probability that the jury would reach a different verdict. We are likely to judge a defendant's prospects in a hypothetical trial in light of the observed outcome even though a new trial would be an independent event.[150] Hindsight bias thus causes us to overestimate the probability of guilt in a hypothetical trial.[151]

---

145. *See Id*.
146. *See* Eisenberg, et al., *supra* note 139, at 181 (19 percent and 6 percent, respectively).
147. *Id*. Given the reported prediction error rates of 22 percent and 25 percent, a judge's opinion about the likely verdict of a reasonable jury does not meet the conventional standard for scientific certainty (95 percent) nor the standard thought to dispel reasonable doubt (90 percent).
148. *See* Nassim Nicholas Taleb, Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets 55-56 (2005) ("Things are always obvious after the fact."); Kahneman, *supra* note 122, at 202-204. For implications of rationalization and overconfidence to law, *see* Cass R. Sunstein, *Behavioral Analysis of Law*, 64 U. Chi. L. Rev. 1175 (1997); Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 Ohio St. L.J. 1391 (2006).
149. *See* Jonathan Baron & John C Hershey, *Outcome Bias in Decision Evaluation*, 54 J. of Personality & Soc. Psychol. 569 (1988). On the effect of hindsight bias on trials generally, *see* Michael J Saks & Robert F Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 Law & Soc'y Rev. 123 (1981).
150. Cooper, *supra* note 79, at 343 ("[T]he fact of the jury's guilty verdict is likely to affect the appellate judge's perception of the weight of the evidence against the defendant.").
151. *See* Keith A Findley & Michael S Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, Wis. L. Rev. 291, 321 (2006) ("With hindsight knowledge that a jury found the defendant guilty beyond a reasonable doubt, judges are likely to be predisposed to view the conviction as both inevitable and a sound decision, despite a procedural or constitutional error in the proceedings.").

Judges may feel that scientific analysis of the effect of trial errors intrudes on their own reasoned judgment and is unnecessary. Although judges' subjective estimations of harmfulness may be better than a layperson's, experts of all stripes tend to be overconfident in their predictions and fail to appreciate the limits of their knowledge.[152]

The analysis of harm caused by a trial error requires humility and respect for the vagaries of personal opinions. Who has not been surprised to discover what the public really thinks and likes? While we may know our own thoughts, it is difficult, perhaps impossible, to know what others are thinking, much less know what a group of unknown jurors would think about a trial that has never happened. No one needs a meteorologist to know if the sun is out, but a weather report allows us to see beyond our immediate horizon, know what the weather is like elsewhere, and forecast the chance of rain in the future. Likewise, scientific analysis of potential jurors' opinions about actual and hypothetical trials can help understand the effects of trial errors. The researcher can analyze jurors' opinions about hypothetical trials that have never occurred; the research can ask potential jurors why they think the defendant is guilty or not guilty, what they thought of particular items of evidence, and what would change their minds about the case. Careful research can allow us to better understand other people's opinions, beliefs, and attitudes, provided we are willing to admit that our own insights into others' minds are inherently limited.

IV. Framework for Analyzing Harm to the Defendant

The preceding Parts addressed the scope and purpose of harmless error analysis in criminal appeals, the litigants' burdens of proof, and the need for scientific analysis of trial errors. In this Part, I translate the harmless error doctrine into a framework for assessing the harmfulness of trial errors. More specifically, this Article's framing of the harmless error rule will include some technical terminology and notation, which may seem esoteric to some readers, but this framework will help us analyze the harmfulness of trial errors in a precise and rigorous manner.

The framework developed in this Part does not propose a new harmless error test, nor does it favor the effect-on-trial approach or the effect-on-defendant approach. Instead, I describe the legal doctrine of harmless error in a framework that generates testable hypotheses about the harmfulness of trial errors. I begin with a framework for analyzing

152. *See, e.g.,* Erik Angner, *Economists as Experts: Overconfidence in Theory and Practice*, 13 J. Econ. Methodology (2006); Jayashree Mahajan, *The Overconfidence Effect in Marketing Management Predictions*, 29 J. of Marketing Rsch. (1992). (The judge or lawyer who thinks they understand how juries think and can predict jury verdicts by virtue of personal experience with jury trials is similar to the talent evaluator who thinks they can identify talent without analytics or data. Predicting outcomes is notoriously difficult. To avoid overconfidence, this article focusing on the probability of a guilty verdict, rather than predicting whether a jury would find a defendant guilty.).

whether a trial error harmed the defendant and then articulate a framework for testing whether a trial error harmed the trial.

A. *Quantifying the Effect-on-Defendant*

Let G signify a guilty verdict.[153] P(G), then, is the probability that a jury returns a guilty verdict after jurors have heard the entire trial.[154] As a probability, the value of P(G) is bound between 0 and 1.[155]

P(G|actual) has the same meaning as P(G), except it is a conditional probability statement. P(G|actual) is the probability of a guilty verdict given the actual trial condition that was flawed by trial error. P(G|hypothetical) is also a conditional probability statement; it is the probability of a guilty verdict given the hypothetical trial condition. The hypothetical trial condition is the error-free trial that defendant seeks.

The effect of a trial error on the defendant, herein referred to as "effect-on-defendant," is the difference between P(G|actual), the probability of a guilty verdict given the defendant's actual trial, and P(G|hypothetical), the probability of a guilty verdict following a hypothetical, errorless trial.

$$Effect\ of\ error\ on\ defendant = P(G|actual) - P(hypothetical)$$

Subtracting P(G|hypothetical) from P(G|actual) yields an estimate of the harm to defendant caused by the trial error. If the trial error made a guilty verdict more likely, the estimate of harm is a positive number because P(G|actual) is greater than P(G|hypothetical).[156] If the trial error did not increase the probability of a guilty verdict, the difference between P(G|actual) and P(G|hypothetical) will be close to zero, or perhaps even a negative number.

Calculating the effect-on-defendant does not require finding that a harmful error occurred whenever P(G|actual) is greater than

153. One can also express the quantities discussed in this Part in terms of the probability that a jury returns a not guilty verdict, P(NG). Because P(NG) = 1 - P(G), the choice between the two outcomes of interest is arbitrary. I focus on the probability of guilty verdict to avoid writing double negatives when discussing results.

154. Trial errors are not limited to cases where jurors are asked to decide if the defendant is guilty. If a defendant is found guilty of a capital offense, jurors may be asked to recommend a death sentence or life imprisonment. In these cases, the quantities of interest may be expressed as P(D), the probability a jury recommends death. For simplicity, this Article will use P(G).

