3-1-2018

Non-Prosecution of Corporations: Toward a Model of Cooperation and Leniency

Cindy R. Alexander

Yoon-Ho Alex Lee

Follow this and additional works at: http://scholarship.law.unc.edu/nclr

Part of the Law Commons

NON-PROSECUTION OF CORPORATIONS: TOWARD A MODEL OF COOPERATION AND LENIENCY[*]

CINDY R. ALEXANDER AND YOON-HO ALEX LEE[**]

We apply the Kaplow-Shavell model of optimal law enforcement to study the effects of prosecutors' use of non-prosecution agreements (NPAs) to obtain cooperation on broader enforcement objectives, including deterrence of crime. The NPA policy of the Department of Justice is documented in a series of memos that provide guidance to federal prosecutors on the charging of corporations. Prosecutors may offer NPAs as alternatives to a plea agreement in exchange for the company's authentic cooperation. A benefit of authentic cooperation is to reduce the prosecutors' costs of case development, both post-referral and pre-trial. But in order for the NPA to induce cooperation, the company must regard it as more lenient (or no less lenient) than the plea settlement. Thus, one concern regarding the use of a NPA is that the leniency it provides may, if anticipated, undermine general deterrence. For this reason, the prosecutor who seeks to maximize general deterrence may be more cautious in offering NPAs than one who is primarily concerned about minimizing the use of federal resources. A closer look at the tradeoffs reveals strategic benefits of the use of NPAs beyond the resource savings from cooperation. Using our basic model application as a guide, we conclude that the policy of limiting the use of NPAs to cases where the company provides authentic cooperation serves several enforcement objectives. From a traditional social welfare perspective, the efficiency of the

NPA relative to a plea depends on whether the value of the resources saved through cooperation—in the form of increased ex ante probability of sanction faced by the offender—exceed the direct loss of deterrence due to the leniency of the sanction needed to obtain cooperation under the policy, other things equal. We also conclude that the use of NPAs—with or without cooperation—can facilitate efficient substitution between the informal sanctions that attach to criminal conviction, which can be socially costly (a deadweight loss), and the monetary sanction (a transfer). We suggest extensions of the model in which the effect of the NPA is to facilitate substitution into more efficient forms of settlement than those that would occur through plea agreements alone.

INTRODUCTION

In 1994, John Hancock Mutual Life Insurance ("John Hancock") settled charges that it had "repeatedly violat[ed] Massachusetts state law by giving legislators gifts worth more than $50," such as theater and sports tickets and, in one instance, an expense-paid trip to the Super Bowl.[1] After a lengthy investigation, the federal government entered into a non-prosecution agreement ("NPA") with the company in which it admitted to violations of the state criminal law, "paid a substantial fine, and implemented structural, management and policy changes to prevent the recurrence of the matters under investigation."[2] "[T]he government [also] levied a $1.01 million fine against John Hancock, [and] mandated that [it] implement a more stringent corporate compliance policy, fire or reassign[] two employees … , and continue[] to cooperate."[3] In other words, the government and John Hancock had a formal agreement, not filed with or supervised by any court, which stipulated that the government would drop the charge in exchange for John Hancock's ongoing cooperation and acceptance of the terms of settlement. Although it may have been costly for John Hancock to fulfill the terms of the agreement, at the end of the day, it avoided having to plead guilty to the charges and avoided further sanctions and adverse consequences. The government, on the other hand, was able to avoid the expenditure of resources needed to convince John Hancock to enter a guilty plea.

What happened between the government and John Hancock is an early example of the growing numbers of agreements that have

---

1. Ralph F. Hall, *Deferred Prosecution and Non-Prosecution Agreements*, *in* PUNISHING CORPORATE CRIME: LEGAL PENALTIES FOR CRIMINAL & REGULATORY VIOLATIONS 119, 129 (James T. O'Reilly et al. eds., 2009); s*ee* John Hancock Mut. Life Ins. Co., 1994 Rulings 646 (Mass. State Ethics Comm'n 1994), https://www.mass.gov/files /documents/2017/11/15/1994%20State%20Ethics%20Commission%20Rulings%20final.pdf [https://perma.cc/AG2Y-B7B6]. The Chief of the Criminal Division of the United States Attorney's Office noted that Hancock paid the same level of penalty it would have paid under a criminal prosecution, that Hancock's exceptional cooperation was a factor in the decision to enter into the civil agreement rather than pursue a criminal charge, and that the agreement did not prevent prosecution of employees for the misconduct. *See John Hancock Mutual Life Insurance Company Enters Into Civil Settlement, Announces U.S. Attorney*, PR NEWSWIRE, Mar. 22, 1994, Infotrac Newsstand, p0322NE009. NPAs were also entered into with Salomon Brothers in 1992 and with Sequa and Aetna in 1993. Hall, *supra*, at 128–31.

2. *Id.* at 129 (internal citation omitted).

3. *Id.*

since taken place between federal prosecutors and corporations that are suspected of committing misconduct. A NPA is an agreement between the prosecutor and the company, without any direct judicial supervision, in which the prosecutor agrees not to prosecute in return for cooperation and other concessions.[4] In some instances, an agreement between the prosecutor and the company may take the form of a deferred prosecution agreement ("DPA"). A DPA is filed with a court; the prosecutor offers to defer any prosecution until a certain date and to drop the case if the company fulfills some obligations by that date.[5] Such was the case with Prudential Securities, which settled with the U.S. Attorney in Manhattan through a DPA in 1994.[6] Because neither the NPA nor the DPA entails the corporate defendant pleading guilty, we refer to them as *non-plea settlements*.

The prosecutor's growing use of non-plea settlements raises a number of questions as a matter of law and policy. This Article sidesteps the legal questions and seeks to examine the practice from a policy perspective using an economic approach. For example, what would motivate the prosecutor to seek a NPA as opposed to a plea? What tradeoffs would the prosecutor face in choosing between traditional and alternative forms of settlement? Should the public be concerned that the practice of non-plea settlements may undermine the deterrent effect of enforcement and thereby lead to more crimes? How might the prosecutor's choice affect social welfare? To answer these questions, this Article considers the parties' incentives in reaching this type of non-plea criminal settlement and considers the benefits and costs the prosecutor may face. Our understanding of non-plea settlements and their operation derives from a series of memoranda issued by the Department of Justice ("DOJ"), which provide prosecutors with specific guidance in bringing criminal

---

4. *See* Cindy R. Alexander & Mark A. Cohen, *The Evolution of Corporate Criminal Settlements: An Empirical Perspective on Non-Prosecution, Deferred Prosecution, and Plea Agreements*, 52 AM. CRIM. L. REV. 537, 545 (2015)

5. *Id.*

6. *See* Kurt Eichenwald, *Brokerage Firm Admits Crimes in Energy Deals*, N.Y. TIMES, Oct. 28, 1994, http://www.nytimes.com/1994/10/28/us/brokerage-firm-admits-crimes-in-energy-deals.html?pagewanted=all [https://perma.cc/JX33-SCTG (dark archive)] (noting that Prudential Securities admitted to wrongdoing and cooperated with investigators, the U.S. Attorney in Manhattan was confident the government could have obtained a conviction, and factors including cooperation led to a decision to resolve the matter by deferring prosecution); *Interview with Mary Jo White*, *Partner*, *Debevoise & Plimpton LLP*, *New York*, *New York,* CORP. CRIME REP. (Dec. 12, 2005), http://www.corporatecrimereporter.com/news/200/interview-with-mary-jo-white-debevoise -new-york-new-york/ [https://perma.cc/3WFD-S9ES].

charges against corporations. We then employ a standard economic model of settlement as a guide to the analysis.

To highlight the overall effect of having prosecutors confer leniency in exchange for obtaining the firm's cooperation, we compare two extreme forms of criminal settlement—NPA versus traditional plea. We view the role of NPA as designed primarily to reduce prosecutors' costs of case development—post-referral and pre-trial—by way of seeking cooperation from corporate defendants. This makes sense. A NPA can shelter the firm from costs of sanction beyond the formal monetary sanction and thereby confer leniency relative to what would occur under a plea, which would bring a criminal conviction as well as other related risks to the company. The central focus of this Article is the question of whether, and under what conditions, the use of the NPA to obtain cooperation promotes deterrence and achieves other distinct enforcement objectives.

Our approach is to explore the enforcement authority's ("EA") incentive to allow or encourage prosecutors to rely on non-plea settlements to close criminal investigations of corporations across various scenarios.[7] In the simplest case, we consider a prosecutor who is myopic and cares only about the budget. In the most nuanced setting, we consider a prosecutor who regards today's decisions as a signal of tomorrow's policy, bringing both the budget and intangible effects of the enforcement decision to the forefront. In evaluating the practical relevance of the implications of the model, we consider the institutional setting and some of the evidence that has begun to emerge regarding NPA versus plea settlements. We identify questions for future empirical research.

The outline of the Article is as follows. Part I contains institutional background and reviews the practical differences between traditional and alternative forms of settlement. It also introduces cooperation and discusses NPA as a form of leniency that, when offered in exchange for cooperation, can promote the enforcement objectives outlined in the series of memoranda issued by the DOJ. In Parts II and III, we use a formal model to investigate the effects of changes in the objectives of the prosecutor on the choices that affect the form of settlement. Part IV discusses other possible

---

7. We use the terms "prosecutor" and "enforcement authority" interchangeably in this Article except when referring to decisions that necessarily commit resources beyond the disposition of an individual case and would be made at the level of the Department of Justice or government (enforcement authority) rather than, for example, the Assistant U.S. Attorney (prosecutor). We recognize, however, that the locus of authority is not always clear-cut and can vary depending on the resource question or the case being resolved.

effects of the use of NPAs, as alternatives to plea agreements. Part V concludes. The Appendix includes technical proofs.

I. INSTITUTIONAL BACKGROUND

Debate over best practices for penalizing corporations for criminal misconduct dates back more than a century. In *New York Central & Hudson River Railroad Co v. United States*,[8] the Supreme Court established that corporations can be held criminally liable for actions of their employees that occur in the scope of their employment with at least some intent to benefit the employer.[9] Because a corporation cannot go to jail and lacks a human mind, the extension of this principle from the law governing civil liability to criminal law has led to debate over the proper form of the sanction that remains an active focus of legal scholars and practitioners.[10] A criminal proceeding involving a corporation is a multi-stage process that begins with the detection of misconduct. Detection itself may be a result of an inspection, initial inquiry, or preliminary investigation. The case may then be referred to a prosecutor for a full investigation. Although investigations can be undertaken by a number of different government agencies, the DOJ has the sole authority to prosecute offenders under federal criminal law.[11] Depending on the strength and the nature of the misconduct, other agencies can decide to seek a civil or administrative sanction, and can refer the cases to the DOJ for criminal prosecution. Close coordination between the DOJ and other agencies of the government became commonplace in the investigation of corporate criminal misconduct after the scandals of 2002.[12] The DOJ investigation of the firm may be expedited by the early cooperation on the part of the target firm in providing facts or access to evidence, should the firm choose to cooperate. Ultimately, the

---

8. 212 U.S. 481 (1909).

9. *Id.* at 493–94 (noting public policy reasons for imposing a fine on a corporation that profits from the action of an agent when the action is taken using authority delegated by the corporation to the agent).

10. *See* Conceição Soares, *Can Corporations Be Criminally Responsible?*, 3 INT'L J. HUMAN. & SOC. SCI. 45, 46–50 (2013); David McCluskey, *Corporate Liability: What Exactly Does It Mean to Prove a Company 'Guilty'?*, TAYLOR WESSING (Jan. 26, 2016), https://united-kingdom.taylorwessing.com/en/corporate-liability-what-exactly-does-it-mean-to-prove-a-company-guilty [https://perma.cc/JS7X-TBC5].

11. *See* 28 U.S.C. § 547 (2012).

