Constitutional Court Review 2019
Volume 9, 81–111
https://doi.org/10.2989/CCR.2019.0004

Plea Bargaining in South Africa: An Economic Perspective

RICHARD ADELSTEIN

ABSTRACT: This essay applies a simple economic model of the plea bargaining process to the two-tiered structure of negotiated pleas and sentences in South Africa. Bargaining in South Africa proceeds along one of two tracks. A formal procedure, authorised and regulated by s 105A of the Criminal Procedure Act, gives defendants represented by counsel access to precise information about the terms of the bargain before the plea is made and permits them to withdraw their pleas should the sentencing judge reject the agreement. In contrast, an older, informal procedure, governed by s 112, applies to defendants without counsel and grants them significantly less information and agency in the bargaining process than does the s 105A procedure. The model illuminates the central role of information and uncertainty in the defendant's decision to plead guilty or insist on a full trial, and suggests that if all plea bargains were governed by the formal procedures of s 105A, the system would more effectively represent the interests of both prosecutors and defendants by producing more bargains, and fewer trials, in cases where both sides want to avoid trials and consummate plea bargains. The final section considers the constitutionality of plea bargaining under ss 35 and 36 of the Constitution of the Republic of South Africa. It reviews the constitutional history of plea bargaining in the United States to emphasise the differing perspectives on the constitutionality of plea bargaining demanded by significant variations in substance and interpretative style in the two constitutions. The essay concludes by briefly suggesting the arguments that might be made against the constitutionality of plea bargaining under s 35 and the corresponding contentions that might be raised during limitations analysis under s 35 to justify the practice should it be found to violate s 35.

KEYWORDS: guilty pleas, sentence agreements, formal/informal plea bargains, information, uncertainty, s 35 right to trial, limitations of s 35 right to trial

AUTHOR: Woodhouse/Sysco Professor of Economics, Wesleyan University.
Email: radelstein@wesleyan.edu

ACKNOWLEDGEMENTS: I am grateful to Stu Woolman, Bryant Jackson-Green, and two anonymous referees for helpful comments on an earlier draft. Remaining errors are my own.

I TWO TRACKS TO AGREEMENT

A *negotiated guilty plea*, or *plea bargain*, takes place when a defendant agrees to forego his right to a full criminal trial by pleading guilty to some offence in exchange for a favourable concession by the prosecution or, less frequently, the court. Where prosecutors exercise substantial discretion in the selection of charges, the range of possible concessions is wide, particularly where the crime is complex or involves several perpetrators. To secure a guilty plea in the circumstances of the case, prosecutors might, among other things, agree to reduce or to drop the charges faced by the defendant; a co-defendant or some third party, protect the defendant from imprisonment or other onerous sanctions; limit the evidence or facts that might be led or revealed to the court; or stay their hand in some other way that might induce, or pressure, defendants to surrender their right to trial.[1] However, in practice, in the many and varied forms of plea bargaining that have emerged in recent decades in both adversarial and inquisitorial systems around the world,[2] the concession ultimately sought by almost every defendant is a reduction in the fine or prison sentence that he will receive upon conviction, punishments whose relative severity can meaningfully be represented as objective numbers. In these cases, a defendant pleads guilty to some offence in exchange for a measurably more lenient sentence than he expects to receive after conviction at trial. For the better part of a century plea bargains of this kind have accounted for at least 90 per cent of all criminal convictions in the state and federal courts of the United States.[3] To focus attention on the essential aspects of the exchange, I assume here that every plea bargain takes this straightforward form.

Guilty pleas have been negotiated in South Africa for several decades.[4] After the passage of the Criminal Procedure Act 51 of 1977 ('CPA'), what came to be called 'informal' plea bargains appear to have been routinely concluded under s 112 of the CPA, which governs guilty pleas at summary trials. Section 112(1)(a) provides that, where a defendant pleads guilty to an offence that does not merit imprisonment or a large fine, and the prosecutor has accepted the plea, the presiding judge may convict the defendant of the offence to which he has pleaded and impose any lawful sentence, without the taking of any evidence beyond the guilty plea itself. For more serious offences involving imprisonment or large fines, s 112(1)(b) still allows the judge to convict and sentence a defendant without trial, provided that she questions the defendant in court to establish a factual basis for the plea, an admission by the accused of facts sufficient to prove the charge to which he has pleaded guilty. While s 112(2) provides for the submission of a written statement of admission in lieu of the oral colloquy, the judge remains free to question the defendant in court in addition to receiving the statement. Nothing in the

[1] Compare E Steyn 'Plea-bargaining in South Africa: Current Concerns and Future Prospects' (2007) 20 *South African Journal of Criminal Justice* 206, 210–211.

[2] See, generally, JI Turner *Plea Bargaining Across Borders: Criminal Procedure* (2009) and the papers collected in JS Hodgson (ed) *Comparative Perspectives on Plea Bargaining: A Conference for Young Scholar*s (2013), available at http://www2.warwick.ac.uk/fac/soc/law/research/centres/cjc/conferences/outandabout/adversarial/youngscholars.

[3] P Bekker 'Plea Bargaining in the United States of America and South Africa' (1996) 29 *Comparative and International Law Journal of Southern Africa* 168, 168–172; AW Alschuler 'Plea Bargaining and Its History' (1979) 79 *Columbia Law Review* 1, 26–32.

[4] D van Zyl Smit & NM Isakow 'The Decision on How to Plead: A Study of Plea Negotiations in Supreme Court Criminal Matters' (1986) 10 *South African Journal of Criminal Law and Criminology* 3; CT Clarke 'Message in a Bottle for Unknowing Defenders: Strategic Plea Negotiations Persist in South African Criminal Courts' (1999) 32 *Comparative and International Law Journal of Southern Africa* 141.

section explicitly contemplates a plea bargain, and while the prosecutor may recommend a specific sentence to the court, the judge retains full discretion either to reject the plea under s 113 and force the parties to a full trial, or accept the guilty plea, convict the defendant, and impose any lawful sentence for the offence to which the plea has been entered. The defendant does not hear the sentence until the plea has been entered and the factual basis established. At this point, the defendant may no longer withdraw the plea and demand a full trial.[5]

Despite its apparent indifference to the practice, once the final procedural piece is in place, s 112 becomes, as Uijs AJ put it in 1999, 'virtually tailor-made for plea bargains'.[6] This was provided in 1985 by *State v Ngubane*.[7] The defendant was charged with murder and, after discussions with the prosecutor, pleaded guilty under s 112 to the lesser offence of culpable homicide. The presiding judge nonetheless convicted Ngubane of murder with extenuating circumstances. On appeal the court overturned the conviction, reading ss 112 and 113 together to conclude that:

> [O]nce the defendant has entered a guilty plea that the prosecutor agrees to accept, the prosecutor's acceptance limits the ambit of the *lis* between the state and the accused in accordance with the accused's plea. . . . That the *lis* is restricted appears from ss 112 and 113. The proceedings under the former are restricted to the offence "to which he has pleaded guilty" and the latter must be read within that frame.[8]

This established holding ensures that, irrespective of what the defendant has actually done and what charge might accurately describe it, once the prosecutor and the defendant have agreed on the charge to which the latter will plead guilty, and the defendant admits facts sufficient to support a conviction on that charge, the presiding judge is bound either to reject the plea or sentence the defendant in accord with the statutorily authorised punishments for the offence to which he has pleaded guilty. Even if Ngubane had actually committed murder, that is, once the prosecutor agreed to accept his plea to culpable homicide and Ngubane admitted 'facts' sufficient to support this lesser charge, however true or false those admissions might be, he could no longer be sentenced to more than the maximum sentence for culpable homicide on his plea. In this way, the prosecutor can control the maximum sentence to which the defendant will be liable by agreeing to accept a guilty plea for an offence less serious than what he may believe the defendant actually committed. If the difference between the penalty for murder and that for culpable homicide is, say, fifteen years in prison, a prosecutor can offer a defendant whom he thinks is actually guilty of murder the chance to plead guilty to culpable homicide. 'Save me the expense and uncertainty of a trial for murder', he says to the defendant, 'and I'll save you fifteen years'.[9]

Given the heavy caseloads faced by criminal courts in South Africa,[10] it is not surprising that bargains of this sort quickly became common. In *North Western Dense Concrete CC v Director*

---

5 Criminal Procedure Act 51 of 1977, ss 112 and 113; W de Villiers 'Plea and Sentence Agreements in Terms of Section 105A of the Criminal Procedure Act: A Step Forward?' (2004) 37 *De Jure* 244, 253; MB Rodgers 'The Development and Operation of Negotiated Justice in the South African Criminal Justice System' (2010) 23 *South African Journal of Criminal Justice* 239, 255–256.

6 *North Western Dense Concrete CC v Director of Public Prosecutions (Western Cape)* 1999 (2) SACR 669, 677c (C) ('*North Western Dense Concrete CC*').

7 *State v Ngubane* 1985 (3) SA 677 (A) 617, 683. Cf Clarke (note 4 above) at 164–165.

8 Cf Clarke (note 4 above) at 160–162.

9 Cf Clarke (note 4 above) at 160–162.

10 Steyn (note 1 above) at 207; de Villiers (note 5 above) at 245; Clarke (note 4 above) at 142.

of Public Prosecutions (Western Cape), a case to which we will return, the High Court bowed to the inevitable. The court not only acknowledged the pervasiveness of plea bargaining in South African courts but, where the Director of Public Prosecutions (DPP) had reinstituted charges against the applicant despite a plea agreement that had provided that they would not be brought, enforced the terms of the bargain and ordered the charges dismissed. Though he was not 'filled with joy' to say so, Uijs J recognised that plea bargains had become an 'entrenched, accepted and acceptable part' of South African law, taking place daily at every level of criminal justice. He compared them to civil settlements, which also adjust the *lis* to the needs of the parties and whose validity is unquestioned, and suggested that the entire system of criminal justice might break down if the courts refused to enforce individual bargains or struck down the practice altogether.[11]

Despite the High Court's resigned legitimisation of the practice, uncertainty as to the legality of plea bargains remained. On its face, s 112 is about guilty pleas, not negotiations and agreements over sentencing. It enables the bargaining system itself to remain underground, hidden from public view; indeed, had one of the parties not broken the agreement in *North Western Dense Concrete*, the case and the bargain at its core would never have been subject to judicial scrutiny at all.[12] Soon after this decision, as part of a larger inquiry into South African criminal procedure, the South African Law Commission (as it was then called) issued a report recommending that the CPA be amended to provide explicitly for plea bargains, and thus exchanges of guilty pleas for leniency in sentencing. So they could be regularised and regulated, the South African Law Commission offered a statute embodying its proposals for legislators to consider. With some modifications, this recommendation was promptly followed. The amended CPA contained a new s 105A – governing plea and sentence agreements.[13]

Section 105A is a lengthy provision. It regulates its subject in minute detail, ostensibly to protect defendants from overzealous prosecutors who might deprive them of their right to a full trial. In many respects, it is a model statute. It allows only prosecutors who are authorised by the National Director of Public Prosecutions to do so to negotiate agreements with defendants, who must be legally represented. Before entering into an agreement, the prosecutor must consult both the investigating officer in the case and the complainant, advising the latter on the possibility of compensation as part of the agreement. Any agreement on plea and sentence must be reached before the defendant is asked to plead, and the specific terms of the agreement, the charge to which the defendant will plead, the facts to be admitted by the defendant in support of the plea, and the sentence to be imposed by the court after the plea, must all be made explicit in a detailed written document, signed by both parties, that affirms that the defendant's agreement was given voluntarily and in full knowledge of the rights that he would surrender with the guilty plea. The agreement is then put before the presiding judge, who may hear evidence and take statements from the defendant or others before passing sentence. Once all these procedural requirements have been met, the judge, nominally still in full possession of the authority to sentence, considers the sentence proposed in the agreement. If she finds the sentence to be 'just', she then informs both sides of this finding, convicts the defendant on the guilty plea, and imposes the sentence provided in the agreement. If she finds the sentence

[11] *North Western Dense Concrete CC* (note 6 above) at 683f, 674e, 676f and 678c.