155. One can think of P(G) as the proportion of guilty verdicts one would obtain if the case were tried many times before different juries. P(G) cannot be less than 0 nor greater than 1. If P(G) = 0, there is no chance of a guilty verdict; if P(G) = 1, a guilty verdict is absolutely certain. P(G) can be expressed as a proportion or as a percentage.

156. The order of terms in the effect of error on defendant equation is somewhat arbitrary. This framework makes the results more intuitive and easier to interpret. The analysis is designed to measure the harmfulness of a trial error. Positive values correspond to greater harm to the defendant.

P(G|hypothetical). Based on the prevailing legal rule, some trial errors that increase the probability the defendant would be found guilty are tolerable and do not necessitate a new trial.[157] A trial error is intolerable only when it causes us to lose confidence in the fairness of defendant's trial.[158]

Without specifying a numeric value, let T represent the amount of trial error that is tolerable. A trial error should be classified as harmful if its effect on the defendant is greater than the amount of tolerable error. If the trial error's effect on the defendant does not exceed the amount of tolerable error, the error should be classified as harmless.[159] To summarize formally:

$$\text{Trial error is } \begin{cases} \text{harmful,} & \text{if } effect\ of\ error\ on\ defendant > T \\ \text{harmless,} & \text{if } effect\ of\ error\ on\ defendant \leq T \end{cases}$$

Some comments on this definition of harmful and harmless effects on defendants may help avoid confusion and misinterpretation. Although the observed outcome of the actual trial was a guilty verdict, P(G|actual) is a quantity to be estimated rather than a known constant.[160] P(G|actual) does not equal 1. If one (mistakenly) assumes that the actual trial was certain to yield a guilty verdict, that is P(G|actual) = 1, one will overstate the effect-on-defendant.[161] While the jury may have rendered a guilty verdict, it is not clear what the probability of that outcome was.[162] It is tempting to think of observed events as inevitable, but they are not.[163] Since the

157. *See* Strickland v. Washington, 466 U.S. 688, 694 (1984).

158. Social scientists will understand this as a requirement to demonstrate substantive significance in addition to statistical significance.

159. It does not matter whether we say an error is harmful if it exceeds the amount of error we can tolerate (defined here as T) or we say an error is harmful if it is greater than or equal to the amount of error we find intolerable (an alternative formulation using ≥ and < instead of > and ≤). If T is an exact number, a precisely estimated harm should not be exactly equal to T because there are an infinite number of points between any two different points. I let T signify tolerance for error to keep this framework simple.

160. The probability that the defendant would be found guilty following the trial he received can be estimated by surveying potential jurors. The probability of a guilty verdict is a function of the jury pool's verdict preferences. If, for example, 50 percent of the jury pool would find the defendant guilty based on the trial evidence, one would expect P(G) to equal, perhaps 40 percent because the defendant is entitled to a presumption of innocence. *See* MacCoun & Kerr, *supra* note 141, at 21; 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, 585 (2012).

161. For someone who has already been convicted, the hypothetical trial always offers a better opportunity to be acquitted than the status quo, no matter how slim the odds of acquittal in a new trial.

162. If we knew the actual jurors' individual preferences at the end of the trial, we would have a better idea of what the probability of a guilty verdict was, but those six or twelve jurors offer a very small sample with which to estimate the expected probability of guilt in the actual trial condition.

163. Once the outcome is known, it takes on an air of false inevitability, like it was certain to happen. We can construct a satisfactory narrative to explain what

defendant was, in fact, found guilty in the actual trial condition, the value of P(G|actual) should be relatively high, though it will be less than 1.[164] The probability of a guilty verdict in both the actual and hypothetical trial conditions, P(G|actual) and P(G|hypothetical), must be estimated.

One cannot assess the effect a trial error had on a defendant based on P(G|hypothetical) or P(G|actual) individually. If there is a great deal of incriminating evidence apart from mistakenly admitted evidence, the value of P(G|hypothetical) may be greater than .50, suggesting the defendant would probably be found guilty in a new trial; however, focusing only on the value of P(G|hypothetical) is problematic. According to the Supreme Court, the defendant does not need to prove the verdict would likely have been different in a fair trial.[165] The defendant should not automatically lose if he would probably be found guilty in a new trial.[166] Similarly, he should not automatically win if he would probably be acquitted in a new trial. If a defendant had a relatively high probability of acquittal in his actual trial, a P(G|hypothetical) value less than .50 does not necessarily show the trial error caused intolerable harm.[167] The level of tolerable harm is thus somewhere in between, and setting this threshold depends on normative judgments about fairness and justice.

---

happened, but before the outcome is observed we know that more than one outcome was possible. The world is probabilistic, not deterministic. If I flip a fair coin and it lands on heads, the probability that coin would land on heads when it left my hand does not change, nor does the probability of it landing on heads if I flip it again. *See generally* Taleb, *supra* note 148.

The observed trial should be viewed as one instance of the trial. This line of thinking does not require great imagination. There are cases that get tried more than once after a mistrial is declared or a new trial is ordered. The parties may make the same arguments and put on the same evidence but get a different result because of random variation in the composition of the jury.

164. We should expect P(G|actual) to be greater than 0.5 given one observation of a guilty verdict from the universe of possibilities. Additionally, we might have the prior belief that the probability of a guilty verdict is greater than 0.5 because the prosecutor made an informed decision to take the case to trial. These prior beliefs do not affect the quantities we estimate in controlled experiments but can help assess the validity of the analysis.

165. *Strickland*, 466 U.S. at 693 (stating that the defendant is not required to show that he is likely to prevail in a new trial to show harm).

166. Some probabilities of a guilty verdict over .50, while making guilt the more likely verdict, are not cases where the evidence of guilt was overwhelming and dictated one result from reasonable juries. If P(G|hypothetical) > .90, however, the effect-on-defendant will necessarily be less than .10.

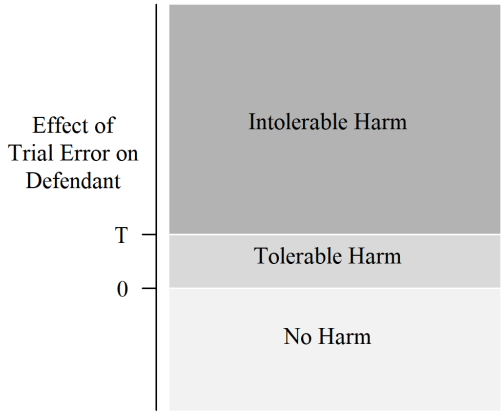
167. Additionally, this interpretation of overwhelming evidence of guilt allows for the perverse result than defendants who suffer less injury are treated more favorably than those who suffered more injury as the result of trial error. For example, a defendant who sees the probability of acquittal drop from 51 percent to 49 percent as the result of error would get a new trial while a defendant who sees the probability of acquittal drop from 75 percent to 55 percent would be denied relief even though the latter's injury is ten times greater than the former's injury (a loss of 20 percent compared to a loss of 2 percent).

