From:	Jacob Schuman
To:	Public Comment
Subject:	Comment on Proposed Changes to Drug Quantity Table
Date:	Tuesday, March 18, 2014 1:19:25 PM
Attachments:	Probability and Punishment by Jacob Schuman.pdf

Hello,

My name is Jacob Schuman and I am a currently a law clerk for Judge James E. Boasberg on the District Court for the District of Columbia. I am sending this email in response to the Commission's request for public comment on its proposed changes to the Drug Quantity Table in § 2D1.1 of the United States Sentencing Guidelines.

The Commission asked for public comment about changes it should make to the Drug Quantity Table. I recently published a law review article on this very subject, in which I suggest that the Guidelines, especially those for drug-trafficking crimes, should do more to incorporate probability into recommended sentences.

In short, my article demonstrates that because the Guidelines only require a fact be proved by a "preponderance of the evidence" in order to trigger an offense level increase/decrease, they recommend sentences that are both inefficient and unfair. This is a particular problem in drug sentencing, where, as my article shows, the Guidelines often recommend lengthy sentences based on drug quantity calculations that carry a high risk of error. My article suggests changes to the Guidelines that would address this problem by incorporating probability into punishment.

I have attached a copy of my article to this email. It will be published by The New Criminal Law Review in the coming months. I hope this comment is helpful for you. Thank you very much for your time and consideration.

Best,

Jacob Schuman

--
Jacob Schuman
Harvard Law School | J.D. 2012
Brown University | A.B. 2008

Probability and Punishment:
How to Improve Sentencing by Taking Account of Probability

Jacob Schuman*

INTRODUCTION

Imagine two defendants, A and B, who have each been convicted of drug trafficking. Defendant A was caught with 1,000 grams of crack cocaine. Defendant B was only caught with 100 grams of crack cocaine, but he also had a large sum of cash, which he probably – though not certainly – earned by selling an additional 900 grams of crack just before his arrest. When the time comes for sentencing, should A and B receive the same punishment?

The federal criminal justice system says that they should.[1] This Article will argue that they should not. The probability that A trafficked 1,000 grams of drugs is higher than the probability that B did, so B deserves the lighter sentence.

Calls for sentencing reform – especially drug sentencing reform – are growing louder. One so-far overlooked way to improve the efficiency and fairness of the criminal justice system is to vary punishments based on the probability of the underlying facts. Although probability estimations regarding past events are fundamental to the administration of criminal justice, no scholar has ever examined the role that probability plays in sentencing, nor has anyone ever explored how decision-makers in the justice system can account for their level of certainty when they impose punishment.

The United States Sentencing Commission, for example, recently sought comment on a proposal to make drug sentencing less punitive by reducing the recommended sentences associated with trafficking various quantities of drugs.[2] This Article will show that, beyond simply reducing all prison sentences for drug offenders, drug sentences could be made shorter, fairer, and more efficient by varying the punishment imposed based on the

---

* Jacob Schuman is a Law Clerk for Judge James E. Boasberg on the U.S. District Court for the District of Columbia. He received his J.D. from Harvard Law School in 2012. He thanks Judge Boasberg, Geoffrey Derrick, Daniel Hemel, Professor Roger J.R. Levesque, Jeff Love, and Caroline Van Zile for their helpful comments, which vastly improved the article.

[1] *Compare* United States v. Lucas, 282 F.3d 414, 417-18 (6th Cir. 2002), *overruled on other grounds by* United States v. Leachman, 309 F.3d 377 (6th Cir. 2002) (defendant caught with 595.8 grams of crack-cocaine received 210-month sentence), *with* United States v. Gardner, 417 F.3d 541, 543 (6th Cir. 2005) (defendant caught with 72 grams of crack-cocaine and $16,000 cash, believed to represent proceeds from sale of an additional 598.74 grams of crack-cocaine, received 210 month sentence). In both cases, the defendants' sentences were also enhanced for their possession of firearms. *See Gardner*, 417 F.3d at 543; *Lucas*, 282 F.3d at 418.

[2] *See* U.S. Sentencing Commission, News Release: U.S. Sentencing Commission Seeks Comment on Potential Reduction to Drug Trafficking Sentences 1, available at http://www.ussc.gov/Legislative_and_Public_Affairs/Newsroom/Press_Releases/20140109_Press_Release.pdf. This is a rather dramatic oversimplification of the reform proposed by the Commission, which is explored in greater detail below. *See supra*, Part VI.C.3.

probability that the offender trafficked a particular quantity of drugs. This is a reform, moreover, that district courts can already begin to implement using their sentencing discretion,[3] while policymakers consider more systemic changes, several of which are suggested later on below.[4]

There is a special relationship between probability and punishment because the criminal justice system is inherently fallible. A trial can never determine with absolute certainty that an accused defendant committed a particular crime – some margin of doubt will always remain. To render judgment, therefore, the criminal law must estimate the probability that each defendant is guilty of the offense charged and then translate that probability into specific penal consequences.

As this Article will explain, there is more than one way to translate probability into punishment. The guilt stage of criminal proceedings – the criminal trial – places little emphasis on probability. Trials use a "threshold model" of decision-making, in which the prosecution convicts the defendant by establishing that the likelihood that he committed the crime charged exceeds a certain "threshold" level of probability. If the jury believes that it is "beyond a reasonable doubt" that the defendant did the deed – a level of proof typically quantified as 95% probability – it will return a guilty verdict. If not, then the defendant will walk free. Neither outcome will reflect a precise measure of the odds of the defendant's guilt.

The threshold model of conviction is so basic to American criminal justice that it may seem inevitable. But in fact, there is another way to translate probability estimations into punishment: a "probabilistic model" of decision-making. The probabilistic model places far more emphasis on probability, directly incorporating it into legal outcomes. A probabilistic model of conviction, for example, would vary the outcome of each trial based on the probability of the defendant's guilt. If the defendant were more likely to be guilty, he would receive a harsher verdict; if he were less likely, he would get a lighter one; and if he were almost certainly innocent, he would be exonerated.

This Article begins with the as-yet unappreciated observation that the penalty stage of criminal proceedings – the sentencing hearing – also uses a "threshold model" of decision-making that largely ignores probability. The United States Sentencing Guidelines instruct federal district judges to make a series of factual findings related to the offender and his offense, which either add to or subtract from an ultimate recommended sentence. Just like with the threshold model of conviction, a sentencing judge determines the applicability of these sentence adjustments by deciding whether it is "more likely than not" – more than 50% likely – that the factual predicate for an adjustment has been fulfilled. If it is, then that adjustment will apply in full. Otherwise, it will not. Once again, neither outcome will reflect the actual probability that the sentence adjustment is appropriate.

---

[3] *See infra* Part VI.B.
[4] *See infra* Part VI.C.

This Article then breaks new ground by demonstrating that the justifications for the threshold model of conviction do not hold up at sentencing. Moreover, the two flaws identified with the threshold model of conviction – inefficiency and unfairness – are not only present at sentencing, but in fact are exacerbated by several unique features of this stage of the proceedings.

Finally, this Article applies these insights to drug sentencing, and demonstrates that the threshold model of sentencing is especially problematic when it comes to determinations of drug quantity. Under the Sentencing Guidelines, drug offenders receive longer sentences if they trafficked in larger quantities of contraband. But district judges often must rely on extrapolation and inference to make such findings. As a result, courts frequently mete out lengthy prison terms based on quantity determinations that carry a high risk of error. This Article will argue that courts and policymakers should mitigate the inefficiencies and injustices that result from these fact-findings by incorporating probability into drug quantity determinations at sentencing.

I. THE THRESHOLD MODEL OF CONVICTION

This Part will show that a criminal trial can never determine with absolute certainty whether a particular defendant committed a particular crime, which means judges and juries can only estimate the probability of a defendant's guilt when they render judgment upon him. There are two ways that scholars have identified to translate these probability estimations into trial outcomes. The "threshold" model turns on a single probability threshold, while the "probabilistic" model incorporates many levels of probability. Federal criminal trials, for example, use a "threshold model" of conviction – the jury or judge will convict the defendant if the probability that he committed the crime charged is "beyond a reasonable doubt." This burden of proof is typically quantified as 95% likelihood of guilt. As an alternative to the threshold approach, scholars have proposed a "probabilistic" model of conviction that would use multiple trial outcomes to more precisely approximate the probability of the defendant's guilt.

A. Threshold and Probabilistic Models of Decision-Making

1. The Impossibility of Absolute Certainty

"Certainty, absolute certainty, is a satisfaction which … we are continually grasping at," lamented Jeremy Bentham in his *Rationale of Judicial Evidence*, "but which the inexorable nature of things has placed forever out of reach."[5] Over two centuries later, "absolute certainty" remains outside the grasp of the criminal law. No judge or jury can ever know with absolute certainty that a defendant committed the criminal act of which he is accused. A specter of doubt haunts every verdict, even if it is an unreasonable, or a fantastical, doubt.

[5] 5 JEREMY BENTHAM, RATIONALE OF JUDICIAL EVIDENCE 351 (J.S. Mill ed. 1827).

No matter how strong the prosecution's evidence, a clever criminal defense attorney can always find cracks in the case against his client. Imagine, for example, that Brutus is on trial for murdering Caesar. The prosecution might offer the testimony of two eyewitnesses who say that Brutus did the deed. Yet Brutus's lawyer, in response, can argue that both witnesses have misremembered and mistaken his client for the real killer. Perhaps the government will present DNA evidence linking Brutus to the crime scene. Yet there is also the chance that the blood samples were accidentally switched in the lab. Maybe the police even physically arrested Brutus at the scene of the crime, knife in hand. Still, at trial, his attorney might argue that Brutus had been framed by law enforcement.[6]

In short, whenever the prosecution argues that the evidence before the court reflects a certain narrative about the past, the defendant can always present a counter-narrative – an alternative possible story. That alternative may be quite implausible, but it will always enjoy some degree of likelihood, no matter how slim.

2. Two Ways to Translate Practical Certainty

Even though absolute certainty is unattainable in the courtroom, legal fact-finders can still approach the bounds of 100% confidence in their conclusions. In other words, while a judge can never know that a fact about the past is 100% likely to be true, she can estimate that it is 51% likely to be true, or 95% likely, or maybe even 99% likely.[7] Indeed, the legal system has codified certain levels of probability as standards of proof: "proof by a preponderance of the evidence" is just over 50% probability of truth, "proof by clear and convincing evidence" is roughly equal to 70% probability of truth, and "proof beyond a reasonable doubt" is commonly quantified as 95% probability of truth.[8] This is the "[p]ractical certainty" in which Bentham sought solace, "a degree of assurance sufficient for practice … the attainment of which … may be sufficient to console us under the want of any … superfluous and unattainable acquisitions."[9]

The impossibility of absolute certainty raises a fundamental question of legal epistemology: When a judgment turns on a specific fact about the past, how should the justice system translate the probability of that fact's truth into legal consequences? In other words, how likely must it be that Brutus killed

---

[6] Professor Eugene Volokh makes a similar version of this argument when he speculates that in criminal cases involving a "word against word situation," the "omnipresent" risk that the prosecution's witness is lying may actually make fact-finders *more* likely to convict. Eugene Volokh, *Wrongful Convictions and Proof Beyond a Reasonable Doubt*, THE VOLOKH CONSPIRACY (Jan. 2, 2014; 1:11 PM), http://www.volokh.com/2014/01/02/wrongful-convictions-proof-beyond-reasonable-doubt/.

[7] *See* MAGUIRE, ET AL., CASES AND MATERIALS ON EVIDENCE 1 (6th ed. 1973).

[8] *See, e.g.*, United States v. Fatico, 458 F. Supp. 388, 403-06 (E.D.N.Y. 1978) (Weinstein, J.). The conversion of standards of proof into levels of probability is sometimes controversial – for more on the quantification of the beyond-a-reasonable-doubt and preponderance-of-the-evidence standards, see *infra* notes 14 and 35.

[9] BENTHAM, *supra* note 5, at 351.

Caesar in order for the criminal justice system to hold him responsible for the crime? And what relevance should that probability judgment have for the severity of the punishment he receives?

Scholars have identified two possible answers to these questions: the "threshold" model of decision-making and the "probabilistic" model of decision-making.[10] What follows is a brief, general explanation of how these models work in theory. It will help inform the more concrete discussion of how they work at trial and at sentencing later on.

According to the threshold model of decision-making, only one "threshold" level of probability matters. A court using the threshold model decides that a fact about the past is either true or untrue based on whether the probability of its truth exceeds a specific level of likelihood. If the probability does exceed that specific "threshold" level, then the court will declare the fact "true." The court will then apply all the legal consequences for the truth of that fact. If the probability does not cross the threshold, then the court will decide that the fact is "untrue," and no consequences will follow. The threshold model is therefore "all-or-nothing."[11] It declares facts about the past to be either true or not true, and the consequences for those facts either apply in full or not at all.

By contrast, the "probabilistic" model of decision-making incorporates the probability of a fact's truth *into* the application of its legal consequences. Under the probabilistic model, the court estimates the odds of a fact's truth across an open range of probabilities. The court then imposes legal consequences in proportion to the probability that the fact is true. As the probability of the fact's truth increases, the consequences applied will also increase. As the probability decreases, so too will the consequences. The probabilistic model therefore never definitively decides one way or the other whether a fact about the past is "true" or "untrue." Instead, the probability that a fact is true determines the magnitude of its legal consequences.

B. Threshold and Probabilistic Models of Conviction

1. The Threshold Model of Conviction

Federal criminal trials use a threshold model of decision-making. Outcomes are limited to two verdicts: guilty or not guilty.[12] The presumption

---

[10] *See, e.g.*, Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254 (2013); Talia Fisher, *Conviction Without Conviction*, 96 MINN. L. REV. 833, 834-35 (2012); Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385, 398-404 (1985); Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1361-62 (1985); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV L. REV. 1329 (1971).

[11] Fisher, *supra* note 10, at 834-35.

[12] *See* Samuel Bray, *Not Proven: Introducing a Third Verdict*, 72 U. CHI. L. REV. 1299, 1299 (2005). There are a few exceptions on the margins, such as "not guilty by reason of insanity," but by and large the standard criminal case is limited to the guilty/not guilty binary. *See id.* at 1299 n.4.

of innocence sets the default outcome at "not guilty," but that will switch to "guilty" if the prosecution persuades a jury that it is "beyond a reasonable doubt" that the defendant committed the crime charged.[13] This burden of proof is, in essence, a threshold level of probability, usually quantified as 95% likelihood that the defendant did the deed.[14] Proof above or below that specific degree of certainty is irrelevant. If the odds that the defendant is guilty exceed the 95% threshold of probability, then he will be convicted. If they are 94% or less, then he will be "categorically acquitted."[15]

Criminal liability under the threshold model of conviction is "all or nothing."[16] Once the case against the defendant crosses the threshold level of certainty, the defendant will receive the same conviction as any other offender who committed the same crime, although in some cases the prosecution will have had a slam-dunk case (100% certainty of guilt) and in others it will have just barely outstripped any reasonable doubts (95% certainty of guilt). And, conversely, if the government does not meet its burden of proof, then the offender will not suffer any legal consequences at all, no matter whether his innocence was obvious (0% certainty of guilt), or only the narrowest sliver of a reasonable doubt remained (94% certainty of guilt).

