

3-1-2022

Using Burdens of Proof to Allocate the Risk of Error when Assessing Developmental Maturity of Youthful Offenders

David L. Faigman

Kelsey Geiser

Follow this and additional works at: <https://scholarship.law.wm.edu/wmlr>



Part of the [Criminal Law Commons](#), [Criminal Procedure Commons](#), and the [Neuroscience and Neurobiology Commons](#)

Repository Citation

David L. Faigman and Kelsey Geiser, *Using Burdens of Proof to Allocate the Risk of Error when Assessing Developmental Maturity of Youthful Offenders*, 63 Wm. & Mary L. Rev. 1289 (2022), <https://scholarship.law.wm.edu/wmlr/vol63/iss4/6>

Copyright c 2022 by the authors. This article is brought to you by the William & Mary Law School Scholarship Repository.
<https://scholarship.law.wm.edu/wmlr>

USING BURDENS OF PROOF TO ALLOCATE THE RISK OF
ERROR WHEN ASSESSING DEVELOPMENTAL MATURITY
OF YOUTHFUL OFFENDERS

DAVID L. FAIGMAN* & KELSEY GEISER**

ABSTRACT

Behavioral and neuroscientific research provides a relatively clear window into the timing of developmental maturity from adolescence to early adulthood. We know with considerable confidence that, on average, sixteen-year-olds are less developmentally mature than nineteen-year-olds, who are less developmentally mature than twenty-three-year-olds, who are less developmentally mature than twenty-six-year-olds. However, in the context of a given case, the question presented might be whether a particular seventeen-year-old defendant convicted of murder is “developmentally mature enough” that a sentence of life without parole can be constitutionally imposed on him or her. While developmental maturity can be accurately measured in group data, it cannot be assessed in individuals with confidence. This fact is an instance of a fundamental disconnect that occurs at the intersection of science and law between what scientists

* Chancellor & Dean and John F. Digardi Distinguished Professor of Law, University of California Hastings College of the Law.

** 2022 J.D. Candidate, University of California Hastings College of the Law.

This Article arises out of work done by the first author as a member of the MacArthur Foundation’s Research Network on Law and Neuroscience. We are indebted in particular to the members of the Group to Individual (G2i) working group, including Paul Appelbaum, Josh Buckholtz, Andre Davis, Philip Dawid, Nancy Gertner, Peter Imrey, Russell Poldrack, Marcus Raichle, Valerie Reyna, and Tracey Woodruff. Additionally, special thanks are owed to Laurence Steinberg and Elizabeth Scott, members of the Network, for their early conversations over lunch regarding the feasibility of employing burdens of proof as described here. Nonetheless, the contents of this Article reflect the views of the authors and do not necessarily represent the official views of either the John D. and Catherine T. MacArthur Foundation or the MacArthur Foundation Research Network on Law and Neuroscience (www.lawneuro.org).

study and what courts ordinarily need to know. Scientists typically study phenomena at the group or population level, whereas courts usually need to determine whether a particular case is an instance of some known phenomenon. This is called the group to individual (G2i) problem.

Although the G2i problem cannot be fully resolved, it can be managed by using the base-rate data available in the research literature to set the burden of proof. Setting the burden of proof is a classic mechanism for allocating the risks of making a mistake. Two factors in particular inform judgments about allocating risk of error, with the first being the likelihood or frequency of the fact in question and the second being the costs associated with the error. The rarer the fact and the larger the cost of a mistake, the greater the burden of proof should be. The latter factor, the costs associated with error, lies behind the traditional burdens of proof of preponderance of evidence and proof beyond a reasonable doubt in civil and criminal cases, respectively. In contrast, while the former factor, the frequency of the fact in question, is used regularly in areas of applied science, it has generally not informed allocations of burdens of proof in court.

This Article sets forth a framework of shifting burdens of proof grounded in the research literature that can be employed to allocate the risk of error when assessing developmental maturity in the sentencing of offenders across the age spectrum.

TABLE OF CONTENTS

INTRODUCTION	1292
I. THE GROUP TO INDIVIDUAL PROBLEM	1293
II. SENTENCING YOUTHFUL OFFENDERS UNDER THE EIGHTH AMENDMENT	1297
III. BURDENS OF PROOF AND DEVELOPMENTAL MATURITY ...	1305
IV. EMPLOYING BURDENS OF PROOF FOR SENTENCING YOUTHFUL OFFENDERS	1310
V. THE SENTENCING DECISION	1314
CONCLUSION	1315

INTRODUCTION

In *Roper v. Simmons*, the United States Supreme Court held that the Eighth Amendment prohibited executing youthful offenders who committed capital crimes before the age of eighteen due to the inherent immaturity of adolescents relative to adults.¹ Although the Court recognized that this categorical rule was both over- and under-inclusive because some offenders younger than eighteen are considered developmentally mature and some over that age are not, the difficulty lay in identifying which were which.² This is a classic example of a core challenge that lies at the intersection of law and science. Whereas scientists primarily research phenomena at the population or group level, courts typically seek to answer whether a particular case is an instance of some phenomenon.³ This is referred to as the “group to individual” (G2i) problem, an issue of statistical inference that plagues most uses of scientific research in court.⁴ Although the problem appears in virtually all situations in which scientific research is used to inform individual courtroom decisions, there is no one approach that can be used across legal contexts to manage it.⁵ In this Article, we consider an approach to meeting the G2i problem in the context of sentencing youthful offenders in which the constitutional rule was set at the group level—age of offender at time of offense—but its application inevitably must be done at the individual level. We propose using burdens of proof, a classic mechanism for balancing the cost of

1. 543 U.S. 551, 573-75 (2005).

2. *Id.* at 574. There is some inconsistency as to how to define the stages of adolescence and adulthood because brain development does not follow a single chronological trajectory. See Laurence Steinberg et al., *Are Adolescents Less Mature Than Adults? Minors' Access to Abortion, the Juvenile Death Penalty, and the Alleged APA "Flip-Flop,"* 64 AM. PSYCH. 583, 592-93 (2009). For clarity, we define early adolescence as ten to thirteen, middle adolescence as fourteen to seventeen, late adolescence as eighteen to twenty-one, and young adulthood as twenty-two to twenty-five.

3. See David L. Faigman et al., *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417, 419 (2014).

4. *Id.* at 417, 420.

5. See *id.* at 419-20.

making errors under conditions of uncertainty, to best manage the G2i problem in the sentencing of youthful offenders.⁶

We begin in Part I with a brief background on the limitations posed by the group to individual problem and an introduction to how it applies to sentencing youthful offenders. Parts II, III, and IV focus exclusively on the law and science surrounding the issue of assessing developmental maturity in different age cohorts. Part V returns to the question of sentencing in light of the fear of “permanent incorrigibility.”

I. THE GROUP TO INDIVIDUAL PROBLEM

The law has largely failed to recognize the fundamental disconnect between the usual level of study in science and what courts typically expect when they employ scientific evidence to decide cases. Although courts decide cases based on the individual or situation at hand, scientists primarily conduct studies at the group level.⁷ Unfortunately, group observations rarely apply universally to their individual members, meaning that group- or population-level findings may only provide weak support for individual determinations.⁸ For example, scientists may have considerable confidence that cross-racial identifications are less accurate than same-race identifications or that the chemical Benzene causes leukemia, but this research cannot show with confidence that a particular identification is accurate or that Benzene caused a particular person’s leukemia.