This formal statement of the harmless error rule distinguishes an error's effect-on-defendant, P(G|actual) - P(G|hypothetical), which can be estimated empirically, from the error tolerance term, T, which represents the normative judgment about fairness and justice. When there are disputes over the harmfulness of trial errors, it is helpful to know if there is disagreement over the effect-on-defendant, the amount of tolerable error, or possibly both terms. Knowing the source of disagreement may not always resolve the dispute, but it may at least reveal where the disagreement lies.

B. *Hypotheses about Harm to Defendant*

Some trial errors do not increase the probability of a guilty verdict; these errors cause no harm and are clearly harmless. Some trial errors harm the defendant by increasing the probability of a guilty verdict, but the magnitude of harm is tolerable; these are also considered harmless errors in the eyes of the law.[168] When a trial error harms the defendant to an intolerable degree, the error is considered harmful.[169] This framework can be depicted visually. Effects greater than T are intolerable, while effects less than or equal to T are either tolerable or not harmful at all.

Figure 1. Levels of Harm to the Defendant



If one could measure the effect-on-defendant precisely and knew the value of T, judging harmfulness would be as simple as comparing the effect size to T. One could pinpoint the effect as a point on Figure 1 and see which zone it falls into. In practice, however, it is not so simple. One can only estimate values of P(G|actual) and P(G|hypothetical). This means one can only estimate the effect-on-defendant. Estimates are inherently uncertain, but fortunately, we can quantify this uncertainty and conduct rigorous hypothesis tests. In the next Part, I use confidence

168. *See supra* discussion at Part II.B of the threshold of tolerable harm.
169. *See supra* discussion at Part II.B of *Strickland*, 466 U.S. at 668.

intervals to illustrate effects on defendants estimated with uncertainty from null hypothesis significance testing.

Null hypothesis significance testing is a common scientific practice. It is analogous to testing the presumption of innocence in a criminal trial. The researcher asserts a hypothesis to be tested (H_a), representing the assertion that an error was or was not harmless.[170] Like a defendant in a criminal trial is presumed innocent, the null hypothesis (H_0), the nullification of the researcher's hypothesis, is presumed to be true. The researcher must demonstrate, to a reasonable degree of certainty, that data compel rejection of the null hypothesis. The statistical significance of a scientific finding is akin to proof beyond a reasonable doubt. We presume H_0 is true but reject it if the data show H_0 is implausible. If the data are inconclusive or merely suggestive, H_0 retains the presumption of truth.

The hypothesis about trial error to be tested depends on which party bears the burden of proof. When the prosecution bears the burden of proving that the trial error was harmless, the analyst should test H_a: the effect of the error was less than or equal to T, against H_0: the effect of the error was greater than T.

When Prosecution Bears Burden of Proof

H_a: $[P(actual) - P(hypothetical)] \leq T$

H_0: $[P(actual) - P(hypothetical)] > T$

To properly test hypotheses about the harmfulness of trial errors, it is important to understand the difference between substantive and statistical significance of estimated effects. Substantive significance refers to the trial error's practical implications for the defendant in the courtroom; a large increase in the probability of conviction is substantively significant while a slight increase is not. Statistical significance refers to our certainty about the estimated effect; an estimate is statistically significant if it cannot be reasonably attributed to random chance; a statistically significant effect is a real signal and not just noise.

Even if one is reasonably certain the trial error increased the probability of conviction, the trial error may be deemed harmless as a matter of law if one is confident that the increase in the probability of conviction was tolerable. In other words, the effect-on-defendant may be positive and statistically significant (it does not fall into Figure 1's "no harm" zone), but not substantively significant because the error's effect on the defendant is tolerable (it does not cross into Figure 1's "intolerable harm" zone). If the data are not sufficiently conclusive, one fails to reject the H_0 and the prosecution fails to carry the burden of proof. (These propositions are illustrated by Figure 2.)

When instead it is the defendant who bears the burden of proving that the trial error was *harmful*, as is the case in a post-conviction

170. The subscript "a" means alternative. The research hypothesis, H_a, is the alternative to the null hypothesis, H_0.

proceeding, the analyst should test H_a: the effect-on-defendant was greater than T, against H_0: the effect-on-defendant was less than or equal to T.

When Defendant Bears Burden of Proof

H_a: $[P(actual) - P(hypothetical)] > T$

H_0: $[P(actual) - P(hypothetical)] \leq T$

The hypothesis testing frameworks for the prosecution and defendant are nearly mirror images, but there is an important asymmetry. If the analyst's estimate of the effect-on-defendant is inconclusive in the sense that it shows the effect of the error could have been tolerable or could have been intolerable, the party with the burden of proof should lose on the issue of the trial error's harmfulness. Put another way, if the prosecution's hypothesis fails, the defendant's hypothesis does not automatically prevail (and vice versa). If the trial error cannot be classified as harmful or harmless with certainty, the analysis fails to support either the prosecution or defendant's hypothesis. (This situation is illustrated by the middle panel in Figure 2.) Because the burden of proof belongs to one party or the other, the analyst should test either the prosecution or defendant's hypothesis, but not both at the same time.

C. *Testing Hypotheses about Harm to Defendant*

The first step of null hypothesis significance testing is stating the hypothesis to be tested as clearly as possible. As discussed in the prior Part, the hypothesis to be tested about trial error's effect on the defendant depends on whether the prosecution must prove the error was harmless or it is the defendant's burden to prove the error was harmful. The second step, which should precede any data collection or empirical analysis, is specifying the standard for statistical significance. By convention, the benchmark for statistical significance is .05.[171] For null hypothesis-testing purposes, we reject H_0 when an effect-on-defendant estimate would occur less than .05 of the time by chance assuming H_0 is true. In applied research, when researchers test a null hypothesis, they do not know whether they are making an inferential error,[172] but they can and should specify the threshold for reasonable scientific certainty before collecting and analyzing data. Specifying the statistical significance level prior to empirical analysis helps us test null hypotheses in a transparent and ethical manner because researchers declare the statistical significance

171. The .05 significance level is widely used, but subject to criticism as is the null hypothesis significance testing framework. *See* Philip H. Pollock III & Barry C. Edwards, The Essentials of Political Analysis 231-32 (6th ed. 2019). These criticisms are made possible by the primary benefit of the scientific methods: the analytical framework and its assumptions may be explicitly stated and evaluated.