2. The Probabilistic Model of Conviction

The threshold model of conviction is fundamental to American criminal justice,[17] but it is not the only option. Michel Foucault has shown, for example, that culpability in medieval European law ran along a spectrum. Under the Ancien Régime, "partial[] punish[ment]" could be imposed on a defendant based on partial suspicion:

> The different pieces of evidence did not constitute so many neutral elements, until such time as they could be gathered together into a single body of evidence that would bring the final certainty of guilt. Each piece of evidence aroused a particular degree of abomination…Thus a semi proof did not leave the suspect innocent until such time as it was completed; it made him semi-guilty…In short, penal demonstration did not obey a dualistic system: true or false; but a principal of

[13] *See* United States v. Haudin, 515 U.S. 506, 510 (1995); In re Winship, 397 U.S. 358, 364 (1970).
[14] *See, e.g.*, David Kaye, *Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34, 40 (1979). There is some disagreement over the 95% figure. *See, e.g.*, United States v. Fatico, 458 F.Supp. 388, 410 (E.D.N.Y. 1978) (Weinstein, J); Laurence H. Tribe, *An Ounce of Detention: Preventive Justice in the World of John Mitchell*, 56 VA. L. REV. 371, 388 (1970). This Article, however, does not hang on that quantification, which is merely offered as convenient shorthand for "some very high level of probability below 100%."
[15] Fisher, *supra* note 10, at 834-35.
[16] *Id.*
[17] *Id.* at 835.

> continuous gradation; a degree reached in the demonstration formed a degree of guilt and consequently involved a degree of punishment.[18]

Times have long since changed, but several scholars have called for a return to a non-binary system for criminal verdicts that would account for the probability judgments that underlie every decision to convict or acquit.[19]

A "probabilistic" model of conviction[20] would differ from the threshold model by incorporating the probability of the defendant's guilt into the outcome of each trial. This approach would add a number of new verdicts to the jury's arsenal that would represent waypoints along the probability spectrum – for example, "not proven,"[21] "blameless violation,"[22] "guilty, but not punishable,"[23] or "innocent"[24] – so that the jury could more precisely express its estimation of the likelihood that the defendant committed the crime charged. Later, at the penalty stage of the proceedings, these new verdicts would help determine the severity of the defendant's punishment.[25] Punishment would increase as the jury's confidence in the defendant's guilt rose, and fall as the jury's level of certainty fell. At very low levels of probable guilt, the jury might officially exonerate the accused by expressing a total lack of confidence in his culpability.[26] One scholar has even suggested a return to the medieval model, in which partial punishments would attach to partway levels of guilt,[27] though that would almost certainly violate the Constitution's Due Process Clause by permitting punishment on less than proof beyond a reasonable doubt.[28]

II. THE THRESHOLD MODEL OF SENTENCING

This Part will argue that the penalty stage of criminal proceedings in the federal court system also uses a threshold model of decision-making. Although a number of scholars have already discussed the threshold model of *conviction*, this Article breaks new ground with a critical analysis of the

---

[18] MICHEL FOUCAULT, DISCIPLINE AND PUNISH 42 (Alan Sheridan trans., Vintage Books 2d ed. 1995).

[19] *See, e.g.*, Talia Fisher, *Constitutionalism and the Criminal Law: Rethinking Criminal Trial Bifurcation*, 61 U. TORONTO L.J. 811, 814 (2011); Bray, *supra* note 12, at 1304-1307; Andrew D. Leipold, *The Problem of the Innocent, Acquitted Defendant*, 94 NW. U. L. REV. 1297, 1314-26 (2000); Paul H. Robinson, *Rules of Conduct and Principles of Adjudication*, 57 U. CHI. L. REV. 729, 766-67 (1990).

[20] This label is borrowed from the one used by Professor Fisher for her own "probabilistic model." Fisher, *supra* note 10, at 836.

[21] Bray, *supra* note 12, at 1304-1307.

[22] Robinson, *supra* note 19, at 766-67.

[23] Paul H. Robinson, *Criminal Law Defenses: A Systematic Analysis*, 82 COLUM. L. REV. 199, 290 (1982).

[24] Leipold, *supra* note 19, at 1314-26.

[25] Fisher, *supra* note 19, at 814.

[26] *See, e.g.*, Bray, *supra* note 12, at 1304-1307; Leipold, *supra* note 19, at 1314-26; Robinson, *supra* note 19, at 766-67.

[27] Fisher, *supra* note 19, at 814.

[28] *See* In re Winship, 397 U.S. 358, 362-64 (1970) (quoting LeLand v. Oregon, 343 U.S. 790, 802 (1952) (Frankfurter, J., dissenting)).

threshold model of *sentencing*. The United States Sentencing Guidelines instruct districts court to make a series of factual findings about each offender and his crime, which are then plugged into an equation that calculates a recommended sentence for each offender. This approach reflects a threshold model of decision-making: the Guidelines require district courts to make a series of all-or-nothing judgments about whether certain facts are either true or untrue, and each judgment turns on whether the probability of each fact's truth exceeds a certain threshold level of likelihood. There are, of course, some differences between trials and sentencing hearings, but fundamentally, both rely on a threshold approach.

A. The Federal Law of Sentencing

1. The Sentencing Hearing

Once a defendant has been convicted at trial, he is subject to a penalty specified by statute. Federal criminal statutes usually provide for fines along with a broad range of possible prison terms. For instance, the punishment for physically assaulting a federal officer is a fine and "imprison[ment] of not more than 20 years," or both.[29] At a sentencing hearing held after the trial, both the prosecution and the convicted defendant have the chance to argue for an appropriate sentence from within that range.[30] A district court judge will then make the final decision. Before she does so, however, the judge must consider a list of factors prescribed by Congress, including the sentence recommended by the United States Sentencing Guidelines. These factors are discussed in greater detail below.[31]

It is important to remember that when a federal judge sentences a defendant, she considers his "real offense," rather than his "charged offense."[32] What this means is that the sentencing judge makes her own findings of fact about what the offender "really" did – she is not limited to the allegations listed in the indictment or proved to the jury at trial. In fact, sentencing courts can even consider criminal conduct for which the offender was specifically acquitted.[33]

It is also important to remember that a sentencing hearing is governed by rules of evidence very different from those at trial. One significant change is that the standard of proof at sentencing is much lower than it is at trial – a "preponderance of the evidence" rather than proof beyond a reasonable doubt.[34] In other words, for a fact to be established at sentencing, it must be "more likely than not" to be true, a level of proof quantified as just over 50%

[29] *See* 18 U.S.C. § 111; *id.* at § 1114.
[30] *See* Gall v. United States, 552 U.S. 38, 49 (2007); FED. R. CRIM. P. 32(i)(4).
[31] *See supra*, Section II.A.2.
[32] U.S.S.G. ch. ONE, Pt. A, Sbpt 1(4)(a); *see also* Deborah Young, *Fact-Finding at Federal Sentencing: Why the Guidelines Should Meet the Rules*, 79 CORNELL L. REV. 299, 323 (1994).
[33] *See* United States v. White, 551 F.3d 381, 383-84 (6th Cir. 2008) (collecting cases).
[34] *See* McMillan v. Pennsylvania, 477 U.S. 79, 84-91 (1986); Young, *supra* note 32, at 335-38.

probability of truth.[35] The standards for the admissibility of evidence are also less restrictive at sentencing than they are at trial, so that, for example, hearsay evidence is admissible,[36] as is evidence seized in violation of the Fourth Amendment.[37]

2. The § 3553(a) Factors

In Title 18, Section 3553(a) of the U.S. Code, Congress provides federal district courts with a list of factors that they must consider before sentencing a convicted defendant.[38] Those factors are:

> (1) the nature and circumstances of the offense and the history and characteristics of the defendant;
> (2) the need for the sentence imposed--
> > (A) to reflect the seriousness of the offense, to promote respect for the law, and to provide just punishment for the offense;
> > (B) to afford adequate deterrence to criminal conduct;
> > (C) to protect the public from further crimes of the defendant; and
> > (D) to provide the defendant with needed educational or vocational training, medical care, or other correctional treatment in the most effective manner; [and]
> (3) the kinds of sentences available;
> (4) the kinds of sentence and the sentencing range … set forth in the guidelines.[39]

Translated into plain English, § 3553(a)(1) instructs courts to individually review the unique circumstances of the offender and his crime. Next, §§ 3553(a)(2)-(4) says that courts must consider the retributive, deterrent, incapacitative, and rehabilitative theories of punishment.[40] After that, § 3553(a)(3) directs courts to take into account all the possible sentences

---

[35] *See* Vern R. Walker, *Preponderance, Probability, and Warranted Factfinding*, 62 BROOK. L. REV. 1075, 1076 n.5 (1996) (collecting sources). In contrast to the debate over quantifying the beyond-a-reasonable-doubt standard, *see* supra note 14, it seems difficult to dispute that the preponderance standard is equivalent to 50% likelihood. Still, Professor Walker has argued on policy grounds against quantification even of the preponderance standard. *See generally* Walker, supra. But even from Professor Walker's perspective, the argument in this Article will still hold if one simply uses the 50% figure as shorthand for "some level of probability below proof beyond a reasonable doubt." *Cf.* supra, note 14.

[36] *See* Williams v. Oklahoma, 358 U.S. 576, 584 (1959).

[37] *See* United States v. Brimah, 214 F.3d 854, 857 (7th Cir. 2000) (collecting cases).

[38] *See* 18 U.S.C. § 3553.

[39] *See id.*

[40] *See* Mistretta v. United States, 488 U.S. 361, 367 (1989).

available. Finally, and most importantly,[41] § 3553(a)(4) instructs courts to consider the sentence recommended by the United States Sentencing Guidelines, discussed in greater detail in the next section.

3. The United States Sentencing Guidelines

The United States Sentencing Guidelines are published annually by the United States Sentencing Commission, an agency created by Sentencing Reform Act of 1984. The Guidelines recommend a standardized sentence based on the specific facts of each case in order to "eliminate wide disparity" in the punishments imposed on similarly situated offenders.[42]

At the outset, the Guidelines assign each generic crime a different "base offense level." Two offenders convicted of the same crime will therefore start at the same offense level. For example, the Sentencing Guidelines assign the crime of aggravated assault a base offense level of 14.[43] The district judge can then adjust that level according to a specific set of rules tied to the unique characteristics of the particular act of wrongdoing at issue.[44] At the end of the process, therefore, two defendants convicted of the same offense may have very different offense levels, depending on the facts of their particular cases. For instance, the base level of 14 for aggravated assault can be adjusted as follows:

> (1) If the assault involved more than minimal planning, increase by 2 levels.
> (2) If (A) a firearm was discharged, increase by 5 levels; (B) a dangerous weapon (including a firearm) was otherwise used, increase by 4 levels; (C) a dangerous weapon (including a firearm) was brandished or its use was threatened, increase by 3 levels…
> (4) If the assault was motivated by a payment or offer of money or other thing of value, increase by 2 levels…[45]

These are just a small sample. The aggravated assault Guideline also provides for adjustment of the offense level depending on the injury suffered by the

[41] *See* Gall v. United States, 552 U.S. 38, 49 (2007).
[42] U.S.S.G. Ch. ONE, Pt. A, Subpt. 1(g); *see also* 28 U.S.C. § 991(b)(1)(B)(1988); *Mistretta*, 488 U.S. at 362.
[43] *See* U.S.S.G. § 2A2.2.
[44] The base offense level itself may also depend on a predicate factual finding. For example, the crime of involuntary manslaughter has a base offense level of 12 if the offense involved criminally negligent conduct, 18 if the offense involved reckless conduct, or 22 if the offense involved the reckless operation of a means of transportation. U.S.S.G. §2A1.4(a). This scaling of the base offense level is effectively the same as making adjustments to the base offense level, the only difference being that the adjustment is made before the base level is chosen, rather than after the fact. Therefore, this article will analyze both the initial selection of the base offense level and subsequent adjustments to that level under the same rubric.
[45] *See id.* at § 2A2.2.

victim, the offender's financial motivation, and whether the offender violated a court protection order when committing the assault.[46]

More generally, the Sentencing Guidelines also include a set of upward and downward adjustments that can apply to any kind of crime, based on facts about the victim and the offender. An offense level may be adjusted upward, for example, if the victim was particularly vulnerable,[47] if the defendant had an aggravating role in the offense,[48] or if the defendant attempted to obstruct justice after the crime.[49] Conversely, the offense level can be lowered if the defendant had a mitigating role in the offense or if he accepted responsibility for his wrongdoing.[50]

The Sentencing Guidelines tie each upward or downward adjustment to the base offense level to a particular factual predicate. In order to adjust the offense level, therefore, the sentencing judge must make a factual finding on the record. Because the standard of proof at sentencing is a "preponderance of the evidence," the prosecution has the burden to establish that each fact central to an upward adjustment of the offense level is "more likely than not" to be true.[51] Similarly, to obtain a downward adjustment, the offender must prove that the facts supporting that reduction are more than 50% likely to be true.[52]

After the court has finished adjusting the offense level for the particular crime at issue, that final number is called the "total offense level." The sentencing court then plugs that number into a two-axis Sentencing Table,[53] comparing it against the offender's "criminal history category."[54] The intersection of these two numbers across the Table yields a narrow range of months in prison, which constitutes the Guidelines' recommended term of incarceration for that case.[55] For example, an offense level of 5 and criminal history category of II would yield a recommended sentence of zero to six months. A much higher offense level of 39 and criminal history category III would result in a recommended sentence of 324 to 405 months.[56]

According to the Supreme Court, this recommended sentence is a "starting point and the initial benchmark," but it is not binding on the district court – the sentencing judge may still select any appropriate punishment from within the statutory range.[57] The Sentencing Commission initially intended the Guidelines to be mandatory, but the Supreme Court's 2005 decision in *United*

---

[46] *See id.*
[47] *See* U.S.S.G. Ch. 3, Pt. A.
[48] *See* U.S.S.G. Ch. 3, Pt. B.
[49] *See* U.S.S.G. Ch. 3, Pt. C.
[50] *See* U.S.S.G. Ch. 3, Pt. E.
[51] *See, e.g.*, United States v. Sklar, 920 F.2d 107, 112 (1st Cir. 1990).
[52] *See, e.g.*, United States v. Morillo, 8 F.3d 864, 871 (1st Cir. 1993).
[53] *See* U.S.S.G. Sentencing Table.
[54] *See* U.S.S.G. Ch. 4.
[55] *See* U.S.S.G. Ch. 5, Pt. A.
[56] *See* U.S.S.G. Sentencing Table
[57] Gall v. United States, 552 U.S. 38, 49 (2007); *see also* United States v. Booker, 543 U.S. 220, 244-58 (2005). Indeed, the district court's sentence may be reversed if it treats the Guidelines as mandatory, and it is even prohibited from presuming that the Guidelines range is reasonable. *See* Gall v. United States, 552 U.S. 38, 50-51 (2007).

States v. Booker,[58] rendered them merely advisory. A district judge is therefore free to vary from the Guidelines's recommendation if the case presents circumstances "not adequately taken into consideration by the Sentencing Commission,"[59] for example, if the recommended sentence does not fit the circumstances of the case or the goals of sentencing laid out in §3553(a), or if the judge simply disagrees with the policy views of the Sentencing Commission on the matter.[60]

B. The Sentencing Guidelines Use a Threshold Model of Decision-Making

1. The Threshold Nature of Offense Level Adjustments

The Sentencing Guidelines reflect a threshold model of decision-making. When a district judge uses the Guidelines to calculate a recommended sentence, she makes a series of binary decisions about whether or not various enhancements or reductions should apply to the offender's base offense level. These decisions depend on whether the facts supporting each enhancement or reduction exceed a 50% threshold level of probability. If the probability of a factual predicate for a sentence adjustment exceeds 50%, then the sentencing court will regard that fact as true and apply the corresponding adjustment. If the probability is less than 50%, then the court will treat the fact as untrue and the adjustment will not apply. The final recommended Guidelines sentence, therefore, is the outcome of a string of threshold decisions about the defendant's culpability.