[12] Rodgers (note 5 above) at 242–243.

[13] South African Law Commission 'Simplification of Criminal Procedure (Sentence Agreements)' (2001). See also Rodgers (note 5 above) at 241; de Villiers (note 5 above) at 244; Criminal Procedure Second Amendment Act 62 of 2001, s 105A.

'unjust', either too harsh or too lenient, then she must inform the two sides of what sentence she would consider just, and offer them the chance to revise the bargain on her new, preferred terms. If either side objects to the judge's terms, then they may withdraw from the agreement, after which the defendant is entitled to a new trial before a different presiding judge.[14]

In a perfect world, as discussed below, CPA s 105A might operate to guarantee all defendants a procedure for the inducement of guilty pleas that ensures that every plea is made fully voluntarily, in the defendant's own interests, and with full knowledge of the sentence to be imposed upon the plea and the specific consequences that flow from the surrender of the right to trial. But it is easy to see how the imperfect world of criminal justice might frustrate this purpose. For one, s 105A applies only to defendants able to afford legal representation, or to whom representation is given. Sections 35(2)(c) and 35(3)(f)–(h) of the Constitution of the Republic of South Africa, 1996, guarantee legal representation at state expense to defendants who are detained and those being tried 'if substantial injustice would otherwise result'. But fulfilling this promise even for these defendants has proven difficult. And it is not clear that it extends to defendants not under detention in the period between arrest and trial, when most plea bargaining takes place. Since 1996, South Africa has struggled to provide representation to poor defendants in the face of very high rates of crime, a lack of trained attorneys, inadequate funding and poor administration of the state's legal aid services.[15] The great majority of defendants are thus unable to avail themselves of the protections of the formal procedure, and must rely, as they always have, on the informal procedure of CPA s 112. On the other hand, the time- and labour-consuming requirements of formal authorisation from supervisors and extensive consultation with victims and the police often deter harried prosecutors facing unmanageable caseloads from employing the s 105A procedure when s 112 proceedings will do equally well. While there is evidence that s 105A procedures are being employed in some courts, the effect of these conditions, as Steyn observes, has been to confine bargaining in the great majority of cases to the informal, shadowy realm of s 112 rather than allowing plea agreements to migrate to the formal, rule-governed domain of s 105A. This confinement frustrates precisely those reforms the new law was designed to implement.[16] The same pressures of time and material cost that propelled the replacement of costly trials by cheaper plea bargains in the first place continue to bear heavily on South African criminal justice, forcing the replacement of more costly forms of plea bargaining with cheaper ones.

The result is a two-tiered system of plea bargaining in South Africa, two alternative tracks to negotiated sentences on which passengers travel first- or second-class. The first-class track is governed by s 105A, with its close supervision of prosecutors by administrative superiors, its attention to the position of the police and the interests of victims, and the care it takes to ensure both that defendants are fully informed of the precise terms of the agreement and the rights they will surrender by pleading guilty and that their consent to the agreement is given freely. But travel on this track is very expensive, not just for defendants, few of whom can afford the legal representation that s 105A makes the ticket for first-class travel, but for the state as well,

[14] Criminal Procedure Second Amendment Act 62 of 2001, s 105A. See also de Villiers (note 5 above) at 245–250; Rodgers (note 5 above) at 244–255.

[15] CJ Ogletree Jr 'The Challenge of Providing "Legal Representation" in the United States, South Africa, and China' (2001) 7 *Washington University Journal of Law & Policy* 47, 49–54.

[16] M Watney 'Judicial Scrutiny of Plea and Sentence Agreements' (2006) 2006 *Journal of South African Law* 224, 224–225; Steyn (note 1 above) at 215–217. On the use of s 105A procedures, see also M Kerscher *Plea Bargaining in South Africa and Germany* (2013, unpublished LLM thesis, Faculty of Law, Stellenbosch University) 10–11.

which must make good the prosecutorial time and effort required to certify a formal bargain. As a result, the great majority of South African defendants must travel second-class, on the informal track governed by s 112. Here prosecutors operate with far less oversight than on the first-class track, the information defendants have at the moment they must decide how to plead is significantly less complete and reliable, judges retain a freer hand in sentencing in cases where an agreement has been reached, and pleas remain binding even where the information on which a defendant has relied turns out to be faulty.

My purpose here is to examine the operation of this system by considering it in light of some very simple economic logic. If there is any legal institution that might usefully be viewed from an economic perspective, one might think, it would be plea bargaining, the forthright exchange of valuable concessions between parties who each stand to gain from the trade. I hope to suggest more precisely how this might be true, but not by assaulting the problem with complex theory or mounds of statistical data. Instead, I propose a simple, intuitive, empirically plausible depiction, a *model*, of how plea bargaining works that allows us to see more clearly how the behaviour of prosecutors, defendants and judges and the existence of various conditions in the bargaining environment affect the system's outcomes, the individual agreements that result in guilty pleas. It identifies the factors that move both prosecutors and defendants to seek bargains, shows how these factors affect their decisions, and examines the problems bargainers face when the information upon which they must base their decisions is unreliable or unavailable. It illuminates the differences between formal plea bargains under s 105A and informal bargains under s 112, showing how imperfections in the latter environment can lead to dysfunctional outcomes that are less likely to occur in the case of s 105A bargains. To the extent it succeeds, the model may be a valuable conceptual tool for South African scholars and policymakers, a fruitful way to understand plea bargaining, the problems it solves for the criminal process, and the new problems it creates at the same time.

But it is important to be clear as to what analysis of this sort can and cannot do. Though it does enable us to ask more salient questions about how the plea bargaining system operates, how to improve its outcomes, whether on balance it is a good or a bad thing, and what its constitutional implications might be, it cannot answer many of these questions at all, and can answer others only in a conditional or qualified way that leaves ample room for disagreement on political, moral or jurisprudential grounds. Economics, like every social science, has both a *positive* and a *normative* dimension, questions about *what is* and questions about *what ought to be*. The former suggests that economics is a science, a way to understand how the social world actually works, how relationships between individuals, groups and institutions are in fact conducted, and how social outcomes that we can see are produced or created. In practising positive economics, economists are like physicists. They try to provide answers to questions about how things really are in terms that all rational people are bound to accept, just as they are bound to accept that the earth is round, whether they like it or not. But physicists are typically not concerned with the rights of electrons or asteroids, or whether it is a good thing that force and mass are related as they are, or whether we should try to change the laws of thermodynamics because they lead to unjust results. All physics is positive science; it does not ask, or answer, normative questions.

The subject matter of economics, in contrast, is living people and their relations to one another. In this realm of perpetual political and moral controversy, normative concerns, ancient, ultimately unresolvable questions about what justice demands, whether the way things

are is the way they ought to be, and what the object of public policy, or social life itself, should be, irrepressibly jostle with positive questions for our attention. Positive economics can inform normative debate, by showing how and with what consequences various normative positions and the policies that flow from them might play out in practice. But positive economics alone can never answer or determine normative questions. How plea bargaining works, how its outcomes are produced, and how various conditions affect its operation are positive questions, ones about which economic reasoning may have much of value to say. What it would mean to make the plea bargaining system work 'better', or whether it should exist at all, are normative questions, and in addressing them, economists have no claim to greater wisdom or expertise than anyone else. But if it cannot answer them, economic reasoning may serve to sharpen these questions, to show just what is at stake and how normative or constitutional concerns can be identified and addressed, and with what consequences. I hope here to show the value of an economic perspective on plea bargaining in both these contexts.

In part II, I start from the assumption that plea bargaining is a desirable, or at least uncontroversial, institutional innovation that raises no normative or constitutional issues. This approach makes it meaningful, after sketching how the system actually works, to ask how to make it work better, a question sharply posed by the two-track South African system. With the help of a general overview of the plea bargaining system and a simple algebraic depiction of the defendant's problem in deciding whether to plead guilty or insist on a trial, I argue that bargaining under s 112 is fraught with problems stemming from uncertainty. Because defendants do not, or do not always, have accurate or reliable information about the factors that most strongly influence their decision to plead guilty or insist on a trial, the plea bargaining system is likely to fail in a specific way. Plea bargains that would have been made had defendants had the information that they needed will not be made, and plea bargains that would not have been made had defendants had this information will be made, because they do not. 'Making the system work better' thus means facilitating the former type of bargain and impeding the latter, a problem of institutional design and performance. If a criminal process commits itself, reluctantly or enthusiastically, to plea bargaining as a legitimate means of resolving criminal cases, it ought to prefer a bargaining system that works well at identifying and completing genuinely consensual agreements to a system that works poorly. I argue that the procedure of s 105A, cumbersome as it is, addresses most of the informational problems that arise under s 112. That is, cost considerations aside, the system's operation would be improved – more bargains that should take place, and fewer that should not, will take place – if all South African plea bargains were formal and governed by s 105A.

The superiority of s 105A bargaining to its informal counterpart on various grounds has been widely observed. While the courts have as yet had nothing to say on the question, several scholars have noted that this disparity raises a serious problem under s 9 of the Constitution, which guarantees equality before the law. The majority of defendants, who are poor, cannot take advantage of the formal protections and guarantees of s 105A because they, unlike wealthier defendants, generally have no legal representation.[17] In the final part of this article, following

[17] De Villiers (note 5 above) at 253–255; Steyn (note 1 above) at 217–218; Rodgers (note 5 above) at 261; A Botman *An Evaluation of the Benefit of Plea and Sentence Agreements to the Unrepresented Accused* (2016, unpublished LLM thesis, Faculty of Law, University of the Western Cape) 82–91. A rather different array of jurisprudential objections to plea bargaining are discussed in M Bennun 'Negotiated Pleas: Policy and Purposes' (2007) 20 *South African Journal of Criminal Justice* 17. I consider some of these objections below.

these scholars, I relax the working assumption of part II and ask not whether plea bargaining is a good thing, but how it comports with the structure and interpretive values of the South African Constitution. A brief look at the constitutional history of plea bargaining in the United States raises a range of normative questions about the system of plea bargaining rather different from the one that concerned these scholars, and engages the interpretive strategies employed by the American federal courts to address them. I then compare these constitutional questions and interpretive strategies to the alternative approaches suggested by the particulars of South African law and institutional development, and conclude with a brief sketch of the issues that might frame a constitutional challenge to plea bargaining in South Africa. I hope in both these ways to advance the important emerging debate over plea bargaining in South Africa.

II MAKING PLEA BARGAINING WORK BETTER

A How plea bargaining works

Challenging Judge Uijs's analogy of plea bargaining to civil settlements in *North Western Dense Concrete*, Mervyn Bennun called the differences between the two so sharp that 'the comparison verges on the cynical'. But it was, he went on, 'useful' in pointing out the resemblance of plea bargaining to 'the purchase of a rug at a Lebanese bazaar':

> To move away from the outcome of a criminal trial as being unpleasantly redolent of the marketplace, a plea and sentence agreement negotiated under s 105A should result in the same verdict and sentence which would have followed had the matter gone to a full trial, and the court should satisfy itself that this is the case. . . . [A] negotiated agreement under s 105A [surely] should propose the sentence which would have been imposed after a full trial.[18]

Perhaps Bennun, a sharp critic of the practice, proposed this interpretation knowing that it would, if adopted, largely eliminate plea bargains under s 105A and render them available only under s 112. Under s 112, he thought, leniency might still be justified if the presiding judge were satisfied that the defendant's plea showed genuine remorse.[19] But his interpretation of s 105A plea bargains would render them pointless, and thus make them disappear.[20]

The crux of every consensual plea bargain is the sentencing discount given in exchange for the guilty plea. As elaborated below, if there is no difference between the sentence the defendant would receive after conviction at trial and the sentence proposed in the agreement, then there is little reason for a defendant with even a very slim chance of acquittal to accept an agreement, convict himself without putting the prosecution to the proof, and surrender his right to take the small chance that he will somehow avoid conviction. Even if the plea is motivated, as some are, not by fear of greater punishment after trial but by the adverse consequences of bringing the facts of the matter to light in a public trial, s 105A's requirement that the facts be stipulated in writing in the agreement largely removes this advantage of guilty pleas.