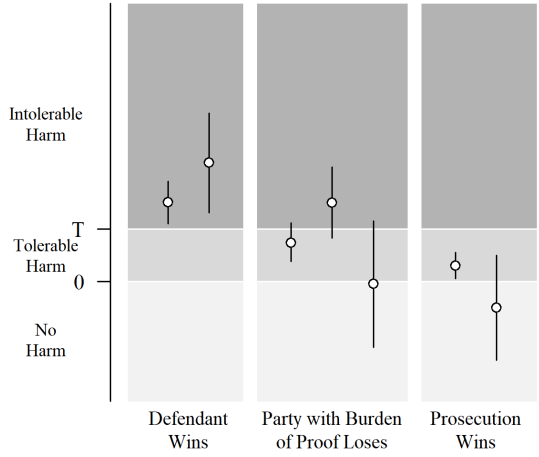
172. The "true" value of the parameter is unknown. If we knew the value, we would not need to estimate it.

threshold before conducting analysis and do not raise or lower it to favor a desirable conclusion.

As a visual aid, the analyst can show the estimated effect-on-defendant as a confidence interval, rather than simply reporting the numbers. Confidence intervals (CIs) are a useful data visualization tool because they show the estimated value along with its range of plausible values. The estimated effect is plotted as a point and the range of plausible values, the margin of error of the estimate, is identified with line segments on either side of the point. The CI of a harm estimate conveys both the substantive and statistical significance of the analysis; the CI's midpoint conveys substantive significance while the line segments on either side of the point represent certainty about the estimate.

Plotting the effect-on-defendant's CI relative to 0 and T is particularly useful.[173] Figure 2 shows some CIs for effect-on-defendant estimates and, below each panel, states how each estimated trial error should be classified according to the null hypothesis testing framework outlined above.

Figure 2. Evaluating Hypotheses about Effect-on-Defendant with Confidence Intervals



The midpoint of the effect-on-defendant CI is difference between P(G|actual) and P(G|hypothetical), a single number. The upper and lower boundaries of the interval identify the extent of reasonable

173. To conduct null hypothesis tests based on the .05 statistical significance level, the researcher should plot 90 percent CI because the research hypotheses are directional and, therefore, evaluated with one-tailed significance tests. If, for example, the prosecution bears burden of proving the effect-on-defendant was less than T, it does so when the 95th percentile of estimates (the upper boundary of a 90 percent CI) is less than T. If the defendant must show the harm was intolerable, he does so by showing that the 5th percentile of estimates (the lower boundary of a 90 percent CI) is greater than T. CIs facilitate one-tail hypothesis tests at the .05 significance level.

certainty. Margin of error calculations are somewhat technical but can be understood as the plausible range of the effect-on-defendant given the data and research design.[174] The larger the interval around the point, the less certain the estimate of harm; the point shows the estimated effect but the true amount of harm to defendant could plausibly fall anywhere on the interval. The point of padding the estimate is so that one is reasonably confident the interval contains the true value.

When the prosecution has the burden to prove a trial error was harmless, the prosecution meets that burden when the upper boundary of the CI is less than or equal to T. When the upper boundary of the CI is less than or equal to T, the entire CI lies below T and one is confident that the trial error did not cause intolerable harm to the defendant (see right panel of Figure 2). The prosecution prevails in these cases because the largest plausible effect-of-defendant is still tolerable. If, however, the upper boundary of the CI is greater than T, the prosecution fails to carry its burden of proof with respect to the harmlessness of the trial error (see left and middle panels of Figure 2).

The null hypothesis test employed when the prosecution bears the burden of proving that an error was harmless does not require the prosecution to prove a negative, a non-effect, or statistical insignificance. The prosecution is not required to prove that the trial error did not increase the probability of a defendant's conviction. Indeed, every facet of the trial may have some effect and, given sufficient sample size, all effects may be statistically significant. The prosecution simply must demonstrate, to a reasonable degree of certainty, that the effect-on-defendant was less than T. Since $T > 0$, the prosecution's hypothesis does not require proving the effect-on-defendant was less than or equal to 0.[175]

When the defendant must demonstrate that the trial error was harmful, the defendant does so by showing that the lower boundary of the effect-on-defendant's CI is greater than T. When the lower boundary of the CI is greater than T, the entire CI is greater than T and one can confidently conclude that the trial error caused intolerable harm to the defendant (see left panel of Figure 2). The defendant prevails in these cases because the least plausible amount of harm, based on the data, is still intolerable harm. If, however, the lower boundary of the CI is less than or equal to T (see Fig. 2, middle and right panels), the defendant fails to carry its burden of proof. To meet his burden of proof, the defendant must demonstrate, to a reasonable degree of certainty, that the effect-on-defendant is significantly greater than T.[176] It is not enough

174. The margin of error is equal to the product of a critical value (set before analysis based on the threshold for statistical significance) and the standard error of the sample statistic (calculated from statistical formula). *See* Pollock III & Edwards, *supra* note 171, at 184-88.

175. Social scientists are used to evaluating causal effects relative to zero, but the applicable legal rule suggests comparing causal effects to T instead.

176. For this hypothesis testing framework, I assumed that the defendant in post-conviction proceedings must prove a trial error was harmful to a reasonable

to establish some positive harm with certainty, the defendant must show substantively significant harm with certainty.[177]

The distinction between substantive and statistical significance is important because, given enough data, nearly all causal factors become statistically significant because the standard error of measurement approaches zero as the sample size approaches infinity.[178] If an analyst is willing to collect enough data, the analyst can estimate the effect-on-defendant with more confidence; however, the substantive significance of an effect estimate is not affected by sample size. Collecting more data would narrow Figure 2's confidence intervals, but the points should not move up or down. Substantive significance turns on the practical difference between defendant's actual and hypothetical trials. The estimated effect of improperly admitted evidence, for example, depends on the nature of that evidence as well as the strength of other evidence; larger samples

degree of certainty, just as a prosecutor in direct appeals must prove a trial error was harmless to a reasonable degree of certainty. As discussed, reasonable scientific certainty corresponds to a 5 percent threshold for statistical significance (which may be a more demanding standard than proof beyond a reasonable doubt). A close reading of caselaw suggests, however, the defendant's burden of proof may be less demanding. In *Brecht*, the Court opined the defendant must demonstrate more than a "reasonable possibility" of harm; in *O'Neal*, the Court held that the defendant's burden is satisfied if the evidence is "evenly balanced" and the judge is in "virtual equipoise" as to the harmfulness of the error. Brecht v. Abrahamson, 507 U.S. 619, 653 (1993); O'Neal v. McAninch, 513 U.S. 432, 435 (1995). One might infer from these cases that the defendant may prove harm at a more lenient significance level than 5 percent. If the analyst adopts a more lenient significance level, such as 10 percent or possibly a 50 percent standard, the confidence intervals buffeting the estimate of harm are narrowed because less certainty is demanded. Referring to Figure 2, if the defendant's burden of proof is less demanding, the line segments around the point estimates will narrow and it is possible that some estimates that were inconclusive now are sufficient to prove harmfulness. For simplicity, this Article uses the same significance threshold whether the burden of proof falls on the prosecutor or the defendant. An analyst could easily apply a more lenient significance threshold to better reflect legal standards. From a practical standpoint, the adjustment in proof standards suggested by *Brecht* and *O'Neal* simply allows the defendant to use smaller samples in his analysis compared to the prosecution. *Id*.