Like the threshold model of conviction, the threshold model of sentencing is an "all-or-nothing" system. If a sentencing judge is more than 50% sure that the factual predicate for an adjustment is true, then the adjustment will apply in full – and the offender's recommended sentence will be increased – without regard to whether the judge was 51% or 100% convinced of the matter. Similarly, if the judge is less than 50% sure that the adjustment should apply, then it will not apply at all, even if the judge's degree of certainty fell just short of the requisite level. Only the 50% threshold level of probability matters.

2. Differences Between Threshold Conviction and Sentencing

There are two obvious differences between the threshold model of conviction and the Guidelines' threshold model of sentencing. Yet on closer inspection, each of these differences is actually less significant than it initially appears.

(a) The Gradated Nature of the Sentencing Guidelines. – First, the calculation of a Guidelines sentence is more gradated, and less binary, than the determination of guilt at trial. A guilty verdict is a categorical declaration of the defendant's culpability – two different defendants convicted of the same

[58] 543 U.S. 220 (2005).
[59] 18 U.S.C. § 3553(b)(1); *see also* U.S.S.G. § 5K2.0(2) & (3).
[60] *See Gall*, 552 U.S. at 49-50; Kimbrough v. United States, 552 U.S. 85, 101, 109-110 (2007)

offense will receive identical convictions, even if the details of their actual crimes varied significantly. By contrast, each Guidelines sentence reflects a precise measure of the offender's particular degree of wrongdoing. The total offense level rises or falls based on the details of each crime, so that two different defendants convicted of aggravated assault may receive very different recommended sentences depending on the severity of their transgressions.[61]

But while the calculation of a Guidelines sentence is more gradated than the binary choice of whether or not to convict a defendant, it is still a *threshold* model of decision-making, not a probabilistic approach. Each adjustment to the offense level depends only on whether the factual predicate for that adjustment is more than 50% likely to be true. Once that probability threshold has been crossed, the adjustment applies in full. Levels of probability above or below that threshold are therefore irrelevant.

The difference between a "gradated" and a "probabilistic" model of decision-making is most obvious with the largest sentence adjustments. For example, the Sentencing Guidelines assign a base offense level of 9 to the crime of "possessing dangerous materials on an aircraft." The Guidelines then instruct that the sentencing judge should increase that level by *15 points* if the defendant committed the offense "willfully and without regard for the safety of human life."[62] The effect of that upward adjustment is to raise the recommended sentence for a first-time offender from 4 to 10 months up to 51 to 63 months.[63] In other words, the Guidelines instruct that, so long as it is more than 50% likely that the offender's possession of dangerous materials on the airplane was willful and without regard for human life, his prison term should be increased *five-fold*, without regard to the court's precise level of confidence in that conclusion. Therefore, while the final outcome of a Guidelines sentence calculation does reflect a more gradated judgment of the offender's wrongdoing, each step in that calculation still employs a threshold model of decision-making.

(b) The Non-Binding Nature of the Sentencing Guidelines. – The second difference between the threshold models of conviction and sentencing is that the recommended Guidelines sentence is not officially binding on the district court. While the Sentencing Guidelines are formally non-binding, however, they still exert a strong gravitational pull on district courts.[64]

Despite their advisory nature, the federal law of sentencing uses several sources of pressure to encourage district courts to adhere to the Guidelines'

---

[61] *See* Stephen Breyer, *The Federal Sentencing Guidelines and the Key Compromises Upon Which They Rest*, 17 HOFSTRA L. REV. 1, 9 (1988). In this way, the Guidelines sentence calculation recalls the medieval model of partial punishment described by Foucault: each factual predicate for an upward adjustment "arouse[s] a particular degree of abomination…and consequently involve[s] a degree of punishment." FOUCAULT, *supra* note 18, at 42.

[62] U.S.S.G. § 2K1.5(a) & (b)(1). For another, slightly less dramatic, example, the crime of mishandling environmental pollutants has a base offense level of 6, but that level can be adjusted upward by 11 points if the district court finds that "the offense resulted in a substantial likelihood of death or serious bodily injury." *Id.* at § 2Q1.3(a) & (b)(2).

[63] U.S.S.G., Sentencing Table.

[64] *See* United States v. Ingram, 721 F.3d 35, 40-41(2d Cir. 2013).

recommendations. District courts must calculate a recommended Guidelines sentence in every single case, and they are required to both "begin their analysis with the Guidelines" and to "remain cognizant of them throughout the sentencing process."[65] It is reversible error for a district court to improperly calculate the Guidelines range.[66] The more a district judge varies from the Guidelines recommendation, the more thorough a justification for that variance she must provide.[67] Furthermore, only within-Guidelines sentences may be presumed reasonable on appeal.[68]

The empirical evidence shows that this pressure has a significant effect on sentencing outcomes. The Supreme Court has observed that "[i]n the usual sentencing,…the judge … use[s] the Guidelines range as the starting point in the analysis and impose[s] a sentence within [that] range."[69] "In *less than one-fifth* of cases since 2007 have district courts imposed above- or below-Guidelines sentences absent a Government motion"[70] and more than half of all sentences handed down in 2012 followed the recommendations of the Sentencing Guidelines.[71] "[T]he Sentencing Commission's data indicate that when a Guidelines range moves up or down, offenders' sentences move with it."[72] Indeed, an in-depth statistical study of pre- and post-*Booker* federal sentencing practices revealed that, for the most part, district judges "continue[] to adhere to the Guidelines to a striking degree,"[73] leading the author of the study to conclude that "the most surprising fact about *Booker* is just how small an effect it actually had….The Guidelines still matter. They still matter nearly as much as they did on the day before *Booker* was decided."[74] In sum, while district judges are not officially required to use a threshold model of sentencing, the powerful influence of the Sentencing Guidelines ensures that in most cases, they do.

III. FLAWED JUSTIFICATIONS FOR THE THRESHOLD MODEL OF SENTENCING

This Part will review three justifications that courts and commentators have offered for the threshold model of conviction. It will show that these justifications do not hold in the unique context of a sentencing hearing. First, the threshold model of conviction is said to protect the presumption of

---

[65] *See Gall*, 552 U.S. at 50 n.6.
[66] *See id.* at 51.
[67] *See Gall*, 552 U.S. at 50.
[68] *See* Rita v. United States, 551 U.S. 338, 347 (2007).
[69] Freeman v. United States, 564 U. S. ___, ___ (2011) (plurality opinion) (slip op., at 5).
[70] Peugh v. United States, 569 U.S. ___, ___ (2013) (slip. op. at 12-13). Furthermore, while criminal convictions rendered by a jury are generally treated as binding, federal district judges technically retain the power to acquit convicted defendants "notwithstanding the verdict." *See* Fed. R. Crim. P. 29(c)(2).
[71] U.S. Sentencing Commission, Final Quarterly Data Report, Fiscal Year 2012, http://www.ussc.gov/Data_and_Statistics/Federal_Sentencing_Statistics/Quarterly_Sentencing_Updates/USSC_2012_Quarter_Report_Final.pdf.
[72] Peugh v. United States, 569 U.S. ___, ___ (2013) (slip. op. at 12-13).
[73] Frank O. Bowman III, *Dead Law Walking: The Surprising Tenacity of the Federal Sentencing Guidelines*, 51 Houston L. Rev. ____ (2014) (forthcoming), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2374910.
[74] *Id.*

innocence. That presumption, however, would not bar a district judge from varying an offender's punishment based on probabilities *above* the requisite threshold level of proof at sentencing. Second, the threshold model of conviction sends a substantive, rather than evidentiary, message, which may assist the deterrent function of punishment. But if conviction sends a substantive message, then sentencing need not do so, and moreover, deterrence is not the only reason for criminal punishment. Finally, the threshold model of conviction streamlines the decision-making process for the jury. The federal law of sentencing, however, is already fairly streamlined, and so an even more simplified process is not necessary.

A. Protecting the Presumption of Innocence

First, the threshold model helps to protect the presumption of innocence by setting a hard floor of certainty below which the government may not punish an accused defendant. The presumption of innocence, however, only requires that punishment be prohibited *below* a certain level of probable guilt. The presumption would not be threatened if courts took account of probabilities *above* the requisite threshold level. At trial, this insight does little good, because there is so little room to maneuver above the threshold level of proof required for conviction. But at sentencing, there is a large range of probabilities available that courts could consider when calculating a recommended Guidelines sentence.

The presumption of innocence is "axiomatic and elementary" to American criminal justice.[75] It provides that a criminal defendant is assumed innocent unless and until the government can prove him guilty.[76] The threshold model of decision-making guarantees this protection against arbitrary punishment by establishing a minimum level of proof that the government must meet in order to convict a defendant – 95% probability of guilt.[77] A return to the medieval model of conviction, where "partial punishments" were imposed on findings of "partial guilt," would dilute the presumption of innocence by making it easier for the government to establish that a person accused of a crime deserves punishment.[78]

At sentencing, the presumption of innocence is not quite so axiomatic, but there is still a small presumption in favor of offenders' liberty. Officially, once a defendant has been convicted of a crime, the presumption of innocence is said to "disappear[]."[79] In practice, however, the preponderance-of-the-evidence standard used at sentencing still presumes the defendant's innocence in cases where the evidence supporting and opposing a sentence enhancement is equally split. In other words, when the probability of the factual predicate for an increase to the offender's sentence is exactly 50%, the offender prevails

[75] Coffin v. United States, 156 U.S. 432, 452-53 (1895); *see also* In re Winship, 397 U.S. 358, 363 (1970).
[76] *See Coffin*, 156 U.S. at 452-53.
[77] *See* Fisher, *supra* note 10, at 882 & n.170.
[78] *See* Bray, *supra* note 12, at 1312.
[79] Delo v. Lashley, 507 U.S. 272, 278 (1993); *see also* Young, *supra* note 32, at 357.

and the enhancement does not apply.[80] A model of decision-making that permitted sentence enhancements based on levels of probability lower than 50% would dilute even this minimal presumption of innocence.

The justice system could, however, take account of probabilities *above* the requisite threshold level of proof without endangering the presumption of innocence. Put another way, the presumption of innocence only forbids "partial punishments" for "partial guilt" because they would make it too easy for the government to punish suspected wrongdoers. *Higher* punishments for *more certain* guilt, however, would not violate the presumption of innocence. Nor would *lower* punishments for *less certain* guilt, so long as the minimum threshold of certainty had been met.

At trial, for example, it would be consistent with the presumption of innocence to impose a more damning conviction when the probability of the offender's guilt was over 95%. Alternatively, the jury could mitigate the offender's guilty verdict the farther the probability of his guilt fell from 100%, with zero punishment for levels of certainty below 95%. Of course, the range of probabilities at issue here are so small as to be practically meaningless. Professor Larry Tribe argues, for example, that the beyond-a-reasonable-doubt standard of proof "come[s] as close to certainty as human knowledge allows."[81] It would be impossible, according to Tribe, to vary punishments based on levels of certainty above that level. Even if one disagreed with Professor Tribe, moreover, the tiny epistemic difference between 95% and 100% probability of guilt would make it very difficult to accurately and consistently differentiate between guilty offenders.

At sentencing, however, the lower standard of proof leaves open a much wider range of epistemic space above the threshold level of certainty. The justice system commonly recognizes at least two levels of probability above the preponderance-of-the-evidence standard – "clear and convincing evidence" and proof "beyond a reasonable doubt," which correspond roughly to 70% and 95% probability.[82] A district judge could thus easily vary the size of an upward adjustment to an offender's offense level based on her level of confidence that its factual predicate had been fulfilled. For instance, she might increase the size of the sentence enhancement as the odds of its factual predicate rose above 50%. Or, she could decrease it the further the odds fell farther from 100%, with zero enhancement for levels of certainty below 50%.

In sum, the threshold model of decision-making protects the presumption of innocence, but that justification for the threshold model only prohibits the consideration of probabilities below the threshold level of certainty, not above that level. While this observation does little good at trial, where the threshold level of certainty is set extremely high, it is quite relevant

---

[80] *See* United States v. Gigante, 94 F.3d 53, 55 (1996). For sentence reductions, the presumption is placed slightly in favor of the *government*, since when the evidence in favor of reducing an offender's sentence is balanced, the law of sentencing holds that the prosecution should prevail. *See, e.g.*, United States v. Morillo, 8 F.3d 864, 871 (1st Cir. 1993).

[81] Tribe, *supra* note 14, at 388.

[82] *See, e.g.*, United States v. Fatico, 458 F. Supp. 388, 403-06 (E.D.N.Y. 1978) (Weinstein, J.).

at sentencing, where the burden of proof is lower and the epistemic range in play much greater. As this Article will show later on,[83] the Sentencing Guidelines's failure to take account of variations in probability above the 50% threshold creates serious problems in the administration of criminal justice, and may in fact implicate the very Due Process concerns that animate the presumption of innocence itself.

B. Sending a Substantive, Rather than Evidentiary, Message

Second, the threshold model facilitates the deterrent function of the criminal law by sending a substantive message about the outcome of the proceedings. That function, however, need not be served at every stage of the judicial process, especially not at sentencing. Moreover, deterrence is not the only purpose of criminal punishment.

Professor Charles Nesson argues that for a conviction to have a deterrent effect, the guilty verdict must send a "substantive" message about the defendant's conduct, rather than an "evidentiary" one.[84] In other words, the guilty verdict must declare that the defendant committed a crime, not that the prosecution established with 95% certainty that the defendant committed a crime. This helps foster public trust in the criminal justice system by presenting trial outcomes as statements of fact, rather than best guesses.[85] It also "forges a link between the judicial account of the defendant's transgression and [the public's] own behavior," communicating that bad *acts* lead to punishment, rather than bad evidence.[86] According to Nesson, future offenders will feel more obligated to respect this substantive, rather than evidentiary, conception of criminal justice. Professor Nesson therefore endorses the threshold model of conviction, which presents the outcome of a criminal trial as a "statement about what happened," rather than a "statement about the evidence."[87]

Nesson's argument, while persuasive, also overstates the importance of the deterrent message in criminal law. First, even if one agrees that the justice system should send a "substantive" message about criminal trials, Professor Nesson does not explain why it is necessary to communicate such a message at *every* stage of the proceedings. Sentencing comes after conviction, so the guilty verdict will already have expressed a substantive condemnation of the defendant's conduct. Moreover, although sentencing hearings are open to the public, they do not involve juries of attentive community members, nor do they commonly attract the same attention as trials. Therefore, even accepting Nesson's argument, there is still less of a need at sentencing to attach a substantive meaning to the outcome of the proceedings.

---

[83] *See infra*, Part IV.
[84] *See* Nesson, *supra* note 10, at 1361-62.
[85] *See id.* at 1362.
[86] *Id.*
[87] *Id.*

Second, deterrence is only one of the purposes of criminal punishment.[88] Three other reasons for punishing wrongdoers, equally endorsed by Congress in § 3553(a), are the retributive, incapacitative, and rehabilitative theories of punishment.[89] These approaches do not incarcerate simply to send a warning to future wrongdoers; they use prison as an end in itself. The retributive model, for instance, uses prison as "just punishment for the offense."[90] The incapacitative approach incarcerates as a means "protect the public from further crimes of the defendant."[91] Finally, the rehabilitative theory seeks to provide "the defendant … needed educational or vocational training, medical care, or other correctional treatment."[92] Because these theories do not depend on communicating any message, they are effective whether or not the justice system attaches a "substantive" or an "evidentiary" meaning to criminal punishment.