12. For example, the Corporate Fraud Task Force was established by executive order on July 9, 2002, to facilitate coordination among prosecutors and between prosecutors and regulators affected by the financial reporting scandals of the era. *See* Exec. Order No. 13,271, 67 C.F.R. § 46091 (2002), *terminated by* Exec. Order No. 1,351,974, 74 C.F.R. § 60123 (2009).

process leads from detection to a DOJ investigation to determine whether the facts are sufficient to support the assignment of criminal liability and impose criminal sanction on the offending firm (and any culpable individuals).[13]

After the investigation by the government, if there is insufficient evidence to move forward with the charge, the prosecutor may decline to prosecute. Otherwise, the process may end with a settlement or a trial. In practice, criminal charges against public companies are nearly always settled rather than taken to trial.[14] This is not necessarily a bad practice. Legal scholars have previously noted the economic benefits of settling criminal charges in the case of corporations.[15] The important question is the terms of settlement and, in particular, how they might affect would-be offenders' incentives.

In deciding on terms of settlement, prosecutors, alongside company management, exercise conditional discretion. Notwithstanding the two examples mentioned in the opening paragraphs, non-plea settlements were rare prior to 2003.[16] The standard form of settlement was a plea agreement that involved the filing of an information or indictment, with a guilty plea and thus a criminal conviction for the defendant company.[17] In negotiating a plea agreement, the prosecutor is constrained by the strength of the case based on the available evidence, which affects the credibility of the threat that a jury would rule in favor of the defendant if the case were to go to trial.

Against this backdrop, the DOJ released a series of memoranda to prosecutors on best practices for resolving criminal investigations

---

13. We use the term "inspection" to refer to the prosecutor's act of following up on leads and tips to determine whether it is unlawful, and we use the term "investigation" to refer to the collection of facts about an action that the prosecutor believes to be unlawful for the purpose of determining whether there is enough evidence to justify a criminal sanction.

14. For example, according to the U.S. Sentencing Commission's sourcebook, in 2016, a guilty plea was entered in 97.7% of all cases sentenced under the Chapter Eight Organizational Sentencing Guidelines, with 2.3% going to trial. *See* U.S. SENTENCING COMM'N, 2016 SOURCEBOOK OF FEDERAL SENTENCING STATISTICS, *at* S-131 (Table 53) (July 2016), https://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2016/Table53.pdf [https://perma.cc/FVW4-YCCJ].

15. *See, e.g.*, Bruce H. Kobayashi, *Case Selection, External Effects, and the Trial/Settlement Decision*, *in* DISPUTE RESOLUTION: BRIDGING THE SETTLEMENT GAP 17, 17–31 (David A. Anderson ed., 1996) (highlighting the efficiency gains in settling criminal charges against corporations).

16. A study of criminal settlement agreements entered by public companies, pre-2003 versus post-2003, found only two non-plea settlements in 1997–2002 (2%); higher shares were found after 2003, with 40 (37%) being non-plea agreements in 2003–07, and 115 (44%) entered in 2007–11. *See* Alexander & Cohen, *supra* note 4, at 569.

17. *Id.* at 543.

866   *NORTH CAROLINA LAW REVIEW*   [Vol. 96

of business organizations. The series began with a 1999 memorandum from then deputy attorney general Eric Holder ("Holder Memo").[18] This memo predates the collapse of Arthur Anderson in 2002[19] and the current debate over the form of sanction. It sets forth a framework to guide the prosecution of criminal conduct by corporations regardless of whether the charging decision is being made by a U.S. Attorney's Office or in the main building of the

18. Memorandum from Eric H. Holder, Jr., Deputy Attorney Gen., U.S. Dep't of Justice, to All Component Heads & U.S. Attorneys, Bringing Criminal Charges Against Corporations (June 16, 1999) [hereinafter Holder Memo], http://www.justice.gov/criminal/fraud/fcpa/docs/response2-appx-k.pdf [https://perma.cc/T8UR-EG3V]. This Memo has been revised by subsequent Deputy Attorney Generals. *See* Memorandum from Mark R. Filip, Deputy Attorney Gen., U.S. Dep't of Justice, to Heads of Dep't Components & U.S. Attorneys, Principles of Federal Prosecution of Business Organizations (Aug. 28, 2008) [hereinafter Filip Memo], http://www.justice.gov/dag/readingroom/dag-memo-08282008.pdf [https://perma.cc/88WG-3S57]; Memorandum from Gary G. Grindler, Acting Deputy Attorney Gen., U.S. Dep't of Justice, for Heads of Dep't Components & U.S. Attorneys, Additional Guidance on the Use of Monitors in Deferred Prosecution Agreements and Non-Prosecution Agreements with Corporations (May 25, 2010), https://www.justice.gov/usam/criminal-resource-manual-166-additional-guidance-use-monitors-dpas-and-npas [https://perma.cc/5E5V-4EY4]; Robert D. McCallum, Jr., Acting Deputy Attorney Gen., U.S. Dep't of Justice, to Heads of Dep't Components & U.S. Attorneys, Waiver of Corporate Attorney-Client and Work Product Protection (Oct. 21, 2005) [hereinafter McCallum Memo], http://federalevidence.com/pdf/Corp_Prosec/McCallum_Memo_10_21_05.pdf [https://perma.cc/4GDY-PPFK]; Memorandum from Paul J. McNulty, Deputy Attorney Gen., U.S. Dep't of Justice, to Heads of Dep't Components & U.S. Attorneys, Principles of Federal Prosecution of Business Organizations (Dec. 12, 2006) [hereinafter McNulty Memo], https://www.justice.gov/sites/default/files/dag/legacy/2007/07/05/mcnulty_memo.pdf [https://perma.cc/5BAV-UQQK]; Memorandum from; Memorandum from Craig S. Morford, Acting Deputy Attorney Gen., U.S. Dep't of Justice, to Heads of Dep't Components & U.S. Attorneys, Selection and Use of Monitors in Deferred Prosecution Agreements and Non-Prosecution Agreements with Corporations (Mar. 7, 2008) [hereinafter Morford Memo], https://www.justice.gov/sites/default/files/dag/legacy/2008/03/20/morford-useofmonitorsmemo-03072008.pdf [https://perma.cc/Q6VF-3PK2]; Memorandum from David W. Ogden, Deputy Attorney Gen., U.S. Dep't of Justice, for Dep't Prosecutors, Issuance of Guidance and Summary of Actions Taken in Response to the Report of the Department of Justice Criminal Discovery and Case Management Working Group (Jan. 4, 2010), https://www.justice.gov/archives/dag/memorandum-department-prosecutors [https://perma.cc/5NJ4-7MXB]; Memorandum from Larry D. Thompson, Deputy Attorney Gen., U.S. Dep't of Justice, to Heads of Dep't Components & U.S. Attorneys, Principles of Federal Prosecution of Business Organizations (Jan. 20, 2003) [hereinafter Thompson Memo], http://federalevidence.com/pdf/Corp_Prosec/Thompson_Memo_1-20-03.pdf [https://perma.cc/GM9K-PZRJ]; Memorandum from Sally Quillian Yates, Deputy Attorney Gen., U.S. Dep't of Justice, to U.S. Attorneys et al., Individual Accountability for Corporate Wrongdoing (Sept. 9, 2015) [hereinafter Yates Memo], https://www.justice.gov/archives/dag/file/769036/download [https://perma.cc/9LH8-SGRQ].

19. Carrie Johnson, *Arthur Andersen to be Sentenced Today*, WASH. POST (Oct. 16, 2002) https://www.washingtonpost.com/archive/business/2002/10/16/arthur-andersen-to-be-sentenced-today/c9b6095e-36ec-48d5-962c-06a454bef645/?utm_term=.82be0da9f090 [https://perma.cc/92W7-D583].

DOJ.[20] Further, the framework applies regardless of the applicability of the organizational sentencing guidelines of the U.S. Sentencing Commission. As the framework has evolved, the factors serve mostly as guidance to reflect the varying cultures and conditions of the different departments and offices of the DOJ that may apply them.[21]

As we explain below, all memoranda beginning with the Holder Memo emphasize *cooperation* as a consideration in the charging of corporations.

A.   *Guidance for Prosecutors: Valuing Cooperation by Corporations*

The recent history of public statements on best practice for charging corporations for criminal misconduct begins with the Holder Memo in 1999.[22] The Holder Memo set forth eight factors for prosecutors to consider in deciding whether to prosecute a case, including the "corporation's timely and voluntary disclosure of wrongdoing and its willingness to cooperate in the investigation of its agents."[23] It also reviews various forms of cooperation that a corporation might offer the prosecutor. For example, "the prosecutor may consider the company's willingness to identify culprits within the corporation," "make witnesses available," "disclose . . . results of internal investigations" and "waive the attorney-client . . . privilege[]."[24] The prosecutor also may consider "whether the corporation is seeking immunity for its employees and officers," "whether the corporation is willing to cooperate in the investigation of culpable employees," and, more generally, whether "the cooperation is complete and truthful."[25]

The prosecutor may, in exchange for these various forms of

---

20. *See* Holder Memo, *supra* note 18, at 3.

21. The cultural reach of the most recent memorandum to prosecutors is apparent in the inclusion of divisions that prosecute cases to which the Organizational Sentencing Guidelines do not apply. Specifically, the Yates Memo is addressed not just to the division and offices that were part of the original corporate fraud task force, but also to the Assistant Attorneys General of the Antitrust Division and the Environmental and Natural Resources Division. *See* Yates Memo, *supra* note 18.

22. *See* Holder Memo, *supra* note 18, at 1.

23. *Id.* at 3 (instructing prosecutors to consider the following eight factors: (1) "nature and seriousness of the offense, including the risk of harm to the public"; (2) "pervasiveness of wrongdoing within the corporation"; (3) "corporation's history of similar conduct"; (4) "corporation's timely and voluntary disclosure of wrongdoing and its willingness to cooperate"; (5) "existence and adequacy of the corporation's compliance program"; (6) "corporation's remedial actions"; (7) "[c]ollateral consequences"; and (8) "adequacy of non-criminal remedies").

24. *Id.* at 6.

25. *Id.* at 12.

cooperation, grant a corporation immunity or amnesty.[26] It is in this context that the Holder Memo refers the prosecutor to principles under U.S. Attorney's Manual ("USAM") Sections 9-27.600 to .650[27] that "'permit a *non-prosecution agreement* in exchange for cooperation' when a corporation's timely cooperation appears necessary to the public interest and other means of obtaining the desired cooperation are unavailable or would not be effective."[28] Thus, from the very first of the memoranda regarding Department policy on the charging of corporations, prosecutors are advised on the use of NPAs in exchange for cooperation.

Four years later, then deputy attorney general Larry D. Thompson issued another memorandum ("Thompson Memo").[29] The Thompson Memo places a greater emphasis on the "authenticity" of a corporation's cooperation when considering leniency and encourages the use of alternative resolutions to seek greater cooperation from corporate defendants.[30] The Thompson Memo refers to offers of "pretrial diversion" in addition to amnesty in exchange for cooperation.[31] As before, the prosecutor is referred to the general principles governing NPAs, citing the USAM.[32] The Thompson

---

26. *See id.* at 6.

27. U.S. Dep't of Justice, U.S. Attorneys' Manual, §§ 9-27.600 to .650 (2017) https://www.justice.gov/usam/usam-9-27000-principles-federal-prosecution [https://perma.cc/VH2T-FRHT].

28. *See* Holder Memo, *supra* note 18, at 6 (emphasis added).

29. *See* Thompson Memo, *supra* note 18.

30. *See id.* at 1 ("The main focus of the revisions [set forth in this memorandum] is increased emphasis on and scrutiny of the authenticity of a corporation's cooperation."). The Thompson Memo introduces a new "factor to be considered" in charging a corporation, specifically, "the adequacy of the prosecution of individuals responsible for the corporation's malfeasance." *Id.* at 3; *see also* Christopher A. Wray & Robert K. Hur, *Corporate Criminal Prosecution in a Post-Enron World: The Thompson Memo in Theory and Practice*, 43 Am. Crim. L. Rev. 1095, 1103 (2006) ("In conducting a Thompson Memo analysis, prosecutors examine where a company's response to a government investigation falls on a continuum between genuinely assisting the government and affirmatively impeding it.").

31. *See* Thompson Memo, *supra* note 18, at 5–6.