177. As a practical matter, the effect-on-defendant be significantly greater than T for the defendant to prevail. If the estimated effect-on-defendant is 0.11, for example, the defendant needs an extremely large sample size to rule out a reasonable possibility of tolerable harm.

178. The standard error of a sample statistic quantifies how much the statistic is expected to vary on repeated measures. If the standard error of a public opinion poll is 1 percent, one would expect the poll result to vary by 1 percent, on average, if it is repeated. If the standard error of a poll is 1 percent, an observed gender different of 1 percentage is not statistically significant as it could result from random error. Standard error declines as sample size increases. With sufficient sample size, a 1 percent point difference is statistically significant. Indeed, as sample size approached infinity, all differences, however slight, are statistically significant. In terms of Figure 2, this is equivalent to saying no point estimate falls exactly on the zero (0) line. Given enough data, random noise is eliminated and variables with slight effects become statistically significant.

can estimate that effect with greater precision, but sample size does not make the improper evidence any more or less damaging. The defendant cannot increase the substantive significance of an effect-on-defendant estimate, and the prosecution cannot lower it, simply by increasing the sample size.[179]

D. *Defining Tolerable Harm to Defendant*

To develop a general framework for analyzing harmfulness, I have expressed the amount of tolerable error, T, as an unknown constant term. The precise value of T is presently undefined. We could simply plot confidence intervals for effects without specifying the value of T and leave it to others to decide whether the harm is tolerable. As stated previously, quantifying T requires a normative judgment while quantifying the effect-on-defendant requires an empirical analysis. These are distinct tasks. Nevertheless, it is helpful to assert a plausible value of T to fully describe the framework for analyzing harmfulness.[180] The final part of Part IV offers some thoughts about specifying the value of T.

How much error should be tolerated? According to the Supreme Court's opinion in *Strickland*, the effect of an error is intolerable if it creates a "reasonable probability" of changing the result of the proceeding.[181] "A reasonable probability," according to the Court, "is a probability sufficient to undermine confidence in the outcome."[182] The Court has never quantified this value, but its loose description in *Strickland* helps us bracket T's possible values.[183] T is greater than 0 because some harm to the defendant is tolerable. At the same time, T is less than .50 because the defendant is not required to show that the trial error changed the outcome from likely conviction to likely acquittal.[184] Therefore, T is greater than 0 and less than .50.

---

179. Larger sample sizes do not make a causal variable's effect larger or smaller, but they do allow us to be more certain about the magnitude of the effect. One can think of enlarging sample size to reduce the margin of error of an estimate like zooming in to see a location of interest on a map in greater detail. Zooming in on a point of interest does not change its location but does allow one to pinpoint the location more clearly relative to roads and other lines on the map.
180. The task is analogous to calling balls and strikes in a baseball game. Even if the umpire knows where a pitch crossed the plate, the umpire needs to define the strike zone to judge whether the pitch crossed strike zone. Defining the strike zone's upper boundary is a different type of problem than locating where a pitch crossed the plate. The upper boundary of the strike zone is baseball is notoriously difficult to define. Nevertheless, some definitive boundaries must be established to consistently call balls and strikes.
181. *See Strickland*, 466 U.S. at 694.
182. *Id.*
183. Error tolerance will vary considerably among individuals. Certainly, disinterested members of the public will be more tolerant of trial errors than those with incarcerated friends or family members. The Court does not say whose confidence must be maintained.
184. .50 is the lowest increase in probability of conviction that always changes the most likely verdict from not guilty to guilty.

A trial error that undermines confidence in a jury's verdict causes reasonable doubts about the outcome; reasonable doubt about guilt is sufficient to undermine confidence in a guilty verdict. Phrases like "beyond a reasonable doubt" and "reasonable degree of scientific certainty" speak to the reliability of evidence and statistical significance, but they also embody a normative judgment about tolerance for error.[185] Defining T as the threshold of reasonable doubt is useful because reasonable doubt has been studied and estimated.[186] The consensus among judges and legal scholars appears to be that "beyond a reasonable doubt" should correspond to a level of certainty of .90 or more.[187] If finding a defendant guilty beyond a reasonable doubt requires at least .90 certainty, up to .10 doubt is tolerable. Accordingly, I suggest that T = .10. If a trial error's effect on the probability of conviction is greater than .10, one loses confidence and has reasonable doubts about the trial outcome. If, on the other hand, the effect-on-defendant is less than .10, one has confidence in the guilty verdict and does not have reasonable doubts about it.[188]

185. Setting the statistical significance threshold at .05 for null hypothesis significance testing is a value judgment about the risk of mistakenly rejecting a null hypothesis. The threshold for statistical significance corresponds to our Type I error tolerance; the 5 percent threshold is a standard level of tolerable error, but researchers may use other values depending on the research application.

186. *See* James Franklin, *Case Comment: United States v. Copeland, 369 F.Supp.2d 275 (EDNY 2005): Quantification of the 'Proof Beyond Reasonable Doubt' Standard*, 5 L., Probability and Risk 159, 159 (2006) (advocating the minimal, imprecise quantitative standard of above 80 percent); Alec Walen, *Proof Beyond a Reasonable Doubt: A Balances Retributive Account*, 76 La. L. Rev. 355, 356–57 (2015) (90 percent certainty customary definition of proof beyond reasonable doubt); Jack B. Weinstein & Ian Dewsbury, *Comment on the Meaning of 'Proof Beyond a Reasonable Doubt'*, 5 L., Probability and Risk 167, 169 (2006) (a probability of guilt no less than 95 percent). Some have argued that attempts to quantify the "beyond a reasonable doubt" standard will produce more confusion than clarity. *See* Catherine MA McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees*, 35 Vand. L. Rev. 1293 (1982).