6. Professor William Berry recently published an excellent article in which he sought to solve the same issue considered here, the under- and over-inclusivity problem created by *Roper*’s bright-line rule at eighteen years of age. See William W. Berry III, *Eighth Amendment Presumptive Penumbra (and Juvenile Offenders)*, 106 IOWA L. REV. 1 (2020). Similarly relying on the developmental literature, he compellingly argues that courts should extend the insight that youthful offenders are largely still immature beyond the age of eighteen. See *id.* at 6 n.23. However, Professor Berry solves this problem using the notion of “penumbra,” thus creating a gray area around the age of maturity. See *id.* at 6. While interesting, we believe that employing burdens of proof will accomplish the objective that we share with Professor Berry, but in a way that provides greater guidance to the sentencer.

7. See Russell A. Poldrack et al., *Predicting Violent Behavior: What Can Neuroscience Add?*, 22 TRENDS COGNITIVE SCI. 111, 115 (2018).

8. See *id.*

In the eyewitness accuracy example, researchers study accuracy rates between groups of eyewitnesses. A particularly robust finding in this literature comes from comparisons of same-race identifications to cross-race identifications. Typically in this research, one group of subjects is given a scenario that requires same-race identifications and another group different-race identifications.⁹ This research finds that, on average, same-race identifications are more accurate than different-race identifications.¹⁰ However, nothing in the research literature indicates that an eyewitness researcher could reliably determine whether a particular cross-race identification was accurate.¹¹ As a consequence of this inherent limitation of the research, eyewitness experts ordinarily are permitted to testify only to the general phenomenon and do not offer a clinical opinion regarding an individual case.¹²

Eyewitness identification experts generally conform their testimony to how most scientists would approach application of their findings. Scientists ordinarily describe their research probabilistically and apply it to some group or population.¹³ Any application of the findings to an individual is usually done probabilistically as well.¹⁴ In forecasting whether it will rain this afternoon, for instance, meteorologists estimate the likelihood it will rain based on models built from group data and speak to individual cases using likelihood estimates.¹⁵ Similarly, when doctors provide informed consent regarding whether a drug causes

9. See Christian A. Meissner & John C. Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review*, 7 PSYCH. PUB. POL'Y & L. 3, 13 (2001).

10. See, e.g., John P. Rutledge, *They All Look Alike: The Inaccuracy of Cross-Racial Identifications*, 28 AM. J. CRIM. L. 207, 211 (2001); John C. Brigham et al., *Accuracy of Eyewitness Identifications in a Field Setting*, 42 J. PERSONALITY & SOC. PSYCH. 673, 674 (1982).

11. See, e.g., Rutledge, *supra* note 10, at 211.

12. See, e.g., *United States v. Hines*, 55 F. Supp. 2d 62, 72 (D. Mass. 1999) (“The function of the expert here is not to say to the jury—‘you should believe or not believe the eyewitness’.... All that the expert does is provide the jury with more information with which the jury can then make a more informed decision.”).

13. See Faigman et al., *supra* note 3, at 419.

14. See *id.* at 421-22.

15. See Nat'l Ctrs. for Env't Info., *Numerical Weather Prediction*, <https://www.ncei.noaa.gov/products/weather-climate-models/numerical-weather-prediction#:~:text=The%20GFS%20model%20is%20a> [<https://perma.cc/AK8L-QMTX>].

side effects, the information is provided statistically, not as a statement of scientific certainty. In other words, scientists ordinarily retain their group perspective even when describing individual cases.

In the example of Benzene, however, courts expect a different approach to the scientific evidence, despite the similar limitations of the underlying science. In the area of medical causation, typically arising in cases involving toxic torts or medical malpractice, courts insist on expert testimony regarding the individual case.¹⁶ The issue of “general causation,” that is, can Benzene cause leukemia, remains a threshold issue that must be supported by valid research.¹⁷ However, courts also require proof of “specific causation” in such cases—in other words, sufficient proof that Benzene caused the plaintiff’s leukemia.¹⁸

Over the years, courts and the experts that appear before them in cases involving disputes over medical causation have jerry-rigged a method to overcome the disconnect between what science can confidently do and what the law demands of it. This method is referred to as “differential etiology,” and essentially calls on clinical judgment to “rule-in” the plaintiff’s claimed cause of the injury and “rule-out” all alternative causes.¹⁹ This process of logical deduction is largely ill-defined,²⁰ yet inevitably leads to statements from experts that the plaintiff’s condition was caused by a particular

16. See, e.g., *Milward v. Acuity Specialty Prods. Grp., Inc.*, 969 F. Supp. 2d 101, 115 (D. Mass. 2013), *aff’d sub nom. Milward v. Rust-Oleum Corp.*, 820 F.3d 469 (1st Cir. 2016).

17. See *id.*; see also U.S. DEP’T OF HEALTH & HUM. SERVS., REPORT ON CARCINOGENS: BENZENE (14th ed. 2016), <https://ntp.niehs.nih.gov/ntp/roc/content/profiles/benzene.pdf> [<https://perma.cc/AT6D-G7PE>].

18. See, e.g., *Milward*, 969 F. Supp. 2d at 115 (holding that without proof of specific causation, summary judgment in favor of defendant was appropriate); see also *In re Aredia & Zometa Prods. Liab. Litig.*, 483 F. App’x 182, 191 (6th Cir. 2012) (“Because Plaintiff failed to demonstrate an essential element of her case, specific causation, the grant of summary judgment was appropriate.”).

19. FED. JUD. CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 617-18 (3d ed. 2011) (“In a differential etiology, an expert first determines other known causes of the disease in question and then attempts to ascertain whether those competing causes can be ‘ruled out’ as a cause of plaintiff’s disease By ruling out (or ruling in) the possibility of other causes, the probability that a given agent was the cause of an individual’s disease can be refined.”); see Joseph Sanders et al., *Differential Etiology: Inferring Specific Causation in the Law from Group Data in Science*, 63 ARIZ. L. REV. 851, 853-54 (2021).

20. In a recent article, one of us (Faigman), along with several coauthors—all trained as scientists, sought to remedy this deficiency. See generally Sanders et al., *supra* note 19, at 6.

substance “to a reasonable degree of [scientific] certainty.”²¹ Such conclusions appear to be little more than oft-repeated mantras that have little basis in either science or law.²²

Hence, although science almost invariably operates at the group level, its use in court depends on the legal framework in which it is being fitted. Returning to the above examples, research on eyewitness identification is relevant and admissible at the group level because it educates the fact-finder regarding factors that might interfere with accuracy.²³ Research on Benzene is also relevant and admissible at the group level, but courts have deemed it necessary that experts opine on the individual case as well, even if such opinions have little empirical support.²⁴

The lesson here is that, despite the limitations inherent in the G2i problem, the law determines the necessary level at which expert testimony must be presented—group or group and individual.²⁵ Although different legal or scientific contexts might demand different kinds of presentation, the goal for the law should be to resolve the G2i problem in ways that are conducive to obtaining the most valid scientific opinions possible while also serving the practical demands involved in trying individual cases. Arguably, this is what has occurred in both the eyewitness and Benzene examples, albeit in different ways. In the former, individualized expert opinions are not necessary for the fact-finders to do their jobs.²⁶ The insight that eyewitnesses are not as accurate under certain specific conditions as perhaps generally supposed is helpful

21. Jules M. Epstein, *Reasonable Certainty: A Term It Is Certainly Reasonable to Repudiate*, 33 CRIM. JUST. 39, 39 (2018) (“Virtually every expert witness examination concludes with the question ‘and do you hold those beliefs/opinions/conclusions to a reasonable degree of [discipline] certainty.’ The phrase, like the term ‘insane,’ is a legal construct; but in science and elsewhere the words have no meaning, as science may include a measure of uncertainty, scientific knowledge evolves, and there is no agreed metric within or across disciplines for what degree of certainty—32 percent, 45 percent, 97 percent—is ‘reasonable.’ The standard is subjective to the declarant.” (alteration in original)).