32. *See id.* at 6 ("[P]rosecutors should refer to the principles governing non-prosecution agreements generally."); *see also* U.S. Dep't of Justice, *supra* note 27, at § 9-27.600. These principles permit a non-prosecution agreement in exchange for cooperation when a corporation's "timely cooperation appears to be necessary to the public interest and other means of obtaining the desired cooperation are unavailable or would not be effective." As Professor David M. Uhlmann points out, this language is similar to that of the Holder Memo except that it introduces "pretrial diversion" as an alternative to immunity or amnesty in obtaining cooperation. *See* David M. Uhlmann, *Deferred Prosecution and Non-Prosecution Agreements and the Erosion of Corporate Criminal Liability*, 72 Md. L. Rev. 1295, 1311 (2013) (discussing how the Thompson Memo introduced "pretrial diversion" for corporations and led to the widespread use of non-prosecution agreements for corporations).

Memo might appear to constrain the use of NPAs by introducing a requirement that they may only be entered into with the approval of each affected district or the appropriate Department official.[33] Yet we suggest that the approval process could alternatively facilitate the development of standards within the divisions and offices of the Department regarding the use of NPAs. Indeed, following the Thompson Memo's release, the use of NPAs soon became widespread.[34] Consistent with the increased emphasis on cooperation from the corporation in resolving criminal investigations of alleged misconduct in this era, the McCallum Memo on October 21, 2005 requires heads of Department components and U.S. Attorneys to establish written-waiver review processes.[35]

In 2006, then deputy attorney general Paul J. McNulty issued a memorandum ("McNulty Memo") that makes it mandatory for prosecutors to consider the factors that had been offered as guidance in considering whether to prosecute corporations, while allowing for judgment by prosecutors in the weighing of those factors.[36] As experience using non-plea settlement agreements grew, guidance became more specialized and focused. The McNulty Memo and its successors addressed questions about the form of cooperation and what the prosecutor was getting in exchange for non-plea settlements, accordingly. The McNulty Memo specifically concerns the use of waivers as a channel for obtaining information from within the firm.[37] It sets forth principles for the prosecutor to consider in determining whether to request waivers of attorney-client and work product privileges as well as how the company's response to a waiver request should affect the severity or leniency of any settlement.[38] Throughout the memo, the role of the waiver is to expedite the investigation of the corporation and the identification of culpable individuals within the corporation prior to settlement.[39]

In each instance, the objective of cooperation in exchange for non-prosecution of the corporation is to expedite the investigation of the corporation leading to settlement. Yet a critical part of the investigation of corporate crime from the perspective of both the prosecutor and the corporation is the discovery of the cause so that it may be corrected to avoid a harmful future recurrence of the

---

33. *See* Thompson Memo, *supra* note 18, at 6.
34. *See, e.g.*, Alexander & Cohen, *supra* note 4, at 567 fig.2.
35. *See* McCallum Memo, *supra* note 18, at 1.
36. *See* McNulty Memo, *supra* note 18, at 4.
37. *See id.* at 8–11 (discussing the use of waivers to obtain critical information).
38. *See id*.
39. *See id.*

misconduct. Consistent with this objective, settlement agreements often contain provisions that commit the company to reforms, depending on the type of misconduct.[40] Some of the reforms are mandates to facilitate a commitment by the firm to better internal policing.[41] In 2008, then acting attorney general Craig S. Morford issued a memorandum ("Morford Memo") to address the use of monitors and to explain their role in monitoring and assessing corporate compliance with agreement terms that are designed to reduce the risk of a repeat offense.[42] Unlike previous guidance on the decision to charge a corporation, the Morford Memo highlights the role of cooperation in making it easier for the prosecutor to obtain information about the quality of the corporation's compliance post-settlement.[43]

Later that year, then deputy attorney general Mark Filip issued a memorandum ("Filip Memo") providing further clarification on the use of "cooperation" as a mitigating factor.[44] Prosecutors were specifically instructed to assess whether corporate defendants disclosed "relevant facts" for prosecution.[45] In doing so, the Filip Memo can be seen as calling for greater attention to the substantive impact of the cooperation offered in exchange for a non-plea settlement of a corporate criminal investigation.

In 2015, then deputy attorney general Sally Quillian Yates issued a memorandum ("Yates Memo") providing the most recent refinement to the Department's guidance.[46] While the Filip Memo may have confirmed that there was no requirement for cooperation in the form of waivers,[47] the Yates Memo requires that no weight be assigned to cooperation as a factor in charging the corporation unless the corporation has identified all individuals of the company that were involved in the offense and all facts related to their involvement

---

40. *See, e.g.*, Alexander & Cohen, *supra* note 4, at 589 tbl.13.

41. *See, e.g.*, Jennifer Arlen & Marcel Kahan, *Corporate Governance Regulation Through Nonprosecution*, 84 U. CHI. L. REV. 323, 353 (2017) (finding that traditional liability regimes should be supplemented by mandates when a firm struggles with significant policing agency).

42. *See* Morford Memo, *supra* note 18.

43. *See id.* at 5–6 (discussing the scope of monitor's duties).

44. *See* Filip Memo, *supra* note 18.

45. *Id.* at 9–13.

46. *See* Yates Memo, *supra* note 18, at 1.

47. *See* Filip Memo, *supra* note 18, at 8 ("[W]aiving the attorney-client and work product protections has never been a prerequisite under the Department's prosecution guidelines for a corporation to be viewed as cooperative.").

in the offense.[48] In principle, the effect is to discourage partial cooperation and encourage full cooperation. Thus, companies that find it worthwhile to cooperate fully but would otherwise have cooperated only partly will provide more information to prosecutors under the memo than otherwise. For other companies, the incentive is to provide less cooperation. The intent of the memo is to eliminate obstacles to the assignment of criminal liability on individuals within the corporation and not just on the corporation itself as an investigative outcome.[49]

As this series of memos indicates, the DOJ has long recognized the policy of offering a NPA as a form of leniency in exchange for cooperation as part of best practice. Although the policy on the form of cooperation has varied over time, the objective in each instance is to expedite the investigation of the corporation and the identification of culpable individuals within the corporation. The company may report the offense in advance of its detection by the EA (self-reporting).[50] Alternatively, the company may cooperate by eliminating obstacles that the prosecutor might otherwise face in collecting facts about an already-detected offense, such as by allowing access to witnesses or work product, as opposed to impeding "the quick and effective exposure of the complete scope of wrongdoing under investigation."[51] Further, the company may aid the prosecutor in collecting facts about a known offense, such as by taking steps to identify the culprits within the corporation and by conducting an internal investigation of the alleged misconduct and disclosing the results to the prosecutor.[52] Finally, the corporation may accept terms of settlement that eliminate opportunities to engage in future misconduct or, of importance to the prosecutor, remove obstacles to the future investigation of the corporation should the misconduct recur.[53] In this Article, we focus on the use of leniency to eliminate obstacles to the collection of facts about an already-detected offense.

---

48. *See* Yates Memo, *supra* note 18, at 3 ("To be eligible for any cooperation credit, corporations must provide to the Department all relevant facts about the individuals involved in corporate misconduct.").

49. *See id.* at 2 (discussing the challenges of identifying culpable individuals).

50. *See, e.g.,* Thompson Memo, *supra* note 18, at 6 ("Some agencies . . . have formal voluntary disclosure programs in which self-reporting, coupled with remediation and additional criteria, may qualify the corporation for amnesty or a reduced sanction. Even in the absence of a formal program, prosecutors may consider a corporation's timely and voluntary disclosure . . . .").

51. *Id.* at 1.

52. *Id.* at 5.

53. *See* Morford Memo, *supra* note 18, at 5 (discussing the monitor's activities post-settlement may include reporting to the government on the corporation's compliance).

B.   *Non-Plea Settlements as Instruments for Obtaining Cooperation*

As mentioned in the Introduction, the DPA and NPA are two novel forms of settlement that have arisen after the release of guidance to prosecutors regarding cooperation by corporations.[54] The corporation in each instance avoids a plea agreement and thus felony criminal conviction in a court proceeding.[55] It does not avoid admitting to the wrongful conduct nor the payment of a monetary sanction and other costs of settlement, however.[56] If, at the end of the term of the agreement, the corporation has followed through on its obligations, the prosecutor will dismiss the charges.[57] Under both DPAs and NPAs, the company is released from the obligations of the agreement after a specified period of time.[58] The company may face a lesser risk of costly collateral effects of the sanction, such as delicensing or debarment, under either of these forms of settlement than it faces under a plea agreement with a criminal conviction.[59]

The difference is that, with a NPA, there are no formal charges and there is no court filing of the settlement.[60] There is no obvious channel through which third parties would obtain a copy of the agreement. Whether the agreement becomes public depends on the prosecutor and the company; there is no direct role for the federal courts in the approval or enforcement of a NPA. The absence of a court filing might thus limit the publicness of a NPA relative to a

---

54. Rachel Delaney, *Congressional Legislation: The Next Step for Corporate Deferred Prosecution Agreements*, 93 MARQ. L. REV. 875, 878 (2009).

55. *See* Morford Memo, *supra* note 18, at 1 n.2 (explaining the difference between NPA and DPA and noting that the terms have often been "used loosely by prosecutors, defense counsel, courts, and commenters").

56. Delaney, *supra* note 54, at 878 ("In a DPA, 'the prosecutor files a criminal charge against a company, but agrees not to prosecute the claim so long as the entity complies with the terms of a deferral agreement.' In an NPA, no charges are filed at the outset but may be filed later if the corporation does not fulfill the terms of the agreement.").

57. *Id.*

58. *See, e.g.*, United States v. Biomet Orthopedics, Inc., No. 07-8133, 2007 WL 2964201, at *1 (D.N.J. Sept. 28, 2007) (noting Exhibit A, which defines a period of eighteen months for the DPA to remain in effect).

59. Cindy R. Alexander & Jennifer Arlen, *Does Conviction Matter? The Reputational and Collateral Effects of Corporate Crime* 42–54 (N.Y.U. Law & Economics Research Paper No. 17-34, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3037054 [https://perma.cc/7463-H2R7] (comparing the risks of costly collateral sanctions, such as debarment or delicensing under a plea agreement versus a DPA).

60. *See* Leonard Orland, *The Transformation of Corporate Criminal Law,* 1 BROOK. J. CORP. FIN. & COM. L. 45, 56 n.62 (2006) ("On occasion, it is difficult to determine if an 'agreement' is a deferred prosecution or non-prosecution agreement. ... [In] non-prosecution agreements, no charges were filed or pending against Prudential.").

DPA. To be sure, for a company with public investors, the absence of a court filing may not be sufficient to prevent the settlement from being public, regardless of whether it is a NPA or a DPA.[61]

For these reasons, companies that value non-plea settlement as a means of limiting the cost of the criminal sanction will either be indifferent between NPA and DPA or prefer the NPA form of settlement for its potential for increased privacy and absence of formal charges. Accordingly, we focus on the hypothetical choice by a prosecutor between an offer of a NPA settlement and a plea agreement and consider the effect of using one versus the other on the achievement of various enforcement objectives.

II. AN ECONOMIC MODEL OF NON-PLEA SETTLEMENT WITH COOPERATION

In this Part, we introduce an economic model of settlement that we adapt from the prior literature and use it as a guide to analyze the effects of relying on NPAs versus plea agreements. The difference between the two forms of settlement in the model is that, with a NPA, the firm faces a formal sanction and an obligation to cooperate, while, in a plea agreement, the firm faces a formal sanction and an informal sanction. We assume that the enforcement policy is transparent to each potential offender. Thus, prior to committing an offense, each can anticipate the probability of detection, the sanction if detected, and the alternative forms of sanction that the prosecutor may offer, including any reward for cooperation that might be offered as part of the offer of a NPA settlement (if there is one).