At best, therefore, deterrence is only a partial justification for the threshold model of sentencing. It is less important to send a substantive message at sentencing, since that message will already have been communicated by the defendant's conviction at trial. Moreover, deterrence is only one of the purposes of criminal punishment, and the other three purposes listed in § 3553(a) do not rely on sending a substantive message. The deterrent benefit of threshold sentencing, therefore, must be carefully weighed against the significant problems with the approach, discussed later on below.[93]

C. Streamlining the Decision-Making Process

Finally, the threshold model spares decision-makers the difficult task of making fine-grain probability determinations. However, while it may be important to simplify the decision-making process for juries of twelve attempting to reach consensus about a defendant's guilt, it is much easier for a single district court judge calculating a recommended Guidelines sentence to take more precise account of probability.

The threshold model of conviction asks the jury to answer a simple question: does the prosecution's evidence establish that the likelihood that the defendant committed a crime is "beyond a reasonable doubt," or more than 95% probable? Because it does not demand any more specifics, the threshold model avoids requiring juries "to ascribe accurate probabilities to their findings."[94] That assignment would be particularly difficult at the high degrees of confidence required to obtain a conviction, since it would ask juries to distinguish between cases in which the odds of the defendant's guilt was 95%, 96%, 97%, 98%, and so on. Indeed, the assignment would be a challenge even

[88] *See* 18 U.S.C. § 3553(a)(2)(B).
[89] *See* 18 U.S.C. § 3553(a)(2)(A), (C), & (D); *see also* Mistretta v. United States, 488 U.S. 361, 367 (1989).
[90] *See* 18 U.S.C. § 3553(a)(2)(A).
[91] *See* 18 U.S.C. § 3553(a)(2)(C).
[92] *See* 18 U.S.C. § 3553(a)(2)(D).
[93] *See infra* Part IV.
[94] Harel & Porat, *supra* note 14, at 299.

for an individual juror, let alone twelve working together to agree on a consensus decision.

At sentencing, however, the simplicity of the threshold model is much less attractive, since the epistemic space at issue is larger and the number of decision-makers fewer. As already explained, the probability range above the "preponderance of the evidence" threshold is quite large: between 51% and 100% likelihood that the factual predicate for a sentence enhancement has been satisfied. So, while it might be difficult for a jury at trial to differentiate between cases of 95% and 100% likelihood of guilt, it would be much easier for a sentencing judge to recognize gradations along the spectrum of probability available to her – for example, 51% versus 70% versus 95% probability of guilt (which roughly correlate with the preponderance-of-the-evidence, clear-and-convincing-evidence, and beyond-a-reasonable-doubt standards of proof).[95] Similarly, while it might be difficult for 12 jurors to agree on their shared level of confidence in a defendant's guilt, there would be no such problem at sentencing, where a single federal judge gets to decide on her own the applicability of various sentence enhancements. The threshold model, therefore, is not necessary to streamline the decision-making process at sentencing in the same way that it is at trial.

Even for a single decision-maker, it may be difficult to reliably and consistently estimate probabilities. Still, the pragmatic argument for the threshold model wields far less force at sentencing than it does at trial, given the wider range of probabilities available and the smaller number of decision-makers. Especially given the flaws of threshold sentencing, discussed in the next section,[96] it may well be worth placing an additional burden on district judges in order to take better account of probability at the penalty stage of criminal proceedings.

IV. SIGNIFICANT PROBLEMS WITH THE THRESHOLD MODEL OF SENTENCING

This Part will review the two problems that courts and commentators have observed with the threshold model of conviction. It will demonstrate that these flaws are even more severe in the threshold model of sentencing. First, the threshold model of conviction is inefficient, because it does not prioritize the allocation of punishment resources based on the likelihood that they will be spent on guilty offenders. That inefficiency is particularly bad at sentencing, because the burden of proof on the government is lower and fewer fact-findings are necessary to increase an offender's sentence. Second, the threshold model of conviction is unfair, because it denies innocent offenders the benefit of any doubts about their guilt. Once again, this problem is especially severe at the penalty stage of the proceedings, where the standard of proof is significantly lower than it is at trial and fewer fact-findings are necessary to impose increased punishment.

---

[95] *See* United States v. Fatico, 458 F. Supp. 388, 403-06 (E.D.N.Y. 1978) (Weinstein, J.).
[96] *See infra*, Part IV.

A. The Threshold Model of Decision-Making is Inefficient

1. The Threshold Model Wastes Punishment Resources

The threshold model of decision-making is inefficient. Professors Talia Fisher and Henrick Lando have demonstrated this point in regard to the threshold model of conviction, using an analytical framework that relies on a deterrence theory of punishment.[97]

To begin, each Professor observes that punishment of the innocent does not communicate an effective deterrent warning to future wrongdoers.[98] The deterrent value of a criminal punishment will, therefore, depend on whether it is imposed on a factually guilty or factually innocent defendant. Taking that premise one step further, Fisher and Lando aver that punishment of offenders who are more likely to be guilty will, on the whole, have more deterrent value than the punishment of offenders who are less likely to be guilty. Fisher explains: "Just as punishment of the factually innocent yields a lower deterrent effect than punishment of the factually guilty, punishment of defendants whose probability of guilt is low yields a lower deterrent effect than identical punishment imposed upon defendants whose certainty of guilt is high."[99]

The threshold model of conviction, however, does not distinguish between defendants based on the probability of their guilt. The model condemns equally those who are *without a doubt* guilty (and whose punishment will be most likely to have deterrent value) and those who are only *beyond a reasonable doubt* guilty (and whose punishment will be slightly less likely to have an impact). According to Fisher and Lando's framework, then, the threshold model of conviction will inevitably spend more punishment resources than it needs (and will needlessly extract more of the social costs that accompany incarceration).[100] It would be more efficient, instead, to punish offenders based on the likelihood that those punishment resources would be well spent. "[T]he greater the certainty of the defendant's guilt, the lesser the concern of 'wasting' punishment resources while obtaining a weaker deterrence effect, and vice versa."[101]

This same inefficiency is also present in the threshold model of sentencing. Begin again with Lando and Fisher's initial premise: just like conviction of the innocent does not communicate an effective deterrent warning, so too will an erroneous increase to an offender's sentence fail to deter future wrongdoers. The deterrent value of a sentence enhancement, therefore, depends on whether the factual predicate for that enhancement was actually fulfilled. Or to paraphrase Professor Fisher, a sentence enhancement

---

[97] *See* Fisher, *supra* note 10, at 855-56 & n.86; Henrik Lando, *The Size of the Sanction Should Depend on the Weight of the Evidence*, 1 REV. L. & ECON. 277, 282 (2005).
[98] *See* Fisher, *supra* note 10, at 855-56.
[99] *See id.*; *see also* Michael T. Cahill, *Punishment Pluralism*, *in* RETRIBUTIVISM: ESSAYS ON THEORY AND POLICY 36 (Mark D. White ed., 2011).
[100] *See* Fisher, *supra* note 10, at 856 n.86. For an in-depth discussion of the costs and benefits of punishment, see generally Cahill, *supra* note 99.
[101] *See id.*

applied when its factual predicate is less probably fulfilled will have a lower deterrent effect than when the enhancement's factual predicate is more probably fulfilled.[102]

The threshold model of sentencing, however, applies sentence enhancements without regard to the probability that their factual predicates have been fulfilled. The model increases sentences equally in cases where the factual predicate has almost *certainly* been fulfilled and those in which the predicate has only *more than likely* been fulfilled. As a result, the Sentencing Guidelines will inevitably recommend longer sentences than necessary. Just like with convictions, therefore, it would be more efficient for sentencing judges to vary punishments based on the probability that they will have an effective deterrent impact.

2. The Threshold Model Wastes More Resources at Sentencing than at Trial

Two unique features of the law of sentencing make the threshold model of decision-making even more inefficient at sentencing than it is at trial.

First, the lower burden of proof at sentencing means that errors are more common at this stage of the proceedings. As explained earlier, the epistemic range above the threshold level of probability required for a conviction – proof "beyond a reasonable doubt" – is very narrow.[103] As a result, the threshold model of conviction will only waste punishment resources on innocent offenders in a very small number of cases: 5% of those decided on 95% probability of guilt.

At sentencing, however, the threshold standard of proof drops from 95% to 50%. This lower threshold opens up a much larger epistemic bandwidth, which leaves much more room for error. In roughly one out of every *two* upward adjustments decided on 51% probability, the factual predicate for the adjustment will not actually have been fulfilled, and so the expense of increasing the offender's sentence will have been wasted. The threshold model of sentencing will therefore waste punishment resources in many more cases than will the threshold model of conviction.

Second, sentences are enhanced based on a disjunctive series of factual predicates, whereas a guilty verdict depends on a conjunctive group of fact-findings. This makes it much easier for the threshold model of sentencing more to waste punishment resources. To convict a defendant at trial, the jury must find that he "is guilty of *every* element of the crime with which he is charged, beyond a reasonable doubt."[104] The odds of an erroneous conviction, therefore, are quite low. For instance, the federal crime of kidnapping requires the government to prove four facts: that the defendant (1) knowingly transported (2) an unconsenting person (3) in interstate commerce (4) in order

[102] *Cf. id.* at 855-56.
[103] *See supra*, Part III.A.
[104] United States v. Gaudin, 515 U.S. 506, 510 (1995) (emphasis added).

to hold him for ransom, reward, or otherwise.[105] If a jury convicts a defendant of kidnapping based on 95% probability that each of these four facts is true, there will be an 81% chance that its decision to return guilty verdict will be the correct one and a 19% chance that punishment resources will have been wasted on an erroneous conviction.[106]

By contrast, a sentencing judge may increase an offender's punishment based on her finding that *any* of the multiple possible factual predicates for a sentence enhancement is more than 50% likely to have been fulfilled. As explained earlier, the Sentencing Guidelines enumerate specific upward adjustments for each crime, with each adjustment triggered by a particular predicate fact.[107] The Guidelines also include general sentence adjustments that may apply to all crimes.[108] Each time a judge finds that one of the factual predicates for an adjustment has been satisfied, the associated sentence enhancement immediately applies, making it much more likely that a judge will wrongly extend a sentence than that a jury will wrongly convict.

For example, the offense-specific adjustments for kidnapping include:

> (1) If a ransom demand or a demand upon government was made, increase by 6 levels.
> (2) (A) If the victim sustained permanent or life-threatening bodily injury, increase by 4 levels; (B) if the victim sustained serious bodily injury, increase by 2 levels; or (C) if the degree of injury is between that specified in subdivisions (A) and (B), increase by 3 levels.
> (3) If a dangerous weapon was used, increase by 2 levels.
> (4)(A) If the victim was not released before thirty days had elapsed, increase by 2 levels.
> (B) If the victim was not released before seven days had elapsed, increase by 1 level.[109]

So, if a judge increases a kidnapper's sentence by 2 levels based on a 51% probability that he used a dangerous weapon,[110] there will be a 51% chance that the decision to increase the sentence will be the correct one and a 49% chance that punishment resources will have been wasted on an erroneous sentence enhancement. Even worse, if a judge decides to apply multiple sentence enhancements at that same level of confidence – say, five different

---

[105] *See* United States v. Barton, 257 F.3d 433, 439 (5th Cir. 2001); *see also* 18 U.S.C. § 1201.
[106] *Cf.* Ariel Porat & Eric A. Posner, *Aggregation and the Law*, 122 YALE L.J. 3, 38 (2012) (noting that defendants are convicted when juries are convinced that the defendant's guilt is 95% probable with respect to each element, even though aggregating those probabilities would yield a level of certainty less than 95%); *see also* Model Penal Code § 1.12(1) (1962) ("No person may be convicted of an offense unless *each* element of such offense is proved beyond a reasonable doubt.") (emphasis added).
[107] *See supra*, Part II.A.3.
[108] *See generally* U.S.S.G. ch. 3.
[109] U.S.S.G. § 2A4.1(b).
[110] *See* U.S.S.G. § 2A4.1(b)(3).

offense-specific adjustments and three general adjustments – there will be only a *.4%* chance that the resulting total offense level is factually accurate, and a *99.6%* chance that a sentence based on the total offense level will be longer than necessary.

Because conviction depends on a conjunctive set of fact-findings but a sentence increase relies on a disjunctive group of possible factual predicates, the odds that a sentencing judge will impose too harsh a sentence are much higher than the odds that a jury will hold the wrong person responsible for a crime. The threshold model is therefore more likely to inefficiently allocate punishment resources at sentencing than it is at trial.[111]

3. Problems with the Deterrence-Based Inefficiency Critique

The deterrence-based inefficiency argument is instructive, but it also suffers from two important problems.

First, effective deterrence requires that the justice system send a warning about future punishments strong enough to dissuade future wrongdoers from committing crimes. That warning will only deter a potential wrongdoer if the expected severity of the punishment he will receive, discounted by the probability that he will not get caught, exceeds his expected gains from the crime. Reducing the severity of the punishment that the wrongdoer will receive based on the court's judgment of the probability of his guilt, therefore, might also reduce its deterrent effect.[112] In other words, a wrongdoer may be more willing to commit a crime if he thinks that the possible punishment he will face, if he gets caught, will be reduced in proportion to the doubts about his guilt.

Second, as explained earlier,[113] the deterrent impact of punishment depends in part on the sending of a "substantive," rather than an "evidentiary," message about the convicted defendant's conduct. Remember that according to Professor Nesson, in order for a punishment to effectively deter bad behavior, it must announce definitively that the defendant committed a crime, thereby communicating a moral warning against bad behavior that "inculcates the behavioral message associated with the applicable legal rule."[114] Scaling punishment to the probability of the defendant's guilt, however, would undermine this moral warning, since it would not declare that the offender

[111] The fact that a wrongful conviction leads to an entirely wasted prison term, while a wrongful upward adjustment leads to a wasted increase in the length of the term, means that on a macro scale, the threshold model of conviction may still be more inefficient than the threshold model of sentencing.

[112] Put another way, punishment-as-deterrence requires the imposition of a punishment of severity P, such that the probability of being punished, x, is greater than the expected gain from the crime, G. Therefore, $xP > G$. Fisher and Lando suggest that it would be more efficient to reduce P based on the probability of the defendant's guilt, b. According to that approach, the calculation of an effective deterrent would become $xbP > G$. To maintain the same deterrent effect, therefore, P would have to increase, which may well undo any of the efficiency gains that Fisher and Lando seek.

[113] *See supra*, Part III.B.

[114] Nesson, *supra* note 10, at 1362-64.

committed a crime, but merely state that the evidence suggests with a high degree of certainty that he did so. This would change the law's "substantive message from one of morality ('feel guilty if you do wrong') to one of crude risk calculation ('estimate what you can do without getting caught')."[115] It may therefore undermine punishment's deterrent effect, since future offenders might not feel as obligated to respect the latter conception of criminal justice.

4. Non-Deterrence Inefficiency Critiques of the Threshold Model

Applying Fisher and Lando's framework to non-deterrent theories of punishment resolves these two problems. In other words, although Fisher and Lando focus solely on deterrence, ironically, their inefficiency critique of the threshold model is much more persuasive from the perspective of the retributive, incapacitative, and rehabilitative theories of punishment.