22. *See id.*

23. *See* Thomas D. Albright, *Why Eyewitnesses Fail*, 114 PROC. NAT’L ACAD. SCIS. 7758, 7759 (2017).

24. *See, e.g., Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 150-51 (1999) (stressing the importance of proffered experts identifying “the particular circumstances of the particular case at issue”).

25. *See* Faigman et al., *supra* note 3, at 419.

26. *See id.* at 432-33.

to the fact-finder.²⁷ In the latter, fact-finders would be lost without expert assistance in applying the general science to an individual case, even if its reliability is uncertain.²⁸ Most fact-finders would have considerable difficulty reasoning from general toxicological and epidemiological evidence to determine causation in a specific case. Experts are presumed to be needed, and they are thought to be capable of making this inferential deduction.²⁹

The situation of sentencing youthful offenders is something of a blend of these two examples. First, as with all applied science, the G2i problem is inherent in the issue of developmental maturity.³⁰ Second, similar to eyewitness identification, group data about the relative maturity of different age cohorts is relevant for judges and fact-finders in considering appropriate penalties.³¹ Third, similar to the Benzene example, sentencing of a youthful offender is inherently an individualized determination, one in which fact-finders might benefit from guidance regarding how to apply the general research literature to individual cases.³² Finally, as the next Part details, unlike many interactions between law and science, the Supreme Court has explicitly recognized the presence of the G2i issue in sentencing youthful offenders, though the Court has yet to identify an appropriate method to manage it.

II. SENTENCING YOUTHFUL OFFENDERS UNDER THE EIGHTH AMENDMENT

In a series of cases, the Supreme Court has limited courts' ability to impose extreme sentences on juvenile and youthful offenders. The Court has repeatedly recognized that children are different from adults, so subjecting them to the same extreme punishments is inconsistent with Eighth Amendment principles.³³ In *Roper v. Simmons*, the Court held that executing minors under the death penalty is "cruel and unusual" punishment prohibited by the

27. *See id.* at 433.

28. *See id.* at 425.

29. *See id.* at 435.

30. *See id.* at 421.

31. *See id.* at 433.

32. *See id.* at 425.

33. *See, e.g., Roper v. Simmons*, 543 U.S. 551, 572-73 (2005).

Constitution.³⁴ In *Graham v. Florida*, the Court held that the same Eighth Amendment clause prohibited a juvenile offender from being sentenced to life without parole (LWOP) for a non-homicide offense.³⁵ Together, *Roper* and *Graham* set forth the “children are different” or “youth matters” jurisprudence that is used today.³⁶ These cases relied on differences between juveniles and adults that inform the key Eighth Amendment principles of culpability and deterrence, such as lack of maturity, increased risk-taking behaviors, susceptibility to outside influences, and inability to appreciate long-term consequences.³⁷ These cases imposed categorical bans on extreme sentences for juveniles largely on the basis that a defendant’s immaturity at the time of committing a crime meant that he or she was less culpable and not amenable to being deterred by extreme punishment.³⁸ Subsequent to *Graham*, the Court turned to the necessary task of how youthful offenders should be sentenced given the categorical restrictions of *Roper* and *Graham*.

In *Miller v. Alabama*, the Court held that mandatory LWOP for a juvenile who committed a homicide was unconstitutional, but held that this extreme sentence could be imposed on a case-by-case basis.³⁹ And in *Jones v. Mississippi*, the Court wrestled with what showing was necessary to sentence a juvenile who committed a homicide to LWOP.⁴⁰

In these cases, however, the Court has tended to confound the issue of maturity at the time of the offense with the separate issue of the likelihood the defendant will mature out of his or her youthful incorrigibility. Both factors—current immaturity and future likelihood of maturing out of incorrigibility—are empirical questions subject to scientific research. The issue of assessing developmental stages in youth has been the subject of extensive work.⁴¹ As a result, we have considerable knowledge about the developmental capacities

34. *Id.* at 575.

35. 560 U.S. 48, 74 (2010).

36. 543 U.S. at 569-70; 560 U.S. at 68.

37. *Roper*, 543 U.S. at 569-70; *Graham*, 560 U.S. at 68.

38. *Roper*, 543 U.S. at 570; *Graham*, 560 U.S. at 68, 74.

39. 567 U.S. 460, 479-80 (2012).

40. 141 S. Ct. 1307 (2021).

41. *See, e.g.*, Steinberg et al., *supra* note 2.

of different age cohorts.⁴² Hence, we know a lot of general information, or “G.” The difficulty lies in assessing a particular youth’s development, or “i.” In regard to predicting future behavior, however, the research literature is much more problematic. That research indicates that predictions regarding future dangerousness cannot be done well at the group level, much less at the individual level.⁴³

In *Roper*, Justice Kennedy held that executing minors was unconstitutional and cited the behavioral science literature to support the finding that juveniles are less developmentally mature than adults.⁴⁴ These differences in maturity levels translated to lesser culpability for juveniles, with corresponding implications for sentencing them to death.⁴⁵ The Court explained that “[o]nce the diminished culpability of juveniles is recognized, it is evident that the penological justifications for the death penalty apply to them with lesser force than to adults.”⁴⁶

The Court reasoned that the death penalty primarily serves two social functions: retribution and deterrence.⁴⁷ However, these functions manifest differently in regard to youthful offenders as compared to adults.⁴⁸ The Court stated, “[r]etribution is not proportional if the law’s most severe penalty is imposed on one whose culpability or blameworthiness is diminished, to a substantial degree, by reason of youth and immaturity.”⁴⁹ And the same logic applies to deterrence: “The likelihood that the teenage offender has made the

42. *See id.* at 592-93.

43. *See* John Monahan, *The Prediction of Violent Behavior: Toward a Second Generation of Theory and Policy*, 141 AM. J. PSYCHIATRY 10, 10 (1984); *see also* TEX. DEF. SERV., DEADLY SPECULATION: MISLEADING TEXAS CAPITAL JURIES WITH FALSE PREDICTIONS OF FUTURE DANGEROUSNESS 23 (2004), https://www.prisonlegalnews.org/media/publications/tx_defender_service_subj_deadly_speculation_misleading_tx_capital_juries_with_false_predictions_of_future_dangerousness.pdf [<https://perma.cc/36KG-DKRB>] (analyzing the disciplinary records of 155 capital defendants in Texas and finding that expert predictions of future dangerousness were wrong 95 percent of the time).

44. 543 U.S. 551, 569-70 (2005).

45. *See id.* at 569-71.

46. *Id.* at 571.