187. *See* Devine, *supra* note 52. According to Francis C Dane, *In Search of Reasonable Doubt*, 9 L. and Hum. Behav. 141, 152-53 (1985), the Supreme Court's description of the reasonable doubt standard in Johnson v. Louisiana, 406 U.S. 356 (1972), can be interpreted to require 88 percent certainty. Jurors appear to have greater tolerance for uncertainty with certainties in the 75 percent-85 percent range crossing the threshold of reasonable doubt. *Id*. at 153. For a comparison of varied approaches to quantifying reasonable doubt, *see Id.*; Irwin A Horowitz & Laird C Kirkpatrick, *A Concept in Search of a Definition: The Effects of Reasonable Doubt Instructions on Certainty of Guilt Standards and Jury Verdicts*, 20 L. and Hum. Behav. 655, 666 (1996) (jurors operationalize the beyond a reasonable doubt standard as requiring surprisingly little certainty, no more than 80 percent across several definitions provided). *See also* Rita James Simon, *"Beyond a Reasonable Doubt" An Experimental Attempt at Quantification*, 6 J. of Applied Behav. Sci. 203, 206 (1970) (jurors have reasonable doubt when they believe there is only .72 to .76 probability defendant committed the crime). Interestingly, the confidence required to surpass reasonable doubt, .90, is lower than the conventional level used for reasonable certainty in social sciences, .95.

188. To be clear, I assert that T = .10 because some error tolerance must be specified

Parties can use scientific analysis to show whether trial errors are harmful or harmless to a reasonable degree of certainty. In some cases, however, the analysis will be inconclusive because the CI of the harm estimate crosses over T =. 10 (*see* Figure 2, middle panel). Inconclusive results are mostly likely to happen when the point estimate of harm is close to .10. It is possible to reduce uncertainty by increasing sample size, but estimates are inherently uncertain. As a practical matter, then, even if we precisely define T, the certainty demanded by null hypothesis testing prevents us from definitively classifying effect-on-defendant estimates in the vicinity of T.[189]

V. Framework for Analyzing Harm to the Trial Process

The effect-on-trial approach does not focus on trial outcomes. Instead, it focuses on the trial process itself. The relevant inquiry is on jurors' internal thought processes and the extent to which the trial error shaped their impressions of the defendant and understanding of the events at issue in the trial. A trial error is harmful if it significantly distorts the narratives and accounts presented to the jury.

As discussed in Part III.A, one cannot interrogate a trial's actual jurors to learn how a trial error affected their impressions of the defendant and their understanding of the events at issue. The researcher, however, can accurately describe trials to human research subjects and study how a trial error affects their impressions of the defendant and understanding of trial events. Primarily, the researcher can ask human subjects the key question that courts cannot ask trial jurors: Does the evidence in question shape your thoughts about the trial? If, for example, the trial error was mistakenly admitted evidence, one can ask potential jurors how much that evidence affected their thoughts about the trial; if the error was one of wrongly omitted evidence, one can ask how much that evidence would change potential jurors' thoughts about the trial. The researcher can even ask respondents how much the difference between actual and hypothetical trials affects their view of the trial.[190] One could

---

to assess whether a trial error's effect-on-defendant is tolerable and this value of T is plausible based on existing research quantifying reasonable doubt in criminal trials. It is reasonable to suggest that the value of T depends on the severity of sentence involved. As a society, we might be confident that verdicts in misdemeanor trials are reliable even though there is more than a 10 percent chance of wrongful conviction but lose confidence in capital punishment even though there is less than a 10 percent chance of wrongful conviction.

189. I make this point to address potential concern that setting T = .10 creates an arbitrary bright line that would award a defendant a new trial based on an .101 estimate but compel the opposite result with a .099 estimate. Because statistical estimates are inherently uncertain, the CI of estimates in the vicinity of .10, like .101 and .099, will overlap .10. Estimates around T = .10 will be inconclusive and insufficient to reject a null hypothesis (see middle panel of Figure 2).

190. There could be cases where people are deeply divided over how much difference the change in evidence makes to the trial. If most people think it makes no difference at all, but some think it makes a great deal of different, do we average

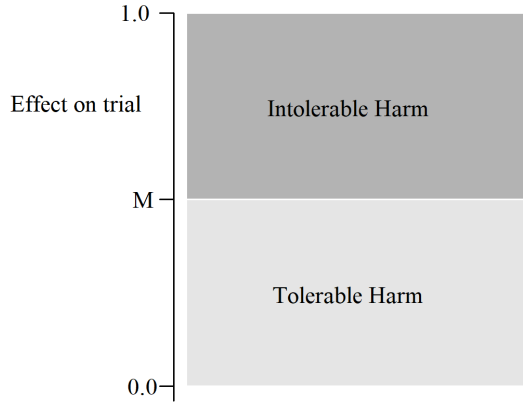
then estimate the effect-on-trial as the proportion of respondents who report the change in evidence makes a moderate to great deal of difference in their impressions of the trial.

The threshold of harm under the effect-on-trial approach has not been described as clearly as tolerance for harm to the defendant, but the discussion in Part IV.D offers some general guiding principles.[191] A trial error is not harmful simply because it has some definite impact on the trial. At the same, a trial error can be considered harmful even though it does not completely overturn everything jurors thought about a case, or significantly affect every juror. A trial error may affect how jurors think about a trial even if it did not change their verdict.

At what point is the effect-on-trial large enough to deem the trial error harmful? Let M represent the threshold for tolerable effect-on-trial. Effects up to M are tolerable, but effects greater than M are intolerable. M is analogous to T in the effect-on-defendant analysis, but M is used to signify harm tolerance for effect-on-trial analysis to distinguish two different types of harm which can be evaluated relative to different thresholds.

The harmfulness of the effect-on-trial can be visualized in much the same way as the effect-on-defendant. If the effect-on-trial is greater than M, the trial error's effect-on-trial was intolerable. If the effect-on-trial is less than or equal to M, the trial error was tolerable and should be considered harmless.

Figure 3. Levels of Harm to the Trial



their responses and say the change in evidence only made a little difference or focus on the percentage who say it makes a great deal of difference? This is an interesting question, not clearly answered by case law, but as an empirical matter should be rare. When people are asked to measure something, their responses tend to be distributed in a bell-curve shaped distribution around a modal response.

191. *See* discussion *supra* Part II.A (discussing the dearth of practical guidelines for implementing the effect-on-trial approach to harmless error analysis).

When the prosecution bears the burden of proving that the trial error was harmless, the analyst should test the research hypothesis (H_a) that the effect-on-trial was less than or equal to M. The null hypothesis (H_0) is that the effect-on-trial was greater than M.

When Prosecution Bears Burden of Proof

H_a: effect on trial $\leq$ M

H_0: effect on trial $>$ M

When the defendant bears the burden of proving that the trial error was harmful, the analyst should test H_a: the effect-on-trial was greater than M, against H_0: the effect of the error was less than or equal to M.

