Apprendi and the Dynamics of Guilty Pleas

Stephanos Bibas

University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship

Part of the Criminal Law Commons, Criminal Procedure Commons, Criminology Commons, Public Law and Legal Theory Commons, and the Social Control, Law, Crime, and Deviance Commons

Repository Citation
https://scholarship.law.upenn.edu/faculty_scholarship/825

This Response or Comment is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.
Apprendi and the Dynamics of Guilty Pleas

Stephanos Bibas*

Professors Nancy King and Susan Klein devote most of their Commentary to a single subsection of my recent article. My entire article argued that Apprendi v. New Jersey exemplified criminal procedure’s misguided focus on jury trials at the expense of the real world of guilty pleas. Professors King and Klein focus on my narrower point that Apprendi undercuts due process by making it harder for many defendants to secure judicial hearings after they plead guilty. In summary, I argued that defendants used to be able to get the massive benefits of pleading guilty while still enjoying enhancement hearings at sentencing. Now that enhancements are issues for jury trials, defendants cannot gain both benefits. They must either allocate to and concede these enhancement issues to gain guilty-plea benefits or go to trial on enhancement issues and forfeit these plea benefits. Professors King and Klein claim that defendants face no additional pressure to give up hearings under this scheme. But they fail to see how prosecutorial and judicial behavior reinforce the pressures to plead guilty, making hearings harder to secure for many defendants.

First, they note that even before Apprendi prosecutors had the same bargaining chips to induce guilty pleas. Both before and after Apprendi, defendants who pleaded guilty enjoyed 35% sentence reductions for accepting responsibility, could avoid recidivism and perjury enhancements, and could gain other benefits as well. What Professors King and Klein miss is that Apprendi has changed the worth of these bargaining chips. Before Apprendi, prosecutors could use these bargaining chips to force guilty pleas. But

---

* Associate Professor, University of Iowa College of Law (bibas@philo.org). B.A., Columbia; B.A., M.A., Oxford; J.D., Yale. I am indebted to George Fisher and Dan Richman for their comments on an earlier draft.

2. 530 U.S. 466 (2000).
3. King & Klein, supra note 1, at 397 & n.16.
4. See Bibas, supra note 1, at 1153-54.
5. Professors King and Klein assume that rational prosecutors would simply use these chips to push for the highest possible sentences in all cases. See King & Klein, supra note 1.
defendants could reap the benefits of pleading guilty while still getting enhancement hearings at sentencing. Now, however, enhancements are elements of the offense. Pressure to plead guilty simultaneously pressures defendants to give up enhancement issues. To this extent, defendants lose hearing rights and are worse off.

Professors King and Klein object to my considering the possibility of guilty pleas, without plea agreements, followed by sentencing hearings, as most

at 296. This assumption ignores my point that prosecutors seek not only to maximize sentences, but also to minimize trials and workloads by trading lower sentences for pleas. See George Fisher, *Plea Bargaining’s Triumph*, 109 YALE L.J. 857, 865, 882-83, 893-903 (2000). Thus, even if the prosecution has a good shot of winning at trial, it may not press certain enhancements against a defendant who would otherwise plead guilty because the threat of a huge enhancement may induce the defendant to try to avoid the enhancement by rolling the dice at trial. And in many cases, the prosecution’s only option is to press the entire enhancement (forcing a trial) or drop it entirely (in return for a plea). Massive enhancements (such as recidivism enhancements) are so large and discrete that they operate as sledgehammers, not scalpels. They can be traded off to prevent trials but cannot be parceled out more finely to tailor the terms of a particular plea. See Bibas, *supra* note 1, at 1153-54 n.342.

Professors King and Klein further suggest that the parties may avoid most of the costs of trial by agreeing to expedited bench trials on enhancements. King & Klein, *supra* note 1, at 306 n.42. While this procedural vehicle may eventually evolve, by and large it has not done so yet, and prosecutors have little incentive to make it easier for defendants to secure hearings. More hearings would give judges more opportunities to check prosecutorial charging and plea decisions, which judges and defendants might favor but prosecutors would not. In the federal system, 28 states, and the District of Columbia, prosecutors can unilaterally veto bench trials and thwart this maneuver. Bibas, *supra* note 1, at 1155 n.346 (also noting that a 29th state forbids bench trials entirely regardless of the parties’ consent).