As Bennun suggests, the exchange of sentencing discounts for guilty pleas is not pretty, but if the system of plea bargaining is to work at all, this is the way it has to work. Everywhere it has taken hold, plea bargaining has been an institutional response to the needs of the state, not the interests of defendants. When prosecutors and administrative judges face caseloads so heavy that only a small sliver of cases can possibly be afforded a full trial given the resources at hand,

[18] Bennun (note 17 above) at 29.

[19] Ibid.

[20] Compare Watney (note 16 above) at 226.

an environment in which the American criminal process has laboured since at least 1900[21] and, as Steyn suggests,[22] has characterised South Africa for at least several decades as well, prosecutors have little choice but to somehow persuade large numbers of defendants to surrender their right to a trial and agree to convict themselves. This deeply entrenched process obviates the need to mount a trial in those cases and allows the state, as the provider of criminal justice, to put its extremely scarce prosecutorial and judicial resources to better use elsewhere. Absent the public expenditure necessary to provide a full trial for every defendant, the alternative, as Uijs J recognised, is the collapse of the system itself. Because so few defendants are so remorseful or so unwilling to endure the publicity of a trial to agree to plead guilty without some material inducement – something they want and deem sufficient to compensate for the loss of rights and the possibility of acquittal their plea entails – prosecutors must offer something in return for the guilty pleas they need to keep the criminal process afloat. In plea bargaining systems everywhere, sentencing discounts are the only answer.

One way bargains might be achieved, *sentence bargaining*, would have the defendant or his legal representative bargain directly with the sentencer, the trial judge. But concerns about judicial neutrality and dignity, in a process most observers (and participants) see as demeaning, keep judges in almost every adversarial system, including South Africa and the United States, from intervening directly in plea negotiations. The alternative, *charge bargaining*, is almost universally employed because it is so well adapted to a peculiar feature of the adversarial system, one that prosecutors facing severely overcrowded dockets exploit to the fullest. In inquisitorial systems, like those of continental Europe, prosecutors are generally bound by a *rule of compulsory prosecution*. This type of rule requires them to try every case on the highest possible charge the evidence will support, nothing more or less, so as to produce a judgment that, to the greatest extent possible, accurately reflects the actual facts and moral weight of the case. If a defendant, like Ngubane, is actually guilty of murder, then the prosecutor must charge him specifically with and try him for murder.

But subject to a provision for private prosecution, at common law and under s 6 of the CPA, South African prosecutors have almost complete discretion to decide whether or not to prosecute any defendant and, if so, for what crimes.[23] At the same time, the substantive criminal law establishes a large repertory of crimes covering a broad range of behaviours, most modified by degrees, many overlapping or governing activities typically undertaken as part of a single criminal transaction.[24] This profusion of distinct offences gives prosecutors a great deal of sentencing power. Subject to whatever constitutional or administrative prohibitions of vindictive overcharging might exist, a prosecutor can typically charge a suspect with several independent crimes committed in the same course of action, exposing him to a potential punishment much greater than that prescribed for the principal crime in the case. Most importantly, as in *Ngubane*, since there is no rule of compulsory prosecution, the charge the prosecutor selects need not accurately describe what she believes the suspect has actually done. So she can charge a suspect with any crime or combination of crimes on the list that he *has* committed, or with some lesser offence that he *has not* actually committed. This latitude

[21] LM Friedman *Crime and Punishment in American History* (1993) 235–323; JH Langbein 'On the Myth of Written Constitutions: The Disappearance of Criminal Jury Trial' (1992) 15 *Harvard Journal of Law and Public Policy* 119; Alschuler (note 3 above).

[22] Steyn (note 1 above) at 207–208, 218.

[23] Criminal Procedure Act 51 of 1977, s 6; Steyn (note 1 above) at 206–207.

[24] G Kemp (ed) *Criminal Law in South Africa* (2nd Ed, 2015).

forces the judge to sentence the defendant to the punishment the prosecutor has promised in the bargain.

Plea bargaining in this form is made *possible* by two aspects of adversarial procedure that are absent in inquisitorial systems. First, defendants have the right to plead guilty and abort trials before they occur. Section 35 of the Constitution guarantees every accused the right to a fair trial, but it does not command anyone to exercise it. Defendants may have reasons other than, or in addition to, an offer of leniency to avoid a trial, and pleading guilty allows them to do it. A guilty plea is a conviction, an agreement by the defendant to submit to punishment for the offence to which he has pleaded guilty, and it ends the criminal case. Typically, the only consequence of the plea to the defendant is the sentence to which he has agreed. Inquisitorial defendants cannot plead guilty or abort their trials, because the point of inquisitorial trials is to provide an accurate account of what actually happened and establish an appropriate punishment for what the perpetrator has done. But adversarial trials, adorned as they are with complex rules of evidence and procedural rights, like the right against self-incrimination, designed to protect defendants from the consequences of the truth, are not primarily about discovering the truth of the matter. They are, ideally, about finding guilty defendants criminally liable for something that allows appropriate punishments to be imposed on them. So the right to plead guilty, the power to subject the prosecutor to a costly trial or make it unnecessary, becomes a valuable bargaining chip for defendants in negotiations. Defendants have something to give that prosecutors want.

The second enabling feature of the adversarial system is that prosecutors, because of their broad charging discretion, have something of value to give them in return. By selecting charges and combining them with sentencing recommendations that, at least in the United States, they know judges, who operate under the same pressures they do, will respect, prosecutors can calibrate punishments as finely as necessary to win guilty pleas from defendants with penalties they consider in each case to be acceptable discounts for the defendant's saving the government the cost and uncertainty of trying them. 'Save us this much money', the prosecutor says, 'and we'll save you this much time'. Again, this requires that defendants, in the last analysis, be the ones who bear the state's costs of convicting them at trial, in the form of heavier punishments upon conviction at trial, so as to preserve their incentive to bargain. Plea bargaining is only possible where defendants who do insist on a full trial are consistently sentenced upon conviction with more severe punishments than they would have been had they accepted a plea agreement. In this way, prosecutors can make surrendering the right to the trial attractive to all defendants. Only then will a proffered sentencing discount be seen by defendants as credible.[25]

Where defendants are denied their freedom before trial, and the direct and collateral consequences of detention weigh heavily on them during bargaining, the prosecutor's hand is greatly strengthened, and the threat of prolonged pre-trial detention can be employed strategically to induce pleas from defendants on terms more favourable to the prosecution than would otherwise be the case. Much the same effect is created by statutes that attach mandatory minimum sentences to conviction for specific offences. The political pressures that produce these statutes generally ensure that the minimum sentences are relatively harsh compared to the perceived inclinations of sentencing judges. If so, and judges apply these elevated sentences consistently, without 'nullifying' them by refusing to find defendants guilty of these offences

[25] R Adelstein *The Exchange Order: Property and Liability as an Economic System* (2017) 206–214.

because judges believe the statutory minimum is inappropriate, then mandatory sentencing statutes allow prosecutors to threaten defendants with very severe punishments if they are convicted of these crimes at trial. As we will see, this increase in the punishments imposed after conviction at trial means both fewer trials at which to impose them, because more bargains will be struck by defendants eager to avoid trials that might result in the statutory minimum, and bargains more favourable to the prosecution. In other words, mandatory minimum sentences will extract guilty pleas from defendants to lesser offences without mandatory minimum sentences at punishments greater than would have resulted without the minimum sentences in the background. In this way, the effects of the mandatory sentences ramify throughout the criminal process, increasing punishments for the crimes to which they do not apply as well as those to which they do.[26]

This much is true of every mandatory sentencing law. But South Africa's mandatory sentencing provisions, s 51 of the Criminal Law Amendment Act 105 of 1997 ('CLAA'), intended to be a temporary response to a perceived wave of violent crime across the country,[27] add a further twist. The statute makes life imprisonment mandatory for defendants convicted of murder and rape under specific aggravating circumstances, mandates fifteen-year sentences for a range of other crimes, requires sentences from fifteen to twenty-five years for repeated offences in various situations, and allows judges to impose lesser punishments only where they are 'satisfied that substantial compelling circumstances exist which justify' them, though that exception was construed quite narrowly in *State v Malgas*.[28] However, the mandatory sentencing provisions apply only to cases tried in a High Court or a regional court. They do not bind district magistrates' courts, where most cases involving offences listed in ss 51(1) and 51(2) are tried. So in addition to the bargaining power conferred by their broad discretion in selecting charges, sometimes further enhanced by extensive pre-trial detention of the defendant, prosecutors can increase the pressure on defendants charged with crimes subject to s 51 by threatening to try their cases in a High or regional court rather than a district court. All of this, combined with the general intrusion on the traditional prerogatives of sentencing judges represented by the statutory mandates, has made s 51 deeply unpopular among South African judges and commentators.[29] But their disapproval has not dislodged it; the temporary legislation was upheld by the Constitutional Court in *State v Dodo* (2001) and extended every two years until 2007, when it was made permanent by the Criminal Law (Sentencing) Amendment Act 38 of 2007.[30]

The institutions of adversarial procedure make plea bargaining possible. What makes it *necessary* in the systems that have come to rely on it is a combination of two factors common to most of them: (i) the sheer volume of crime and correspondingly large caseloads these

[26] Cf SS Terblanche (2003) 'Mandatory and Minimum Sentences: Considering s 51 of the Criminal Law Amendment Act 1997' (2003) *Acta Juridica* 194, 216, 218.

[27] D van Zyl Swit 'Mandatory Sentences: A Conundrum for the New South Africa?' (2000) 2 *Punishment and Society* 197, 200–205; E Cameron (2017) 'Imprisoning the Nation: Minimum Sentences in South Africa' Faculty of Law, University of the Western Cape, Dean's Distinguished Lecture 5–13, available at https://www.groundup.org.za/media/uploads/documents/UWCImprisoningThe%20Nation19October2017.pdf.

[28] Criminal Law Amendment Act 105 of 1997, s 51(1–3); Terblanche (note 26 above) at 197–215; *State v Malgas* 2001 (2) SA 1222 (SCA).

[29] Van Zyl Smit (note 27 above) at 205-208; Terblanche (note 26 above) at 218-220; Cameron (note 27 above) *passim*.

[30] *State v Dodo* [2001] ZACC 16, 2001 (3) SA 382 (CC).

chronically underfunded systems must confront, and (ii) the elaboration of the trial process itself, the complexity of many modern statutes, the difficulties of proof they pose, the evolution of procedural safeguards for defendants and rules of evidence that increase the length, cost and uncertainty of full trials. The proportion of convictions in a system that result from plea bargains, 90 per cent in the United States, and the extent of the sentencing discounts that must be offered to secure the needed rate of guilty pleas, are measures of the pressure exerted by these factors on the system. In South Africa, though trial procedures are simpler than in the United States, the incidence of crime is comparable. As a result, prosecutors and courts still do not have sufficient resources to meet the demands placed on them. The absence of data makes the effect of the above factors harder to determine. However, if Uijs J is right, and the South African criminal justice system, like its American counterpart, could not withstand the abolition of plea bargaining, the practice must be very widespread, and the discounts correspondingly deep.

B The defendant's problem

Although the prosecutor may control the charge and, to a large extent, the sentence, there can be no bargain without the defendant's consent. Thus, if guilty pleas are entered knowingly and voluntarily, the defendant controls the plea bargain. Prosecutors for whom inducing a guilty plea in a given case is not especially urgent, perhaps because they have some special reason to take the case to trial, or because the volume of crime generally does not press too heavily on their resources, may offer only low prices for a plea. Put another way, defendants in these cases will receive smaller sentence discounts relative to the sentences these defendants could expect to receive after conviction at trial. Where, as in the United States, the need to induce pleas is acute because there are far too many cases to try with the available resources, prosecutors may offer very high prices for guilty pleas. Such deep discounts make the negotiated sentence a mere fraction of the expected sentence after a trial. But whether the proffered discount is high or low, it is the defendant who must agree to accept it. Only he can make the trial unnecessary.