47. *Id.*

48. *See id.*

49. *Id.*

kind of cost-benefit analysis that attaches any weight to the possibility of execution is so remote as to be virtually nonexistent.”⁵⁰

A universal measure for “cognitive capacities or psychosocial maturity” remains elusive.⁵¹ Even if the concept of developmental maturity could be clearly defined, it is not a notion in which a threshold is crossed at some specific moment in time.⁵² Moreover, while maturity is correlated with age, members of different cohorts reach cognitive and physical maturity at varying times.⁵³ Given the variability of the construct of developmental maturity, both in terms of definition and onset, one might argue that fact-finders should conduct individualized assessments. As Justice Scalia argued in dissent in *Roper*, this would require “the sentencer to make an individualized determination, which includes weighing aggravating factors and mitigating factors, such as youth.”⁵⁴

The *Roper* Court rejected this ad hoc approach and rested its decision on the group data regarding the general immaturity of offenders under the age of eighteen.⁵⁵ Although the scientific literature makes clear that most juveniles are less mature than most adults, there is no method that allows sentencers to validly determine whether a particular defendant is as mature as an adult.⁵⁶

The Court did not reach the inevitable issue of what penalty young Mr. Simmons should receive on remand if, indeed, “[i]t is difficult even for expert psychologists to differentiate between the juvenile offender whose crime reflects unfortunate yet transient immaturity, and the rare juvenile offender whose crime reflects irreparable corruption.”⁵⁷ Although the Court did not describe it in such terms, this is as good a description of the G2i problem as is to

50. *Id.* at 572 (alteration in original) (quoting *Thompson v. Oklahoma*, 487 U.S. 815, 837 (1988)).

51. Grace Icenogle et al., *Adolescents' Cognitive Capacity Reaches Adult Levels Prior to Their Psychological Maturity: Evidence for a "Maturity Gap" in a Multinational, Cross-Sectional Sample*, 43 *LAW & HUM. BEHAV.* 69, 71 (2019).

52. See Laurence Steinberg, *Should the Science of Adolescent Brain Development Inform Public Policy?*, 28 *ISSUES SCI. & TECH.*, Spring 2012, at 67, 67-70.

53. See *id.* at 70.

54. *Roper*, 543 U.S. at 620 (Scalia, J., dissenting) (citing *Eddings v. Oklahoma*, 455 U.S. 104, 115-17 (1982)).

55. See *id.* at 569-70 (majority opinion).

56. See Steinberg et al., *supra* note 2, at 591-93.

57. *Roper*, 543 U.S. at 573.

be found in a Supreme Court decision. The Court acknowledged that some juveniles may exhibit behavior that is “irretrievably depraved” or that shows “irreparable corruption” as to justify a death sentence, but failed to define these terms.⁵⁸ Despite this possibility, the Court still held that a total ban was appropriate.⁵⁹ In the next case on the issue of developmental maturity, the Court again focused on the “G” of G2i, though it was becoming evident that the issue of “i” was looming.

In *Graham v. Florida*, the Court applied the logic of lessened juvenile culpability from *Roper* to LWOP for juvenile offenders.⁶⁰ The Court again rejected the case-by-case approach in holding that the Eighth Amendment does not permit a juvenile offender to be sentenced to LWOP for a non-homicide offense.⁶¹ The Court’s reasoning closely tracked that of *Roper*:

The case-by-case approach to sentencing must, however, be confined by some boundaries. The dilemma of juvenile sentencing demonstrates this. For even if we were to assume that some juvenile nonhomicide offenders might have “sufficient psychological maturity, and at the same time demonstrat[e] sufficient depravity,” to merit a life without parole sentence, it does not follow that courts taking a case-by-case proportionality approach could with sufficient accuracy distinguish the few incorrigible juvenile offenders from the many that have the capacity for change.⁶²

According to the *Graham* Court, there is no legitimate penological justification for condemning a juvenile to die in prison, and the Court reasoned that such a sentence would deny a juvenile the ability to eventually realize the extent of “human worth and potential.”⁶³

Like *Roper*, the *Graham* approach has much to recommend it, especially as a matter of establishing a categorical constitutional standard marking the line between immaturity and maturity in

58. *Id.* at 570, 573.

59. *Id.* at 568.

60. *See* 560 U.S. 48, 74 (2010).

61. *Id.* at 68, 74.

62. *Id.* at 77 (alteration in original) (citation omitted) (quoting *Roper*, 543 U.S. at 572).

63. *Id.* at 77, 79.

regard to the principles embedded in the Eighth Amendment.⁶⁴ However, also like *Roper*, *Graham* avoids the inevitable issue of how the lower court should sentence Mr. Graham if “incorrigible juvenile offenders” cannot be distinguished from “the many that have the capacity for change.”⁶⁵

In *Miller v. Alabama*, the Court finally reached the issue that it had largely avoided in *Roper* and *Graham*.⁶⁶ The *Miller* Court held that mandatory LWOP for those under eighteen at the time of committing a homicide violates the Eighth Amendment’s prohibition on cruel and unusual punishments.⁶⁷ This holding, of course, contemplates that some juveniles who have committed homicide might be sentenced to LWOP. The Court stated that “appropriate occasions for sentencing juveniles to this harshest possible penalty will be uncommon.”⁶⁸ “That is especially so,” the Court said, because of the difficulty of distinguishing “the juvenile offender whose crime reflects unfortunate yet transient immaturity, and the rare juvenile offender whose crime reflects irreparable corruption.”⁶⁹ In other words, the *Miller* Court suggested that LWOP should be barred for juvenile offenders whose crimes reflect the transient immaturity of youth in contrast to those whose crimes reflect permanent incorrigibility.⁷⁰

The Court stated that lower courts must “take into account how children are different, and how those differences counsel against irrevocably sentencing them to a lifetime in prison.”⁷¹ The Court

64. *See id.* at 82.

65. *Id.* at 77.

66. *See* 567 U.S. 460, 465 (2012).

67. *Id.*

68. *Id.* at 479.

69. *Id.* at 479-80 (quoting *Roper v. Simmons*, 543 U.S. 551, 573 (2005); *Graham*, 560 U.S. at 68). The Court gave no exact definition of what would make a juvenile irreparably corrupt, and the phrase does not come from the scientific literature. *See id.*

70. *See id.* at 479-80. *But see* *Jones v. Mississippi*, 141 S. Ct. 1307, 1317-18 (2021) (interpreting *Montgomery v. Louisiana*, 577 U.S. 190 (2016), and *Miller*, 567 U.S. 460, as not requiring “a separate factual finding of permanent incorrigibility”). The amici in *Miller* emphasized the difficulty in determining permanent incorrigibility, stating “there is *no reliable* way to determine that a juvenile’s offenses are the result of an irredeemably corrupt character.” Brief for the American Psychological Ass’n et al. as Amici Curiae in Support of Petitioners at 25, *Miller*, 567 U.S. 460 (2012) (Nos. 10-9646, 10-9647), 2012 WL 174239, at *25 (emphasis added).