When Defendant Bears Burden of Proof

H_a: effect on trial $>$ M

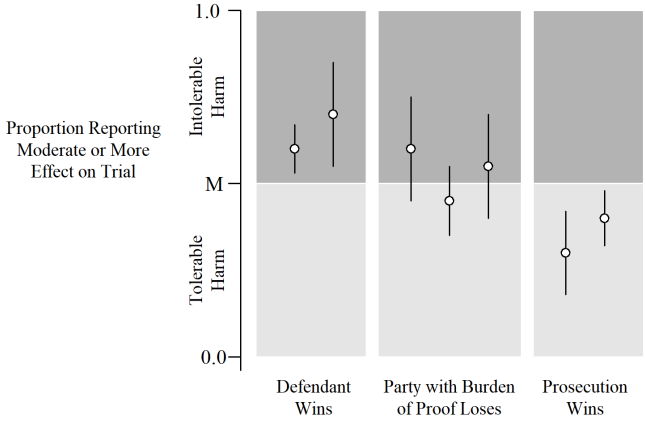
H_0: effect on trial $\leq$ M

We can assess the harmfulness of the effect-on-trial using hypothesis tests and confidence interval in much the same way we did to evaluate trial errors using the effect-on-defendant approach. Plotting the confidence interval against M is particularly useful for evaluating the harmfulness of the trial error. As was the case with evaluating the harmfulness of an effect-on-defendant, one could simply leave the value of M undefined, but it is more instructive to assert a plausible value of M to clarify the framework of analysis. How much distortion of the trial narrative presented to jurors is tolerable?

For present purposes, I posit that M = .50. This is, admittedly, somewhat arbitrary, but setting M = .50 has intuitive appeal. With M = .50, a trial error is classified as harmful when most jurors say the trial error had a moderate to great deal of effect on their impression of the trial. If the estimated effect-on-trial is greater than M = .50, the analysis indicates the trial error would affect most jurors. If the disputed evidence matters to most jurors, they will discuss it during deliberation, think about it, and rely on it to a harmful extent.

Figure 4 illustrates hypothesis testing with confidence intervals using the framework suggested here.

Figure 4. Evaluating Hypotheses about Effect-on-Trial with Confidence Intervals



When the prosecution has the burden to prove a trial error was harmless, the prosecution meets that burden when the upper boundary of the CI is less than M. When the entire CI is less than M, one can be confident that the effect-on-trial was not intolerable (see Fig. 4, right panel). If, however, the upper boundary of the CI is greater than M (the CI either overlaps M or is entirely above M), the prosecution fails to carry its burden of proof with respect to the harmlessness of the trial error (see Fig. 4, left and middle panels).

Conversely, when the defendant must prove the trial error was harmful, the defendant does so by showing that the lower boundary of the CI is greater than or equal to M. When the entire CI is greater than or equal to M, one can confidently conclude that the effect-on-trial was intolerable (see Fig. 4, left panel). If, however, the lower boundary of the CI is less than M (including where the CI intersects M or is less than M), the defendant fails to carry its burden of proof with respect to the harmfulness of the trial error (see Fig. 4, middle and right panels).

Overall, the framework for evaluating the effect-on-trial is similar to the framework for evaluating the effect-on-defendant, but there are some important differences. Primarily, the effect-on-trial approach differs because the inquiry is whether the trial error made a difference in the jurors' perception of the trial without consideration of whether the difference favored the prosecution or defendant.[192] Therefore, the effect-on-trial estimate cannot be a negative value. Conversely, under the effect-on-defendant approach, the inquiry is whether the defendant's chances for a guilty verdict would be more or less likely in a hypothetical

192. One could imagine scenarios where a trial error affects the jurors' understanding of the narrative of events without affecting how they feel about a particular defendant. For example, in a case involving multiple defendants, a trial error may affect jurors' thoughts about some defendants but not others.

trial. If a guilty verdict would be more likely in the hypothetical trial, the result would be negative.

The effect-on-trial approach is more sensitive to differences between actual and trial conditions than the effect-on-defendant approach because a change in the trial can affect jurors' thoughts and impressions without changing their vote on the verdict. The effect-on-trial approach is more likely to find a trial error harmful. This helps explain why M should be set at a higher threshold (.50) than the value of T (.10) in the effect-on-defendant analysis.[193]

An interesting situation arises under the effect-on-trial approach if the disputed evidence makes a big difference in the trial, but the difference favors the prosecution in the appellate proceedings. During trial, some reasonable litigation strategies backfire. This situation could arise when the prosecution introduces a coerced confession that creates doubts about the state's evidence in an otherwise solid case, or the defense fails to introduce character evidence that would have opened the door to evidence of prior convictions.[194] Under the effect-on-trial approach, one could argue these types of errors affect the overall narrative of the trial in favor of the defendant. Is the truth-seeking function of juries so important that appeals courts should order new trials to provide a community an opportunity to learn that the case against the defendant is even stronger than they thought it was? To some, the truth-seeking function of juries may be so important that this kind of harm justifies reversal. It seems appropriate that the disputed evidence's effect on the jury must also be directed against the defendant (even though it does not have to significantly affect the outcome).[195]

193. Additionally, the effect-on-defendant is measured as the difference between two quantities: P(G|actual) and P(G|hypothetical) whereas the effect-on-trial may be measured as the proportion of jurors who believe the trial error had a moderate to great deal of effect on their impression of the trial.

194. This situation is suggested by the Supreme Court's analysis of defense counsel's decision not to introduce character witnesses in the sentencing phase of David Washington's trial. See *Strickland*, 466 U.S. at 700 ("[A]dmission of the evidence respondent now offers might even have been harmful to his case.").

195. In a case where disputed evidence has a significant effect-on-trial but significantly helped the defendant, the defendant's interests should prevail over the community's interest in knowing the unblemished truth. Before their trials, defendants have the right to plead guilty and avoid trial. Defendants are not compelled to stand trial to satisfy the community's interest in knowing the truth. In this situation, the trial error had a substantial effect on the jury's understanding and impressions of the trial, but not a harmful effect on the defendant. While the effect-on-trial approach should not require proof or disproof of intolerable harm to the defendant, it seems appropriate to consider whether analysis of trial error indicates some harm to the defendant.

Conclusion

It is important to assess the harmfulness of trial errors accurately, as judgments about the harmfulness of errors have profound consequences for defendants, crime victims, the families of victims and defendants, the criminal justice system, and society. Appellate judges are expected to apply the rule correctly, but they cannot learn what the jurors thought and are forced to speculate about the difference between the trial defendant received and an error-free fair trial. Making predictions of this sort is extremely difficult. Empirical research suggests that judges should have serious doubts about their assessments of the harmfulness of trial errors.