So it is useful to examine the defendant's problem in a plea bargain more closely. Should he accept the sentence proposed by the prosecutor, or should he retain his right to trial by refusing the bargain and pleading not guilty? On the one hand, we will assume for now that the bargain represents a sentence the defendant can count on, that he knows for certain. On the other, going to trial presents the prospect of two possible outcomes, an acquittal, which for simplicity we will represent as a punishment of zero, and a conviction, which brings a punishment that, again for now, we will assume the defendant knows, along with some perceived likelihood of one or the other. Suppose, for example, I am accused of aggravated robbery, for which, if I am convicted at trial, I am certain to receive a ten-year sentence. My defence is that I did not do it, which may or may not be true. But there is evidence against me: my gun was used in the robbery, and my fingerprints were found on it afterward. All things considered, my attorney tells me, the chance of my being convicted at trial is 70 per cent. So the *expected punishment* I face in going to trial, defined as the sentence that I would receive on conviction multiplied by the likelihood that I will in fact be convicted, is $0.7 \times 10 = 7$ years. This calculation does not mean that I will be sentenced to seven years, since the only possible outcomes of my trial are ten years or none. It means that if, hypothetically, my trial could be conducted identically a million times before a million different impartial judges, and my attorney is right that 70 per cent of these trials would result in a punishment of ten years and 30 per cent in a punishment

of zero, the 'average' punishment I would receive in the million trials would be seven years per trial. Going to trial is thus like being subjected to a punishment lottery that, 'on average', costs me seven years per trial.

Suppose the prosecutor tells me that if I plead guilty to simple robbery, he will guarantee a sentence of seven years. Then one could say that the 'objective values' of the plea bargain and the punishment lottery are the same: take the deal and get seven years for certain, or enter the lottery and expect to pay seven years for an 'average' trial. Whether I take the deal or not depends on how I feel about the risk posed by the lottery. If, for whatever reason, I am especially terrified by the thought of going to prison, I might look at the lottery pessimistically and focus on the awful prospect of ten years behind bars. Seven years for certain seems a better outcome for me now than the prospect of serving even more if I am convicted at trial. If so, I am *risk averse*, and will take the deal. Because the 70 per cent chance of a sentence as long as ten years seems worse to me than a certain sentence of just seven, I would even be willing to accept an even lower price for my plea, a bargained sentence a little *greater* than seven years, just to be relieved of the possibility of serving ten. But if prison does not hold all that much terror for me, perhaps because I have been there before and think of it as part of criminal life, or because I can pursue valuable relationships inside (or outside) the prison while I am there, I will enter the punishment lottery by refusing the bargain and going to trial. Because I look optimistically at the trial as a chance to walk free rather than pessimistically as a chance to serve ten years, I *prefer the risk*. To get me to relinquish my chance for an acquittal, despite the odds against it, the prosecutor will have to offer a deeper discount, a higher price, for my guilty plea, a certain sentence of *less* than seven years.[31]

It is thus not enough to depict the defendant as concerned solely with choosing the option that yields the smallest possible 'objective' sentence, in which case every defendant faced with this situation would accept a bargain at a day under seven years or less and refuse it at a day over seven years or more. Different defendants value a year in prison differently. A year in prison, or each year of a sentence of more than a year, imposes a *cost*, a quantum of distress or suffering, on defendants that differs from defendant to defendant, and, for any given defendant, from year to year. These differences produce different attitudes toward the risks of the punishment lottery. Suppose we define the variable p as the likelihood, or *probability*, that any defendant will in fact be convicted at a full trial. This probability can vary from zero, which means that the defendant is certain to be acquitted at trial, to one, which means that he is certain to be convicted. In the example above, then, the defendant's estimate that the chance that he will be convicted at trial is 70 per cent would mean that he estimates the probability p as 0.7. Now define the variable B as the sentence that the prosecutor offers the defendant in exchange for his guilty plea, and the variable T as the sentence the defendant would receive were he to be convicted after a full trial. B is always less than T, and the difference between them is the sentence discount associated with the negotiated sentence B. Then, as above, the defendant faces an expected punishment of $p \times T$ should he choose to exercise his right to a full trial.

If all that a defendant cared about was the severity of the punishment he would have to serve, the choice between a negotiated punishment B and the prospect of a full trial with an expected punishment $p \times T$ would turn on the relative values of the two punishments: if B were smaller than $p \times T$, the defendant would accept the bargain and the punishment B, but if B were greater than $p \times T$, he would reject the bargain and take his chances at a trial. But

[31] Ibid at 216–217.

because different defendants value the consequences of punishments differently, the relevant comparison is between the *cost* imposed on the defendant by the bargained sentence *B* and the *expected cost* of going to trial and risking the sentence *T*, the defendant's 'average cost per trial'. We can represent the way any given defendant experiences a particular punishment *X* mathematically by defining a *cost function C(X)* for that defendant. This mathematical relationship, which differs from defendant to defendant, transforms the objective punishment *X* into a number, *C*, that represents the defendant's subjective experience of distress or suffering at being forced to endure *X*. If we assume, reasonably, that *C* increases as *X* increases, that is, that defendants see more punishment as worse than less, then the question actually posed by the plea offer is whether *C(B)* is less than $p \times C(T)$. If so, the defendant will agree to plead guilty and accept the sentence *B*. If not, he will reject the agreement and enter the punishment lottery.

This formulation suggests three distinct ways that a prosecutor can induce a defendant to plead guilty, that is, make the left side of the inequality smaller, or the right side larger, for that defendant. All other things being equal, the trial alternative can be made less attractive relative to the bargain by increasing either *T*, the sentence on conviction at trial (perhaps by invoking a mandatory minimum sentence statute), or *p*, the likelihood the defendant will be convicted, or both. Or the bargain can be made more attractive relative to the trial by lowering *B*, the sentence the defendant will serve if he pleads guilty. These variables correspond to the three principal bargaining tactics employed by prosecutors in adversarial systems everywhere, with each raising the question of constitutional propriety in a different way. Most commonly, prosecutors attempt to induce guilty pleas simply by deepening the sentencing discount ('sweetening the deal'), paying high prices for pleas by offering bargains at low values of *B*. This is the form taken by the archetypical plea bargain, so deciding on the propriety of this tactic is tantamount to deciding on the propriety of the practice itself. But even if this propriety is conceded in general, if prosecutors are so desperate for pleas that they offer extremely low values of *B* relative to *T*, say a three-month sentence for a weapons offence for a defendant facing the possibility of ten years for aggravated robbery after a trial, the very sweetness of the deal might induce even a truly innocent defendant who thinks there is even a small chance that he will be convicted at trial to plead guilty, whether he prefers risk or is averse to it.[32]

Alternatively, prosecutors can increase the likelihood of conviction by devoting more resources to investigation or, as is often done in cases involving more than one defendant, securing the testimony of one defendant against another by offering the defector a chance to plead guilty to a lesser offence in exchange for significantly increasing the likelihood that his co-defendant will be convicted at trial, and thus helping persuade the co-defendant to plead as well. It is precisely this situation that economists describe as *the prisoner's dilemma*. Given the propriety of bargaining itself, this common tactic creates no problem of convicting the innocent. Why? Because in any adversarial system, including South Africa's, the only authoritative institution that exists for discovering, or pronouncing, anyone guilty or not guilty is the criminal process itself. If the evidence is strong enough to convince a trial judge beyond a reasonable doubt that a defendant is guilty of a crime, he *is* guilty, and subject to punishment, even if he really did not do it. It is the conviction that creates the legal fact of guilt, irrespective

[32] R Adelstein 'The Negotiated Guilty Plea: A Framework for Analysis' 53 *New York University Law Review* (1978) 783, 820–823.

of what the defendant or anyone else may believe.[33] Subjecting defendants to the prisoner's dilemma thus cannot be unfair to the co-defendants who become more likely to be convicted, because finding defendants guilty or not as a legal fact, not finding authoritatively exactly who did or did not do what, is what adversary trials are about, and increasing the likelihood that a defendant will be found guilty at a fair trial is what prosecutors are supposed to do.

The third strategy, increasing *T*, is more problematic. In the United States, this is typically achieved by 'overcharging', piling a host of distinct charges on a defendant who has committed several offences in the course of a single criminal transaction. In *Brady v United States* (1970),[34] in which the Supreme Court first gave its approval to plea bargaining in general and the first two tactics in particular, the Court stressed that the defendant's participation in a plea bargain must be truly voluntary, and hinted that egregious overcharging might so 'overbear the will of the defendant' as to effectively coerce him into pleading guilty to lesser charges. But eight years later, in *Bordenkircher v Hayes*,[35] the Court was forced to face the harsh realities of American plea bargaining, conducted in an environment of high crime, inadequate resources to prosecute it and, by international standards, very severe prison sentences for almost every serious offence. Hayes, who had two prior felony convictions, was charged with passing a bad cheque for less than $90, an offence punishable by imprisonment of two to ten years. The prosecutor offered Hayes a five-year sentence for his guilty plea, and to pressure him to accept, told him that if Hayes were convicted at trial on the cheque charge, he would invoke the state's habitual offender act, a mandatory sentencing statute that provided for an automatic sentence of life imprisonment after a third felony conviction. Hayes, fully aware of this threat, nonetheless refused the bargain, was convicted at trial and, when the prosecutor invoked the habitual offender act as threatened, duly sentenced to life imprisonment.

But because it had to in order to preserve the plea bargaining system and, with it, the entire American criminal process, the Court held that the threat and eventual invocation of the habitual offender act was an acceptable prosecutorial bargaining tactic. If an injustice was done to Hayes in this case, it suggested, it was not the prosecutor's threat that did it, but the fact that the state's habitual offender act, which was not challenged in the case and which all sides agreed Hayes had violated, permitted a life sentence in these circumstances. If plea bargaining is to work, defendants must know that the sentencing discount is real. The only way any defendant can be induced to plead guilty in exchange for a proffered discount is to turn those, like Hayes, who do insist on a trial into admonitory examples for the rest, by following through on the threat to sentence them more harshly when they are convicted than they would have been had they pleaded guilty. Prosecutors who need to induce pleas, the Court reluctantly recognised, must be permitted to credibly threaten recalcitrant defendants in this way and, as in *Bordenkircher*, to use existing statutes that defendants have in fact violated to do so:

> Indeed, acceptance of the basic legitimacy of plea bargaining necessarily implies rejection of any notion that a guilty plea is involuntary in the constitutional sense simply because it is the end result of the bargaining process . . . . [By] tolerating and encouraging the negotiation of pleas this Court has necessarily accepted as constitutionally legitimate the simple reality that the prosecutor's interest at the bargaining table is to persuade the defendant to forego his right to plead not guilty.[36]

[33] Cf *Herrera v Collins* 560 US 390 (1993)(Petitioner convicted of capital murder is not entitled to *habeas corpus* relief simply because, years after his conviction, new evidence tending to prove his innocence is discovered).

[34] *Brady v United States* 397 US 742, 750 (1970); Adelstein (note 32 above) at 823–827.

[35] *Bordenkircher v Hayes* 434 US 357 (1978); Adelstein (note 25 above) at 227–229.

[36] *Bordenkircher v Hayes* (note 35 above) at 361–364.