71. *Miller*, 567 U.S. at 480. The Court identified five factors to be considered in determining whether to impose a LWOP sentence on a juvenile: (1) his “immaturity, impetuosity,

highlighted juveniles' heightened capacity for change and held that judges and juries should take this into account when determining whether a juvenile LWOP sentence is appropriate.⁷² Essentially, the Court intended sentencers to make a prediction about how a particular juvenile will behave decades in the future—a decision that research has shown is impossible to make with confidence.⁷³

Predictions of an adult's future dangerousness, let alone a juvenile's, are unreliable, inaccurate, and pose a particular concern with regard to racial bias.⁷⁴ Racial bias plays a significant role in predictions of future dangerousness as adults and children of color are more likely to be viewed as more violent and more likely to reoffend.⁷⁵

This case-by-case assessment of youthful defendants was exactly the sort of determination the Court found problematic in *Roper* and *Graham*.⁷⁶ In fact, the *Miller* Court understood that the decision stepped back from a core insight of these earlier cases.⁷⁷ It observed: “Our decision does not categorically bar a penalty for a class of

and failure to appreciate risks and consequences;” (2) his “family and home environment;” (3) “the circumstances of the homicide offense, including the extent of his participation in the conduct and the way familial and peer pressures may have affected him;” (4) “that he might have been charged and convicted of a lesser offense if not for incompetencies associated with youth—for example, his inability to deal with police officers or prosecutors ... or his incapacity to assist his own attorneys;” and (5) “the possibility of rehabilitation.” *Id.* at 477-78.

72. *See id.* at 479-80.

73. Ample research demonstrates that predicting juvenile and adult behavior, especially that far into the future, is incredibly difficult and tends to be overinclusive, also known as the “false positive problem.” *See* Kimberly Larson et al., *Miller v. Alabama: Implications for Forensic Mental Health Assessment at the Intersection of Social Science and the Law*, 39 *NEW ENG. J. ON CRIM. & CIV. CONFINEMENT* 319, 335-36 (2013) (“[T]here is currently no basis in current behavioral science nor well-informed professional knowledge that can support any reliable forensic expert opinion on the relative likelihood of a specific adolescent’s prospects for rehabilitation at a date that may be years to decades in the future.”); Alex R. Piquero, *Youth Matters: The Meaning of Miller for Theory, Research, and Policy Regarding Developmental/Life-Course Criminology*, 39 *NEW ENG. J. ON CRIM. & CIV. CONFINEMENT* 347, 355 (2013) (“[I]t is very difficult to predict early in the life-course which individual juvenile offender will go on to become a recidivistic adult offender.”).

74. *See* Piquero, *supra* note 73, at 355; JOHN H. LAUB & ROBERT J. SAMPSON, *SHARED BEGINNINGS, DIVERGENT LIVES: DELINQUENT BOYS TO AGE 70* 289-90 (2003) (explaining the limitations of using juvenile risk factors to attempt to predict future criminal behavior); Jennifer L. Skeem & Christopher T. Lowenkamp, *Risk, Race, and Recidivism: Predictive Bias and Disparate Impact*, 54 *CRIMINOLOGY* 680, 681 (2016).

75. *See* Skeem & Lowenkamp, *supra* note 74, at 681.

76. *See* 543 U.S. 551, 572-73 (2005); 560 U.S. 48, 75, 77-78 (2010).

77. *See Miller*, 567 U.S. at 483.

offenders or type of crime—as, for example, we did in *Roper* or *Graham*. Instead, it mandates only that a sentencer follow a certain process—considering an offender’s youth and attendant characteristics—before imposing a particular penalty.”⁷⁸

Brief reflection, however, reveals that *Miller* is not analytically different from *Roper* and *Graham*. In all three contexts, lower courts are in the position of having “a sentencer follow a certain process—considering an offender’s youth and attendant characteristics—before imposing a particular penalty.”⁷⁹ Under *Roper*, defendants are not subject to the death penalty, but they still need to be sentenced and could possibly be sentenced to LWOP.⁸⁰ Under *Graham*, non-homicide defendants may not have been sentenced to LWOP, but they were still sentenced to some period of time.⁸¹ Finally, under *Miller*, juvenile homicide offenders are not subject to *mandatory* LWOP, but they still may receive LWOP or a lesser sentence.⁸² Inevitably, whatever value there might be in setting clear lines by age-cohort, individual sentencing cannot be avoided.

Most recently, the Court appeared to reverse course in its treatment of extreme sentences for juvenile and youthful offenders in *Jones v. Mississippi*.⁸³ The *Jones* Court held that a sentencer is not required to make a separate factual finding of “permanent incorrigibility” before sentencing a juvenile to LWOP.⁸⁴ The Court held that in such cases, a discretionary sentencing system is “both constitutionally necessary and constitutionally sufficient.”⁸⁵ The Court continued, stating “so long as the sentencer has discretion to ‘consider the mitigating qualities of youth’ and impose a lesser punishment,” any juvenile convicted of homicide may be sentenced to LWOP, even if his or her crime reflects transient immaturity.⁸⁶

While the Court maintained that it did not overrule *Miller*, it is a troubling step away from a line of cases making clear that

78. *Id.*

79. *Id.*

80. 543 U.S. at 572.

81. 560 U.S. at 74.

82. 567 U.S. at 465.

83. *See* 141 S. Ct. 1307 (2021).

84. *Id.* at 1311.

85. *Id.* at 1313.

86. *Id.* at 1314 (quoting *Miller*, 567 U.S. at 476).

rehabilitation is the focus when juveniles are charged with crimes.⁸⁷ Further, the dissent highlighted how prior Supreme Court rulings, including *Miller*, emphasized that “children are constitutionally different from adults for purposes of sentencing.”⁸⁸ Even if we were to accept the idea that some children are “irreparably corrupt”—which is not supported by available medical and psychological evidence⁸⁹—that would indicate a need for rehabilitation rather than punishment given our scientific understanding of behavioral and neurocognitive maturation. But this is not where the *Jones* Court landed. In light of the Court’s apparent course reversal, it is now more important than ever to establish an empirically grounded legal process for handling the G2i problem in juvenile sentencing. The next Part begins to map such a process.

III. BURDENS OF PROOF AND DEVELOPMENTAL MATURITY

While the G2i issue poses barriers to making determinations in a particular case with confidence, that does not mean that the process of making inferences from G to i cannot be guided by rational principles. Science can set the initial presumptions or a priori starting point for evaluating specific cases. Hence, in the eyewitness identification example discussed in Part I, the fact that the witness and perpetrator were of different races, all things being equal, indicates that an identification may be unreliable.⁹⁰ However, rarely are all things equal. Other factors, such as how much time the witness had to observe the individual, the lighting conditions under which the identification was made, the distance between the witness and the individual, among other factors, might indicate otherwise.⁹¹

The general science can provide a framework within which other considerations ought to be taken into account. At the same time, the

87. *See id.* at 1321.

88. *Id.* at 1328 (Sotomayor, J., dissenting) (quoting *Miller*, 567 U.S. at 471).

89. Evidence suggests that most juvenile offenders do not continue such behavior into adulthood, but rather desist as they mature. *See* Terrie E. Moffitt, *Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy*, 100 *PSYCH. REV.* 674, 675 (1993); Edward P. Mulvey et al., *Trajectories of Desistance and Continuity in Antisocial Behavior Following Court Adjudication Among Serious Adolescent Offenders*, 22 *DEV. & PSYCHOPATHOLOGY* 453, 454, 468 (2010).

90. *See* Albright, *supra* note 23, at 7762.