Fisher and Lando's inefficiency argument is perfectly consistent with these other theories of punishment. From a retributive perspective, for example, incarceration is more likely to provide just punishment when it is imposed on defendants who are more likely to have engaged in wrongdoing. From an incapacitative perspective, prison is more likely to protect the public when it is used to confine defendants who are more likely to pose a public safety threat. Finally, from a rehabilitative perspective, penal resources are more likely to provide needed socialization and training when they are expended on defendants who are more likely to be in need of rehabilitation. The converse is of course also true in all cases – punishments are more likely to be wasted when they are imposed on offenders whose guilt is less probable. From each of these perspectives, therefore, it is inefficient to punish offenders equally when the probability of their culpability is unequal.

Moreover, these non-deterrent theories of punishment do not depend on communicating a stern or substantive warning to future offenders. As a result, a probabilistic approach to sentencing would not undermine the purpose of punishment from these perspectives, as it might under the deterrent theory of punishment. The retributive, incapacitative, and rehabilitative purposes of punishment use prison to achieve certain direct ends, rather than to send a message. While the threshold model may bolster the deterrent impact of punishment, therefore, under these non-deterrent theories, punishment would not lose its effectiveness if it were varied based on the probability of the defendant's culpability.[116] In sum, Fisher and Lando's inefficiency critique of the threshold model actually works better when applied in concert with the retributive, incapacitative, and rehabilitative theories of punishment.

[115] *Id.* at 1362.

[116] The threshold model might not be considered inefficient from an expressive perspective, since it projects certainty as to the defendant's guilt and thereby assists the public in "forg[ing] a link between severity of punishment and the force of the moral repudiation of the offense and the offender." Fisher, *supra* note 10, at 864.

B. The Threshold Model of Decision-Making is Unfair

1. The Threshold Model is Unfair to the Factually Innocent

In addition to being inefficient, the threshold model of decision-making is also unfair. Begin with the threshold model of conviction. By setting the threshold level of probability required for a guilty verdict at less than 100%, the criminal justice system expresses a tolerance for the occasional wrongful conviction. Put another way, convicting on less than absolute certainty means taking the chance that that conviction will be erroneous. The 95% threshold level of probability required for conviction, therefore, reflects an acceptance that as many as one out of every twenty convicted defendants may actually be innocent.

Of course, because absolute certainty is unobtainable in the courtroom,[117] it is a practical necessity for the criminal justice system to tolerate the possibility of an occasional wrongful conviction.[118] What is not necessary, however, is that the threshold model denies to convicted defendants the benefit of any lingering doubts about their guilt. In other words, the system could reduce the suffering of defendants erroneously sent to prison by punishing less when the defendants were less likely to be guilty.[119] Instead, the threshold model of conviction subjects all convicted defendants to the same exact punishment, whether they are 95% or 99% certain to be guilty.[120] It therefore allows a rare but significant degree of suffering on the part of the wrongly convicted.

The capital sentencing doctrine of "residual doubt" reflects precisely this concern. The doctrine permits defendants convicted of a capital offense to raise "residual doubts" about their guilt as a mitigating factor in the penalty phase of the proceedings.[121] "'Residual doubt' is … a lingering uncertainty about facts, a state of mind that exists somewhere between 'beyond a reasonable doubt' and 'absolute certainty.'"[122] In other words, even if the defendant is more than 95% likely to have committed a capital crime, he may still seek to avoid the death penalty by pointing out the remaining 5% chance that he is innocent. The residual doubt doctrine suggests that the state should kill only when it is absolutely certain that a defendant is guilty, not merely when it has ruled out all reasonable doubts about his guilt.[123] In order to avoid

---

[117] *See* BENTHAM, *supra* note 5, at 351.
[118] *See* Cahill, *supra* note 99, at 36.
[119] Indeed, this principal has been endorsed by no less a moral authority than Atticus Finch: "There's always a doubt, sometimes only the shadow of a doubt. The law says 'reasonable doubt,' but I think a defendant's entitled to the shadow of a doubt. There's always the possibility, no matter how improbable, that he's innocent." HARPER LEE, TO KILL A MOCKINGBIRD 242 (1988).
[120] Lando, *supra* note 97, at 277.
[121] Franklin v. Lynaugh, 487 U.S. 164, 188 (1988) (O'Connor, J., concurring).
[122] *Id.*; *see also* Lockhart v. McCree, 476 U.S. 162, 181 (1986); Christina S. Pignatelli, *Residual Doubt: It's a Life Saver*, 13 CAP. DEF. J. 307 (2001).
[123] As the Supreme Court has not required jurisdictions to adopt this doctrine, it evidentially does not consider this unfairness to be of constitutional concern. *See Franklin*, 487 U.S. at 172-174.

the gross injustice of executing an innocent man, not all convicted capital offenders should be treated alike – those that are less likely to be guilty should not be sentenced to death. Although the doctrine is unique to capital punishment, the same principle could easily be extended to incarceration. To reduce the injustice of imprisoning an innocent man, not all convicted non-capital offenders should be treated alike. Those that are less likely to be guilty should receive shorter sentences. The threshold model of conviction violates this principal of fairness by condemning offenders equally without regard to the likelihood of their guilt.

The threshold model of sentencing is unfair in the same way. Like the threshold model of conviction, the Sentencing Guidelines set the threshold level of probability required for a sentence enhancement at less than 100%. The Guidelines therefore tolerate the fact that in some cases – just under half of those decided at 51% certainty – the offender will have his prison term extended based on factual findings about him or his crime that are not actually true.

As with the threshold model of conviction, this unfairness is probably a necessary evil, since it would be impractical for the Guidelines to demand absolute certainty before a court could enhance an offender's sentence. Nevertheless, it is not necessary for the Guidelines to deny offenders the benefit of whatever doubts remain about their culpability. The Guidelines could mitigate the suffering of offenders whose sentences are erroneously extended by instructing courts to increase sentences by a lesser amount when they are less certain that the factual predicates for those increases have been fulfilled. Instead, like the threshold model of conviction, the threshold model of sentencing applies all sentence enhancements equally, regardless of whether the factual predicate for those enhancements is 51% or 100% likely to be true. The Guidelines therefore accept that some offenders will serve longer prison terms than they deserve, and yet they do not attempt to reduce those terms for offenders whose culpability is in doubt.

2. The Threshold Model is More Unfair at Sentencing than at Trial

Because the two stages of the proceedings place different burdens of proof on the government, the penalty stage inflicts far more unfairness on innocent offenders than does the guilt sage. The standard of proof used at trial is so high that it likely mitigates the unfairness of the threshold model of conviction. According to Professor Larry Tribe, the "beyond a reasonable doubt" standard does *not* imply an acceptance of occasional erroneous convictions. To the contrary, it reflects a "refus[al] to take a deliberate risk of punishing any innocent man."[124]

Professor Tribe argues that it would be a mistake to infer that the criminal justice system tolerates the conviction of innocent defendants in 5% of cases simply because the "beyond a reasonable doubt" standard is

[124] Tribe, *supra* note 14, at 388.

commonly equated with 95% certainty.[125] Focusing too hard on these numbers misses the entire point of that burden of proof, in which the "fundamental postulate" is "that deliberately … punish[ing] a man of whose guilt we feel unsure is wrong."[126] In other words, Tribe rejects the quantification of the "beyond a reasonable doubt" standard as 95% probability of guilt. Instead, he says that the criminal justice system uses that standard in order to credit accused defendants with any reasonable doubts about their guilt, setting them free if they can make any plausible argument that they did not commit the crime charged. The "beyond a reasonable doubt" standard, therefore, is not "a mere probabilistic device to assure a sufficiently low frequency of erroneous convictions."[127] Although Tribe accepts that absolute certainty is impossible, he asserts that the beyond a reasonable doubt standard of proof attempts to prevent *any* erroneous conviction by "com[ing] as close to certainty as human knowledge allows."[128] It is a "basic security conferred by a system that promises never to punish in the face of real doubt."[129] In short, according to Tribe's analysis, the threshold model of conviction uses the fairest, most favorable standard of proof that could possibly be applied in order to protect innocent defendants.

Unfortunately, however, the threshold model of sentencing uses a much lower "preponderance of the evidence" standard, and so it cannot be defended under Professor Tribe's analysis. Far from a "basic security" against wrongful punishment,[130] the preponderance standard reflects the criminal justice system's "minimal concern"[131] with the decision whether to increase an offender's sentence. Courts have offered several justifications for using this lower standard of proof at sentencing,[132] but the fact remains that proof by a preponderance of the evidence clearly reflects the acceptance of "a deliberate risk"[133] that in just under half of all cases, offenders may spend more time in prison than they deserve.

Because of this lower standard of proof, the unfairness of the threshold model will manifest with far greater frequency at sentencing than it will at trial. At worst, one in every twenty criminal trials will result in a false conviction. But under the preponderance-of-the-evidence standard, approximately one out of every *two* defendants may have their sentences enhanced based culpable conduct that they never actually committed. Remember, as well, that errors are more common at sentencing due to the

---

[125] *See id.* at 385-86.
[126] *Id.* at 386.
[127] *Id.*
[128] *Id.* at 388. Tribe's perspective suggests that, since there is no conceivable epistemic space above proof "beyond a reasonable doubt," the "residual doubt" doctrine, *see* Franklin v. Lynaugh, 487 U.S. 164, 188 (1988) (O'Connor, J., concurring), is a figment of the legal imagination
[129] Tribe, *supra* note 14, at 386.
[130] *Id.*
[131] Addington v. Texas, 441 U.S. 418, 423 (1979).
[132] *See* Young, *supra* note 32, at 335-36.
[133] Tribe, *supra* note 14, at 388.

disjunctive nature of sentence enhancements.[134] As a result, the rate at which offenders are erroneously punished (and then denied the benefit of any doubts about their culpability) is almost certain to be higher at sentencing than it is at trial.

Of course, errors at trial, when they do occur, may still be more unfair than those at sentencing, since the consequences of a wrongful conviction (a prison term) are much more severe than for an wrongful sentence enhancement (a slightly longer prison term). Nevertheless, especially for larger sentence enhancements, an erroneous increase to an offender's punishment can do significant harm. Recall, for example, that the recommended sentence for the crime of "possessing dangerous materials on an aircraft" is *sextupled* if the judge believes that there was a 51% chance that the offender did so "willfully and without regard for the safety of human life."[135] If a court misjudges and erroneously applies that 15-level enhancement – as it will in just under half of all cases decided at the minimum required level of certainty – it will be committing a serious injustice.

The federal courts have recognized that the lower standard of proof at sentencing creates the potential for greater injustice. In *McMillan v. Pennsylvania*, the Supreme Court warned that the Due Process Clause and the Sixth Amendment forbade states from "evad[ing]" the beyond-a-reasonable-doubt standard mandated at trial by "restructuring existing crimes" to make certain sentencing factors exceptionally punitive, such that they would become "a tail which wags the dog of the substantive offense."[136] In other words, a state may not take advantage of the lower standard of proof at sentencing by creating an excessively punitive sentence enhancement, which would allow the prosecution to convict a defendant on more innocuous conduct at trial and then wait until sentencing to bring up his truly culpable behavior. This principal was applied most famously by the Third Circuit in *United States v. Kikumura*, where the panel held that facts that would increase the defendant's offense level by 22 levels had to be found by clear-and-convincing evidence, rather than by a mere preponderance of the evidence.[137] According to the *Kikumura* court, "the potential for significant unfairness" becomes too great "[i]n this extreme context," where the offender's sentence increased "twelve-fold" based only on a 51% probable fact-finding.[138] Although that holding was later reversed *en banc*, it demonstrates that federal courts are well aware of the higher risk of error at sentencing, and that when the penal consequences become large enough, the unfairness of an erroneous 51% probable fact-finding at sentencing may attain constitutional dimensions.

---

[134] *See supra*, Part IV.A.2.
[135] U.S.S.G. § 2K1.5(a) & (b)(1).
[136] McMillan v. Pennsylvania, 477 U.S. 79, 87-89 (1986).
[137] *See* United States v. Kikumura, 918 F.2d 1084, 1101-02 (3d Cir. 1990), *overruled by*, United States v. Grier, 475 F.3d 556 (3d Cir. 2011) (en banc).
[138] *Id.* at 1099-101.

V. THE THRESHOLD MODEL OF DRUG QUANTITY DETERMINATION

This Part will argue that the threshold model of sentencing is particularly problematic when it comes to determinations of drug quantity. Under the Sentencing Guidelines, drug quantity helps determine the offense level for crimes involving the unlawful manufacturing, importing, exporting, or trafficking of a controlled substance (as well as possession-with-intent-to-commit any of those crimes).[139] Like other facts at sentencing, drug quantity is found by a judge using a preponderance-of-the-evidence threshold standard of proof. Three unique features of drug quantity determinations, however, make these fact-findings especially vulnerable to the flaws of the threshold model of sentencing. First, because drug quantity determinations are made at each sentencing for a drug-trafficking offense, they are particularly common, and thus they are more frequently responsible for wasting resources and unfairly extending prison sentences. Second, drug quantity determinations can produce unusually large increases to the offender's sentence, which gives them the potential to create the most inefficiency and unfairness at sentencing. Finally, judges often must estimate drug quantities, and these estimations are notoriously unreliable. As a result, drug quantity determinations are more likely than other fact-findings to erroneously extend a sentence, thereby wasting punishment resources and unfairly punishing less-culpable offenders.

A. Drug Quantity Determinations Are Particularly Common

Because drug quantity determines the base offense level for a drug-trafficking offense, these fact-findings are particularly common. As a result, they are responsible for a large proportion of the inefficiencies and injustices at sentencing.

For most crimes, the Guidelines assign a *specific* base offense level that the court can then adjust upward or downward based on the particular conduct of the offender. The base offense level for first-degree murder, for example, is 43.[140]

For drug crimes, by contrast, the Guidelines provide that the base offense level *depends* on the total quantity of drugs involved in the offense.[141] The Guidelines do not provide a set base offense level for a drug-trafficking crime. Instead, they instruct the court to determine the quantity of drugs involved in the offense, and then to use a "Drug Quantity Table" to convert that quantity into a base offense level.[142] So, for instance, if the court finds that one-kilogram of marijuana was involved in a crime, the base offense level

---

[139] *See* U.S.S.G. § 2D1.1. Calculating the total offense level for other drug-related offenses may also require drug quantity determinations; for example, use of a communication facility in committing a drug offense, *see* U.S.S.G. § 2D1.6, and narco-terrorism, *see* U.S.S.G. § 2D1.14. By contrast, the offense level for mere possession of a controlled substance does not depend on the quantity of drugs involved. *See* U.S.S.G. § 2D2.1.

[140] *See* U.S.S.G. § 2A1.1.

[141] *See* U.S.S.G. § 2D1.1(a)(5) & (c). There is an exception if death or serious bodily injury resulted from the use of the offender's drugs, in which case the Guidelines do prescribe a specific base offense level. *See* § 2D1.1(a)(1)-(4).

[142] *See* U.S.S.G. § 2D1.1(a)(5).

would be 10, but if the court finds that five kilograms of marijuana were involved, the base offense level would be 14.[143] After the base level has been decided, the court may then adjust that level by adding or subtracting other offense-specific and general sentence adjustments as it would for any other crime.