C Information and uncertainty

Defendants, and only they, know their own cost functions and the attitudes toward punishment these reflect. But in order to decide how to plead, a defendant also needs information about three variables, *B*, *p* and *T*. These are 'objective' in the sense that they can be expressed numerically, but the information the defendant has about them may be imperfect, forcing him to make largely subjective estimates on the basis of the information he has about what the true values of the objective variables are. In many situations, defendants do have about as much of the information as it is possible for them to have. If the prosecutor is clear as to what value of *B* he is offering and the defendant knows both that the prosecutor will not renege on the offer in court and that the sentencing judge will respect the agreement and impose the sentence *B*; if the defendant can rely on a well-informed guess (all that is possible in any case), formed either on the basis of his own experience or the advice of counsel, as to the likelihood that he will be convicted; and if he can be confident as to what sentence he would receive if he were convicted at trial, then the decision the defendant makes as to plea can be said to be fully informed, and absent coercion – a question we will soon revisit – a reflection of his best interests as he sees them in a painful predicament. But if not, apart from whatever normative questions of fairness or constitutionality might arise from allowing defendants to plead guilty in the absence of reliable information about *B*, *p* and *T*, where defendants must in fact decide whether to plead guilty in substantial ignorance of these variables, the plea bargaining system itself is likely to perform poorly in the role the criminal process has assigned to it.

How do defendants acquire this information and assess its reliability? Some defendants, as suggested above, will be able to rely on personal experience. Having encountered the criminal process before, they may be aware of the sentencing practices of trial judges, have a good sense of what the likelihood is that they will be convicted on the available evidence, and perhaps even be able to judge whether the prosecutor is offering them the best deal they can get or if a bit of resistance might result in a better one. But whether they are experienced or not, personal knowledge, from whatever source it might be gleaned, and the representations and promises of the prosecutor are all the information that most defendants in CPA s 112 proceedings have. For any defendant, the best source of reliable information during the bargaining process about all the relevant variables, and the person best positioned to wrangle the best bargain possible from the prosecutor, is a competent defence attorney, and counsel is a luxury most CPA s 112 defendants do not enjoy. Without an attorney, defendants, typically at a moment filled with stress and anxiety, must evaluate not just the likelihood of conviction at trial and the sentence that will be imposed thereafter, but the precise terms and legal consequences of the proffered discount and its value relative to the trial sentence, by themselves. Compared to competently represented defendants, these circumstances put them at a substantial disadvantage in identifying and concluding agreements that actually serve their interests. Should they err in estimating the key variables and agree to a bargain they come to regret, once entered and accepted by the judge, the deal becomes final, just as it does for represented defendants. In *State v Armugga*, Msimang J noted as follows in refusing a request to reopen a s 105A bargain after the defendant learned he had agreed to a less favourable deal than he thought he had:

> In the process of bargaining, numerous assumptions are made and mistakes are bound to happen. Provided that a party is found to have acted freely and voluntarily, in his or her sound and sober senses and without having been unduly influenced when conducting a plea bargaining

> agreement, the fact that the assumptions turn out to be false, does not entitle such a party to resile from the agreement.[37]

In addition to an attorney to advise them about T and p, s 105A defendants also have the advantage of having the agreement in all its particulars reduced to writing, so that the true value of the prosecutor's offer B, the magnitude of the discount it represents and the full legal consequences it entails, are made as clear as possible to them. If, having agreed to plead guilty in return for the sentence specified in the agreement, the defendant is surprised at the judge's refusal to impose it in court, he may withdraw the plea and demand a full trial. His s 112 counterpart, in contrast, represented or not, will have no opportunity to withdraw his plea even if, having agreed to an informal deal with the prosecutor and pleaded guilty in reliance on it, he finds the judge unwilling to honour the agreement and determined to impose a sentence greater than B. Even where the information the defendant needs exists in reliable form, where, that is, prosecutors offer clear terms with no intention of reneging and judges are prepared to sentence as the agreement specifies, the procedures of s 105A are clearly superior to the practices of s 112 in providing defendants with the information that they need to ensure that their decisions represent, as far as practicable in a challenging environment, their own interests as they see them. Where information exists but is unreliable, not only are the difficulties compounded for individual defendants already faced with tough decisions – whether they have legal representation or not – but serious questions are raised about the dysfunction and fairness of the plea bargaining system itself.

Because every criminal case is different, and because at the moment defendants must plead, the trial lies in an unknowable future, it is hard to see how an experienced defence attorney's estimate of p, the likelihood of a particular defendant's conviction, can be improved by institutional design. The punishments in the statute books themselves, supplemented perhaps by informal knowledge of the sentencing habits of particular judges, are generally sufficient to give defendants a useful predictor of T. But misinformation, not simply error, about the variable B, the precise value of the sentencing discount the defendant is being offered for his guilty plea, is often the result of specific actions, intentional and unintentional, by prosecutors and judges that can effectively be policed and remedied by statute and judicial oversight. If left unaddressed, these errors and omissions will erode the bargaining system's ability to identify and facilitate mutually beneficial bargains.

Recall that, given B, p and T, a defendant will plead guilty if the subjective cost to him of the bargained sentence is less than the expected cost of going to trial, that is, $C(B) < p \times C(T)$. Suppose, however, the possibility exists either that the prosecutor will renege on his offer when it comes time for him to perform, say by making a different sentencing recommendation to the court than that provided in the agreement, or that the judge, who ultimately remains in possession of the sentencing power, will not abide by the terms of the agreement and sentence the defendant more harshly than the agreement demands. In either case, if the defendant thinks there is a chance that, having been promised a sentence B, he will actually receive a more severe punishment and not be able to extract himself from the bargain, he will have to replace B in his calculations with the higher sentence $(B + h)$, where h represents the expected increase in the sentence created by the possibility that the prosecutor will renege on his promise or the judge will impose a sentence greater than B. If there are enough broken promises or unexpected judicial nullifications of bargains so that h is significant, plea bargains that would have served

[37] *State v Armugga* 2005 (2) SACR 259, 265c–d (N).

not just the interests of the immediate parties to it but the larger interest of the criminal process in diverting trials to less costly procedures as well, and that would have been made without this uncertainty, won't be made because of it. This will occur whenever the cost to the defendant of a bargain at B is smaller than the expected cost of the trial but the cost of a bargain at $(B + h)$ is greater than the expected cost of the trial, that is, when $C(B) < p \times C(T) < C(B + h)$. The existence of even a few unkept bargains will be enough to surround every subsequent negotiation with an uncertainty that keeps at least some otherwise mutually beneficial bargains from being concluded. Expensive trials that no one wants will have to be mounted, and the already burdensome costs of criminal justice will rise still further. Just as bad money drives good money from circulation, bad bargains will drive out good ones, to everyone's detriment.[38]

In 1971, in *Santobello v New York*, the United States Supreme Court addressed the first of these sources of misinformation, the failure of the prosecutor to keep the promises made in the agreement.[39] Santobello, charged with two gambling offences, agreed to plead guilty to one in return for the prosecutor's promise not to recommend a prison sentence. But between the time the agreement was concluded and Santobello was to be sentenced, a new prosecutor had taken over the case and, unaware of the promise his predecessor had made, recommended that Santobello be sentenced to a year in prison on his guilty plea, and the judge complied. Santobello protested, but was not allowed to withdraw his plea. On appeal, the Court, speaking through Burger CJ, found for Santobello. It began by explicitly acknowledging the inability of the American criminal process to survive without plea bargaining, something the Court in *Brady* had refused to do the previous year, grounding its approval of bargaining instead on the 'mutuality of advantage' it offers to both prosecution and defence.[40] Plea bargaining, the *Santobello* Court now said, is 'an essential component of the administration of justice. Properly administered, it is to be encouraged'. But proper administration requires 'fairness in securing agreement', and in ensuring it the Court applied a basic principle of ordinary contract law to hold that, despite the fact that the breach of the agreement was inadvertent and the one-year sentence itself not unjustified, Santobello must receive specific performance of the agreement or be permitted to withdraw his plea: '[W]hen a plea rests in any significant degree on a promise or agreement of the prosecutor, so that it can be said to be part of the inducement or consideration, such promise must be fulfilled'.[41]

For decades before 1970, the Supreme Court turned a blind eye to the unpleasant business of plea bargaining. Only after its hand was forced by some careless language in a 1968 opinion not involving a plea bargain at all[42] did the Court discuss plea bargains in detail and start regulating them in *Brady* and *Santobello*. In South Africa too, plea bargaining under s 112, increasingly acknowledged by scholars and practitioners,[43] remained in the judicial shadows until 1999, when *North Western Dense Concrete CC v Director of Public Prosecutions* was decided by the High Court on facts much like those in *Santobello*. The applicant and its employee were

[38] Adelstein (note 32 above) at 814–816.

[39] *Santobello v New York* 404 US 257 (1971).

[40] Ibid at 260. See also *Brady v United States* (note 34 above) at 752.

[41] *Santobello v New York* (note 39 above) at 260, 261, 262. Note that reaching the 'fair' outcome in this case also had the effect of 'making plea bargaining work better'. This confluence of fairness and efficient operation in the bargaining system is discussed further in R Adelstein 'Institutional Function and Evolution in the Criminal Process' (1981) 76 *Northwestern University Law Review* 1, 71–79.

[42] *United States v Jackson* 390 US 570 (1968). See Adelstein (note 25 above) at 218–220.

[43] Compare Bekker (note 3 above) at 218; Steyn (note 1 above) at 206; Kerscher (note 16 above) at 30.

charged with culpable homicide, and the prosecutor agreed to drop the charge against the applicant if the employee pleaded guilty, which he did in a s 112 proceeding. But after another party had petitioned the respondent for a *nolle prosequi* in the case, the DPP reconsidered its position and reinstituted the charge against the applicant, in violation of the earlier plea agreement.[44]

Uijs J's opinion closely parallels those of the Supreme Court in *Brady* and *Santobello*. He began by acknowledging the long history of informal plea bargaining under s 112 in South Africa, a provision 'tailor-made' for plea bargains, and lamented the judiciary's unwillingness to confront it openly.[45] The court conceded, as the Supreme Court had in *Santobello*, that bargaining under s 112 was an integral component of South African criminal justice, one so entrenched that eliminating it was likely to cause the breakdown of the entire criminal justice system[46] and, like the *Santobello* Court, appealed to ordinary contract law to draw the necessary conclusion. Because it would be unfair to allow the respondent to enjoy the benefits of the agreement without meeting the obligations it created for him, the court granted a permanent stay of prosecution to the applicants, effectively ordering specific performance of the contract. It held that a guilty plea under s 112, if accepted by the prosecutor, binds both the prosecutor and the court on the facts stipulated in the agreement, and requires the prosecutor to honour the terms on which it has been reached.[47] As in *Santobello*, requiring prosecutors to fulfil the agreements on which defendants have relied in pleading guilty serves not just the interest of fairness to defendants, but goes far toward preserving the system's ability to facilitate the completion of agreements that represent the defendant's informed judgment of his own interests in the case.

The second source of uncertainty for defendants is whether the sentencing judge will abide by the agreement and impose the sentence it calls for. If the defendant fears that the sentence he receives will be greater than he bargained for, he will have to add the increment h to his assessment of the bargain's terms, reproducing the problem caused by breach of the agreement by the prosecutor. But the independence of sentencing judges makes securing their cooperation in the completion of plea bargains by sentencing in accord with them a delicate matter. One approach, adopted by the American federal courts for plea agreements couched in carefully specified language, is to explicitly withdraw the sentencing power from the judge in these cases and require the bargained sentence to be imposed.[48] Another, more implicit approach employed successfully by most of the states is not to direct the judge to impose the sentence specified in the agreement but to rely instead on administrative judges, whose interest in relieving the enormous pressure on the criminal process is closely aligned with the prosecutor's, to ensure that sentencing judges do what they must to keep the system afloat.

South Africa's approach differs across the two tracks to agreement. Because judges retain the full authority to impose any lawful sentence after a guilty plea in a s 112 proceeding, informal bargaining in that context can only be charge bargaining, with the parties' agreement on the facts and the charge the principal constraint on the judge's choice of sentence within the statutory range. If a defendant believes there is a significant likelihood that the judge will

[44] *North Western Dense Concrete CC* (note 6 above) at 671–672.