91. *See id.*

stronger the general science, more additional considerations should be needed to say that this case is not an instance of the general phenomenon.⁹² If, for example, 90 percent of cases in which cross-race identification are made result in inaccurate identifications, this would create a strong presumption that the specific identification in question is inaccurate.

An additional and important consideration in deciding whether a particular case is an instance of some general phenomenon is the cost of making a mistake. In general, there are two kinds of errors. False positive errors occur when we mistakenly conclude that a case is an instance of some phenomenon.⁹³ Hence, concluding that an identification was accurate when it was not is a false positive error.⁹⁴ Alternatively, false negative errors occur when we mistakenly conclude that a case is not an instance of some phenomenon.⁹⁵ Hence, concluding that an identification was inaccurate when it was accurate is a false negative error.⁹⁶

Different errors have different consequences. This insight is well-captured by Blackstone's famous statement that it is "better that ten guilty persons escape, than that one innocent suffer."⁹⁷ According to this judgment, false positive convictions are ten times worse than false negative acquittals.⁹⁸

Blackstone's insight is operationalized in the law by the burden of proof employed in criminal trials, which requires proof beyond a reasonable doubt.⁹⁹ The burden of proof in civil cases, in contrast, is a mere preponderance of the evidence,¹⁰⁰ suggesting that we are largely ambivalent between false positive and false negative errors in civil cases. In *Addington v. Texas*, the Supreme Court made this

92. See *id.* at 7762-63.

93. *False Positive*, BLACK'S LAW DICTIONARY (11th ed. 2019).

94. See *id.*

95. *False Negative*, STEDMANS MEDICAL DICTIONARY (2014).

96. See *id.*

97. 4 WILLIAM BLACKSTONE, COMMENTARIES *358. See generally Daniel Epps, *The Consequences of Error in Criminal Justice*, 128 HARV. L. REV. 1065 (2015).

98. See BLACKSTONE, *supra* note 97, at *358.

99. See *Reasonable Doubt*, BLACK'S LAW DICTIONARY (11th ed. 2019).

100. See *Preponderance of the Evidence*, BLACK'S LAW DICTIONARY (11th ed. 2019).

calculation explicit in adopting a clear and convincing evidence standard for civil commitment cases.¹⁰¹ Civil commitment is “quasi-criminal,” in that the defendant stands to lose his or her liberty if committed.¹⁰² At the same time, the *Addington* Court recognized the cost of erring in the other direction would mean that a mentally ill person would not receive treatment.¹⁰³ The Court explained: “In considering what standard should govern in a civil commitment proceeding, we must assess both the extent of the individual’s interest in not being involuntarily confined indefinitely and the state’s interest in committing the emotionally disturbed under a particular standard of proof.”¹⁰⁴ In regard to the consequences of making an error, the Court asserted: “It cannot be said ... that it is much better for a mentally ill person to ‘go free’ than for a mentally normal person to be committed.”¹⁰⁵ This calculation resulted in the adoption of the intermediate burden of clear and convincing evidence, a standard that lies between proof beyond a reasonable doubt and preponderance of the evidence.¹⁰⁶

Burdens of proof, therefore, effectively embody two key considerations, one quantitative and the other normative. The greater the frequency of some general phenomenon, of which a particular case is thought to be an instance, the greater should be the burden to demonstrate that it is not. And the greater the harm that might come from a particular type of error, the greater the burden should be to avoid making that error.

As the Court in *Roper* recognized, selecting eighteen years as the dividing line for developmental maturity was essentially arbitrary.¹⁰⁷ The behavioral and neuroscience literatures bear this out. Maturity, which itself is not a unidimensional construct, occurs over time, with age cohorts varying considerably.¹⁰⁸ The research liter-

101. 441 U.S. 418, 433 (1979).

102. *See id.* at 424. It is worth emphasizing that the “criminal” component of the “quasi-criminal” nature of civil commitment is not based on any culpability of the defendant, but rather the prospect of the defendant’s loss of liberty, which is analogous to the consequence confronted by a criminal defendant.

103. *See id.* at 432.

104. *Id.* at 425.

105. *Id.* at 429.

106. *See id.* at 430-33.

107. *See Roper v. Simmons*, 543 U.S. 551, 574 (2005).

108. *See Steinberg et al.*, *supra* note 2, at 588-91.

ature generally supports the view that brain development continues throughout the lifespan, with dynamic changes taking place during early, middle, and late adolescence through approximately age twenty-five.¹⁰⁹ As different facets of psychological and neural functioning develop along different timelines, there is no scientific evidence of a specific age that humans move from childhood to adulthood.¹¹⁰

Adolescent brains are neither advanced child brains nor immature adult brains—they are specifically tailored to meet the needs of the young-adulthood stage of life.¹¹¹ There is a large and growing body of literature dedicated to better understanding how and which ways children, adolescents, and young adults differ.

Current empirical evidence from the behavioral sciences suggests that adolescents differ from adults and children in three important ways that lead to differences in behavior.

[(1)] adolescents lack mature capacity for self-regulation in emotionally charged contexts, relative to adults and children[;]

109. See COMM. ON IMPROVING THE HEALTH, SAFETY, & WELL-BEING OF YOUNG ADULTS, INST. OF MED. & NAT'L RSCH. COUNCIL, INVESTING IN THE HEALTH AND WELL-BEING OF YOUNG ADULTS 35-42 (2015).

110. Elizabeth S. Scott et al., *Young Adulthood as a Transitional Legal Category: Science, Social Change, and Justice Policy*, 85 *FORDHAM L. REV.* 641, 648 (2016). Recently, “research on adolescent behavior has been increasingly influenced by ... perspectives on the adolescent brain that emphasize the different developmental trajectories of brain systems that govern incentive processing and cognitive control.” Laurence Steinberg et al., *Around the World, Adolescence Is a Time of Heightened Sensation Seeking and Immature Self-Regulation*, 21 *DEVELOPMENTAL SCI.*, No. 2, 2018, at 2. Accordingly, “risky behavior in adolescence is the product of the interaction between changes in two distinct neurobiological systems: a ‘socio-emotional’ system ... and a ‘cognitive control’ system.” Laurence Steinberg, *A Dual Systems Model of Adolescent Risk-Taking*, 52 *DEVELOPMENTAL PSYCHOBIOLOGY* 216, 216 (2010). The maturation of these systems is:

gradual, unfolds over the course of adolescence, and permits more advanced self regulation and impulse control. The temporal gap between the arousal of the socioemotional system, which is an early adolescent development, and the full maturation of the cognitive control system, which occurs later, creates a period of heightened vulnerability to risk-taking during middle adolescence.

Id. Some researchers have described this imbalance as akin to starting a car’s engine before a well-functioning brake system is in place. See, e.g., Monica Payne, “*All Gas and No Brakes!*”: *Helpful Metaphor or Harmful Stereotype?*, 27 *J. ADOLESCENT RSCH.* 3 (2012).

111. See COMM. ON ASSESSING JUV. JUST. REFORM & COMM. ON L. & JUST., NAT'L RSCH. COUNCIL, REFORMING JUVENILE JUSTICE: A DEVELOPMENTAL APPROACH 89-91 (2013).