Sentences for drug-trafficking crimes are therefore "largely quantity-driven."[144] Sentence enhancements typically apply only sporadically – they are only relevant if the specific criminal conduct in question calls for them. So, for example, the enhancement for committing a hate crime will only come into play if the prosecution alleges at sentencing that the crime was motivated by racial animus.[145] But because drug quantity controls the base offense level for a drug-trafficking crime, it must be determined at each and every sentencing hearing for each and every drug-trafficking case.[146]

Keep in mind, as well, that drug offenses account for nearly one-third of all sentencings in the federal court system, which makes them the second-most commonly sentenced offense, just barely behind immigration offenses.[147] So, of all the punishment resources wasted and offenders wrongly incarcerated under the threshold model of sentencing, a large proportion can be traced back to drug offense sentencings, and of those, a large proportion can be traced even further back to the calculations of drug quantities made at those sentencings. To reduce the inefficiency and unfairness of the Sentencing Guidelines on a national scale, therefore, it would make sense to begin by reforming drug quantity determinations.

B. Drug Quantity Determinations Are Particularly Consequential

Because drug quantity determinations can produce the largest sentence variances in the federal Guidelines, they are also particularly consequential. Therefore, they have the potential to produce the most extreme inefficiencies and injustices at sentencing.

Most sentence enhancements, with some exceptions, range between one and ten levels, even for particularly heinous conduct. For example, a kidnapper's offense level will increase by two if he used a dangerous weapon,

---

[143] *See* U.S.S.G. § 2D1.1(c).

[144] United States v. Sepulveda, 15 F.3d 1161, 1196 (1st Cir. 1993); *see also* United States v. Rivera-Maldonado, 194 F.3d 224, 228 (1st Cir. 1999).

[145] *See* U.S.S.G. § 3A1.1(a).

[146] Of course, in many cases the offender will have pled guilty to a specific amount of contraband, which means the court will not need to independently calculate a quantity. However, not every plea includes an agreement on quantity. Moreover, plea negotiations take place in the shadow of the rules for sentencing, and so those rules still have an impact even in cases where the court does not itself perform the drug quantity determination. *See* Stephanos Bibas, *Plea Bargaining Outside the Shadow of Trial*, 117 HARV. L. REV. 2463, 2486-91 (2004).

[147] *U. S. Sentencing Commission's 2012 Sourcebook of Federal Sentencing Statistics, Figure A: Distribution of Offenders in Each Primary Offense Category, United States Sentencing Commission, http://www.ussc.gov/Research_and_Statistics/Annual_Reports_and_Sourcebooks/2012/FigureA.pdf.*

by four if the victim sustained permanent or life-threatening injury, and by six if he sexually exploited the victim.[148]

By contrast, a single drug quantity calculation can change an offender's offense level by up to *32 levels*. The First Circuit has referred to this phenomenon as the "dramatic leveraging effect" of drug quantity determinations.[149] The "Drug Quantity Table," used to convert the sentencing judge's drug-quantity determination into a base offense level, ranges from a minimum level of six (applied, for example, to crimes involving less than 250 grams of marijuana) to a maximum of 38 (applied to crimes involving more than 30,000 kilograms of the same).[150] The court's calculation of drug quantity can therefore swing the offender across 32 offense levels, more than any other factual finding in the federal Guidelines.

Of course, because drug quantity determines the base offense level, there is no reference-point against which to measure the size of this "enhancement." But remember that the Guidelines use a "real offense" approach to sentencing, which means that judges are not bound by the amount of drugs alleged in the indictment or proved to the jury.[151] As a result, an offender may be convicted based on a single transaction involving a relatively small quantity of drugs, but have his sentence dramatically lengthened when the sentencing judge independently finds that a much larger quantity of drugs was actually involved in the scheme.[152]

These stories are commonplace. Emilio Correa-Alicea, for example, was convicted of conspiring to possess-with-intent-to-distribute crack cocaine based on two sales of approximately 40 grams of the drug.[153] At sentencing, however, he was held responsible for 4.5 kilograms of crack, which effectively raised his offense level by 10.[154] Scott Jarvi was found with 22.73 grams of methamphetamine in his home and convicted of possession-with-intent-to-distribute, but he was sentenced based on a drug quantity of 853.05 grams, effectively raising his offense level by 12.[155] Levi Culps was caught distributing one kilogram of marijuana but sentenced based on the court's estimation that he was responsible for selling between 80 to 100 kilograms of the drug, an effective 16-level enhancement.[156] Kevin Townley was convicted of possession-with-intent-to-distribute 27 grams of cocaine but held responsible at sentencing for an additional six kilograms, an effective 18-level enhancement that corresponded to a *tenfold increase* in his recommended sentence.[157]

---

[148] *See* U.S.S.G. § 2A4.1(b)(2)(A) & (5).
[149] *Sepulveda*, 15 F.3d at 1198.
[150] *See* U.S.S.G. § 2D1.1(c).
[151] *See* United States v. Williams, 917 F.2d 112, 114 (3d Cir. 1990) (collecting cases).
[152] *See* United States v. Shonubi, 895 F. Supp. 460, 476 (E.D.N.Y. 1995) (Weinstein, J.).
[153] *See* United States v. Correa-Alicea, 585 F.3d 484, 489-90 (1st Cir. 2009).
[154] *See id.* at 490.
[155] *See* United States v. Jarvi, 537 F.3d 1256, 1258-59 (10th Cir. 2008).
[156] *See* United States v. Culps, 300 F.3d 1069, 1077 (9th Cir. 2002).
[157] *See* United States v. Townley, 929 F.2d 365, 367-68 (8th Cir. 1991).

Drug quantity determinations at sentencing, therefore, have the potential to cause some of the worst inefficiencies and injustices. Recall that the threshold model is inefficient because it extends offenders' sentences without regard to the possibility that that extended sentence will be wasted. The threshold model of sentencing will, accordingly, waste the *most* resources when a drug quantity determination results in the *longest* extension of an offender's sentence. Similarly, the threshold model is unfair because it increases offenders' punishments equally without crediting them for the possibility that they are innocent of the underlying culpable conduct. Once again, the model will cause the *most* unfairness when a drug quantity determination results in the *largest* increase in the offender's punishment. Reforming drug quantity determinations therefore has the potential to mitigate the most serious inefficiencies and injustices that are associated with the threshold model of sentencing.

C. Drug Quantity Determinations Are Particularly Unreliable

Because judges must often extrapolate drug quantities from a few key data points, these fact-findings are particularly unreliable. Consequently, they are more likely to extend offenders' sentences erroneously, resulting in wasted punishment resources and the unfair punishment of less-culpable offenders.

For most crimes, the Guidelines suggest that judges perform a relatively narrow inquiry into the "relative conduct" for which the defendant is responsible. The defendant's total offense level depends on his behavior "*during* the … offense of conviction, in *preparation* for that offense, or in the course of attempting to *avoid detection or responsibility* for that offense."[158] District courts will therefore focus only on what happened before, during, and after the offense.

For drug crimes, however the Guidelines instruct judges to consider a much wider range of conduct. This broader inquiry results from several unique rules. First, the "relevant conduct" inquiry is broader for drug offenders than it is for regular offenders. According to the Guidelines, drug offenders should be sentenced based on "all acts and omissions … that were part of *the same course of conduct or common scheme or plan* as the offense of conviction."[159] The sentencing judge is therefore not strictly limited to what happened before, during, and after the offense of conviction, but may also consider conduct "substantially connected [to that offense] by at least one common factor," as well as conduct "sufficiently connected [to that offense] … as to warrant the conclusion that they are part of a single episode, spree, or ongoing series of offenses."[160]

Second, "in the case of a jointly undertaken criminal activity," the Guidelines hold offenders responsible, for "all reasonably foreseeable acts and omissions of others in furtherance of the jointly undertaken criminal

[158] U.S.S.G. § 1B1.3(a)(1).
[159] U.S.S.G. § 1B1.3(a)(2) (emphasis added).
[160] U.S.S.G. § 1B1.3 n. 9.

activity."[161] Criminal conspiracies are, of course, quite common in drug-trafficking cases. The Guidelines therefore specifically instruct that drug offenders should be held responsible not only for "all quantities of contraband with which he was directly involved" but also "all reasonably foreseeable quantities of contraband that were within the scope of the criminal activity that he jointly undertook."[162]

Finally, and perhaps more importantly, the Guidelines do not limit judges to counting only quantities of drugs that have actually been seized. Instead, "[w]here there is no drug seizure or the amount seized does not reflect the scale of the offense," the Guidelines provide that the court should "approximate the quantity of the controlled substance."[163] As a result of these three rules, judges must perform a much broader inquiry into the quantity of drugs for which an offender is responsible than they do for most other fact-findings at sentencing.

Because "judges are forced to try to estimate the total quantity of drugs handled" by the offender and his co-conspirators, they must often resort to evidence that "differs greatly in both characteristic and quality."[164] This evidence may include inferences, extrapolations, and less-than-credible witness testimony.[165] Specifically, the Guidelines suggest that "the court may consider … the price generally obtained for the controlled substance, financial or other records, similar transactions in controlled substances by the defendant, and the size or capability of any laboratory involved."[166] Though these calculations are practically routine in drug-trafficking cases,[167] they "can be more accurately characterized as educated guesses."[168]

Take the relatively common scenario in which a defendant is caught with only a small amount of drugs but a large amount of cash. In that case, if the court finds that the money represents the proceeds from drug sales, it may convert the cash into an estimated total quantity of contraband sold based on a presumed price-per-unit.[169] Or, if a defendant is caught manufacturing drugs in a laboratory, the court may use the labels of leftover empty bottles in order to calculate the total drug quantity that he likely produced.[170] Even more

---

[161] U.S.S.G. § 1B1.3(1)(B).
[162] U.S.S.G. § 1B1.3 n.2.
[163] U.S.S.G. § 2D1.1, n.5
[164] Johan Bring & Colin Aitken, *Burden of Proof and Estimation of Drug Quantities Under the Federal Sentencing Guidelines*, 18 CARDOZO L. REV. 1987, 1987, 1990 (1997).
[165] *See* United States v. Shonubi, 895 F. Supp. 460, 476 (E.D.N.Y. 1995) (Weinstein, J.); *see also* United States v. Uwaeme, 975 F.2d 1016, 1019 (4th Cir. 1992).
[166] U.S.S.G. § 2D1.1, cmt. n. 5.
[167] *See* United States v. Marquez, 699 F.3d 556, 561 (1st Cir. 2012); *see also* United States v. Rios, 22 F.3d 1024, 1028 (10th Cir.1994); United States v. Colon, 961 F.2d 41, 43 (2d Cir. 1992); *Shonubi*, 895 F. Supp. at 475-78 (collecting cases).
[168] Bring & Aitken, *supra* note 164, at 1987.
[169] *See, e.g.*, United States v. Keszthelyi, 308 F.3d 557, 577 (2002); *Rios*, 22 F.3d at 1028; *see also* United States v. Ortiz–Martinez, 1 F.3d 662, 675 (8th Cir.1993); United States v. Duarte, 950 F.2d 1255, 1265 (7th Cir.1991); United States v. Jackson, 3 F.3d 506, 511 (1st Cir.1993).
[170] *See, e.g.*, United States v. Macklin, 927 F.2d 1272, 1282 (2d Cir. 1991); United States v. Smallwood, 920 F.2d 1231, 1236–37 (5th Cir. 1991); United States v. Evans, 891 F.2d 686, 687–88 (8th Cir.1989).

inferential is the "multiplier method" of drug quantity calculation, endorsed by several courts of appeal, used in cases where the evidence suggests that the defendant sold a certain quantity of drugs at a certain frequency over a certain period of time.[171] In that scenario, the district court may estimate a total drug quantity by "determining a daily or weekly quantity, selecting a time period over which it is more likely than not that the defendant was dealing in that quantity[,] and multiplying these two factors together."[172]

When drug quantities are calculated based on a few key data points, they are particularly vulnerable to error.[173] For instance, the court might misjudge the street price of the drug when converting cash to drug weight or overestimate the capacity of a drug-manufacturing defendant's laboratory.[174] Alternatively, the court might receive bad evidence on the number of drug sales a trafficker typically made, or the quantity of drugs sold in each transaction.[175] In each of these cases, a minor mistake would be multiplied into an enormous miscalculation – a phenomenon described as the "pyramiding [of] unreliable inferences."[176] Accordingly, even when they satisfy the preponderance-of-the-evidence standard of proof, drug quantity estimates based on inference and extrapolation will "inherently possess a degree of uncertainty."[177]

Courts of appeal have therefore repeatedly reversed drug quantity determinations that are overly reliant on extrapolations,[178] although plenty of others have been affirmed.[179] As compared to fact-findings based on hard evidence, then, drug quantity determinations are particularly likely to erroneously extend offenders' sentences, wasting punishment resources and unjustly extending prison terms. They are also particularly well suited to reforms that would reduce the negative impact of any judicial miscalculations.

VI. INCORPORATING PROBABILITY INTO DRUG QUANTITY DETERMINATIONS

This Part will show that, despite all these problems, there is some good news. While the flaws of the threshold model are particularly acute at sentencing, and are even worse when it comes to drug quantity determinations, sentencing is also especially amenable to probabilistic reform. Because the Sentencing Guidelines are not binding, district judges have the discretion to

---

[171] *See* United States v. Culps, 300 F.3d 1069, 1077 (9th Cir. 2002); *see also* United States v. Paulino, 996 F.2d 1541, 1548 (3d Cir.1993); *Colon*, 961 F.2d at 43; *Shonubi*, 895 F. Supp. at 477-78 (collecting cases).
[172] *Culps*, 300 F.3d at 1077.
[173] *See* United States v. Rivera-Maldonado, 194 F.3d 224, 231 (1st Cir. 1999); *see also* Bring & Aitken, *supra* note 164, at 1997.
[174] *See, e.g.*, *Shonubi*, 895 F. Supp. at 476.
[175] *Cf. id.* at 492.
[176] *Rivera-Maldonado*, 194 F.3d at 233.
[177] United States v. D'Anjou, 16 F.3d 604, 614 (4th Cir. 1994).
[178] *See, e.g.*, United States v. Shonubi, 998 F.2d 84, 89-90 (2d Cir. 1993); United States v. Garcia, 994 F.2d 1499, 1509 (10th Cir. 1993); United States v. Hewitt, 942 F.2d 1270, 1274 (8th Cir. 1991).
[179] *See, e.g.*, United States v. Thomas, 12 F.3d 1350, 1369 (5th Cir. 1994); United States v. McMillen, 8 F.3d 1246, 1249-51 (7th Cir. 1993); United States v. Sklar, 920 F.2d 107, 113 (1st Cir. 1990).

select punishments at the low end of the recommended Guidelines range, or even to vary downward from that recommendation, in cases where the recommended Guidelines sentence is not commensurate to the probability of the offender's culpability. And because the law of sentencing is quite flexible, policymakers in the criminal justice system – courts of appeal, the Sentencing Commission, Congress, and even the President – can implement reforms in order to mitigate the negative effects of the threshold model of sentencing.

A. The Promise of Probabilistic Reform at Sentencing

Three features of sentencing make it a better stage than conviction to incorporate probability into punishment. Sentencings are far more common than trials, the Sentencing Guidelines are not binding on district courts, and the law of sentencing is much more flexible than the law of conviction.

First, it is actually quite rare for the government to convict a defendant by proving his guilt at trial beyond a reasonable doubt. More than nine of out ten federal criminal defendants waive their trial right and instead plead guilty, making "[o]ur world … no longer one of trials, but of guilty pleas."[180] Even when they plead guilty, however, defendants still have a right to a sentencing hearing before a judge, who will calculate and impose an appropriate punishment using the Sentencing Guidelines procedure described earlier. Therefore, while changes to the threshold model of conviction would only impact the small percentage of defendants who go to trial, probabilistic reforms implemented at sentencing would affect every single convicted criminal defendant.[181] Both in terms of the number of defendants affected and the total number of years of incarceration at stake, probabilistic reform of the law of sentencing therefore has greater potential to meaningfully impact the administration of American criminal justice.