[45] Ibid at 672–677.

[46] Ibid at 683f, 676f and 678c.

[47] Ibid at 670f-g, 677b. See also South African Law Commission 'Sentence Agreements' (note 12 above) at 3.16.

[48] *Federal Rules of Criminal Procedure*, Rule 11(c)(3).

not impose the sentence he and the prosecutor hope she will, this uncertainty will distort his choice as we have discussed, and his reluctance to bargain will only be increased by his inability to withdraw his plea should he discover after he has entered it that the sentence is greater than he bargained for. But in severely overburdened criminal courts, as in the United States, to the extent that trial judges respond to the pressures that heavy caseloads put on the judiciary and the prosecution alike, they too may tacitly agree to honour the bargains put before them and keep the cases moving along.

Section 105A seems to recognise this problem, and takes steps to alleviate it. The written agreement that manifests the formal bargain must not only 'state fully the terms of the agreement, the substantial facts of the matter, all other facts relevant to the sentence agreement and any admissions made by the accused', it must also propose a specific punishment that, in the view of the parties, would constitute 'a just sentence to be imposed by the court'.[49] First, the judge questions the defendant to determine that he understands the proceedings and the agreement. Second, she establishes a factual basis for the plea.[50] Third, she then considers the sentence agreed upon by the parties. If she finds the agreement 'just', she must then convict the defendant and sentence as the agreement requires. If she finds the agreement 'unjust', she must inform the parties of what alternative sentence she would consider just, and give them an opportunity to revise the agreement on the basis of this sentence and resubmit it to her for sentencing. If either side declines to accept these terms, the plea is withdrawn and the defendant is entitled to a new trial *de novo* before a different judge.[51]

Steyn points out that the statute's use of the word *just* to describe an acceptable sentence differs from the more traditional usage in this context, *appropriate*, and suggests that this shift is meant to signal to judges that they should approach the sentencing of defendants in s 105A proceedings differently than they do the sentencing of defendants convicted after a full trial.[52] Two important cases support this view. In *State v Sassin*, the High Court considered the sentence imposed in a s 105A bargain.[53] Majiedt J refused to interpret 'just' to imply that the sentencing judge is required to agree with the proposed sentence, despite his suggestion that, had he been the judge in the case, he would have imposed a longer sentence. Why? Because the trial judge took proper account of the 'triad' of traditional determinants of appropriate sentencing enunciated in *State v Zinn*:[54] (1) the seriousness of the offence; (2) the interests of society; and (3) the personal circumstances of defendant.[55] In this way, Watney notes, *Sassin* encourages judges to take account of 'the same factors a sentencing court would normally consider but with the proviso that any sentence that could possibly be considered appropriate under the circumstances would suffice'.[56]

Two years after *Sassin*, in *State v Esterhuizen*,[57] the High Court revisited the question of 'justness' in s 105A matters. The sentencing judge's role, wrote Els J, was not merely to consider the 'well-known triad', but also to take 'a broad overview of the facts admitted . . . together

[49] Criminal Procedure Act 51 of 1977, ss 105A(2)(b), 105A(1)(a)(ii)(aa).
[50] Criminal Procedure Act 51 of 1977, ss 105A(5–6).
[51] Criminal Procedure Act 51 of 1977, ss 105A(7–9).
[52] Steyn (note 1 above) at 213–215.
[53] *State v Sassin* 2003 (4) All SA 506 (NC).
[54] *State v Zinn* 1969 (2) SA 537 (A).
[55] *State v Sassin* (note 53 above) at 15.5–15.8, 18.1.
[56] Watney (note 16 above) at 227.
[57] *State v Esterhuizen* 2005 (1) SACR 490 (T).

with the proposed sentence to be imposed, all with a view to establishing whether the sentence agreed upon and its effective content bear an adequate enough relationship to the crimes committed – taking into account all of the agreed facts, both aggravating and mitigating, so that it can be said that justice has been served'.[58]

Like Majiedt J, Els J noted that, as a sentencing judge, he 'would probably have imposed a much heavier sentence under the circumstances', but 'it must be so that the court, in considering the "justness" or "unjustness" of a sentence agreement cannot simply decide for itself *in vacuo* what sentence it would have imposed for crimes to which the accused is pleading guilty'.[59] In a case that is difficult to prove and the likelihood of conviction correspondingly low, but where the prosecutor is convinced of the defendant's guilt, a plea bargain may well result in a smaller sentence than might be imposed after a successful trial. But this alone would not render the sentence unjust:

> In return for the concession of a plea of guilty to a charge difficult to prove, it must be so that the Legislature has envisaged that the bargaining mechanism would bring home a result which satisfies the interests of justice. These would be that where a crime has been committed a conviction has been achieved. The price may be that the sentence which would normally flow from the commission of such a crime is lower than might otherwise have been imposed. This does not mean that justice has not been achieved.[60]

The High Court seems to be treading a very narrow path. At the same time the High Court is preserving nominal judicial discretion in sentencing and attention to the traditional factors of sentencing even in the presence of a s 105A agreement, it is encouraging judges to agree to negotiated sentences that may be far less severe than those that would result after a conviction at trial, even (or perhaps especially) in cases where imperfect evidence makes conviction at trial hard to achieve. It seems to be inviting, or asking, as subtly and delicately as it can, sentencing judges to swallow hard and impose whatever sentence the parties may specify in a plea agreement, so long as it bears a reasonable relationship to the crime the defendant has admitted to committing. This compromise, as Judge Els put it, is the 'price' that must be paid for the system's commitment to plea bargaining.

III IS PLEA BARGAINING CONSTITUTIONAL?

A Fairness and coercion

There are good reasons to think that plea bargaining is a bad thing. For one, it produces inaccurate outcomes, not once in a while, or even inevitably, but *systematically*. Allowing offenders to be punished for crimes less serious than what they have actually done offers prosecutors several advantages. As Els J suggests, it lets them successfully conclude a larger portion of cases with convictions, and deploy their scarce human and material resources to inflict some punishment, even a small amount, on as many offenders as possible, all of which is surely in the interests of the criminal process, if not necessarily justice. But when huge caseloads put their resources under great stress, so the price in punishment prosecutors are willing to pay for a guilty plea is very high, defendants routinely plead guilty to crimes that bear only

[58] Ibid at 495a–b.

[59] Ibid.

[60] Ibid at 494c, 494b, 494h. Cf Watney (note 16 above) at 228–229. But see Bennun (note 17 above) at 32 (Criticising this view as 'a serious misconception of the nature of a criminal trial'.)

faint resemblance to, and suffer punishments far smaller than the law prescribes for, what they have actually done. And because the essence of the bargain is the sentencing discount, the bias is always in the same direction – *every* bargain results in a sentence smaller than what the defendant could expect after a trial on an accurate charge. When discounts are deep, punishments vastly understate the moral weight of the offences that were actually committed, so that when this systematic effect is added to the uncertainty already attached to the law's inability to capture and convict every offender, the expected punishments that prospective offenders actually face for their crimes become very much lower than the moral harm done by their crimes. To the extent that punishment deters crime, this result means that many more crimes will be committed than would be the case than were the punishments authorised by statute for the crime actually committed imposed in every case.

A second objection is that plea bargains remove the doing of criminal justice from public view. Charges and sentences are negotiated informally under s 112, often hastily and under pressure, by prosecutors and defendants, or by prosecutors and defence attorneys, who must then negotiate with their clients, far from public view. Even in a s 105A bargain, the only evidence ever brought to light is what the parties choose to incorporate into the written agreement and the defendant admits in colloquy with the judge, and the only members of the public permitted even indirect access to the proceedings are the victims, whose views prosecutors are required to solicit. In this way, Bennun argues, plea bargaining subverts adversarial norms by supplementing the prosecutor's already substantial charging discretion with the further power to determine the legal fact of guilt or innocence, traditionally the province of the trial court, and denies the public *its* interest, distinct from that of the defendant, in the defendant's right under s 35(3)(c) of the Constitution to 'a public trial before an ordinary court'.[61] In the United States, where public defenders do a great deal of the bargaining for impecunious defendants, defendants may be pressured to accept agreements by their attorneys, hoping to maintain cooperative relationships with prosecutors at the expense of the particular client whose fate they are deciding. It is not a pretty sight, even if we could see it.

'Even if we could see it' – *sub rosa* justice is a third objection. Criminal trials have a kind of majesty, as the full power of the state is deployed against a lone defendant purposely equipped by the law with rights to resist it effectively. Conducted with due regard to these rights, they help make ordinary people into citizens by inviting them to observe how the government treats suspected offenders, what standards it applies to them and itself, and when and how it makes mistakes. Staying its hand against defendants until fair procedures convict them shows the modern state at its best, elevated in its citizens' eyes by the respect it shows them and the law. Plea bargains, as Bennun and many others have pointed out, possess none of this appeal or obvious normative legitimacy. Compared to the moral weightiness of a full trial, they seem shabby, even dirty, like dispensing rough-and-ready justice in a dusty bazaar. Judged against the normative and constitutional standards manifested in a modern criminal trial, the undignified, unlovely haggling of plea bargaining is hard to defend on any ground beyond necessity. Not even its defenders like it much.

But to say that plea bargains are ugly, or even that they produce systematic inaccuracies, is not necessarily to say that they are unconstitutional. The fate of plea bargaining in the American federal courts is instructive in this regard. As early as 1908, in *Twining v New Jersey*, the Supreme Court had recognised that

[61] Bennun (note 17 above) at 32–34, 37–43.

> some of the personal rights safeguarded by the first eight Amendments against National action may also be safeguarded against state action, because a denial of them would be a denial of due process of law. If this is so, it is not because those rights are enumerated in the first eight Amendments, but because they are of such a nature that they are included in the conception of due process [guaranteed by the fourteenth amendment].[62]

After World War II, the Supreme Court applied the criterion of 'fundamental fairness' to various aspects of state criminal procedure as they came under constitutional challenge. The Court acted on *Twining's* promise by gradually 'incorporating' almost the full range of rights guaranteed to suspects and defendants by the first eight amendments into the due process clause of the fourteenth amendment, thus making them binding on the states as well as the federal government.[63] But until *Brady* was decided in 1970, as negotiated pleas increasingly colonised both the state and federal criminal courts and the regime of plea bargaining came under increasing constitutional pressure in the lower federal courts, the Supreme Court refused to acknowledge the existence of bargaining or test its constitutionality against the fifth amendment right against self-incrimination and the sixth amendment right to a criminal trial before a jury.[64] It regulated the taking of guilty pleas only by observing that they represented a waiver of the defendant's right to trial and a conviction on the basis of the plea alone, and requiring that they be made 'knowingly', in full awareness of their consequences for the defendant, and 'voluntarily', without physical or mental coercion that might deny a defendant freedom of choice in waiving the right to trial.[65]

The problem of knowledge attracted little judicial attention. It could effectively be addressed by careful design of the rules that govern the bargaining process itself, as indeed it has been, legislatively by s 105A in South Africa and in the United States by judges and judicial administrators after 1970. Instead, the voluntariness of guilty pleas became the principal field on which the constitutionality of negotiated pleas was contested. In one sense, absent physical duress, it is hard to see how a defendant who knowingly trades his right to trial for the advantage he perceives in the lower sentence the prosecutor is offering for his guilty plea might be acting involuntarily. The fact that the defendant must choose the lesser of two evils, pain or the chance of more pain or none, does not make his choice to plead guilty involuntary. But the close relation between the due process clause of the fourteenth amendment and the definition of constitutionally acceptable criminal procedure in American law means that the construction of every right guaranteed by the fourth, fifth, sixth or eighth amendments is strongly inflected by concerns of 'fundamental fairness'. Accordingly, the Supreme Court has taken a subtler view of voluntariness and coercion. As Frankfurter J put it in 1948, holding that a station-house confession by a fifteen-year-old boy was not a voluntary act:

> It would disregard standards that we cherish . . . to hold that a confession is 'voluntary' simply because the confession is the product of a sentient choice. . . . [C]onduct devoid of physical

[62] *Twining v New Jersey* 211 US 78, 99 (1908).