Finally, Professors King and Klein claim that the Department of Justice’s Thornburgh memorandum prevents prosecutors from forgoing readily provable enhancements. King & Klein, *supra* note 1, at 297 n.18. When Janet Reno succeeded William Barr as Attorney General, however, she promulgated further guidance to federal prosecutors that gave them more leeway in deciding which charges to press. See Memorandum from Attorney General Janet Reno to all Holders of U.S. Attorney’s Manual (Reno Bluesheet on Charging and Plea Decisions), Oct. 12, 1993, reprinted in 6 FED. SENTENCING REP. 352 (1994) (endorsing plea bargaining “on the basis of an individualized assessment of the extent to which particular charges fit the specific circumstances of the case, are consistent with the purposes of the federal criminal code, and maximize the impact of federal resources on crime”). Compare this with United States Attorneys’ Manual § 9-27.400 (1999); Professors King and Klein, at p.3 n.18, quote the manual as saying that prosecutors should not bargain away readily provable charges, but they fail to note an exception in the same section that allows supervisors to approve charge bargaining for other reasons, such as lightening a heavy prosecutorial workload. And whatever the formal doctrine on paper, prosecutors do in practice take into account their caseloads and trial burdens in their charging and plea decisions. See Joseph A. Colquitt, *Ad Hoc Plea Bargaining*, 75 TUL. L. REV. 695, 707-10 (2001) (summarizing various empirical studies, all of which found that bans on plea bargaining broke down, and suggesting that plea bargaining may be inevitable); Robert A. Weninger, *The Abolition of Plea Bargaining: A Case Study of El Paso County, Texas*, 35 UCLA L. REV. 265, 265, 311-13 (1987) (noting that empirical study showed that plea bargaining resurfaced in one county despite an official ban and suggesting that evidence shows that bargaining is inevitable regardless of official policy).
issues are resolved as part of plea bargains and so actual sentencing hearings are uncommon. 6 This objection misses my point. Before Apprendi, defendants had the realistic option of pleading guilty without agreements and insisting on enhancement hearings at sentencing. Plea bargaining took place in the shadow of this option, forcing prosecutors to provide additional consideration and strike reasonable deals in return for waivers of these realistic hearing rights. Prosecutors had to purchase these waivers, so the rights led prosecutors to make lower plea offers, even when defendants ultimately did not exercise these rights. This option, however, is foreclosed by Apprendi. In exchange for taking away these realistic hearing rights, it gave defendants theoretical jury-trial rights that they cannot afford to exercise lest they forfeit the benefits of pleading guilty. Prosecutors know that for most defendants the threat of going to trial is implausible, so these trial rights do not cast serious shadows over most plea bargaining. Knowing that judges have much less power to check their bargains, prosecutors can now drive harder bargains. 7

Professors King and Klein respond that after Apprendi defendants could plead guilty to base offenses or offer to do so, reaping the benefits of pleas while still enjoying trials on enhancements. Prosecutors need not charge lesser-included offenses, however, and courts are unlikely to let defendants plead guilty to lesser-included offenses if prosecutors have not charged them. 8 True, one circuit has granted acceptance-of-responsibility credit to a defendant who offered to plead guilty to two base offenses, was convicted at trial, and was sentenced based on exactly the drug quantity to which he had offered to plead guilty. 9 Three other circuits have taken a contrary approach, however. 10 And