Our model of non-plea settlement incorporates cooperation and shares features with Kaplow and Shavell's canonical model of optimal law enforcement with self-reporting.[62] In their model, the prosecutor conserves enforcement resources by offering violators a lenient sanction equal to the expected sanction that they would face otherwise in exchange for self-reporting and all offenders self-report

---

61. Given that the company would have discretion over the release of the information in a NPA, the publicness of the company may be a factor in determining whether a NPA settlement is public. The regulatory status of the company also could be a factor. Thus, a prosecutor might plausibly enter into a private NPA settlement with a private company that has no regulatory requirement to release the contents of the settlement agreement to the public. This is the extreme case where the risk of costly reputational effects of settlement would be lower under a NPA than under a plea, even with other things equal (such as the crime, the severity, and the company characteristics).

62. *See* Louis Kaplow & Steve Shavell, *Optimal Law Enforcement with Self-Reporting of Behavior*, 102 J. POL. ECON. 583, 587–90 (1994).

as a result.[63] We think an analogy between self-reporting and cooperation provides a good starting point for understanding NPAs: each can lower the cost to the prosecutor of reaching a resolution of an offense. Following Kaplow and Shavell, we focus on a given offense and evaluate the effects of allowing all caught firms to cooperate and thus become eligible for the offer of leniency, here, in the form of a NPA. To the best of our knowledge, this is the first model to explore the effects of non-plea settlement.[64]

In our model, the DOJ or other EA publicly announces a policy with respect to cooperating corporate defendants, and firms react in ways that affect the frequency of crime. By changing the policy, the EA can accordingly change the amount of crime that occurs. We begin by describing the basic enforcement program in which the prosecutor is limited to choosing between plea and trial for those cases he does not decline. We then introduce an alternative enforcement program that allows the prosecutor to offer a NPA that confers some leniency (including avoiding conviction) in exchange for cooperation (lower cost of investigation) as an alternative to a plea agreement.

A.   Basic Enforcement Program

We begin with the scenario in which a firm engages in an act that exposes it to criminal liability and the prosecutor has decided to bring the case. The prosecutor can do one of the following: seek a plea settlement or go to trial. In the event the prosecutor seeks a plea settlement, the firm would face the cost of a certain criminal conviction. The company can reject a plea offer, however, and choose to go to trial. Whether it is desirable to do so will depend on the probability that the company will be convicted at trial.

Formally, we assume that each company may commit a single type of crime that in each instance results in a harm to society of $h > 0$. The size of the entire group of companies is normalized to mass 1. The enforcement policy and the prosecutor's actions administering the policy are assumed to be public knowledge among

---

63. To be sure, there are differences, too. Self-reporting helps conserve on inspection costs, whereas cooperation helps conserve on investigation costs.

64. In an independent effort, Murat C. Mungan has recently developed a model of non-plea settlement that explores the effect of privacy of the settlement on the cost of reputational damage to the firm. *See* Murat C. Mungan, *Optimal Non-Prosecution Agreements and the Reputational Effects of Convictions* 4–15 (George Mason Law & Economics Research Paper No. 17-40, 2017), https://papers.ssrn.com/sol3/papers.cfm ?abstract_id=3041967 [https://perma.cc/LDC8-YEVZ].

potential offenders. We think this is a reasonable assumption in the case of corporate crimes, given the publicness of the DOJ enforcement policy (e.g., the USAM, speeches, and settlement press releases).[65]

The prosecutor sets the formal sanction in the plea agreement at a level $\phi > 0$ that reflects the firm's willingness to pay.[66] This depends on the prosecutor's bargaining position. The prosecutor's bargaining position will depend on how likely it is that the company will be convicted if the case were to go trial (which we denote by a fixed probability p) and on the amount of formal sanction (which we denote by $s_0 > 0$) and any informal sanctions that the company might face as a result of a criminal conviction at trial. In this basic scenario, we assume that the informal sanctions are the same regardless whether the company is convicted through a plea agreement or at trial. That is, informal sanctions in this scenario arise entirely from the company having a criminal conviction. We have in mind the costs of unavoidable collateral effects of the conviction that include the risk of current or future debarment, delicensing, or exclusion from government contracts. This can be accompanied by a cost of reputational damage in some instances. We let $s_1 > 0$ denote the cost of the collateral effects of the sanction that arises from the criminal conviction, whether through a plea agreement or at trial, and is absent otherwise. In summary, the caught company is confronted with the choice between a certain sanction of $\phi + s_1$ with a plea agreement and the expected sanction of $p \cdot (s_0 + s_1)$ if the case proceeds to trial.

Deterrence depends on the probability of a sanction, conditional on the occurrence of an offense (represented by probability, ρ). In general, ρ will be determined by a number of other probabilities. Specifically, it will be the product of (i) q_1, the probability that misconduct will get reported (e.g., by tippers, whistleblowers, or witnesses), (ii) q_2, the probability with which the prosecutor will inspect or follow up on the leads it receives, and (iii) q_3, the

---

65. In the model, potential offenders must be aware of the enforcement policy in order for it to have an effect on general deterrence. *Cf.* Aaron Chalfin & Justin McCrary, *Criminal Deterrence: A Review of the Literature*, 55 J. ECON. LIT. 5, 5 (2017).

66. The public release of DOJ guidance on the corporate charging decision enhances transparency and thereby affects the conduct of the prosecutor and the corporations who are, or may be, subject to criminal investigations. *See* Christopher A. Wray & Robert K. Hur, *The Power of the Corporate Charging Decision over Corporate Conduct*, 116 YALE L.J. POCKET PART 306 (2007) (suggesting that the Thompson and McNulty Memos built transparency into the prosecutors' deliberative process and thus increased the "fairness, discipline, and consistency of that process by forcing decision-makers to justify discrepancies more rationally and persuasively").

probability with which the prosecutor will decide to conduct a full investigation upon following up on a lead. We observe two points regarding $\rho = q_1 q_2 q_3$.[67] First, this is the ex ante probability of sanction the offender faces upon committing a misconduct. Second, although q_1 is assumed to be exogenously determined and fixed, the EA (and thus the prosecutor) is assumed to exercise some control over ρ, implicitly by deciding how much time and effort to spend in following up on referrals and leads (i.e., by affecting q_2 and q_3). The EA is assumed to have no control over p, s_0 or s_1 (probabilities of conviction at trial as well as formal and informal sanction upon conviction), which are set as a matter of law, policy, and other factors that lie outside the prosecutor's discretion in this context. Inspection (or alternatively, following up on leads) is assumed to be costly (i.e., $c_{inspect} > 0$), and we assume $c_{inspect}$ is large enough that EA cannot simply set q_2 equal to 1 but faces a constraint in choosing q_2. Because the EA also must allocate resources to investigate detected offenses and build cases, we will consider the allocation of resources between inspection efforts and investigation efforts.

Following Becker,[68] we assume that each company commits a crime if the expected private benefit exceeds the expected private cost from the company's perspective. That is, if company *i* expects to reap a private benefit of b_i by engaging in an activity (while causing society a harm of h) and the expected cost is $\rho(\phi + s_1)$, then it will commit the crime if the benefit exceeds this expected sanction amount (i.e., if $b_i > \rho(\phi + s_1)$). To introduce heterogeneity of firms, we assume that for each firm the private benefit b_i is a random variable distributed according to a probability density function, $f(\cdot)$, which has support over positive real numbers. The offense frequency thus equals the density of firms with positive net benefits from the

---

67. This is a departure from Kaplow & Shavell's original set-up, which models a form of industry-wide inspection by a government actor. Prosecutors, however, do not conduct such general inspections; instead, they typically act upon being tipped by third parties. Kaplow & Shavell, *supra* note 62, at 587.

68. Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 176 (1968). For a more recent discussion, see Chalfin & McCrary, *supra* note 65, at 8. Differences in the potential offender's expected sanction can also be obtained by varying parameters of Becker's model. *See, e.g.*, Justin McCrary, *Dynamic Perspectives on Crime*, in HANDBOOK ON THE ECONOMICS OF CRIME 82, 87–96 (Bruce L. Benson et al. eds., 2010) (considering the time preferences for marginal offenders); Edward M. Iacobucci, *On the Interaction Between Legal and Reputational Sanctions*, 43 J. LEGAL STUD. 189, 189–207 (2014) (discussing outsider beliefs about the offender's type relative to the non-offender). *See generally* Mungan, *supra* note 64, (considering the publicness of the settlement).

offense, i.e., $1 - F\big(\rho(\phi + s_1)\big)$, where $F(\cdot)$ is the corresponding cumulative distribution function. We assume that the magnitude of the social harm from the offense, h, is smaller than $\phi + s_1$. Finally, we also allow for the possibility that in some instances, the private benefit *may* be greater than h.

Without novel forms of criminal settlement, each caught firm faces a choice between accepting a plea and going to trial, with a sanction under the plea agreement that is similar to, or lower than, the expected sanction at trial. For simplicity, we assume that the result of getting caught is a plea bargain with a formal sanction of $\phi = \phi(p, s_0)$. This is in addition to the informal sanction s_1 and consistent with Bar-Gill and Ayal.[69] In this setting, the higher the probability of conviction at trial and the formal sanction in case of a guilty verdict, the greater the expected total sanction amount under the plea.

B.   Alternative Enforcement Program: NPA

From the perspective of the prosecutor, the use of a NPA rewards cooperation by the caught offender and thereby lowers the prosecutor's cost of the investigation needed to justify the sanction s_0 under the existing standards of law and policy. From the perspective of each firm, the NPA can potentially confer greater leniency than under a plea offer net of any cost to the firm of providing the cooperation, which we assume to be zero.[70]

To explore the effects of offering leniency and obtaining cooperation through NPA in more detail, we observe the following. First, if the normal cost of investigation for each case is $c_{\text{invest}} > 0$, the offer of a NPA should change the prosecutor's cost of investigation to some lower amount $\beta_1 c_{\text{invest}}$, where $0 < \beta_1 < 1$. The condition that $\beta_1 < 1$ captures the requirement of the guidance that the prosecutor offers a NPA only if a firm's cooperation is authentic,

69. *See* Oren Bar-Gill & Oren Gazal Ayal, *Plea Bargains Only for the Guilty*, 49 J.L. & ECON. 353, 357 (2006). There is a rich literature examining the dynamics of strategic bargaining. *See* Jennifer F. Reinganum, *Plea Bargaining and Prosecutorial Discretion*, 78 AM. ECON. REV. 713, 713–15 (1988). We simply assume that ϕ is the equilibrium sanction amount for accepting a plea offer (e.g., a Nash bargaining solution), and that the prosecutor and the defendant agree that a sanction of ϕ together with an admission of guilt is mutually beneficial.

70. In this Part, cooperation mainly consists of not putting in place obstacles that might hinder the prosecutor's access to facts about the offense and culpable individuals where it is costless to the company to provide access. If obstacles are costly to put in place, the cost of cooperating may indeed be negative.

timely, and effective enough to reduce the prosecutor's cost of the investigation.[71]

Second, companies that are faced with a choice between a NPA and a plea may be willing to pay a higher monetary sanction and sustain other burdens—beyond what occurs under the plea settlement—to obtain the NPA rather than a plea.[72] This assumes that the company places a value of s_1 on the leniency it obtains by entering into a NPA instead of a plea agreement.

Suppose that a prosecutor is able to impose a monetary sanction under a NPA settlement in the amount of $\beta_2 \phi$, where $0 \leq \beta_2$. From the perspective of an offending firm, the *expected* cost of a NPA is simply $\rho \beta_2 \phi$. Unlike β_1, we do not require that $\beta_2 < 1$. That is, we reserve the possibility that the monetary sanction levied under a NPA may be equal to, or even higher than, one levied under a plea for the same offense.

Note that in practice β_1 and β_2 may vary from one case to another. In this model, since we are dealing with a given type of offense and cooperation for which the DOJ's guidance applies, we assume they are both fixed and exogenously determined, and the prosecutor's choice is allowed to reflect the comparison of the payoffs from NPA and plea.

This leads us to describe the leniency that arises from the substitution of a NPA for a plea agreement. Consider first the case where $\beta_2 = 1$. This indicates that regardless of whether the DOJ pursues a plea or a NPA, the formal sanction (e.g., the monetary fine) the firm would face is the same. Nevertheless, the advantage of a NPA is that no formal conviction attaches to it, and thus, the company does not face s_1 (e.g., the incremental costs of collateral effects, such as debarment, or of any reputational damage from a criminal conviction) as above. Rather than facing the cost of a plea that includes a conviction, $\phi + s_1$, the company faces the cost of a NPA settlement, ϕ. Thus, the NPA is a more lenient settlement even when the formal sanction remains unchanged between plea and NPA.