[(2)] adolescents have a heightened sensitivity to proximal external influences, such as peer pressure and immediate incentives, relative to adults[; and]

[(3)] adolescents show less ability to make judgments and decisions that require future orientation.¹¹²

Undoubtedly, some percentage at each age might be said to be “mature” for legal—that is, Eighth Amendment—purposes, possibly beginning younger than eighteen. That percentage increases with age, so that, on average, the percentage of nineteen-year-olds that are “mature” is greater than the percentage of eighteen-year-olds, and a greater percentage of twenty-year-olds are mature compared to nineteen-year-olds, and so on through age twenty-five.¹¹³ This insight, drawn from a rich developmental literature, suggests that burdens of proof might very well establish threshold standards for factual judgments about maturity. Simply put, the younger the defendant, the greater should be the prosecution’s burden to show maturity.

Additionally, as discussed above, there is a normative component to burdens of proof. The greater the costs of making a mistake, the greater should be the burden established against it.¹¹⁴ In the context of sentencing, and especially in the context of youthful offenders, the prospect of making a false positive error of too long a sentence is considerable.¹¹⁵ Of course, the defendants in these cases have been convicted of serious crimes, so false negative errors are problematic as well.¹¹⁶ On balance, although we believe that the costs associated with false positive errors in sentencing substantially outweigh those associated with false negatives, reasonable people can disagree. However, on the empirical question of developing

112. *Id.* at 91 (citing Leah H. Somerville et al., *Behavioral and Neural Representation of Emotional Facial Expressions Across the Lifespan*, 36 DEVELOPMENTAL NEUROPSYCHOLOGY 408, 420-23 (2011)); Bernd Figner et al., *Affective and Deliberative Processes in Risky Choice: Age Differences in Risk Taking in the Columbia Card Task*, 35 J. EXPERIMENTAL PSYCH. 709, 710-11 (2009); Steinberg, *supra* note 52, at 67-70).

113. See Steinberg et al., *supra* note 2, at 588-91.

114. See BLACKSTONE, *supra* note 97, at *358.

115. See COMM. ON ASSESSING JUV. JUST. REFORM & COMM. ON L. & JUST., *supra* note 111, at 129-30, 186-87.

116. See Epps, *supra* note 97, at 1068-70.

maturity over age-cohorts, there is considerably less room for disagreement. In the following Part, therefore, we propose a regime of shifting burdens of proof for sentencing youthful offenders primarily on the quantitative base-rates for maturity in each cohort.¹¹⁷

IV. EMPLOYING BURDENS OF PROOF FOR SENTENCING YOUTHFUL OFFENDERS

Although the associated subjects of presumptions, burdens of production, and burdens of proof can be complex, for present purposes we focus only on the two related concepts of presumptions and burdens of proof—with the latter sometimes referred to as burdens of persuasion.¹¹⁸ There are two basic types of presumptions: conclusive and rebuttable.¹¹⁹ A conclusive presumption is essentially a rule of law.¹²⁰ The *Roper* Court created a conclusive presumption in holding that a defendant who committed a capital crime before turning eighteen is too immature to be sentenced to death.¹²¹ In the framework of presumptions, finding the defendant to be under eighteen when the crime occurred means that the Eighth Amendment forbids imposing a sentence of death on him.¹²²

A rebuttable presumption is true to its name and can be rebutted with adequate proof.¹²³ For example, the presumption that a person missing for seven years is dead can be rebutted with evidence that he is, in fact, alive. The question of what level of proof is adequate to rebut a presumption can be answered differently and is best

117. To the extent that courts or policymakers disagree regarding the balance of normative factors between false positives and false negatives, they might adjust the proposed burdens of proof accordingly. Such disagreement does not undermine the wisdom, based on solid empirical grounds, to employ burdens of proof in these cases.

118. See GEORGE E. DIX ET AL., MCCORMICK ON EVIDENCE 572 (Kenneth S. Broun ed., 6th ed. 2006) (observing that presumptions are the “slipperiest member of the family of legal terms,” except perhaps for “burden of proof”).

119. See *id.* at 572-73.

120. See *id.*

121. See *Roper v. Simmons*, 543 U.S. 551, 574 (2005). The Court created a similar conclusive presumption in *Graham*, holding that someone who committed a non-homicide offense prior to turning eighteen was too immature to be sentenced to LWOP. See *Graham v. Florida*, 560 U.S. 48, 79 (2010).

122. See *Roper*, 543 U.S. at 574.

123. See DIX ET AL., *supra* note 118, at 572-73.

understood as involving burdens of proof, or burdens of persuasion.¹²⁴ Hence, a criminal defendant is presumed innocent, but that presumption can be overcome with proof beyond a reasonable doubt.¹²⁵

The Supreme Court established the conclusive presumption of immaturity at age eighteen in death penalty and non-homicide LWOP cases.¹²⁶ Yet nothing magical occurs at the moment when someone turns eighteen in regard to brain development. The general phenomenon of developmental maturity is complex and ultimately probabilistic.¹²⁷ People and their brains develop over time and not uniformly.¹²⁸ So few youthful offenders under eighteen are “mature” for Eighth Amendment purposes that the entire cohort is conclusively deemed immature. As offender ages increase, the percentages of offenders that are considered “mature” changes—for example, the vast number of twenty-five-year-olds are “mature” for Eighth Amendment purposes.¹²⁹

This empirical reality suggests a straightforward and rational solution to sentencing youthful offenders. The burden of proof should relate to the base rate of maturity in respective age cohorts between eighteen and twenty-five. Under eighteen, as held in *Roper* and *Graham*, defendants are conclusively presumed to be immature for Eighth Amendment purposes in death penalty and non-homicide LWOP cases.¹³⁰ We would suggest the following burdens of proof thereafter.

124. See DOUGLAS WALTON, BURDEN OF PROOF, PRESUMPTION AND ARGUMENTATION 8 (2014) (“[W]hen burden of proof and presumption are linked together, they function as evidential devices that are useful and even necessary when dealing with defeasible arguments that need to be used under conditions of uncertainty and lack of knowledge.”).

125. Douglas Walton, in his excellent treatment of the subject, explains the insight of burdens of proof as follows:

Generally speaking, the burden of proof tells you how strong an argument needs to be in order to be successful. It represents a description of a task such that if you fail to carry out this task, your argument will fail. Burden of proof rests on the prior notion that there can be different standards of proof appropriate for different contexts of argumentation.

Id. at 8-9.

126. See *Roper*, 543 U.S. at 574; *Graham*, 560 U.S. at 79.

127. See Steinberg et al., *supra* note 2, at 590-93.

128. See *id.*

129. See *id.*; *Graham*, 560 U.S. at 79; *Roper*, 543 U.S. at 574.

130. See *Roper*, 543 U.S. at 574; *Graham*, 560 U.S. at 79.

For defendants who have committed crimes under the age of eighteen, the prosecution should have the burden to prove maturity beyond a reasonable doubt.¹³¹ For defendants accused of committing crimes between eighteen and twenty, the prosecution should have the burden to prove maturity by clear and convincing evidence; for defendants accused of committing crimes between twenty-one and twenty-three, the prosecution should have the burden to show maturity by a preponderance of evidence; and for those defendants accused of committing crimes between twenty-four and twenty-six, the defendant should have the burden to demonstrate immaturity by clear and convincing evidence. After the defendant turns twenty-six, he or she should be presumed mature, though he or she could provide proof to the contrary. For a visual representation of our proposal, see Table 1 below.