6. King & Klein, supra note 1, at 297 n.14, 298 n.19.
7. Bibas, supra note 1, at 1159-60, 1165.
8. Though I have found no law on this precise point, the closest analogue is most courts’ refusal to give lesser-included offense instructions in non-capital cases unless prosecutors have chosen to charge the lesser-included offenses. See Michael G. Patillo, Note. When “Lesser” Is More: The Case for Reviving the Constitutional Right to a Lesser Included Offense. 77 TEX. L. REV. 429, 453-59 (1998) (noting that a plurality of federal circuits have rejected any right to lesser-included-offense instructions in non-capital cases and only one circuit has embraced such a right across the board).
9. See United States v. Guerrero-Cortez, 110 F.3d 647, 653-56 (8th Cir. 1997). The second case cited by Professors King and Klein, United States v. McKinney, 15 F.3d 849, 851-54 (9th Cir. 1994), is inapposite. It involved no issues of base offenses versus enhanced offenses or quantities, but simply a confused, contrite defendant who merited acceptance-of-responsibility credit because (1) the court had prevented him from pleading guilty, (2) he had made a full confession right after arrest, (3) he had assisted the authorities, and (4) he had put on a “minimal and perfunctory” defense at trial. In addition, after McKinney the Sentencing Commission amended the relevant Sentencing Guidelines commentary. The amended language draws a clear distinction between defendants who go to trial and those who plead guilty, and it forbids acceptance-of-responsibility credit for those who contest factual guilt at trial. See United States v. Dia, 69 F.3d 291, 293 (9th Cir. 1995) (distinguishing McKinney on this and other grounds).
10. E.g., United States v. McLaurin, 57 F.3d 823, 827-28 (9th Cir. 1995) (holding that district court did not err in denying acceptance-of-responsibility credit to defendant who had
the federal Sentencing Guidelines commentary denies acceptance-of-responsibility credit to defendants who deny "the essential factual elements of guilt" at trial.11 Perhaps other courts will refuse to apply this plain language to enhancement trials or the Sentencing Commission will amend it; perhaps not. Even if defendants do get acceptance-of-responsibility credit, judges still have discretion to sentence within the resulting ranges. In practice, judges sentence leniently those who plead guilty and spare the courts trials, while being harsher on those who insist on trials.12 In short, defendants must now forfeit at least some of the benefits of pleading guilty if they want hearings on enhancements at trial. Apprendi forces defendants onto the horns of this dilemma.

Professors King and Klein further claim that my example of federal drug sentencing is atypical of the criminal justice system, though more than a third of federal inmates are charged with drug crimes.13 It is true that peculiar features of the federal drug laws and guidelines (such as relevant conduct, mandatory minima, and recidivism enhancements) exacerbate the Apprendi problem. But the root of the problem lies not in these features, but in the

offered to plead guilty to the only count on which he was eventually convicted, where defendant did not actually make a pre-trial confession to all of the elements of that offense; distinguishing McKinney as turning on McKinney’s confession upon arrest and repeated expressions of contrition); United States v. Clark, 25 F.3d 1051 (table), No. 93-1418, 1994 WL 194286, at *3 (6th Cir. May 16, 1994) (unpublished per curiam) (holding that defendant who had offered to plead guilty to all of the charges of which he was ultimately convicted was not entitled to acceptance-of-responsibility credit and would not have been entitled to this credit automatically even if he had entered guilty pleas); see also United States v. Best, 139 F.3d 908 (table), No. 97-30172, 1998 U.S. App. LEXIS 1875, at *8 (9th Cir. Feb. 9, 1998) (unpublished per curiam memorandum opinion) (holding that defendant’s timely offer to plead and eventual guilty plea did not entitle defendant to full three-level reduction for acceptance of responsibility, where defendant balked at allocating to the full loss amount that judge ultimately found at sentencing hearing); United States v. Jones, 899 F.2d 1037, 1100-01 (11th Cir. 1990) (holding that concession of guilt of the only offense of which defendant was ultimately convicted did not entitle defendant to acceptance-of-responsibility reduction, where district court could have discounted this concession as a trial tactic rather than a sincere expression of remorse), overruled in part on other grounds by: United States v. Morrill, 984 F.2d 1136, 1137 (11th Cir. 1993) (en banc; per curiam).