Scientific research can inform harmless error analysis. This Article provides a scientific framework for testing claims about the harmfulness of trial errors. This framework allows social scientists to estimate quantities that matter legally and report findings that help judges evaluate the harmfulness of trial errors. Scientific analysis of trial errors can help courts assess the harmfulness of trial errors more accurately, efficiently, and confidently. Everyone affected by the application of the rule stands to benefit.

Scientific analysis of trial errors should generally support and reinforce the conclusions appellate court judges reach after hearing the parties' arguments and considering the record. Scientific findings on the impact of errors can reassure judges who uphold or reverse convictions. In some cases, however, the results of scientific analysis of trial errors may be inconclusive, inconsistent, or surprising. Inconclusive results, where analysis does not conclusively show the trial error was harmful or harmless, are a necessary byproduct of requiring a high level of certainty and help us avoid overconfidence.

Inconsistent results may occur if effect-on-defendant analysis yields a different conclusion about the harmfulness of an error than effect-on-trial analysis does. What conclusion should a judge reach if analysis shows the effect-on-defendant was harmless but the effect-on-trial was harmful? That is a question judges must answer. Because there are two different approaches, it is helpful to think about the merits of both perspectives and decide which approach to follow if they yield inconsistent results.

Although scientific analysis of trial errors should generally support and reinforce judges' intuitions about trials and trial evidence, there may be surprising findings that contradict and challenge judges' expectations about jurors and juries. In these cases, scientific analysis would provide judges an opportunity to understand a perspective they had not considered and prompt them to reconsider their initial impressions. Nevertheless, judges may be reluctant to have others contradict or challenge their subjective opinions about the harmfulness of errors. Courts need help to analyze the effects of trial errors accurately but are prone to overconfidence so may not think they need help. For these reasons, improved

analysis of trial errors is unlikely to arrive at the request or invitation of judges.[196]

Improved analysis of the effects of trial errors is likely to come at the insistence of parties with objective evidence on their side. It should be used by prosecutors to prove the defendant received a fair trial and does not deserve another one.[197] It should be used by defendants who did not receive fair trials when they petition courts for post-conviction relief.[198] The adversarial process incentivizes innovation, and the work product doctrine should protect the analysis from involuntary discovery. To assist parties in litigation, the analysis needs to be affordable, swift, and accurate; a scientific framework for analyzing the harmfulness of trial errors is a big step in that direction.

When I contemplate future uses of scientific analysis of trial errors, I think about the Innocence Project and the analysis of DNA in post-conviction proceedings to prove innocence. That movement has not overturned hundreds of convictions because courts asked scientists to analyze DNA evidence. Instead, DNA analysis was first used by defendants who were wrongly convicted to prove their innocence; those who had the most to gain from it were the first to use it.[199] Courts were ini-

196. Winkelman et al. suggest that judges appoint analysts as special masters to conduct empirical analysis on the harm caused by trial errors. *See* D. Alex Winkelman et al., *An Empirical Method for Harmless Error*, 46 Ariz. St. L.J. 1405, 1430 (2014).

197. Parties may have some difficulty introducing scientific analysis of trial errors on direct appeal because appellate courts do not entertain new evidence or expert witness testimony. *See generally* Jeffrey C. Dobbins, *New Evidence on Appeal*, 96 Minn. L. Rev. 2016 (2011). That said, analysis of the harmfulness of a trial error is not new evidence that speaks to whether defendant committed a crime, like the discovery of a new eyewitness, and could not possibly be introduced during trial because it is a retrospective evaluation of the trial itself. If the analysis cannot be introduced by the appellee or appellant, it could be submitted by an independent third party as an amicus brief, a common vehicle for supplying technical information and scientific reports to appellate courts. *See Id.*, at 2051-54; The problem with relying on disinterested third-parties to supply the analysis as amicus is the lack of incentive to do so; perhaps such work will one day by supported by public interest organizations or law school clinics.

198. The clearest opportunity for scientific analysis of trial errors would be a habeas corpus petition for post-conviction relief based on constitutional trial error. Federal law greatly limits petitioner's opportunity to present evidence in support of his claim, but scientific analysis of the effect of constitutional errors may fit within the rules. *See* Wood, *supra* note 117. In this type of civil proceeding, filed in trial court, the petitioner (formerly the defendant) must prove the trial error was harmful. Scientific analysis of the trial error's effect is clearly relevant to the central issue in this type of proceeding. Additionally, retrospective analysis of the harmfulness of trial errors cannot be conducted while a trial is still underway. Introducing the analysis in this context would give the analyst an opportunity to carefully plan and conduct research, explain his or her work, or answer questions about it under oath. PWJ Comment: the reasoning in the final two sentences are unclear to me. If others agree, I would suggest a rephrasing.

199. *See* Jim Dwyer, et al., Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted (2000).

tially reluctant to consider DNA post-conviction analysis; proponents had to fight to obtain evidence for testing and have courts consider evidence of actual innocence.[200] Now, DNA analysis is widely accepted and used by both prosecutors and defendants to help judges and juries make better decisions.[201]

This Article does not suggest an alternative to judicial analysis of trial errors, nor does it argue for modification of the harmless error rule. The goal is to analyze the effect of trial errors in a scientific manner to help judges make decisions consistent with the objectives of the harmless error rule. While it is ultimately the judge's responsibility to decide whether the defendant is entitled to a new trial or whether he should remain incarcerated, an analyst applying the foregoing scientific framework can inform the judge's analysis and thereby instill public confidence in the result.

200. See Karen Christian, "*And the DNA Shall Set You Free": Issues Surrounding Postconviction DNA Evidence and the Pursuit of Innocence*, 62 Ohio St. L. J. 1195 (2001); Edward F. Connors et al. Nat'l Inst. of Justice, Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence after TRIAL (1996).

201. While post-conviction DNA testing is readily associated with exonerating innocence men and women, only 40 percent of DNA tests conducted by the Innocence Project come out in the inmate's favor. "Sixty percent of the time the people do not want to admit it, or they are lying or psychopaths." Barry C. Scheck, *Barry Scheck Lectures on Wrongful Convictions*, 54 Drake L. Rev. 597, 601 (2005); *see also* Tonja Jacobi & Gwendolyn Carroll, *Acknowledging Guilt: Forcing Self-Identification in Post-Conviction DNA Testing*, 102 Nw. Univ. L. Rev. 263 (2008). One might similarly expect scientific analysis of the harmfulness of trial errors to disprove an inmate's claim more often than support it.