Second, district court judges are not bound to accept the sentences recommended by the Sentencing Guidelines. District courts must begin their analysis by calculating a recommended sentence through the Guidelines's threshold decision-making process.[182] However, the Supreme Court's decision in *Booker*[183] frees them to vary from that recommendation. As will be explained in a few moments, this flexibility permits district courts to vary from the recommended Guidelines sentence in cases where that recommendation overestimates the probability of the offender's overall culpability.[184] While judges are not free to innovate when it comes to determinations of guilt and innocence, then, they can use their discretion under *Booker* to incorporate probability into sentencing outcomes.

---

[180] Stephanos Bibas, *Judicial Fact Finding and Sentence Enhancements in a World of Guilty Pleas*, 110 YALE L.J. 1097, 1150 (2001).

[181] Although guilty pleas may also include admissions to specific factual predicates required for sentence enhancements, plea negotiations always take place in the shadow of the rules for sentencing. Therefore, reforming those rules along probabilistic lines will still have an impact even in cases where the court does not itself perform the relevant fact-finding. *See* Bibas, *supra* note 146, at 2486-91.

[182] *See* Gall v. United States, 552 U.S. 38, 50 n.6 (2007).

[183] 543 U.S. 220 (2005)

[184] *See infra*, Part VI.B.

Finally, the law of sentencing is more amenable to change. The presumption of innocence, along with the "beyond a reasonable doubt" standard of proof used at trial, are hard-coded into criminal proceedings under the Constitution's Due Process Clause.[185] Only a constitutional amendment could change these rules, which would require action from either two-thirds of both houses of Congress or a constitutional convention called by the states.[186] The law of sentencing, however, is far less constitutionally constrained – sentencing judges do not, in theory, decide guilt, and so the Due Process Clause bears less heavily on the proceedings.[187] In addition, there are several policymakers in the criminal justice system with the power to reform the law of sentencing. Courts of appeal, for instance, could tinker with the standard of proof at sentencing using the Due Process Clause.[188] The United States Sentencing Commission could change the Sentencing Guidelines that take better account of probability, and Congress could do the same by amending the Sentencing Reform Act. The President and Attorney General, too, could play a role, by considering the probability of the various factual predicates at issue when they decide which sentencing enhancements to seek.[189] The incorporation of probability into sentencing is therefore a policy reform within the power of several different actors.

B. Using District Court Discretion to Take Account of Probability

1. Doctrinal Justifications for Discretionary Probabilistic Sentencing

In each individual case, district court judges can use their sentencing discretion in order to take better account of probability when punishing criminal defendants. By sentencing less severely when the probability of the offender's culpability is lower, district courts will mitigate the inefficiency and unfairness of the threshold model of sentencing. Given the manifold problems with the threshold model of drug quantity determination,[190] district courts should especially consider exercising their discretion to vary downward from the Guidelines recommendation when they sentence drug offenders.

First, sentencing judges can simply select sentences at the low-end of the recommended Guidelines range. Remember that the end result of the

---

[185] *See* Apprendi v. New Jersey, 530 U.S. 466, 477 (2000); Sullivan v. Louisiana, 508 U.S. 275, 278, (1993); In re Winship, 397 U.S. 358, 364 (1970).
[186] *See* CONST. ART. V.
[187] *See* Williams v. New York, 337 U.S. 241, 246-52 (1949); *see also* McMillan v. Pennsylvania, 477 U.S. 79, 91 (1986); Spaziano v. Florida, 468 U.S. 447, 459 (1984); United States v. McDowell, 888 F.2d 285, 290–91 (3d Cir. 1989).
[188] *See, e.g.*, United States v. Kikumura, 918 F.2d 1084 (3d Cir. 1990), *overruled by*, United States v. Grier, 449 F.3d 558 (2006); *see also* United States v. Restrepo, 946 F.2d 654, 663-64 (9th Cir. 1991) (Pregerson, J., dissenting).
[189] The Obama Administration recently took similar action to implement reforms of the federal mandatory minimum sentencing regime. *See* Douglas A. Berman, *Some Sentencing-Related Highlights from AG Holder's Remarks Today to the ABA*, SENTENCING LAW AND POLICY BLOG (Aug. 12, 2013, 6:43 PM), http://sentencing.typepad.com/sentencing_law_and_policy/2013/08/some-sentencing-related-highlights-from-ag-holders-remarks-today-to-the-aba.html.
[190] *See supra* Part IV.

Guidelines calculation is not a specific sentence, but a range of months – for instance, a criminal history category of II and an offense level of 20 yields a recommended sentence of 37 to 46 months.[191] Therefore, even if a district court judge wants to stick with the Guidelines recommendation, she can still choose the lowest recommended sentence in a case where the recommendation overestimates the strength of the government's case. Although this will not have a major impact on the efficiency or fairness of the threshold model of sentencing, it will still reflect a small measure of progress.

Second, and more significantly, district court judges have the discretion to vary downward from the recommended Guidelines sentence if it is not justified by the weight of the evidence against the offender. Under the indeterminate sentencing regime that preceded the Sentencing Guidelines, courts could "ameliorate any adverse impact on defendants from unreliable fact-finding" by using their unbridled discretion to give lighter sentences in cases where the evidence of culpability was less persuasive.[192] Under the current system, judges should remember that despite the threshold nature of the Sentencing Guidelines, they still retain the power to vary from those Guidelines if the strength of the evidence against the offender does not support the recommended sentence.

The Second Circuit suggested a doctrinal foundation for this exercise of discretion in *United States v. Gigante*.[193] There, two senior members of the Genovese and Colombo crime organizations had been convicted of extortion and assigned base offense levels of 18, which would have resulted in recommended sentences of 27-33 months.[194] At sentencing, however, the district court applied several sentence enhancements related to the defendants' roles in the offense, the amount of money they extorted, and their attempts to obstruct justice.[195] All told, these upward adjustments raised the defendants' total offense levels to 34 and 35, and the district judge ultimately sentenced them to prison terms of 188 months and 200 months, respectively.[196] The defendants appealed, arguing that upward adjustments this substantial could not be premised on facts found by a mere preponderance-of-the-evidence without violating the Constitution's Due Process Clause.[197]

The Second Circuit rejected that argument, reaffirming that the preponderance standard applied at sentencing.[198] But the court also suggested a way that district courts could use their sentencing discretion to mitigate "the danger of factual error [that] would permeate a substantial upward departure … proven only by a bare preponderance."[199] The panel's logic proceeded as

---

[191] *See* U.S.S.G., Sentencing Table.
[192] Young, *supra* note 32, at 305-06; *see also id.* at 330.
[193] 94 F.3d 53 (1996).
[194] *See* United States v. Gigante, 39 F.3d 42, 44, 46 (2d Cir. 1994), *vacated and superseded in part on denial of rehearing by* United States v. Gigante, 94 F.3d 53 (2d Cir. 1996).
[195] *See id.*
[196] *See id.*
[197] *See id.* at 46-47.
[198] *See* United States v. Gigante, 94 F.3d 53, 56 (2d Cir. 1996).
[199] *Id.*

follows: a district court has the discretion to vary from a recommended Guidelines sentence if it finds that in the case before it "there exists an aggravating or mitigating circumstance of a kind, or to a degree, not adequately taken into considering by the Sentencing Commission in formulating the [G]uidelines."[200] The "risk of factual error" when applying an especially large sentence enhancement at a relatively low level of confidence (though still above 50%) "is a circumstance present at least 'to a degree' not adequately considered by the Commission."[201] In other words, because the Sentencing Commission has not yet taken into account the problems of the threshold model of sentencing, district courts have the power to vary downward from Guidelines' recommendations on this basis.

Indeed, the *Gigante* court did not merely recognize this as a permissible exercise of discretion; it specifically advised district courts to do so:

> In our view, the preponderance standard is no more than a *threshold* basis for adjustments and departures, and the weight of the evidence, at some point along a continuum of sentence severity, should be considered with regard to both upward adjustments and upward departures. With regard to upward adjustments, a sentencing judge should require that the weight of the factual record justify a sentence within the adjusted Guidelines range…Where a higher standard [of proof], appropriate to a substantially enhanced sentence range, is not met, the court should depart downwardly.[202]

The court further held that circuit courts "should take the weight of the evidence into account" when reviewing sentence enhancements on appeal.[203] Upward departures are reviewed for reasonableness, the *Gigante* court explained, and "[t]he reasonableness of substantial upward departures will depend in part on the standard of proof by which the conduct warranting the departure is established."[204]

2. An Example of Probabilistic Sentencing in a Drug Quantity Case

The case of *United States v. Mills*[205] offers a perfect example of a drug quantity determination that could have merited a downward variance for uncertainty. Mr. Mills was caught crossing the border from Canada with 8.5 grams of oxycodone hidden inside his rectum.[206] He was convicted of

---

[200] 18 U.S.C. § 3553(b)(1); *see also* U.S.S.G. § 5K2.0(2) & (3).
[201] *Gigante*, 94 F.3d at 56.
[202] *Id.* (emphasis in original).
[203] *See id.* at 57.
[204] *Id.*
[205] 710 F.3d 5 (1st Cir. 2013).
[206] *See id.* at 7.

unlawfully importing a controlled substance into the United States.[207] At Mills's sentencing hearing, however, the prosecution alleged that he was responsible for trafficking much more than 8.5 grams of contraband.

The prosecution presented evidence showing that over the course of several months before his arrest, Mills had exchanged $369,203 of American currency in Canada and had crossed the border from Canada into Maine over 200 times.[208] The prosecution also presented testimony from several confidential informants alleging that Mills had been smuggling drugs from Canada into the United States and then exchanging the proceeds for Canadian dollars upon his return, with which he bought more product.[209] According to the prosecution, Mills's money exchanges and border-crossings represented additional instances of drug smuggling, which "were part of the same course of conduct or common scheme or plan as [Mills's] offense of conviction."[210] The prosecution therefore asked the district court to hold Mills for an additional $369,203 worth of oxycodone, which at $100 per gram, would be equivalent to 295.4 grams of the drug – a figure 35 times greater than the quantity with which he was actually caught.[211]

Mills denied that he had ever brought drugs into the United States other than that one unlucky time that he had been caught. He also explained that he crossed the border daily in order to visit his longtime girlfriend in Canada, and that he had been exchanging American cash in Canada on behalf of a different drug dealer in return for free drugs for his own personal use, a story substantiated by a recorded jailhouse conversation.[212] Mills asked the court to sentence him based only on the 8.5 grams of oxycodone with which he was caught.[213]

Weighing these arguments, it seems clear that the government has the better of the case. Confidential informants corroborated the physical evidence against Mills, and Mills's story seems far-fetched. Still, Mills offered a plausible explanation for the border crossings, and confidential informants do sometimes misinform the police. Mills also presented extrinsic evidence to support his narrative about the money exchanges. His is not the more persuasive story, but it is certainly conceivable, and it would likely have been sufficient to raise reasonable doubts about the government's argument if presented at trial. If the sentencing court had been asked to attach probabilities to each side's version of the events, it might have concluded that the government's story was 75% likely to be true and that Mills's was 25% likely.[214] Or, if enumerating a precise probability for each side of the story

---

[207] *See id.* at 8; see also 21 U.S.C. 952(a).
[208] *See Mills*, 710 F.3d at 8.
[209] *See id.* at 9.
[210] U.S.S.G. § 1B1.3(a)(2).
[211] *See Mills*, 710 F.3d at 8.
[212] *See id.* at 9.
[213] *See id.*
[214] Even with these odds, moreover, the government might still have been wrong about some of the details of its case. The street price of oxycodone, for instance, might well be higher than $100, as Mills argued, *see id.* at 13, which would decrease the total quantity of drugs that the

would be too difficult, the court might have relied on existing standards of proof – perhaps the government had proved its case by a preponderance of the evidence, but not by clear-and-convincing evidence.

The threshold model of sentencing, however, asked the district court to answer only a single, simple question: was it more than 50% likely that Mills's cash exchanges and border-crossings represented instances of drug-trafficking? After considering the arguments, the court sided with the government. It concluded that the $369,203 reflected Mills's proceeds from selling oxycodone at approximately $140 per gram, and, dividing the money accordingly, found that Mills was responsible for trafficking 210.97 grams of the drug.[215] Between the 8.5 grams of oxycodone and the over 200 grams, Mill's offense level increased by 12 and his recommended sentence more than tripled, from roughly three years in prison to over 10.[216] Because Mills also received a three-level offense level decrease for accepting responsibility for his crime, the court ultimately sentenced him to 9 years in prison.[217]

If the district court had wanted to take a more probabilistic approach to Mills's punishment, it could begin by sentencing him at the low-end of the recommended sentencing range. This would reflect the court's judgment that Mills was more than 50% likely to have smuggled 295.4 grams of oxycodone, but that he had also offered a plausible counter-narrative. The impact would be rather minor – the difference in prison time at issue, not counting the three-level downward adjustment for acceptance of responsibility, would mean that Mills would receive a 121-month rather than 151-month prison sentence (the low and high ends of the recommended sentencing range, respectively). This technique would thus spare Mills a small, though meaningful, degree of punishment.

To take more substantial account of probability, the district court would have to break from the Sentencing Commission and vary downward from the recommended Guidelines sentence. Following the approach in *Gigante*, the judge would conclude that the dramatic increase in Mills' recommended sentence was not warranted in light of the relatively low probability that the prosecution's drug quantity argument was correct. The combination of a large offense level increase and a low level of confidence in the predicate fact is a mitigating circumstance "not adequately taken into consideration by the Sentencing Commission in formulating the [G]uidelines," and so it would

---

$369,203 in drug profits represented. Or, perhaps Mills only earned some of that cash by selling oxycodone.

[215] *See id.* at 13.

[216] As the *Mills* court explained, the Sentencing Guidelines use a formula to convert various drugs into units of marijuana in order to calculate a recommended sentence. *See id.* at 9. (citing U.S.S.G. 2D1.1 cmt. N.8(D)). 1 gram of oxycodone is equivalent to 6,700 grams of marijuana, so 8.5 grams oxycodone equals 56.95 kilograms of marijuana, and 210.97 grams equals 1413.499 kilograms. *See* U.S.S.G. 2D1.1 cmt. N.8(D). The Sentencing Guidelines assign a base offense level of 20 for drug crimes involving between 40 and 60 kilograms of marijuana, and an offense level of 32 for crimes involving between 1,000 and 3,000 kilograms. *See* U.S.S.G. 2D1.1(c); *see also* U.S.S.G., Sentencing Table.

[217] *See Mills*, 710 F.3d at 13.

justify a downward variance from the Commission's recommendation.[218] The court therefore could give Mills a sentence significantly below the one recommended by the Sentencing Guidelines, in order to account for the low probability of the fact-finding that yielded such a high recommendation. This approach would save the federal government's penal resources, and also a meaningful portion of Mills's life outside bars, based on the not-insignificant chance that the court's drug quantity determination was wrong.