12. Gerald W. Heaney, The Reality of Guidelines Sentencing: No End to Disparity, 28 AM. CRIM. L. REV. 161, 170 nn.28, 29 (1991) (summarizing empirical study of Guidelines sentencing, which showed that judges are more likely to sentence those who plead guilty at the bottom of the sentencing range and are more likely to sentence those who go to trial at the top of the sentencing range). Judges might come to see how Apprendi hurts defendants’ hearing rights. They might compensate by signaling that they will be more lenient on defendants who go to trial simply to have hearings on enhancements. Bibas, supra note 1, at 1153 n.341, 1154 n.345. But as long as judges accept the conventional wisdom that Apprendi is an unalloyed good for defendants, they will likely resent defendants who insist on enhancement trials.

13. Bibas, supra note 1, at 1167 & n.385 (collecting sources that show that 36.2% of federal defendants in 1999-2000 were charged with drug crimes, 58.9% of federal inmates in 1998 were imprisoned for drug-related offenses, and almost a quarter of all federal and state inmates are serving sentences for or awaiting trial on drug crimes).
massive discounts for guilty pleas common to most crimes and jurisdictions. Federal and many state defendants must now choose between enhancement hearings and the massive discounts for pleading to every element. This dilemma simply did not exist before Apprendi. 14

Professors King and Klein also overstate how much Apprendi strengthens hearings by requiring proof beyond a reasonable doubt of enhancements. 15 This supposed benefit is largely illusory, as legislatures and prosecutors will circumvent it. Even if prosecutors fail to prove an enhancement beyond a reasonable doubt at trial, the Federal Sentencing Guidelines allow them to try to prove it again at sentencing by only a preponderance of the evidence. This use of relevant conduct at sentencing swallows up the reasonable-doubt safeguard at trial. 16 Judges can do more or less the same thing under state indeterminate-sentencing schemes by using their broad discretion to punish the defendant’s actual conduct over and above the conduct of conviction. In other words, judges who sentence within broad sentencing ranges can look at defendants’ uncharged behavior in deciding where within the range to sentence. For example, a judge sentencing an embezzler within a zero-to-twenty-year range can impose the maximum sentence on finding that the embezzler committed an uncharged murder. 17

The only limit on using uncharged conduct at sentencing to circumvent the reasonable-doubt standard at trial is set by the statutory maximum. So where statutory maxima are already high, this rule does little good. Legislatures can get around Apprendi’s rule simply by raising maxima, which renders Apprendi
toothless. And, as Professors King and Klein suggest in a footnote, prosecutors can get around this rule as well. They need only charge the same transaction as a conspiracy plus multiple substantive counts, then stack maxima by asking judges to impose multiple consecutive sentences.* Any halfway-clever prosecutor can do so.

18. The Apprendi majority recognized that legislatures could “hypothetically” evade its rule by raising statutory maxima. The majority thought that “structural democratic constraints” would stop legislatures from doing so, but this claim is unconvincing for a variety of reasons that I explore in detail in my original article. Apprendi, 530 U.S. at 490 n.16; Bibas, supra note 1, at 1136-38. Indeed, Professors King and Klein have written elsewhere that legislatures are likely to circumvent Apprendi and that the rule in Apprendi itself is open to circumvention. See Nancy J. King & Susan R. Klein, Essential Elements, 54 Vand. L. Rev. 1467, 1488-95 (2001). One need not be a hardened cynic to see that legislatures have some incentive to look tough on crime by trying to counteract Apprendi’s new right, and according to Professors King and Klein they have repeatedly done so in the past. Id. My point is not that legislatures should or should not act this way. Rather, in light of this historical experience, courts should take this legislative propensity into account when considering what form their rules should take. If they are to adopt rules at all, courts should try to structure them in such a way that legislatures cannot circumvent them so easily.