[63] The scholarly literature on incorporation of the first eight amendments into the due process clause of the fourteenth is vast. See, eg, AR Amar 'The Bill of Rights and the Fourteenth Amendment' (1992) 101 *Yale Law Journal* 1193 (Amar offers a perceptive overview of the literature.)

[64] The fifth amendment right against self-incrimination was incorporated into the fourteenth amendment and applied to the states in *Malloy v Hogan* 378 US 1 (1964), as was the sixth amendment right to a jury trial in *Duncan v Louisiana* 391 US 145 (1968).

[65] Adelstein (note 32 above) at 816–820; Alschuler (note 3 above) at 33–38.

> pressure but not leaving a free exercise of choice is the product of duress as much so as choice reflecting physical constraint.[66]

In this pre-1970 tradition, the federal appellate courts confronted a range of issues in considering the voluntariness of guilty pleas. Their primary concern was the danger that innocent defendants would falsely condemn themselves by accepting the offer of a sharply discounted sentence and pleading guilty to offences that they did not in fact commit. The problem of erroneous convictions thus became intertwined with the voluntariness of the plea, and the justiciable question that emerged was whether a given bargaining tactic, or the plea bargaining process itself, created a substantial risk that innocent defendants would waive their right to trial and convict themselves by falsely pleading guilty to some offence.

How might plea bargains be seen as coercive in this sense? The model of part II suggests a useful approach. Recall our earlier example: I am accused of aggravated robbery, which I may or may not have done. The probability that I'll be convicted at trial is 70 per cent, and if I am, I'll be sentenced to ten years in prison, so the expected trial sentence is $0.7 \times 10 = 7$ years. As we saw, if the prosecutor offers me a sentence of seven years in exchange for my guilty plea, the 'objective values' of the bargain and the punishment lottery are the same: accept the bargain and get seven years for certain, or enter the lottery and pay seven years for an 'average' trial. Whether or not I take the deal depends on my attitude toward the risk of the lottery: if I am a risk averter, I will take the deal, but if I am a gambler, I will take the risk of trial. But in neither case am I forced to 'pay a price' for exercising my right to trial. As far as anyone knows before the trial, the 'objective' punishments in the two alternatives are the same, and whether I take the deal or not depends on my preferences and the attitudes toward risk they imply. It is only when the price the prosecutor is willing to pay for the plea, the extent of the sentencing discount, is greater than the objective value of the plea that a price is put on exercising the right to trial. It is the magnitude of this price that raises the question of coercion.

Suppose the prosecutor offers a sentence of five years. Then the price he is willing to pay for the plea, five years less than the trial sentence, is greater than the three-year discount needed to make the definite punishment of the bargain and the expected punishment in the lottery have equivalent objective value. Accepting the bargain means five years for me, when the expected sentence after trial is seven, so exercising my right to trial means paying an objective price of two years. Whether I surrender my right and accept the bargain or pay the two-year price for exercising it depends, again, on my preferences. But in either case, refusing the deal means paying an objective price of two years for exercising the right to a trial.

Now, as prosecutors in large American cities must do every day, turn up the heat. Suppose a hard-pressed prosecutor offers me nine months for my plea. This may be an offer I cannot refuse, a bargain that, given my predicament, is much too good to pass up no matter how I feel about the risk of trial. In light of the possible outcomes, the price I would now have to pay to exercise my right to trial is very high. Even if I know I am not guilty, and even if I am a gambler and time in prison holds less terror for me than it does for risk averters, if there is a 70 per cent chance that I will have to serve ten years if I go to trial and I can avoid this risk by serving just nine months, I am very likely to accept the deal and plead guilty. The overwhelming need of American prosecutors to secure guilty pleas has, for at least a hundred years, meant that terms

[66] *Haley v Ohio* 332 US 596, 606 (1948).

like this are offered to most American defendants, burdening their choice as to plea by making exercise of their right to trial a very costly proposition.[67]

This analysis came to the fore in federal appellate courts throughout the 1960s. In 1969, Bazelon J proposed that courts and legislatures had an obligation to determine when the discounts prosecutors were offering created a significant risk that innocent defendants would plead guilty in response, and to develop guidelines to help prosecutors avoid crossing this threshold of impermissibility.[68] England ultimately adopted this approach. In 1994, it largely obviated the need for explicit bargaining by establishing a fixed price for guilty pleas in every case: a one-third reduction in the statutorily prescribed sentence if the plea is made at the first reasonable opportunity; a one-quarter reduction once the trial date has been set; and a one-tenth reduction if the plea comes after the trial begins.[69] However, in *Brady* and *Bordenkircher*, the US Supreme Court created no such guidelines and put few limits on the discounts prosecutors may offer to secure guilty pleas, even where these are so great as to impose a significant price on a defendant's exercise of the right to trial. And in *Santobello*, the Court made explicit that the reason it was tolerating such a regime was that the alternative, a full criminal trial for every defendant, was far beyond the material means of the criminal process. There was, it acknowledged, no plausible alternative to plea bargaining but chaos.

B Rights and limitations

This history illuminates two important aspects of the question of plea bargaining's constitutionality in South Africa that distinguish it from the American case. The first is what Snyckers and le Roux[70] call the *due process wall*, a conceptual barrier erected by the Constitutional Court to analytically separate the 'residual liberty rights' of s 12 of the Constitution, and in particular the right of every person 'not to be deprived of freedom arbitrarily or without just cause' guaranteed by s 12(1)(a), potentially broad, due process-like rights that might well be informed by judicial notions of justice or fairness, from the specific rights of arrested, detained and accused persons in the criminal process enumerated in s 35. Where the American due process clause has infused the definition of fifth and sixth amendment rights with considerations of fundamental fairness, the wall is intended to prevent the residual rights of s 12(1) from 'seeping' into the definition of the enumerated rights of suspects and defendants and expanding them beyond what the language of s 35 itself, presumably in its traditional or common law usage, requires.

The prosecutor's offer of nine months in our example illustrates the effect this interpretive constraint might have in the context of plea bargaining, an alternative to the full adversarial trial as a means of convicting defendants. In the United States, the seepage of due process rights into the definition of voluntary guilty pleas might lead a judge, like Frankfurter J, to

[67] Adelstein (note 25 above) at 216–218. Compare HM Caldwell 'Coercive Plea Bargaining: The Unrecognized Scourge of the Justice System' (2011) 61 *Catholic University Law Review* 63; A Ashworth 'Four Threats to the Presumption of Innocence' (2006) 10 *International Journal of Evidence and Proof* 241, 256–257; C McCoy 'Plea Bargaining as Coercion: The Trial Penalty and Plea Bargaining Reform' (2005) 50 *Criminal Law Quarterly* 67.

[68] *Scott v United States* 419 F2d 264 (DC Circuit 1969).

[69] UK Sentencing Council 2017 'Reduction in Sentence for a Guilty Plea: Definitive Guideline' at 5, 11, available at https://www.sentencingcouncil.org.uk/wp-content/uploads/Reduction-in-Sentence-for-Guilty-plea-Definitive-Guide.pdf.

[70] F Snyckers & J le Roux 'Criminal Procedure: Rights of Arrested, Detained and Accused Persons' in S Woolman & M Bishop (eds) *Constitutional Law of South Africa* (2nd Ed, 2008) Chapter 51 at 1–19.

look beyond the defendant's ostensible consent to the bargain to conclude that the offer put too great a price on the right to trial and was thus unfairly coercive. But a South African court, forced by the wall to eschew considerations of this sort and consider the claim solely in terms of an accused's rights to be presumed innocent under s 35(3)(h) and not to be compelled to give self-incriminating evidence under s 35(3)(j), rights themselves derived from the common law, might take a narrower version of coercion and sustain the bargain on the basis of the defendant's informed acceptance of the offer. Indeed, Snyckers and le Roux read the series of Constitutional Court decisions that erected the due process wall itself, particularly *Nel v Le Roux NO & Others*,[71] as seeking to narrow the rights enjoyed by suspects and defendants under *both* s 12 and s 35 of the Constitution and establishing a striking principle with clear relevance to a general regime of plea bargaining:

> The sort of 'trial' one is entitled to under FC s 12(1)(b) [the right not to be detained without trial] differs from the sort of trial one is entitled to under FC s 35(3). The former lays down less rigorous requirements, or requirements less generous to the individual concerned, than the latter. . . . The more arbitrary or informal the proceedings in question, ie the less closely they resemble a criminal trial, the less claim the imprisoned individual has to 'full' fair trial rights under FC s 35. *If the state were to embark upon a general practice of instituting summary proceedings which bear little resemblance to criminal trials as its main process to incarcerate those suspected of crimes, the reasoning in Nel would hold that the new class of criminal 'accused' would not be entitled to the right to a fair trial with trimmings*.[72]

The second point is suggested by the Supreme Court's tortured resolution of the plea bargaining question. Though the appellate courts, led by Judge Bazelon, had framed the question for the Supreme Court in terms of the price plea bargains might place on the exercise of the right to trial, the Court in *Brady* declined their invitation to examine the potentially coercive effects of large sentencing discounts and, with one eye clearly fixed on preserving what the justices knew to be an indispensable part of the American criminal process, swept them away with an appeal to the 'mutuality of advantage' offered by plea bargains and a heroic assumption about the state of mind of defendants who accept them:

> Of course, that the prevalence of guilty pleas is explainable does not necessarily validate those pleas or the system which produces them. But we cannot hold that it is unconstitutional for the State to extend a benefit to a defendant who in turn extends a substantial benefit to the State and who demonstrates by his plea that he is ready and willing to admit his crime and to enter the correctional system in a frame of mind that affords hope for success in rehabilitation over a shorter period of time than might otherwise be necessary.[73]

Not until the following year, in *Santobello*, did the Court admit the practical necessity of negotiated pleas, and of adjusting the contours of the fifth and sixth amendments to it, even as it could scarcely bring itself to call the practice by its name:

> The disposition of criminal charges by agreement between the prosecutor and the accused, sometimes loosely called 'plea bargaining,' is an essential component of the administration of justice. Properly administered, it is to be encouraged. If every criminal charge were subjected to a full-scale trial, the States and the Federal Government would need to multiply by many times the numbers of judges and court facilities.[74]

[71] 1996 (3) SA 562 (CC), 1996 (1) SACR 572 (CC).

[72] Snyckers & le Roux (note 70 above) at 51 13–14 (emphasis added).

[73] *Brady v United States* (note 34 above) at 752–753.

[74] *Santobello v United States* (note 39 above) at 260.

In effect, *Brady* and *Santobello* apply a limitations analysis to the sixth amendment right to trial, in the absence of an explicit limitations clause in the American Constitution. Because it was imperative to the continued operation of the criminal process itself that the right to trial be limited by making its exercise costly (sometimes, as in *Bordenkircher*, very costly) to defendants, the Court concluded that it could not find the practice unconstitutional. This conflation in a single constitutional analysis of two distinct questions – how far a given right extends and when it might be infringed, and what practical or policy considerations might call for defining the right in one way or another – invites what Woolman and Botha call 'the worst kind of analytical confusion'.[75] But the South African Constitution *does* have a limitations clause, s 36. It requires analysis of the constitutionality of plea bargaining to proceed in two stages. First, there must be an initial determination of the content of a given right and whether it has been infringed by some specific bargaining tactic or by the plea bargaining system in general, without reference to practical considerations. Then, should the right indeed be found to have been infringed, a second, subsequent inquiry into whether the infringement can be deemed reasonable and justifiable in terms of s 36 must take place.