---

71. In the basic model, we are making a simplifying assumption that each firm provides cooperation. Under this model, every firm can qualify for a NPA. In an extension of the model, we could assume that, once an offense is committed, there is a probability $\alpha \in (0, 1]$ that the offending firm will be able to offer authentic cooperation. Such a set-up would ensure that at equilibrium we observe both NPAs and plea agreements among offending firms. In the Appendix, we establish Propositions 1 and 2 with this extension. The basic model is covered by letting $\alpha = 1$.

72. *See* Mungan, *supra* note 64, at 1 (noting that "firms would be willing to pay an NPA premium to avoid convictions").

The set-up of our model relaxes the assumption that the formal sanction is the same under both regimes, however. We allow β_2 to take on values other than 1. With an adjustment to the monetary sanction, the cost of the sanction to the company with a NPA would be $\beta_2 \phi$ as opposed to $\phi + s_1$ with a plea.[73] One possibility is that the prosecutor can enhance the leniency offer by also lowering the monetary sanction in addition to relieving the offender of the burden of a conviction. Thus, the prosecutor might impose a monetary sanction that is a fraction β_2 of what the offender would face with a plea where $\beta_2 < 1$.

We note that a prosecutor concerned with the possibility of excess loss of deterrence might even choose $\beta_2 > 1$. In other words, it is possible that the firm may be offered a NPA with a higher monetary sanction than one it would face under a plea. Such an offer can still be consistent with the use of a NPA as a lenient alternative to plea if the company might still prefer to avoid a conviction because it would face a rather large informal sanction s_1. The higher the value of the elimination of a criminal conviction to the company, s_1, the greater the room for a partially offsetting increase in the monetary sanction. Separately, for the time being, we assume the cost of cooperation to the company is zero[74] and that if $\beta_2 \phi = \phi + s_1$, the company will still prefer a NPA over a plea. The NPA policies are implementable over a wide range of monetary sanctions, subject to the constraint that $\beta_2 \in (0, 1 + s_1 / \phi)$. Obviously, if $\beta_2 \phi > \phi + s_1$, then the company should prefer a plea agreement, and would never agree to a NPA.

For this reason, in our model, we formally define *leniency* to mean $\beta_2 \phi < \phi + s_1$, the condition under which the company would (weakly) prefer the NPA. The value of leniency to the company is

---

73. Generally, β_2 may be equal to 1, but it may also be greater than or less than 1, and s_1 may be greater than or equal to the expected cost of collateral sanctions arising from the conviction, such as exclusion from future dealings with the government. For example, if the settlement is more public under the plea than under the NPA, s_1 may be greater than the cost of the collateral sanction by an amount equal to the cost of reputational damage from the reactions of outsiders to the plea relative to the NPA settlement.

74. That is, we consider cooperation as an action or process that is costless to the firm and informative to the prosecutor. Relaxing this assumption has the effect of reducing the expected cost savings to the company from the NPA. If the cost of cooperation to the company is high enough, the leniency value of the NPA may disappear, and the company may end up preferring the plea agreement as a way to avoid paying not just the penalty but also the cost of collecting the facts (or other effort required) to justify the penalty under existing legal or policy standards. In terms of our model, if the cost to the company of cooperation is $c_{\text{coop}} > 0$, then the firm would consider the NPA more cost-effective (or no less cost-effective) as long as $\beta_2 \phi + c_{\text{coop}} \leq \phi + s_1$.

NORTH CAROLINA LAW REVIEW [Vol. 96

greatest in the case of complete amnesty ($\beta_2 = 0$) and is smallest when the company is indifferent between the NPA and plea i.e., when $\beta_2 \phi = \phi + s_1$.

Table 1 below lists the variables and parameters in this model. Part III examines the choice of the prosecutor who is able to substitute a NPA for a plea agreement to obtain cooperation, after which we turn to the bundling of the use of a NPA with other enforcement strategies.

$h > 0$	Harm to society of a single occurrence of crime
$b_i > 0$	Private benefit to firm i of committing the crime
$f(\cdot)$	Probability density function of firms' benefits from committing the crime
$p \in (0,1)$	(Fixed) probability of conviction in case of trial
$\rho \in (0,1)$	*Ex ante* probability of sanction for the offender upon committing the crime
$q_1 \in (0,1]$	Probability that misconduct will get reported by tippers, whistleblowers, or witnesses
$q_2 \in (0,1]$	Probability that the prosecutor will inspect or follow up on the leads it receives
$q_3 \in (0,1]$	Probability that the prosecutor will decide to conduct a full investigation upon following up on a lead
$\phi > 0$	Formal sanction specified in a plea settlement
$s_0 > 0$	Formal sanction in case of conviction at trial
$s_1 > 0$	Informal sanction in case of conviction by plea or trial (i.e., cost of reputational or collateral effects, such as exclusion from government dealings)
$c_{\text{inspect}} > 0$	Cost of inspection for the prosecutor
$c_{\text{invest}} > 0$	Cost of investigation for the prosecutor
$B > 0$	The prosecutor's total budget
$\beta_1 \in (0,1)$	Cost of investigation with the firm's cooperation as a fraction of the cost of investigation without any cooperation
$\beta_2 \geq 0$	Monetary sanction under NPA as a fraction of the monetary sanction under a plea

Table 1: Variables and Parameters

III. APPLYING THE BASIC MODEL: PROSECUTOR CHOICE UNDER ALTERNATIVE ENFORCEMENT OBJECTIVES

Next, we consider how the ability to offer leniency in exchange for cooperation through a NPA affects the incentives of a prosecutor who faces not only a single hypothetical case but a population of cases

of a given type, each calling for the use of resources to develop the case for trial. Whether the prosecutor would want to offer a NPA to every offending firm depends on the enforcement objective. To illustrate this, we consider a series of scenarios in which the prosecutor is confronted with distinct objectives. The point is that the number of firms that actually get leniency will depend not just on whether use of NPA is part of the policy but also on the incentives of the prosecutor who applies the policy.[75]

A.   *Objective #1: Conserving Enforcement Resources*

As a benchmark for comparison, we consider the hypothetical prosecutor who cares only about freeing up enforcement resources through the more efficient resolution of criminal investigations. From a traditional welfare perspective, the prosecutor may regard the sanction as welfare-neutral, as it would be if it were a wealth transfer between each firm and the government with no other consequences. Such an objective may be difficult to sustain in practice: it would imply that the prosecutor is excluding from consideration the long-run outcome of the settlement policy.

Nevertheless, in this scenario, the value to the prosecutor of being able to offer non-plea settlements to all firms that provide authentic cooperation is simply the value of the saved enforcement resources. This is the product of the expected cost of investigating firms whose offenses are referred to the prosecutor after being detected, on the one hand, and the proportion of the investigation cost saved, on the other. A prosecutor who faces an objective of minimizing the use of resources to settle cases, and only that objective, would thus prefer to use a NPA rather than a plea in each instance. This is without any adjustment to the monetary sanction (i.e., with $\beta_2 = 1$).

In summary, the prosecutor can maximize the savings of enforcement resources by offering a NPA to each company that is willing to cooperate. This behavior is not what we observe in practice, however. Prosecutors use a variety of forms of settlement. Thus, we turn to other objectives that might more realistically characterize the incentive environment of the prosecutor in choosing between NPA and plea.

---

75. The incentives of the prosecutor are considered here as distinct from the objective of the law, such as that of incapacitating the offender or deterring a future offense. *See, e.g.*, Chalfin & McCrary, *supra* note 65, at 12 ("Generally speaking, there are two mechanisms through which criminal-justice policy reduces crime: deterrence and incapacitation.").

B.   *Objective #2: Maximizing Proceeds and Resource Savings*

The prosecutor may seek to increase the aggregate proceeds that accrue from settlement while limiting the use of enforcement resources. The assumption that prosecutors care about increasing their aggregate proceeds is consistent with the models of plea bargaining by Landes and Bar-Gill and Ayal.[76] There are two ways to use a NPA to obtain an increase in the aggregate settlement proceeds, net of the enforcement resource cost.

First, the prosecutor could monetize the value that firms place on having a NPA rather than a plea settlement. This could be done by raising the monetary sanction imposed in the NPA above what it would be in the plea agreement for the same offender and offense. In other words, if $\beta_2 > 1$, the prosecutor would strictly prefer a NPA to obtain both the higher proceeds (e.g., higher revenue from fines) and to save on investigation costs.

Second, even without a higher monetary sanction, the use of a NPA should enable the prosecutor to increase settlement proceeds in the aggregate—by settling each case more quickly and with lesser demand on the prosecutor's resources. In other words, even if $\beta_2 \leq 1$, the prosecutor would prefer a NPA if the savings from reduced investigation costs were sufficiently high to offset the loss in proceeds.

In summary, the NPA policy can pay off from the perspective of both proceeds and budgetary resources. The prosecutor may increase the monetary sanction to capture some of the gain to each firm that is relieved of criminal conviction and also capture the resource savings that come from the decline in case-preparation costs that is the direct result of authentic cooperation. In practice, the prosecutor need not raise the monetary sanction under the NPA to a higher level than what it would be under the traditional plea regime, however. The prosecutor could use the resources that are freed up by cooperation to generate more settlements, and thus more settlement proceeds, even if the monetary sanction remains the same under the NPA as in the plea agreement. Indeed, in the case where cooperation is costly to the offending firm, a strategy of lowering the monetary sanction as an increased reward for cooperation could increase in the resulting resource savings and proceeds above what occurs under the traditional plea settlement.

All of this assumes no effect on general deterrence, as would occur if offending firms were unable to anticipate the ability to obtain

---

76.   *See* Bar Gill & Ayal, *supra* note 69, at 357; William M. Landes, *An Economic Analysis of the Courts*, 14 J. L. & ECON. 61, 63–64 (1971).

leniency through a NPA. The resource savings and increased proceeds are entirely a result of the increased bargaining power of the prosecutor under the NPA policy. Enforcement policies, however, are generally designed to improve deterrence.

C. Objective #3: Maximizing Deterrence with a Fixed Budget

Turning to the effect of the NPA policy when the objective is to maximize general deterrence, we suppose now that the prosecutor is interested in deterring occurrences of misconduct as much as possible. Suppose in this hypothetical that the government controls the actions of the prosecutor by allocating a fixed budget to cover both its inspection costs and investigation costs. This introduces a constrained optimization problem for the prosecutor with a solution that requires finding the best mix of effort—here, between inspection and investigation—to achieve the enforcement objectives. Such an optimization problem appears to be consistent with the U.S. Attorney Manual's directive to "maximize the impact of federal resources on crime."[77]

To understand how the NPA influences the prosecutor's preferred choice when the objective is to maximize deterrence, we need to consider how companies are likely to behave in anticipation of the possibility of leniency. Formally, the direct effect of the NPA on deterrence can be seen by comparing the company's total expected sanction under a plea regime with that under the NPA regime. If the introduction of the NPA were to cause companies to anticipate a lower cost of crime in the form of more lenient settlements, the effect would be to increase the net benefit of committing a crime to the company; deterrence would suffer accordingly.

To be more specific, recall the default condition for violation under the basic enforcement program: $b_i > \rho_{\text{plea}}(\phi + s_1)$. Here, ρ_{plea} denotes the ex ante probability of sanction faced by the offender under the plea regime. In other words, each firm is assumed to commit a crime when the benefit it reaps from doing so exceeds the total expected sanction it will face in the future, which is the product of the total sanctions it will face upon getting caught and the

---

77. U.S. DEP'T OF JUSTICE, *supra* note 27, at § 9-27.300; *see also* Holder Memo, *supra* note 18, at 11 ("Once the decision to charge is made, the same rules as govern charging natural persons apply. These rules require 'a faithful and honest application of the Sentencing Guidelines' and 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.'" (citing U.S. DEP'T OF JUSTICE, *supra* note 27, at § 9-27.300)).

probability that it will be caught. This threshold condition changes in case of a NPA as follows: $b_i > \rho_{\text{NPA}}(\beta_2\phi)$, where ρ_{NPA} is the ex ante probability of sanction faced by the offender under the NPA regime. Under this condition, the informal sanction, s_1, is eliminated from the inequality to reflect that the NPA does not result in the informal sanctions for the company that come with a public felony conviction. On the other hand, the formal sanction is multiplied by a factor of β_2 to indicate that the defendant would be subject to a sanction amount corresponding to a NPA rather than a plea.