19. Bibas, supra note 1, at 1157 & n.353; King & Klein, supra note 1, at 302 n.29. Professors King and Klein note some dissension among lower courts on the propriety of this tactic, but most reported cases rely on the availability of consecutive sentences to cure any Apprendi error. See, e.g., United States v. Duarte, 246 F.3d 56, 62 n.4 (1st Cir. 2001) (noting that courts have held that any Apprendi errors were not plain where consecutive sentences were available); United States v. Smith, 240 F.3d 927, 930 (11th Cir. 2001) (per curiam) (finding that Apprendi error did not prejudice defendant where consecutive sentences could have reached the same result), United States v. White, 240 F.3d 127, 135 (4th Cir. 2001); United States v. Angle, 254 F.3d 514, 518 (4th Cir. 2001) (en banc) (holding that any Apprendi error did not affect substantial rights because Guidelines would have produced same result via consecutive sentences in any event); United States v. Parolin, 239 F.3d 922, 930 (7th Cir. 2001) (holding that stacked consecutive sentences that effectively raised the statutory maximum did not violate Apprendi), United States v. White, 238 F.3d 537, 542-43 (4th Cir. 2001) (same as Angle); United States v. Sturgis 238 F.3d 956, 960-61 (8th Cir. 2001); United States v. Page, 232 F.3d 536, 543-45 (6th Cir. 2000) (declining to notice unpreserved Apprendi error where defendants in any event would have been imprisoned for the same period through the imposition of consecutive sentences), cert. denied, 121 S. Ct. 1389 (2001); State v. Gambrel, No. 2000-CA-29, 2001 WL 85793, at *5-*7 (Ohio App. 2d Dist. Feb. 2, 2001), People v. Wagener, No. 88843, 2001 WL 587044, at *10 (Ill. June 1, 2001); People v. Caruth, 322 Ill. App. 3d 226, 2001 WL 599716, at *5 (Ill. App. Ct. May 31, 2001) (holding that Apprendi does not restrict judges’ power to determine at sentencing whether to run sentences concurrently or consecutively); see also People v. Cleveland, 104 Cal. Rptr. 2d 641, 645-46 (Cal. App. 2001) (holding that Apprendi does not forbid stacking sentences for separate counts consecutively rather than concurrently); People v. Martinez, 2001 WL 360836, at *8-10 (Colo. Ct. App. Apr. 12, 2001) (holding that imposition of consecutive sentences did not violate Apprendi). But cf. United States v. Vasquez-Zamora, No. 99-51182, 2001 WL 585127, at *2 (5th Cir. May 31, 2001) (holding that district court could have reached same result via consecutive sentences but that appellate court could not do so in the first instance); United States v. Jones, 235 F.3d 1231, 1238 (10th Cir. 2000) (holding that appellate court could not treat Apprendi error as harmless simply because trial court could have reached same result via consecutive sentences, but not holding that district court could not achieve same result via consecutive sentences on remand). The clustering of these cases in federal and a few state courts may indicate that Apprendi has
Professors King and Klein further catalogue a number of cases that have reduced sentences in the wake of *Apprendi*.

These transitional disruptions do not speak to the long-term impact of *Apprendi* on hearings. In other words, this new rule gives a short-term windfall to some defendants whose sentences violated *Apprendi*. This windfall will not continue now that legislatures, prosecutors, and judges know *Apprendi*'s strictures and can circumvent them by raising maxima, charging the same transaction in multiple counts, or using consecutive sentences.

Furthermore, Professors King and Klein’s focus on reported decisions skews the picture. The beneficiaries of *Apprendi*, who persuade courts to reduce sentences, are visible in the appellate reports. The victims, who have to enter guilty pleas and forgo hearings, will not enjoy judicial fora and hearings. The low-visibility world of plea bargaining may fly beneath the radar of reported decisions, but this is no reason to ignore it.

Finally, Professors King and Klein suggest that my article mistakenly exalts judges as “provid[ing] process that is better than the process jury trials could provide.”

In the real world of guilty pleas, however, where fewer than 4% of defendants ever get to juries, this rhetorical invocation of juries is almost pointless. The real institutional competition is among judges, prosecutors, and legislators, not juries. The jury-based system that grew up to check prosecutors no longer works now that juries are all but gone. Prosecutors already have plenty of power in charging and sentencing, and one cannot realistically abolish this power. The only choices are to leave this power more or less unfettered or to give judges countervailing power to check it at sentencing. One does not have to worship judges to see that some check on prosecutors is better than none. This check need not be in the hands of judges alone. Sentencing caused the most disruption in the federal system and a few states such as Illinois, California, Colorado, and Ohio, but it is hard to draw solid conclusions from such a small sample size. See supra note 14 (discussing why *Apprendi* will affect some states more severely than others).