Table 1. Proposal for Shifting Burdens of Proof/Presumptions for Competency of Youthful Offenders

Age at commission of offense	Who bears the burden?	Evidentiary standard	To prove?
Under 18*	State/ Government	Beyond a reasonable doubt	That defendant is developmentally mature
18-20		Clear & convincing evidence	
21-23		Preponderance of the evidence	
24-26	Defendant	Clear & convincing evidence	That defendant is not developmentally mature
Over 26		Beyond a reasonable doubt	
*Except in death penalty or non-homicide LWOP cases, where <i>Roper</i> and <i>Graham</i> impose a conclusive presumption—that is, a rule of law—that the defendant is not developmentally mature.			

131. As noted in Part II, although *Roper* and *Graham* hold categorically that minors cannot be sentenced to death in homicide cases or LWOP in non-homicide cases, the defendant's maturity remains an issue for determining what sentence is appropriate under those ceilings.

At sentencing, therefore, we propose that the sentencer be instructed that when a defendant was under twenty-six when he or she committed the crime, developmental maturity must be determined separately according to the burdens of proof outlined above. Of course, many factors go into sentencing decisions beyond maturity level at the time the crime was committed.¹³² A host of mitigating and aggravating circumstances are considered in determining the appropriate sentence for the crime committed.¹³³ However, as pointed out in the several Supreme Court decisions considered above, youthfulness and the notion of incorrigibility ought to be essential considerations in the decision.¹³⁴

Finally, it should be noted that although the empirical literature can help establish the levels of proof needed to determine maturity of a particular defendant, the assessment under that umbrella is likely to be largely the product of some narrative about the life and times of the defendant and his or her amenability to rehabilitation, either through natural maturation or education. The research, as Justice Kennedy noted in *Roper*, simply cannot distinguish whether an individual defendant's crime was a product of "transient immaturity" or "irreparable corruption."¹³⁵ Proof on this issue is likely to be nonempirical and narrative in content.¹³⁶

For instance, in a particular case, the prosecution might point out the prior intricate planning and the subsequent cover-up of the crime to prove maturity; the defense might point out the extreme

132. See generally Thomas Grisso & Antoinette Kavanaugh, *Prospects for Developmental Evidence in Juvenile Sentencing Based on Miller v. Alabama*, 22 PSYCH. PUB. POL'Y & L. 235 (2016).

133. See *id.*; Steinberg et al., *supra* note 2, at 585.

134. See *supra* Part II; see also Steinberg et al., *supra* note 2, at 585.

135. See *Roper*, 543 U.S. at 573.

136. Although the sentencing decision is likely to be primarily based on normative and nonempirical considerations, the general research literature in developmental psychology might yet provide clinical insights regarding individual offenders. See Grisso & Kavanaugh, *supra* note 132, at 246 ("Developmental science has successfully provided the research evidence that the law needed to make its normative decisions about juveniles' lesser maturity and culpability. We now face the task of creating models and methods to provide relevant developmental and clinical data about individuals in cases involving *Miller* sentencing and resentencing."). Nonetheless, the G2i problem will remain a substantial stumbling block to the reliable application of group data to individual defendants. See Carl E. Fisher et al., *Toward a Jurisprudence of Psychiatric Evidence: Examining the Challenges of Reasoning from Group Data in Psychiatry to Individual Decisions in the Law*, 69 U. MIA. L. REV. 685, 688 (2015).

gullibility and marginal IQ of the defendant to prove immaturity. The ultimate decision is for sentencers, but, as recommended here, they should be guided by burdens of proof that are informed by the base rates for the defendant's age cohort.

V. THE SENTENCING DECISION

As discussed above, the Supreme Court confused two factors in its cases that consider developmental maturity and Eighth Amendment standards.¹³⁷ The principal basis for treating children differently from adults was the empirical finding regarding their levels of immaturity.¹³⁸ Such developmental immaturity is inconsistent with the twin justifications for punishment under the Eighth Amendment: culpability and deterrence.¹³⁹ A developmentally immature child is less culpable for his offense and less likely to be deterred by punishment.¹⁴⁰ Hence, in *Roper* and *Graham*, in which the focus was on the demands of the Eighth Amendment, the principal consideration involved the contemporaneous level of maturity of the youthful offender.¹⁴¹ However, always looming in the background was the separate question of sentencing. If the Eighth Amendment does not permit the death penalty for juvenile murderers or LWOP for non-homicide juvenile offenders, then what sentences should they receive?

Whereas in considering the outer limits of what sentence might be imposed involves an Eighth Amendment analysis, sentencing within those limits might consider a vast number of factors, both mitigating and aggravating.¹⁴² Most of these factors come in narrative form, since they are either not amenable to scientific study or have not been studied extensively.¹⁴³ Sentencing factors might include the defendant's level of remorse, family support network, level of education, and so forth.¹⁴⁴ The question of the defendant's

137. See *supra* Part II.

138. See *supra* Part II.

139. See *Roper*, 543 U.S. at 571.

140. See *id.* at 571-72.

141. See *id.*; *Graham v. Florida*, 560 U.S. 48, 79 (2010).

142. See generally Grisso & Kavanaugh, *supra* note 132.

143. See *id.*

144. See generally *id.*

amenability for rehabilitation will inevitably be part of this narrative.¹⁴⁵

Our proposal to use burdens of proof, therefore, is intended to set the floor on which the ultimate sentencing decision will be made. It involves the pivotal and principal issue of the constitutional justification for punishing the youthful offender. This is an essential first consideration for sentencers and it should be guided, to the extent possible, by the scientific literature. Inevitably, sentencers must make an individual decision about the defendant and this will involve not only considerations of maturity but also sundry other aggravating and mitigating considerations. Ultimately, the sentencing decision will be a product of competing narratives offered by the defense and the prosecution, in a way symbolized conventionally by the scales of justice. The burden of proof simply puts a thumb on the scales, depending on the age cohort to which the defendant belongs.

CONCLUSION

In a series of decisions, the Supreme Court has limited the ability of courts to impose extreme sentences on juvenile offenders based on a substantial research literature demonstrating that juveniles are less culpable than adults due to their lack of maturity, susceptibility to outside influences, and their under-formed characters. The Court, however, has maintained that a youthful offender is still capable of committing offenses deserving of punishments such as LWOP.¹⁴⁶ Such an assessment as to whether one is “developmentally mature enough” to be sentenced to LWOP or other extreme sentences requires a bright-line determination that is fundamentally at odds with an established body of research that juvenile and young-adult development occurs gradually through approximately age twenty-six. As it is extremely difficult, if not impossible, to determine an individual’s level of maturity, let alone to predict their capability for rehabilitation decades in the future, the risk of error

145. *See id.*

146. *See Graham*, 560 U.S. at 96 (Roberts, C.J., concurring) (“Some crimes are so heinous, and some juvenile offenders so highly culpable, that a sentence of life without parole may be entirely justified under the Constitution.”); *Roper*, 543 U.S. at 572.

is extremely high and the associated cost considerable. In this Article, we outline a framework of shifting burdens of proof grounded in the scientific literature that can be employed to allocate this risk of error. This approach would better reflect the well-established science on the trajectory of developmental maturity from childhood through young adulthood while still appreciating that courts require a semblance of a bright-line rule to maintain consistency and predictability.