In light of all this, how might the constitutionality of plea bargaining under s 35 of the Constitution be tested? A brief sketch of two related claims that might be raised against both the informal bargaining system that has grown around ss 112 and 113 of the CPA and the formal bargains governed by the Act's s 105A may be suggestive. Both focus on the adequacy of the process by which guilt is adjudicated in the regime of plea bargaining. As Bennun points out, prosecutors largely determine the ultimate outcomes of bargaining, on the basis of factors that often have little to do with finding the result in each case that most closely conforms to the governing substantive law. Before the decision to prosecute is made, they evaluate whatever evidence exists against defendants and dismiss charges against those whose guilt they might not doubt but cannot prove. Where the evidence is strong enough to proceed, they select the charges for strategic as well as legal reasons, and adjust their offers to reflect the likelihood that defendants can be convicted at all. Once they decide what discount to offer for the defendant's plea, the defendant must accept, which may require negotiation and a further concession. Whether the deal is sealed informally by oral promises under s 112 or recorded as an enforceable agreement under s 105A, once it has been made, and before the case ever appears before a trial court, it is the parties, and primarily the prosecutor, who will have effectively made the decision as to what crimes the defendant will be found to have committed, and what evidence has or has not been sufficient to support that conviction, with little or no independent review of the decisions or the evidence. The outcome of this complex set of behind-closed-door calculations, determined in large part by the prosecutor, is then presented to a judge for ratification as 'just'.

The presumption of innocence and its sibling, the prosecutor's burden to prove guilt beyond a reasonable doubt, are nowhere to be seen. Section 35(3)(h) of the Constitution grants every accused the right 'to be presumed innocent'. Snyckers and le Roux interpret this provision as 'a right to expect the state to bear the full burden of proving the case and therefore not to be allowed to compel assistance from the accused'.[76] The typical plea bargain in a crowded jurisdiction is unlikely to meet such a standard. The admissible evidence in the case is never

[75] S Woolman & H Botha 'Limitations' in S Woolman & M Bishop (eds) *Constitutional Law of South Africa* (2nd Ed, 2008) Chapter 34 at 17–30, quotation at 21.

[76] Snyckers & le Roux (note 70 above) at 51–144.

put before an independent fact-finder in full. The prosecutor determines the probative value of the evidence, what it can or cannot be used to prove. The prosecutor's decisions almost never take into account a presumption of innocence or a felt need to prove guilt beyond a reasonable doubt. The job of adjudicating guilt and fixing punishment is effectively transferred from neutral courts to partisan prosecutors. The judicial process, with its ancient institutions and ideological commitments, is transformed into an administrative system with very different objectives and standards. Because this outcome of plea bargaining is endemic rather than idiosyncratic, a consequence of the system working routinely – as it is supposed to do – in the majority of cases rather than the occasional malfunction or malfeasance, the proper claim would seem to be not that any particular plea bargain subverts the defendant's right to be presumed innocent, a claim that would present formidable problems of proof in any case, but that the system itself operates in violation of s 35(3)(h). Taken together, its structure of powers, incentives, and outcomes is inconsistent with reserving conviction and punishment solely for defendants who have been convicted beyond a reasonable doubt. The remedy would thus not be reopening the occasional bad bargain, as in *Santobello*, but overturning the system of plea bargaining itself.

A reply to this claim might be grounded in Uijs J's analogy of plea bargains to civil settlements. In civil cases, where the stakes for the defendant do not involve incarceration and the plaintiff's burden of proof is less onerous, the parties are given full freedom to determine by negotiation the precise distribution of liability and damages that the court will be asked to ratify, and their decisions in bargaining may be influenced by whatever particular interest or objective they may have in the lawsuit. It is the plaintiff's right to sue, and the defendant's right to choose whether or not to defend in court. Criminal proceedings, the argument might go, should be seen no differently. Inquisitorial criminal processes may have a rule of compulsory prosecution, a requirement that, to the extent possible, every offender should be prosecuted and punished for the highest crime the evidence in the case will support. Adversarial systems like South Africa's do not. The same ancient traditions to which the claim appeals also give prosecutors substantial discretion to select charges not solely on the basis of factual accuracy but on their view of what justice requires as well. More importantly, defendants, who may have many reasons for wanting to settle a criminal case 'out of court', have always had the right to do so if they wish by pleading guilty. To overturn the plea bargaining system would require the withdrawal of both the prosecutor's charging discretion and the defendant's right to plead guilty, and go far toward transforming the adversarial system on which the Constitution's array of powers and rights is based into an inquisitorial system inconsistent with its common law and constitutional values.

Plea bargaining's opponents might seize on this last point, but turn it around. Even where formal agreements are reached and ratified under s 105A, they might claim, the system itself replaces adversarial procedures of accusation and proof before an impartial judge with administrative resolution by prosecutors, informed by evidence gathered by police and endorsed by compliant judges after cursory inquiries into the facts. It is an inquisitorial system in all but name, without the protections afforded by the rule of compulsory prosecution and the conduct of investigative trials by forensically-trained judges professionally committed to finding the truth in every case before them. In the twilight regime of conviction by agreement, halfway between the adversarial and the inquisitorial, plea bargains may be precisely the 'summary proceedings which bear little resemblance to criminal trials as [the] main process to incarcerate

those suspected of crimes' that Snyckers and le Roux feared might follow from *Nel* and afford defendants even fewer protections than those offered by s 35.[77] If so, and an 'ordinary court' is held to be an adversarial court, then the system of plea bargaining might be challenged under s 35(3)(c) of the Constitution for failing to guarantee defendants 'a public trial in an ordinary court'. Despite the reasoning in *Nel*, and despite the legislative approval of plea bargaining signaled by passage of s 105A, Snyckers and le Roux insist that

> The clear aim of this provision is to ensure that normal criminal trials with full fair-trial protection are the only acceptable method whereby the state may prosecute individuals for committing offences. Exceptions to this rule require FC s 36(1) justification.[78]

If this is held to be the case, and the regime of plea bargaining is found to infringe either the right of defendants to the presumption of innocence under s 35(3)(h) of the Constitution or a fair public trial under s 35(3)(c), might these limitations themselves be justified? Once again, no more than a suggestive sketch can be offered here, but the broad outlines of such a justification in the universal problem of high caseloads and scarce judicial resources seem clear enough. Section 36(1) of the Constitution provides that

> The rights in the Bill of Rights may be limited only in terms of law of general application to the extent that the limitation is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom, taking into account all relevant factors, including –
> (a) the nature of the right;
> (b) the importance of the purpose of the limitation;
> (c) the nature and extent of the limitation;
> (d) the relation between the limitation and its purpose; and
> (e) less restrictive means to achieve the purpose.[79]

The limitation in question would be the general infringement of the rights of defendants under s 35(3)(h) or s 35(3)(c) by the system of plea bargaining operating within the procedures provided by ss 112, 113 and 105A of the Criminal Procedure Act 51 of 1977 as amended – all laws of general application. To support the claim that the limitations the system places on these rights are reasonable and justifiable in terms of the underlying values expressed in the section, the state might address each of the five factors in turn.

It cannot be denied that access to a fair trial with a presumption of innocence is a right of first importance to an open and democratic society based on human dignity, equality and freedom. However, a fair adversarial trial 'with trimmings' is an expensive proposition, one that must be paid for in largest part from the public purse. Plea bargaining is an attempt to direct extremely scarce judicial resources toward achieving convictions and some measure of punishment where appropriate in the greatest number of cases possible, without which the public's dissatisfaction with the operation of criminal justice might be substantially aggravated. This limitation, importantly, operates with the lightest possible touch on defendants. It does not *require* them to surrender any rights; it merely tries to make it attractive for them to do so voluntarily. Every defendant who insists on a trial will have one, even if the sentencing discounts that the plea bargaining system offers lead few to want one. The purpose of the limitations that the system of plea bargaining places on the rights of defendants is to conserve not just the material resources of the criminal process, but at the same time to preserve the ability of the process

[77] Snyckers and le Roux (note 70 above), at 51-13.

[78] Ibid at 51–94.

[79] Constitution of the Republic of South Africa, 1996, s 36(1).

itself to provide a satisfactory resolution to as many cases of crime as possible, a public good of highest importance to societies everywhere. A court might well conclude that it achieves all of these ends in a manner that evolves 'naturally' from the adversarial structure itself, one that does less violence to it than would an imperfect implantation of inquisitorial procedures and assumptions. Finally, since both the courts and the legislature have acknowledged that adequate funding cannot be found to meet the cost of providing every defendant with a full criminal trial, plea bargaining can plausibly be said to be the least restrictive means of achieving the goal of justice in an open and democratic society based upon human dignity, equality and freedom.

As the Constitutional Court has repeatedly held, 'the least restrictive means' to limit a constitutional right does not, and must not, oblige the Court to substitute its judgment of what it 'imagines' the fairest possible arrangement to be for the legislature's well-considered judgment regarding the 'least restrictive means'. Woolman and Botha note that '[in *State v*] *Manamela,* the Court displays [an] awareness of the difficulties associated with determining the availability of less restrictive means, [and] arrives at a somewhat crisper articulation of the problem' that the phrase suggests.[80] 'The problem for the Court is to give meaning and effect to the factor of less restrictive means without unduly narrowing the range of policy choices available to the Legislature in a specific area'.[81] The *Manamela* Court goes on to observe:

> It will often be possible for a court to conceive of less restrictive means, as Blackmun J has tellingly observed: 'And for me, "least drastic means" is a slippery slope ... A judge would be unimaginative indeed if he could not come up with something a little less "drastic" or a little less "restrictive" in almost any situation, and thereby enable himself to vote to strike legislation down'.[82]

The *Manamela* Court's fear that judges may 'annihilate the range of choices' available to the legislature flows from its recognition that the phrase 'less restrictive means' often raises difficult issues relating to 'cost, practical implementation, the prioritisation of certain social demands and the need to reconcile conflicting interests'.[83] In *State v Mamabolo*, Kriegler J addresses these concerns in the following manner:

> Where section 36(1)*(e)* speaks of less restrictive means it does not postulate an unattainable norm of perfection. The standard is reasonableness. And in any event, in theory less restrictive means can almost invariably be imagined without necessarily precluding a finding of justification under the section. It is but one of the enumerated considerations which have to be weighed in conjunction with one another, and with any others that may be relevant.[84]

The same fears underlie the concerns that many have expressed about the use of plea bargains in an adversarial system constitutionally committed to fair trials. However, as I have tried to show here, depending on the circumstances that make them necessary, the range of 'less restrictive' institutional means to limit a defendant's constitutional right 'to a public trial before an ordinary court' may be quite wide. In South Africa, both the courts and the legislature have attempted to ensure that defendants in the nation's criminal courts are able to make reasonably informed decisions at the same time that the public sees justice as being broadly served in the dispensation of punishments. In this way, plea bargains, in all their complexity,

[80] Woolman and Botha (note 75 above) at 88–89.

[81] *S v Manamela & Another (Director-General of Justice Intervening)* 2000 (3) SA 1 (CC), 2000 (5) BCLR 491 (CC) at para 95.

[82] Ibid at para 94 quoting Blackmun J in *Illinois State Board of Elections v Socialist Workers Party* 440 US 173, 188–189 (1979).

[83] Ibid at para 95.

[84] *S v Mamabolo* 2001 (3) SA 409 (CC), 2001 (5) BCLR 449 (CC) at para 49.

can persuasively be said to preserve the system of justice upon which South Africa's open and democratic society relies. They will almost certainly continue to do so until the nation possesses the resources necessary to offer a full trial to every criminal defendant.

Economic ways of thinking do not provide answers to questions such as these so much as they provide useful ways for people to think, and disagree, about them. I do not presume to have offered more than this, or to urge any particular resolution of the constitutional issues raised by plea bargaining. Making and defending the kinds of constitutional claims sketched here would surely require not just more imaginative lawyering than this, but a good deal of sophisticated empirical research on the environmental and material conditions of criminal justice in South Africa, the incidence of guilty pleas, both formal and informal, in the system, and the sentences actually being imposed after both guilty pleas and convictions at trial. All of these are tasks for another time. I hope here simply to have opened new avenues of study and debate about a problem of great importance to the law in South Africa, and around the world.