If the probability of sanction were held constant under these two regimes (i.e., $\rho_{\text{plea}} = \rho_{\text{NPA}}$), there would potentially be *less* deterrence under the NPA regime: since leniency implies that $\beta_2\phi < \phi + s_1$, each firm may face a lower expected sanction under the NPA regime. This suggests a tendency for the use of a NPA to reduce the deterrence effect due to leniency. Nevertheless, we should not expect the probability of sanction to be held constant. Because the NPAs are granted only to offenders who assist in their own investigation (i.e., $\beta_1 < 1$), the prosecutor will be left with more resources to devote to inspection effort or investigation of other cases (i.e., q_2 and/or q_3 will increase), and thus, we can expect ρ_{NPA} to be greater than ρ_{plea}.

The effect of offering leniency to cooperating offenders is the subject of the model that Kaplow and Shavell develop to study the use of self-reporting to reduce costs of inspection by a government administrator. They show how a decline in the sanction imposed on those who self-report their violations can nonetheless improve social welfare because the EA may conserve on resources in return, which can be devoted to increasing inspection.[78] A similar argument applies in our model of non-plea settlement. In the absence of non-plea settlement, the EA does not have the benefit of cooperation and thus must bear the full investigation cost. With a NPA, the EA has the possibility of incurring a lower cost of investigation to build her case and thus has more resources left for inspection or for more investigations. The effect is to increase the frequency of settlement and thus the probability that an offender will receive a sanction.

If the prosecutor can obtain cooperation for free—in the sense of not having to offer any leniency to achieve the lower investigation cost (i.e., if $\beta_2\phi = \phi + s_1$—there is no loss in deterrence, and the use of non-plea settlement to obtain cooperation would have no adverse

---

78. *See* Kaplow & Shavell, *supra* note 62 at 585–97 (discussing how an enforcement model with a self-reporting regime—in which those who report their own crimes face reduced fines—can lead to a reduction in enforcement resources and is ultimately more efficient).

effects. The puzzle of whether non-plea settlement is in the social interest emerges in the case where the prosecutor offers leniency—here, $\beta_2 \phi \leq \phi + s_1$—to ensure a sufficient incentive for the firm to cooperate. The answer is that since $\beta_1 < 1$, the prosecutor will have more resources to devote to detection efforts, which raises the probability of a sanction for other offenders.

In this setting, there is always some price the prosecutor is willing to pay to get a given amount of cooperation and in turn the ability to settle more cases on the fixed budget. In any case, it is clear that whatever the budget is, a prosecutor seeking to maximize deterrence will spend it all. Thus, if the budget is fixed at $B > 0$, under the basic program, the prosecutor will choose the inspection and investigation frequency q_2 and q_3 so that

$$B = q_1 q_2 (c_{\text{inspect}} + q_3 c_{\text{invest}})(1 - F(\rho(\phi + s_1))).$$

In the extreme alternative scenario of a prosecutor who opts to settle all cases with cooperation under the NPA policy, he will choose q_2 and q_3 so that

$$B = q_1 q_2 (c_{\text{inspect}} + q_3 \beta_1 c_{\text{invest}})(1 - F(\beta_2 \rho \phi)).$$

A comparison of these alternatives leads to the following result under the assumptions of our model:

PROPOSITION 1 (NPA AND DETERRENCE UNDER A FIXED BUDGET). Deterrence can improve by the prosecutor's reliance on NPAs if the increase in cases resolved is high relative to the value to the offender of the leniency conferred. This can happen in one of two ways. First, if the prosecutor imposes the same amount of formal sanction under the NPA as under the plea (i.e., $\beta_2 = 1$), the NPA policy can enhance deterrence if the informal sanction attaching to the plea is sufficiently small. Second, if the prosecutor can impose a different amount of formal sanction under the NPA, then for each level of cooperation required by the prosecutor ($\beta_1 < 1$) and any informal sanction $s_1 > 0$, there exists a sanction sufficiently high that deterrence can improve.

The proof is included in the Appendix. The intuition of the proof is that when there are no informal sanctions and the formal sanctions must remain unchanged between a NPA and a plea, the probability of

sanction ρ_{NPA} will be greater than ρ_{plea} since the resources saved from cooperation can be used to increase the inspection or investigation frequency. Therefore, the expected sanctions will be greater under the NPA, and the result will continue to hold as long as informal sanctions remain small. Second, suppose the prosecutor can impose a higher sanction under a NPA than under a plea. Then if the monetary sanction with a NPA is characterized by β_2 arbitrarily close to $1 + s_1/\phi$, the company would then enjoy minimal leniency, and yet firms would still prefer a NPA over a plea. In that case, the direct effect of using a NPA is negligible: even without any change in ρ, the prosecutor can achieve nearly the same amount of deterrence with a NPA as he can with a plea. Since a NPA is assumed to reduce investigative costs, the prosecutor can then divert his saved resources to increase ρ. Thus, ρ_{NPA} will also be greater than ρ_{plea}, and the result will continue to hold as long as β_2 is sufficiently close to $1 + s_1/\phi$.

This result has implications for the design of the NPA policy. In order to avoid the unintended consequence of a loss in general deterrence as a result of the offer of leniency to obtain cooperation, the prosecutor can ensure that the monetary sanction imposed under the NPA is not too low. If the prosecutor is unable to adjust the monetary sanction to compensate for the adverse effect of leniency on general deterrence, the prosecutor seeking to maximize general deterrence on a fixed budget should offer NPAs more sparingly.

Proposition 1 joins the ongoing debate regarding the overall effect on deterrence of NPAs in those instances where NPA would eliminate a risk of collateral sanctions and other costs that a firm would face under a plea agreement. One view is that, by conferring leniency, NPAs will not have the same deterrent effect as a plea agreement. Missing from this argument is the observation that prosecutors require authentic cooperation from offenders as a condition for conferring leniency through a NPA. If the leniency discount is bundled with a sufficient requirement for cooperation (or an adjustment to the monetary sanction), the net effect of the NPA policy is to enhance and not diminish general deterrence. This applies even when the NPA policy leads each firm to face a lower sanction than it would otherwise. Formally, the question of whether the NPA diminishes or enhances general deterrence hinges not on whether s_1 is present or absent, but instead on whether $\rho_{\text{NPA}}(\beta_2\phi) > \rho_{\text{plea}}(\phi + s_1)$.

D.   *Objective #4: Enforcement Resources and Social Welfare*

Now consider the prosecutor's choices in the absence of a formal budget constraint. The prosecutor may regard cooperation as a means of lowering the cost of the investigations that are required to justify the collection of settlement proceeds from offending companies. Formally, the prosecutor would solve an unconstrained optimization problem in which the enforcement resource constraint is incorporated into the objective function.

We consider social welfare following Kaplow and Shavell's formulation, in which the objective function includes the private benefit accruing to corporate defendants from misconduct along with the cost to the victims of the misconduct and the cost of inspection.[79] One important difference is that we have informal sanctions in the mix, which is assumed to be a net cost to society.

To summarize, we study the choice problem of the prosecutor who has some control over the ex ante probability of sanction for the offender, ρ, and seeks to maximize the overall social welfare, defined as the sum of the private benefits reaped by firms less the costs faced firms, the harm to society incurred by the offenses, and the prosecutor's resource costs. The prosecutor can offer leniency (i.e., $\beta_2\phi \leq \phi + s_1$) in exchange for cooperation, which in turn reduces the prosecutor's investigation costs (i.e., $\beta_1 < 1$). If the sanction probability that maximizes social welfare under the plea regime is nonzero,[80] we have the following result under the assumptions of our model, which is similar to Proposition 1:

> PROPOSITION 2 (NPA AND SOCIAL WELFARE). The net social welfare can increase by the prosecutor's reliance on NPA if the increase in cases resolved is high relative to the value to the offender of the leniency conferred. This can happen in one of two ways. First, if the prosecutor must impose the same amount of formal sanction under the NPA as it would under the plea (i.e., $\beta_2 = 1$), then the use of NPA can enhance net social welfare if the informal sanction attaching to a plea is sufficiently small. Second, if the prosecutor can impose a different formal sanction

---

79. *Id.* at 586.
80. If the inspection cost is high, for example, it is possible for the optimal inspection frequency to be zero. *See id.* (discussing the same possibility). We do not, however, consider such a scenario to be realistic.

> under the NPA, for each level of cooperation required
> by the prosecutor $(\beta_1 < 1)$, there is a sanction
> sufficiently high that net social welfare can improve.

The proof is included in the Appendix. The basic idea is similar to Proposition 1. When β_2 is very close to $1 + s_1/\phi$, then the adverse effect of leniency on deterrence (if any) is negligible. In this case, the NPA is socially beneficial (and preferred to a plea) simply because even at the same detection probability, the deterrence effect will be nearly identical, but the prosecutor in turn will save resources through cooperation, and the NPA involves no informal sanctions. But there is a second important effect of a NPA. Specifically, when β_2 is close to $1 + s_1/\phi$, the NPA policy effectively facilitates an efficient substitution between informal sanctions (which can be net costly to society) and a fine (which is a transfer). This way a NPA policy can ensure that the deterrence effect remains nearly identical, but society can avoid the cost arising from the informal sanctions. This is an additional benefit of the NPA, apart from the budgetary resource savings.

Meanwhile, we note that, if the NPA is bundled with a penalty β_2 that is arbitrarily close to 0, each firm can anticipate amnesty, and there would effectively be no deterrence effect even if the probability ρ is high. A plea agreement would then be the preferred choice.

IV. OTHER MOTIVATIONS FOR NPA?: THE DISCRIMINATING PROSECUTOR WITH BUNDLING

While our objective is to examine the effect of using NPAs on broader enforcement objectives, the series of DOJ policy statements points to further effects of their use. In particular, the prosecutor may bundle reform requirements and mandates into the NPA as a condition for leniency, and offending firms may react in ways that may have implications for the achievement of the enforcement objective. We consider three possibilities. First, prosecutors can use the added leniency that is possible in a NPA to settle cases that would otherwise be declined. The effect is to extend the scope of criminal enforcement and the frequency with which misconduct leads to a criminal sanction. Second, prosecutors may use the added leniency to convince the offending firm to accept provisions—such as commitments to self-reporting or future cooperation—that increase its expected sanction for a future offense. The prosecutor would then face lower costs of future detection and investigation of high-risk offenders. The effect is to improve general deterrence. Third, we note

that the prosecutor's use of a NPA may have the effect, in the short term, of changing the profile of the corporation that pleads guilty to an offense, and in the long term, of encouraging firms to invest more fully in robust compliance programs specifically to prepare themselves to provide authentic cooperation in the event of a criminal investigation.

A.   *Use of NPA in Complex Cases to Extend the Scope of Criminal Enforcement*

One possible benefit of using NPAs is that such agreements can enable prosecutors to resolve weaker cases—for example, complex cases that have merit yet face a lower probability of conviction at trial—the resolution of which would not be cost-effective under the basic enforcement set-up. For example, in some instances, the firm and the prosecutor may obtain facts early in an investigation that lead both to recognize that a case has merit despite a low probability of conviction at trial. Of course, the prosecutor can allocate more resources to the case to improve the chance of prevailing at trial. Ultimately, however, some such cases will be declined. But as an alternative to declining the case, the prosecutor may offer a NPA to obtain a timely resolution (with the cooperation of the firm) and a smaller sanction (than under a plea agreement or at trial) in exchange for walking away without a criminal conviction. From this perspective, a NPA would allow the prosecutor to impose a criminal sanction in the marginal case that would otherwise be declined. Given the option of proposing a NPA, the same prosecutor who would not develop a case due to its complexity might seek to build that same case. The effect is not just to increase the number of cases that are settled but also to expand the scope of settled cases to include cases that would otherwise be declined.