These decisions did not turn on the procedural obstacles to raising post-conviction claims. Rather, they reasoned that appellate judges should affirm sentences where trial judges could and would have used consecutive sentences to achieve the same result in the first instance. So these cases pave the way for future prosecutors to charge multiple counts and for future trial judges to impose consecutive sentences after *Apprendi*, allowing them to reach the same results as enhancements without using the forbidden enhancement mechanism. It is hard to see how the Court could or would extend *Apprendi*’s limit on sentencing-enhancement procedures to plug this loophole by regulating charging practices and concurrent and consecutive sentences.

20. King & Klein, supra note 1, at 300 n.23.

21. Perhaps the Court will at some point in the future come up with a rule that regulates these practices, but *Apprendi* itself does not do so. See *Apprendi*, 530 U.S. at 490 n.16 (acknowledging that legislatures could “hypothetically” evade the *Apprendi* rule by raising maxima). Given the slender one-vote majority in *Apprendi*, it seems unlikely that the Court will extend *Apprendi* far enough to regulate prosecutors’ traditionally unreviewable charging discretion and legislatures’ almost unfettered latitude to set maximum sentences.

22. King & Klein, supra note 1, at 310.
commissions, probation and parole officers, and appellate courts can also contribute to real-offense sentencing, keeping both prosecutors and judges honest. To make this scheme work, legislatures should simplify the thicket of duplicative statutory offenses, minima, and maxima—not the venerable gradations set up by the common law, such as the ancient distinction between murder and manslaughter, but the fine and somewhat arbitrary distinctions drawn by thousands of modern statutes. There is plenty of room to prune back the complexity of the criminal law toward the simplicity it had decades or centuries ago. This step would limit prosecutors’ power to dictate sentences by charge bargaining and give judges more power to check them. In the real world of guilty pleas, the key is to give judges and other actors enough room to counterbalance prosecutors, lest prosecutors alone run the show.

23. See Ashe v. Swenson, 397 U.S. 436, 445 n. 10 (1970) (bemoaning the move from a few distinct categories of offenses at common law to the recent “extraordinary proliferation of overlapping and related statutory offenses, [which has allowed] prosecutors to spin out a startlingly numerous series of offenses from a single alleged criminal transaction”); see also Albert W. Alschuler, The Prosecutor’s Role in Plea Bargaining, 36 U. CHI. L. REV. 50, 77-78 (1968) (acknowledging that “the distinctions drawn by the criminal code . . . sometimes prove too fine for workable, everyday application” and suggesting that the solution is “simplification of the criminal code to reflect ‘everyday reality’ rather than common-law refinement”). Indeed, Professor Klein herself has criticized the “drastic” growth of “thousands” of overlapping criminal statutes, which has given prosecutors too much power to overcharge and pursue successive prosecutions. See Susan R. Klein & Katherine P. Chiarello, Successive Prosecutions and Compound Criminal Statutes: A Functional Test, 77 TEX. L. REV. 333, 358-60 (1998).

24. Professors King and Klein point to benefits that flow from grading offenses, but we could reap most of these benefits by having legislatures outline culpability factors rather than rigid maxima, or having sentencing commissions set up more flexible grades as part of a real-offense system. There are many ways to achieve these benefits, and in this footnote I can sketch only a few suggested methods. First, legislatures could draft sentencing enhancements designed to evade Apprendi, by for example setting high maxima and then designating facts that, if found by sentencing judges, would lower those maxima. Second, legislatures (or sentencing commissions) could set forth factors and criteria that should guide sentencing discretion, much as they already do in capital sentencing. Judges, rather than applying mathematical formulae, would weigh and balance these criteria and give reasoned explanations of why the factors and criteria led them to particular sentences. Third, unless and until the Court extends Apprendi to sentencing guidelines, sentencing commissions can continue to use guideline enhancements to tailor punishment to defendants’ real offenses, checking prosecutors’ power to dictate sentences through their charging and plea-bargaining decisions.