To generalize, after a district court judge calculates the Guidelines sentence for a drug offender, she should make sure to reflect on her level of confidence in the drug quantity that she has attributed to that defendant. This is especially important when the quantity determination has significantly increased the defendant's total offense level. The judge need not compute a precise percentage level of confidence, since she can also rely on the existing standards of proof – proof by a preponderance of the evidence, proof by clear and convincing evidence, proof beyond a reasonable doubt – to guide her judgment. If, in the end, the judge is not totally confident in her quantity determination, she should remember, first, that she has the discretion to select a sentence on the low-end of the recommended Guidelines range. Second, she should consider that her low level of confidence in the quantity determination, combined with the size of the resulting offense level increase, may result in an unacceptable risk of error that is a "mitigating circumstance of a kind, or to a degree, not adequately taken into considering by the Sentencing Commission in formulating the [G]uidelines."[219] Therefore, if the district court judge believes that her level of confidence in her quantity determination is out of proportion to its corresponding effect on the defendant's offense level, she should feel free to vary downward from the recommended Guidelines sentence. By exercising this power, district judges can begin to immediately take better account of probability at sentencing.

C. Policy Reforms that Take Account of Probability

On a more systemic level, policymakers can reform the law of sentencing in order to incorporate probability into criminal punishment. The rules for drug quantity determinations, in particular, call out for change. A few courts and commentators have already hinted at possible approaches to this problem. Each approach, however, comes with its own drawbacks, and all will require clearer parameters in order to ensure that they are applied with precision and consistency.

1. Raising the Standard of Proof for Drug Quantity Determinations

First, the standard of proof at sentencing could be raised for drug quantity determinations. Several circuit courts have effectively done so, although they describe it as "err[ing] on the side of caution" in the calculation

[218] 18 U.S.C. § 3553(b)(1); *see also* U.S.S.G. § 5K2.0(2) & (3).
[219] 18 U.S.C. § 3553(b)(1); *see also* U.S.S.G. § 5K2.0(2) & (3).

of drug quantity.[220] In practice, "erring on the side of caution" is equivalent to raising the burden of proof for drug quantity, since when a district court favors lower quantity estimations, it effectively requires the prosecution to provide a higher quantum of evidence in order to prove a higher quantity of drugs. Courts that have suggested this approach have expressly linked it to the dramatic consequences of drug quantity determinations[221] and to due process concerns.[222] Alternatively, a few appellate courts have suggested more targeted approaches to adjusting the burden of proof in drug quantity calculations, by raising it only where a court is relying on uncertain witness testimony[223] or where the court is considering the offender's "relevant conduct" rather than the offense of conviction.[224]

In order to avoid sentencing disparities between judges who might adopt different understandings of the rather vague instruction to "err on the side of caution," courts of appeal would have to abandon euphemism and clearly instruct district courts to raise the standard of proof for drug quantity determinations a specific higher quantum – for instance, clear-and-convincing evidence or proof beyond-a-reasonable-doubt.[225] Of course, applying a higher standard of proof would not change the "threshold" nature of these decisions, but it would mitigate their inefficiency and unfairness. A raised standard of proof would reduce the frequency at which sentencing courts erroneously hold defendants responsible for excessive quantities of contraband, thus reducing the waste of punishment resources and the incarceration of less-culpable offenders.[226]

On the downside, a higher standard would also make it more likely that courts would underestimate the scale of offenders' drug-trafficking operations and impose too lenient punishments.[227] Still, the Guidelines provide a number of other upward adjustments that prosecutors could use against particularly dangerous or large-scale drug dealers,[228] and there is an emerging consensus in

---

[220] United States v. Walton, 908 F.2d 1289, 1302 (6th Cir. 1990); *see also* United States v. Culps, 300 F.3d 1069, 1076 (9th Cir. 2002); United States v. Mahaffey, 53 F.3d 128, 131-32 (6th Cir. 1995); United States v. Sepulveda, 15 F.3d 1161, 1198 (1st Cir. 1993). *But see* United States v. Kiulin, 360 F.3d 456, 461 (4th Cir. 2004).

[221] *Sepulveda*, 15 F.3d at 1198.

[222] United States v. Zimmer, 14 F.3d 286, 290 (6th Cir. 1994).

[223] *See* United States v. Sampson, 140 F.3d 585, 592 (4th Cir. 1998).

[224] *See* United States v. Restrepo, 946 F.2d 654, 663-64 (9th Cir. 1991) (Pregerson, J., dissenting); *see also* Young, *supra* note 32, at 354.

[225] The rather vague instruction that courts should "err on the side of caution" risks sentencing disparities between judges who adopt different understandings of that phrase.

[226] *See* Young, *supra* note 32, at 355.

[227] *See* United States v. Walton, 908 F.2d 1289, 1302 (6th Cir. 1990)s.

[228] *See, e.g.*, U.S.S.G. § 2D1.1(b)(1) ("If a dangerous weapon (including a firearm) was possessed, increase by 2 levels."); U.S.S.G. § 2D1.1(b)(2) ("If the defendant used violence, made a credible threat to use violence, or directed the use of violence, increase by 2 levels."); U.S.S.G. § 2D1.1(b)(11) ("If the defendant maintained a premises for the purpose of manufacturing or distributing a controlled substance, increase by 2 levels."); U.S.S.G. § 2D1.1(b)(14) ("If the defendant receives an adjustment under § 3B1.1 (Aggravating Role) and … [t]he defendant committed the offense as part of a pattern of criminal conduct engaged in as a livelihood[,] [i]ncrease by 2 levels.").

the United States that federal drug sentences are already overly punitive.[229] Accordingly, this could well be a trade-off worth making.

2. Deemphasizing Drug Quantity Determinations

Second, Professors Johan Bring and Colin Aitken have proposed that sentences for drug-traffickers should be less reliant on the total quantity of drugs involved in the offense.[230] To implement their proposal, Bring and Aitken suggest "a compromise between a charged-offense sentencing and a real-offense sentencing."[231] First, the offender's base offense level would be determined by the drug quantity proven to the jury at trial beyond a reasonable doubt. Second, the sentencing judge would enhance the base offense level by a fixed number of steps "based on whether the convict has been involved in similar [drug-dealing] activities besides those proven at trial or not."[232] Bring and Aitken make clear that "the exact quantity [of drugs] handled in these previous occasions" would not be relevant.[233] In other words, the increase in the offender's sentence would depend "on a qualitative variable (previous drug handling or not) rather than a quantitative (quantity previously handled)."[234] The remainder of the sentencing process would proceed as usual.

This approach has all the advantages of raising the standard of proof for drug quantity – since it includes that change as its first step – and also avoids some of its disadvantages. Large-scale drug-traffickers would not be able escape punishment as easily as they might if the standard of proof were raised alone, since sentencing courts could also enhance their sentences based on their involvement in related drug-dealing activities.

However, by delinking the offender's punishment from the quantity of drugs he handled, Bring and Aitken's proposal would permit two offenders who trafficked in vastly different quantities of drugs to receive the same sentence. Bring and Aitken acknowledge this as a problem, but they contend that there is "not much [moral] difference between smuggling, say, three or ten times."[235] That notion is certainly up for debate: if drug dealing is immoral because of the harm it causes to drug-users,[236] then dealing more drugs will cause even more harm and therefore be even more immoral. But even accepting Bing and Aitken's argument, moral retribution is only one of the reasons for criminal punishment. Deterrence, too, is an important consideration.[237] And in order to deter traffickers from seeking the larger

---

[229] *See, e.g.*, Douglas A. Berman, *Federal sentencing reform: an unlikely Senatorial love story and a Booker double-dose?*, SENTENCING LAW AND POLICY BLOG (Oct. 23, 2013, 6:43 PM), http://sentencing.typepad.com/sentencing_law_and_policy/2013/08/some-sentencing-related-highlights-from-ag-holders-remarks-today-to-the-aba.html.
[230] Bring & Aitken, *supra* note 164, at 1998.
[231] *Id.*
[232] *Id.*
[233] *Id.*
[234] *Id.*
[235] *Id.*
[236] *See* Terrebonne v. Butler, 848 F.2d 500, 504 (5th Cir. 1988).
[237] *See* 18 U.S.C. § 3553(a)(2)(B).

profits that come from selling larger hauls of contraband, sentences do need to have some relation to the quantity of drugs involved in the offense.

3. Incorporating Probability into Drug Quantity Determinations

Finally, probability could be directly incorporated into the Drug Quantity Table itself, so that the length of a drug offender's sentence would depend on the court's level of confidence in its drug quantity calculation. There are multiple possible ways to implement this reform. For instance, the Sentence Guidelines could provide an instruction that when judges consider uncharged quantities of contraband, they should scale the additional offense levels that those drugs would add based on their degree of certainty in the underlying quantity estimation.[238] Alternatively, the Guidelines might include a special downward adjustment for uncertainty, which would apply in cases where the sentencing court believes that the base offense level prescribed by the Drug Quantity Table overestimates the strength of the government's case.

A few district courts have already improvised a "discount" approach that is roughly similar to these suggestions. In so-called "multiplier cases," courts in the Third, Sixth, and Ninth Circuits sometimes "discount" their estimated drug quantity by some percentage – say, 50% – in order "to account for uncertainties and satisfy [their] duty to err on the side of caution."[239] As the Third Circuit explained in one such case: "The halving of [the total drug quantity calculated through the multiplier method] … is … a reasonable calculation by the district court, erring on the side of caution, to take into consideration 'off' days and days in which perhaps lesser sales occurred."[240] This method effectively correlates the total drug quantity to the sentencing judge's level of confidence in the underlying facts.

The Sentencing Commission has just recently voted to seek comment on "a proposed amendment to lower by two levels the base offense levels in the Drug Quantity Table across drug types," in order to make drug sentencing less punitive.[241] In addition to lowering the base offense levels for these crimes, the Commission should also follow the lead of the Third, Sixth, and Ninth Circuits and use probability to reduce drug sentences. The Commission could do so by adopting an official, and clear, "discount" rule that would decrease the offender's base offense level depending on the court's level of confidence in its drug quantity determination.

This "uncertainty discount" should be available in all drug-trafficking cases, not just those that use the "multiplier method" for estimating drug

---

[238] According to this approach, for instance, if an offender's base offense level was 10 according to the drug quantity proved at trial, and would be 20 according to the additional drug quantity found by the sentencing court at 60% certainty, then the additional 10 levels would be reduced by 40%, resulting in a final base offense level of 16.

[239] United States v. Culps, 300 F.3d 1069, 1081 (9th Cir. 2002); *see also* United States v. Paulino, 996 F.2d 1541, 1548 (3d Cir. 1993); United States v. Walton, 908 F.2d 1289, 1302 (6th Cir. 1990).

[240] *Paulino*, 996 F.2d at 1548.

[241] U.S. Sentencing Commission, News Release: U.S. Sentencing Commission Seeks Comment on Potential Reduction to Drug Trafficking Sentences, *supra* note 2, at 1.

quantity. After all, uncertainty about the court's drug quantity calculation is not unique to multiplier cases.[242] Furthermore, the Guidelines should define a precise trigger for the discount, so that it will be applied consistently. For instance, the adjustment could apply in cases where the court's drug finding would increase the offender's sentence by more than 8 levels as compared to the quantity proved at trial. In such cases, the "uncertainty discount" would instruct the district court to reduce the offender's offense level by some number depending on its level of confidence in the underlying drug quantity estimation. The Commission should again ensure that courts consistently apply these downward adjustments by keying them to the various burdens of proof. For example, the "discount" adjustment might instruct the court to reduce the offense level by 6 if it was only 50% sure the calculation is correct (a bare preponderance of the evidence), by 4 if it was 70% sure (clear and convincing evidence), by 2 if it was 95% sure (just over beyond-a-reasonable-doubt), and by none if it was over 95% certain (approaching absolute certainty).[243] This downward adjustment for uncertainty would blunt the impact of drug quantity sentence enhancements in cases where the incriminating facts were less likely to be true, and thereby make the drug trafficking Guideline more efficient and more just.

VII. CONCLUSION

Criminal trials use a "threshold model" of decision-making. If the likelihood of the defendant's guilt crosses a 95% "threshold" level of probability, then he is convicted without regard to whether the evidence against him just barely met that threshold or vastly exceeded it. By contrast, if the case against the defendant does not meet the threshold, then he is entirely acquitted, even if is still very probable that he committed the crime.

The law of sentencing also uses a threshold model of decision-making. If the likelihood that a factual predicate for an enhancement to the offender's sentence has been fulfilled is more than 50%, then the enhancement applies in full. If it is not, then the enhancement does not apply at all. The traditional justifications for the threshold model of conviction, however, do not apply to the threshold model of sentencing. Moreover, the threshold model of sentencing suffers from the same two flaws as the threshold model of conviction – inefficiency and unfairness – and each to a greater degree.

One area of sentencing especially in need of change is the rules for drug quantity determinations. Because these fact-findings are particularly frequent, consequential, and unreliable, they are responsible for some of the most severe inefficiencies and inequities at sentencing. The criminal justice system can begin to address these concerns if district courts exercise their discretion to depart downward from recommended sentences that overestimate the strength of the evidence against the defendant and if policymakers

[242] In a cash-conversion case, for example, some of the money in the offender's possession might have come from sources other than drug sales, or the offender might have occasionally charged a higher-than-average price.

[243] These particular numbers are merely offered for proof of concept, and are, of course, up for debate.

implement reforms that better incorporate probability into sentencing outcomes.

Of course, the flaws of the threshold model of sentencing are not unique to drug quantity determinations – the inefficiency and unfairness of this approach is a problem for every fact-finding made at every sentencing. Although drug quantity presents the most pernicious problem, each time a sentencing judge enhances an offender's sentence without regard to actual probability of the underlying fact-finding, she risks wasting punishment resources and unfairly extending the offender's time in prison. District judges should therefore consider using the probabilistic approach to drug sentencing outlined above[244] in all cases, particularly those involving especially large sentence enhancements. Criminal justice policymakers, too, should consider broader reforms of the kind described above[245] that would incorporate probability into the Guidelines calculations for all kinds of offenses, not just drug-trafficking ones. Ultimately, the failure to acknowledge the role of probability reflects a dangerous overconfidence that must be chastened at the heart of all legal decision-making, where present judgments about the past are treated as sacrosanct and the consequences of human fallibility are ignored.

Hoping to encourage a more effective and humane penal system, Professor Bill Stuntz once called for a restoration of "the quality of mercy" to American criminal justice.[246] He asked Americans to remember that "idea that once was so well understood that it needed never be expressed, yet now is all but forgotten: the idea that legal condemnation is a necessary but terrible thing – to be used sparingly, not promiscuously."[247] The threshold model of sentencing, unfortunately, reflects a proudly promiscuous philosophy of criminal punishment. It presents an image of a penal system that is incapable of error, where as soon as an offender's culpability has been proved in court, he is subject to the full punishment, without regard to any uncertainty about whether the conduct actually occurred. One way to build a more merciful model of criminal justice would be to recognize that the system is fallible, that probability matters, and that while the image of certainty may contribute to the esteem of the justice system, real people in the real world still suffer because of sentencing mistakes. If the penal system could learn to punish while also restraining itself in cases of doubt, it would operate more effectively and at the same time accommodate its own fallibility. Judge Learned Hand remarked, "The spirit of liberty is the spirit which is not too sure that it is right."[248] The American criminal justice system should honor that spirit by recognizing the relationship between probability and punishment.

---

[244] *See* supra at Part VI.B.
[245] *See* supra at Part VI.C.
[246] WILLIAM J. STUNTZ, THE COLLAPSE OF AMERICAN CRIMINAL JUSTICE 311 (2011).
[247] *Id.*
[248] LEND ME YOUR EARS: GREAT SPEECHES IN AMERICAN HISTORY 63 (William Safire, ed. 1997).