For example, consider the following. Thus far, we have assumed that upon being detected in a violation, each firm faces the same probability p of conviction. We relax this condition here. Suppose that once the government catches an offender through inspection, there are two types of cases: a case with a high probability of conviction at trial (p_H), and a case with a low probability of conviction at trial (p_L), where $p_H > p_L$. Assume these probabilities are common knowledge. The idea is that even among all cases that have merits, there may be a difference in the availability of evidence with which to build the case. Suppose now that the parties face the following costs. If the case goes to trial, both sides need to spend $100 in case preparation costs (e.g., attorney fees for the corporate

defendant). If the case were to end in a guilty plea, both sides need to spend $50. Finally, if the case were to be resolved by way of a NPA, resolution takes place much earlier and both sides need to spend only $12 to come to an agreement.

Under this cost assumption, suppose we have an instance of a caught offender and the offender faces a high probability of conviction, p_H. The monetary sanction in case of a guilty verdict at trial is s_0. The expected cost of going to trial for the offender is $p_H s_0 + 100$. Meanwhile, if we assume the prosecutor seeks to increase the aggregate proceeds that accrue from settlement while limiting the use of its own resources, the expected value of going to trial for the prosecutor is $p_H s_0 - 100$. On the other hand, if both parties were to work toward getting a plea with a sanction amount of $p_H s_0$, then the respective values are $p_H s_0 + 50$ and $p_H s_0 - 50$, which would be better than the trial option for both parties. In addition, the prosecutor will find it worthwhile to seek a plea as long as $p_H s_0 > 50$. The calculation, however, is different when we have a low-type offender. If the offender faces a low probability of conviction, it may be the case that the prosecutor, if he were restricted to a plea or a trial, would find neither worthwhile. Specifically, if $p_L s_0 < 50$ (either because the sanction s_0 is low or the probability p_L is sufficiently low or both), then the prosecutor who compares his own cost against the gain from sanctions collected will bring neither the plea nor seek to go to trial.

From this perspective, the policy reflected in the DOJ's memoranda, starting with the Holder Memo, can be seen as encouraging the use of NPA in cases with p_L (e.g., low culpability). If the prosecutor has the option of seeking a NPA, then even if $p_L s_0 < 50$, as long as if $p_L s_0 > 12$, the prosecutor will find it worthwhile to seek that simple sanction of $p_L s_0$. Therefore, the assumption that a NPA can be negotiated by the prosecutor more quickly and more easily than a plea implies that the prosecutor will be able to resolve more cases criminally, including those cases for which the prosecutor previously lacked any credible threat to pursue. The effect is to increase the number of cases that are settled with a criminal sanction as opposed to a declination. In this manner, use of non-plea settlement can promote greater deterrence.

Consistent with this analysis, Lanny A. Breuer, a former head of the DOJ Criminal Division, remarked as follows:

> When the only tool [the prosecutors] had to use in cases of corporate misconduct was a criminal indictment, prosecutors

sometimes had to use a sledgehammer to crack a nut. More often, they just walked away. In the world we live in now . . . prosecutors have much greater ability to hold companies accountable for misconduct than we used to . . . .[81]

Relatedly, a recent study by Alexander and Cohen notes that there may be systematic differences in the cases that are disposed of under plea agreements as compared to those that are disposed of under NPAs and DPAs. They note in particular that (1) "the average offense scores for DPAs and NPAs are larger than that for pleas" (that is, DPAs and NPAs generally involve more severe crimes); (2) the base fine is higher for DPAs and NPAs, and monetary penalties are also higher;[82] and (3) "mean and median culpability scores are markedly lower for NPAs and DPAs than for plea agreements when averaged across all offense categories."[83] Of these, the third observation is consistent with the analysis in this Section suggesting that the use of NPAs (due to lowered cost of preparation) can incentivize prosecutors to resolve cases that present low probabilities of conviction.

B.   *Use of NPA to Obtain a Commitment to Future Cooperation*

Another possible use of NPAs is that such agreements can commit the firm to reforms and mandates that give the prosecutor easier information access post-settlement and thereby reduce the prosecutor's future costs of inspection and investigation. This is different from the use of a NPA in our basic model, where the effect is to achieve cooperation to lower the current costs of investigation. Instead, the NPA commits the firm to what is essentially future cooperation.

For example, note that a significant number of NPAs involve some type of mandate, governance reform or waiver of privileges.[84]

---

81.   Lanny A. Breuer, Assistant Attorney Gen., U.S. Dep't of Justice, Address at the New York City Bar Association (Sept. 24, 2012), http://www.justice.gov/criminal/pr/speeches/2012/crm-speech-1209131.html [https://perma.cc/BP97-GQ3G].

82.   They note, however, that there is no evidence that the total monetary sanctions under NPAs and DPAs differ relative to the fine that would be assessed under the Guidelines if the offender had instead settled for a plea agreement. *See* Alexander & Cohen, *supra* note 4, at 585.

83.   *Id.* at 576–77

84.   *Id.* at 540–42. Reforms and mandates through a NPA can commit the firm to a more effective regime for self-reporting than under traditional enforcement. *See generally* Jennifer Arlen, *The Potentially Perverse Effects of Corporate Criminal Liability*, 23 J. LEGAL STUD. 833 (1994) (exploring alternatives to strict liability that may increase the efficiency of enforcement against corporate criminal defendants); Jennifer Arlen &

According to Alexander and Cohen, in the sample of ninety NPAs collected between 1997 and 2011, fifty-two (58%) required training, the same number of them required some changes in accounting or auditing reforms, and forty-nine (54%) required compliance-related reforms.[85] Similarly, in the same sample of NPAs, eighty-three (92%) contained a waiver of right to speedy trial and/or statute of limitations, twenty-three (26%) a waiver of attorney-client privilege, and twenty-two (24%) a waiver of attorney work-product privileges.[86] These reform measures do not directly provide any type of tangible evidence that can help the prosecutor build her case in the present. Rather, they can be seen as measures that will increase the probability of conviction and lower the cost of investigation in the case of a future violation.

Now consider the following. Given a case in the current period with a low probability of conviction, the prosecutor may not find it worthwhile to seek a conviction through a plea. He may instead draft a NPA to impose governance reforms and waiver of privileges, such that if the offending firm were to commit a violation (again) in the future, upon its detection, the defendant will face a high probability of conviction (and in equilibrium a plea agreement). This indicates that, in a repeated game, use of NPAs can bring a greater number of detected violations within the pool of credible threat of plea bargains, which can in turn promote deterrence.

This analysis, however, raises a question: given that corporate defendants might see monetary sanctions and costly governance reforms as effective substitutes, why might the prosecutor and defendant have distinctly different preferences for one form over the other? One response is that the parties may have divergent expectations of the benefits of offense that may be realized in the future.[87] More specifically, the defendant and the prosecutor may actually differ in their (current) preferences for monetary sanctions versus governance reforms. For example, if the corporate defendant believes that it was fortunate to get a high realization of benefit from committing the offense in the current period, but it is unlikely that any sizable benefit will again be realized from a future offense, it may readily agree to governance reforms and waiver of privileges. This is

---

Reinier Kraakman, *Controlling Corporate Misconduct: An Analysis of Corporate Liability Regimes*, 72 N.Y.U. L. Rev. 687 (1997) (analyzing a shift from strict vicarious liability to duty-based schemes that mitigate liability for corporate criminal defendants).

85. *See* Alexander & Cohen, *supra* note 4, at 588–89.

86. *See id.* at 586–87.

87. In other words, they may disagree about the underlying distribution of $f(\cdot)$.

because the defendant may expect that it will have no need or desire to commit any similar violation in the future. On the other hand, if the prosecutor believes otherwise (i.e., if the distribution of benefits is fat-tailed in the higher end), then he may prefer the governance reforms and waiver of privileges to levying a sanction in the current period, which will not be very high given the low probability of conviction the defendant faces based on the current period's violation.

In short, a prosecutor who regards the NPA as a means of lowering the current costs of investigation and future costs of detection would include governance and legal reforms as terms in the NPA. The ability to include governance reforms to commit the offending firm to better future compliance is unique to the prosecution of corporations, as distinct from the prosecution of natural persons.

Recall that in the basic model of this paper, NPAs were considered as net substitutes for plea agreements. Alexander and Cohen, however, find that the use of NPAs (and DPAs) is correlated with an increase in plea agreements.[88] This finding is consistent with the notion that NPAs act as complements to plea agreements. Our extension of the basic model, here, to a multi-period model with differentiated case quality points to the use of NPA as a complement to traditional plea settlement that could lead to an overall increase in plea agreements in the long-run.

C. *Use of NPA with Selection Effects That Encourage Investment in Compliance Programs*

Finally, we note that the prosecutor's use of a NPA may have, in the short term, the effect of changing the profile of the corporate offender who pleads guilty to an offense, and in the long term, the effect of encouraging firms to invest more fully in robust compliance programs. This is distinct from self-reporting of an undetected offense. That is, firms may invest in compliance programs that eliminate obstacles within the firm to providing the prosecutor with authentic and timely cooperation, thereby qualifying for the offer of a lenient NPA sanction, should the firm commit an offense. To illustrate, we relax the assumption that all caught firms qualify for leniency and consider the effect on the profile of the firms that are caught.

Recall that, under the USAM, the prosecutor's decision to offer

---

88. *See id.* at 541.

a NPA depends on the authenticity of a firm's cooperation. If we assume that each firm can have greater leniency with the NPA, we might expect every firm whose only other alternative is pleading guilty to cooperate. This was indeed our assumption in Part III. But in practice, there may be a number of different reasons as to why a firm, caught for a violation, would choose not to cooperate with the prosecutor. Here we note that, for a given offense and relation of the firm to the offense, the cost (or ease) of cooperation to the offending firm may be lower if the firm has prior experience as the target of a criminal investigation or has otherwise prepared to be an investigation target. A firm that has committed criminal offenses would be familiar with the procedures of the Department of Justice and the expectations of its prosecutors. Such a firm may know what is required, at minimum, to obtain credit for authentic cooperation and thus a lenient NPA. In contrast, a company caught in its first offense would face greater uncertainty as to what forms of cooperation are sufficiently authentic to qualify for a lenient NPA. The firm may be complex, and the prosecutor may find it difficult to instruct the firm on how to cooperate. Authentic cooperation may require initiative on the part of the company. As in all regulatory compliance, the implicit assumption here is that there is a fixed up-front cost of compliance with the authentic-cooperation requirements of the prosecutor and that firms in their first caught offenses face obstacles to cooperation that other firms do not face.

For these reasons, the willingness and ability of each caught company to offer authentic cooperation may depend on its prior level of preparation for cooperating with the government. To be sure, guilt can provide an incentive not to cooperate. Yet even holding the culpability of the firm constant, it seems that a firm that is unfamiliar with the procedures of the Department of Justice and its prosecutors would face greater obstacles to determining how to cooperate than another firm with a history as an investigation target. Similarly, firms in regulated industries may be better prepared to deal with the government and thus better able to respond to a call for cooperation. This would make them more likely to provide the authentic cooperation needed to be offered a NPA settlement.

An empirical implication is that—holding constant the offense and the severity of the prospective sanction—the frequency of non-cooperation may be higher among firms that have less prior experience dealing with the Department of Justice as an investigation target and less prior experience in dealing with government investigators generally. In addition, we would expect the frequency of

non-cooperation to decline over time as firms become more familiar with the Department's procedures and practices. In the long-run, however, the overall effect may be that firms recognize that investing in robust compliance programs can increase their chances of getting offered a NPA in the event of an offense, and therefore, the prosecutor's selective use of NPAs (for those firms that can provide authentic cooperation) may have the effect of encouraging the firms to invest in compliance programs early on.

V. CONCLUSION

In this Article, we have identified efficient government resource allocation, post-detection and pre-trial, as a beneficial outcome of the introduction of NPAs as an alternative to plea agreements for the purpose of obtaining cooperation from caught firms to resolve corporate criminal investigations. We used a model from the prior literature on enforcement to assess the implications of more efficient investigations of caught offenders, post-detection.

The policy of limiting the use of NPAs to cases where the company provides authentic cooperation can serve several enforcement objectives. First, the use of NPAs to obtain cooperation can lead to enhanced deterrence and improved resource allocation by lowering the cost to prosecutors of investigating companies, post-detection and pre-trial. From a traditional social welfare perspective, the efficiency of a NPA relative to a plea agreement depends on whether the returns to resources saved through cooperation—in the form of increased ex ante probability of facing a sanction for the offender—exceed the loss of deterrence due to the leniency of the sanction provided to obtain cooperation under the policy, other things equal. We also find that the use of NPAs can facilitate efficient substitution between an informal sanction that attaches to criminal conviction (net cost to society) and a monetary fine (a transfer). Put differently, the elimination of a public guilty plea can transform an inefficient informal sanction into a formal monetary sanction in some instances.

We also identified possible extensions of the basic model that incorporate bundling. First, the use of NPAs can be combined with the exercise of prosecutorial discretion to facilitate efficient discrimination between cases with high versus low costs of resolution, allowing more cases to be resolved through criminal sanctions rather than declination. Second, the use of NPAs may be

combined with requirements for enhanced future policing to provide for an efficient substitution between a current sanction and the future probability of conviction. Third, the practice of offering NPAs selectively to those offenders who can provide authentic cooperation may in the long run have the effect of encouraging firms to invest more in robust compliance programs.

APPENDIX

PROOF OF PROPOSITION 1. We prove the results of Proposition 1 for a slightly more general case in which the fraction of the population of caught firms that can provide authentic cooperation is assumed to be $\alpha \in (0, 1]$, which is realized for each firm after the offense is committed. If we let $\alpha = 1$, the result of the original model follows. Under the basic enforcement regime in which only plea agreements are available, the prosecutor who seeks to maximize deterrence solves the following optimization problem:

$$\max_{q_2, q_3} q_1 q_2 q_3 (\phi + s_1) \text{ subject to } q_1 q_2 (c_{\text{inspect}} + q_3 c_{\text{invest}})(1 - F(q_1 q_2 q_3 (\phi + s_1))) \leq B.$$

Maximizing deterrence in this instance is equivalent to maximizing the expected sanction. Under the basic enforcement regime, the potential offender will commit a crime based on whether their realized benefit value is greater than $q_1 q_2 q_3 (\phi + s_1)$. Let D_{plea} be the maximal deterrence achieved under this problem. Under the NPA regime, an offender's ex ante expected sanction is $q_1 q_2 q_3 (\alpha \beta_2 \phi + (1 - \alpha)(\phi + s_1))$. Thus, the prosecutor solves the following optimization problem:

$$\max_{q_2 q_3} q_1 q_2 q_3 (\alpha \beta_2 \phi + (1 - \alpha)(\phi + s_1)) \text{ subject to}$$

$$q_1 q_2 (c_{\text{inspect}} + q_3 (\alpha \beta_1 + (1 - \alpha)) c_{\text{invest}}) \left(1 - F \left(q_1 q_2 q_3 (\alpha \beta_2 \phi + (1 - \alpha)(\phi + s_1)) \right) \right) \leq B.$$

Let D_{NPA} be the maximal expected sanction under the lenient NPA policy. To prove the first part, we show that when $\beta_2 = 1$ and $\beta_1 < 1$, there is $\bar{s} > 0$ such that for all $s_1 \in [0, \bar{s}]$, the prosecutor's objective of maximizing deterrence will improve by his use of NPA. Suppose $\beta_2 = 1$ and $s_2 = 0$. Then note that the maximands are identical under the two optimization problems. But $\rho = q_1 q_2 q_3$ will still be higher under the NPA because $\beta_1 < 1$ and therefore the prosecutor can afford to choose higher q_1 and/or q_3. (If $\beta_1 = 1$, then the two budget constraints will also be identical). Therefore, if $s_1 = 0$, then $D_{\text{NPA}} > D_{\text{plea}}$. Now since $\rho(\phi + s_1)$ and $\rho(\alpha \beta_2 \phi + (1 - \alpha)(\phi + s_1))$ are both continuous in s_1, it must be the case that for some values of $s_1 > 0$, D_{NPA} will continue to remain greater than D_{plea}. Let $\bar{s}$ be the supremum over all such s_1 values (for

which $D_{\text{NPA}} > D_{\text{plea}}$), and we have the first result. To prove the second part, we show that for each $\beta_1 < 1$, there exists $\beta_2^*(\beta_1) < 1 + s_1/\phi$ such that the prosecutor's objective of maximizing deterrence will improve by his use of NPA if $\beta_2 > \beta_2^*(\beta_1)$. Suppose $\beta_2 = 1 + s_1/\phi$. Then D_{NPA} will necessarily be strictly greater than D_{plea} because, again, the maximands would be identical in the two optimization problems, but the budget constraint is more favorable toward yielding a higher ρ in the D_{NPA} case because, again, $\beta_1 < 1$. Since $\rho\big(\alpha\beta_2\phi + (1-\alpha)(\phi + s_1)\big)$ is continuous in β_2, it must be the case that for some values of $\beta_2 = 1 + s_1/\phi$, D_{NPA} will continue to remain greater than D_{plea}. Let β_2^* be the infimum over all such β_2 values (for which $D_{\text{NPA}} > D_{\text{plea}}$), and we are done. It is clear that $\beta_2^* > 0$. *Q.E.D.*

PROOF OF PROPOSITION 2. As with Proposition 1, we prove the results of Proposition 2 for the case in which the fraction of the population of caught firms that can provide authentic cooperation is assumed to be $\alpha \in (0, 1]$, which is realized after the offense is committed. For a given inspection probability $\rho \geq 0$, define

$$
\begin{aligned}
SW_{\text{plea}}(\rho) = SW_{\text{plea}}(q_2, q_3) \\
\equiv \int_{q_1 q_2 q_3 (\phi + s_1)}^{\infty} (b - h) f(b) db \\
- s_1 q_1 q_2 q_3 \left(1 - F(q_1 q_2 q_3 (\phi + s_1))\right) \\
- q_1 q_2 \big(c_{\text{inspect}} + q_3 c_{\text{invest}}\big)(1 - F(q_1 q_2 q_3 (\phi + s_1)))
\end{aligned}
$$

and

$$
\begin{aligned}
SW_{\text{NPA}}(\rho) = SW_{\text{NPA}}(q_2, q_3) \\
\equiv \int_{q_1 q_2 q_3 (\alpha\beta_2\phi + (1-\alpha)(\phi + s_1))}^{\infty} (b - h) f(b) db \\
- q_1 q_2 \big(c_{\text{inspect}} + q_3 (\alpha\beta_1 + (1 - \alpha)) c_{\text{invest}}\big) \\
\times \left(1 - F\left(q_1 q_2 q_3 (\alpha\beta_2\phi + (1 - \alpha)(\phi + s_1))\right)\right).
\end{aligned}
$$

Then $SW_{\text{plea}}(\rho) = SW_{\text{plea}}(q_2, q_3)$ is the value of net social welfare (including enforcement resource costs) under the plea regime with q_2 and q_3, and $SW_{\text{NPA}}(\rho) = SW_{\text{NPA}}(q_2, q_3)$ is the corresponding value under the NPA regime. The set-up here is similar to Kaplow and Shavell's model. In the expression for $SW_{\text{plea}}(q_2, q_3)$, the first term (the integral term) aggregates private benefits and social

harms of all offenses committed, the second term represents the informal sanctions faced by firms who would get caught and thus eventually plead guilty, and the final term represents the aggregate resource costs of inspection and investigation. Note that the formal sanction imposed, ϕ, is not included in this expression as a cost because it is a simple transfer. In the expression for $SW_{\text{NPA}}(q_2, q_3)$, the integral term again aggregates private benefits and social harms of all offenses committed. Note that the threshold for committing an offense is $q_1 q_2 q_3 \big(\alpha \beta_2 \phi + (1 - \alpha)(\phi + s_1)\big)$ under the NPA. The second term represents the aggregate resource costs of inspection and investigation, where the per-offense investigation cost is now reduced to $\beta_1 c_{\text{invest}}$ for the share of firms that cooperate. Finally, there is no term involving the informal sanction, s_1, for firms agreeing to NPA because such informal sanctions only attach to convictions. Now let $SW_{\text{plea}}^* \equiv \max_{q_2, q_3} SW_{\text{plea}}(q_2, q_3)$ and $SW_{\text{NPA}}^* \equiv \max_{q_2, q_3} SW_{\text{NPA}}(q_2, q_3)$. The proof proceeds in the similar manner as in Proposition 1. To prove the first part, we show that when $\beta_2 = 1$ and $\beta_1 < 1$, there is $\bar{s} > 0$ such that for all $s_1 \in (0, \bar{s}]$, the prosecutor's objective of maximizing overall social welfare will improve by his use of NPA. Suppose $\beta_2 = 1$ and $s_1 = 0$. Then it is clear that $SW_{\text{NPA}}^* > SW_{\text{plea}}^*$. This can be seen as follows. Let $\rho_{\text{plea}}^* = q_1 q_2^* q_3^*$ be the optimal ex ante probability of sanction chosen for $SW_{\text{plea}}(q_2, q_3)$ so that $SW_{\text{plea}}^* \equiv SW_{\text{plea}}(\rho_{\text{plea}}^*) = SW_{\text{plea}}(q_2^*, q_3^*)$. By assumption $\rho_{\text{plea}}^* > 0$. Now note that in this case we necessarily have $SW_{\text{NPA}}(\rho_{\text{plea}}^*) \equiv SW_{\text{NPA}}(q_2^*, q_3^*) > SW_{\text{plea}}(q_2^*, q_3^*) \equiv SW_{\text{plea}}(\rho_{\text{plea}}^*)$. This is true since the two integral terms become identical and the resource cost will favor $SW_{\text{NPA}}(\rho_{\text{plea}}^*)$, since the expressions become identical except for β_1, which is smaller than 1. In addition, SW_{NPA} is free of the informal sanction term. Then it follows that $SW_{\text{NPA}}^* \equiv \max_{\rho} SW_{\text{NPA}}(\rho) \geq SW_{\text{NPA}}(q_2^*, q_3^*) = SW_{\text{NPA}}(\rho_{\text{plea}}^*) > SW_{\text{plea}}(\rho_{\text{plea}}^*) = SW_{\text{plea}}^*$. At this point, since both $SW_{\text{NPA}}(\rho)$ and $SW_{\text{plea}}(\rho)$ are continuous in s_1, it must necessarily be the case that for some $s_1 > 0$, $SW_{\text{NPA}}^* > SW_{\text{plea}}^*$ will continue to hold. Let $\bar{s}$ be the supremum over all such s_1 values (for which $SW_{\text{NPA}}^* > SW_{\text{plea}}^*$), and we have shown the first part. To prove the second part, we show that for each $\beta_1 < 1$, there exists $\beta_2^*(\beta_1) < 1 + s_1 / \phi$ such that the prosecutor's objective of maximizing overall social welfare will improve by his use of NPA if $\beta_2 > \beta_2^*(\beta_1)$. Suppose $\beta_2 < 1 + s_1/\phi$. Then a similar argument can establish that

$SW^*_{\text{NPA}} > SW^*_{\text{plea}}$. At this point, since $SW_{\text{NPA}}(\rho)$ is continuous in β_2, it must necessarily be the case that for some $\beta_2 < 1 + s_1/\phi$, $SW^*_{\text{NPA}} > SW^*_{\text{plea}}$ will continue to hold. Let $\underline{\beta_2}$ be the infimum over all such β_2 values, and then we have a positive measure of leniency for which is welfare-improving (for all $\beta_2 \in (\underline{\beta_2}, 1 + s_1/\phi]$). This shows the second part, and we are done. *Q.E.D.*

902   *NORTH CAROLINA LAW REVIEW*   [Vol. 96