THE VANISHING TRIAL: A DYNAMIC MODEL WITH ADAPTIVE AGENTS

Moti Michaeli* and Yosef Zohar†

Abstract: In recent years, Trial Waivers were introduced into many legal systems around the world. Once trial waivers (TW) were introduced, most—if not all—of these legal systems have witnessed a steep increase in the usage of TW, to the extent that trials are virtually disappearing. Our model shows how introducing TW into a legal system inevitably triggers a dynamic process whose ultimate result is that all defendants choose the trial waiver. The power given to prosecutors to decide whether or not to offer such trial waivers—and the negative externality imposed on other defendants when TW are used—further implies that their introduction increases total sanctioning and reduces the welfare of many, if not all, defendants.

Keywords: Trial Waivers, Plea Bargains, Criminal Law.

JEL codes: K14; D04; D62.

*The University of Haifa, Department of Economics. Contact: motimich@econ.haifa.ac.il.

†Post-doctoral Research Fellow, Judicial Conflict Resolution Lab, Faculty of Law, Bar-Ilan University. Contact: yosef.zohar@biu.ac.il.

1 Introduction

The "Vanishing Trial" phenomenon (Galanter 2004)—the decline in the portion and in the absolute number of criminal cases that are terminated by trial—which portrays the adversarial-legal-system countries (Conrad Jr and Clements 2018), continues to spread around the globe to inquisitorial-legal-system countries in Europe and in South America (Langer 2004), reaching Central Asia (Alkon 2007), China (Lynch 2009) and lately penetrating also Japan for the first time (Japan Times 2018). In reality, more and more convictions are imposed today around the world without a full trial ever taking place, as a result of a Trial Waiver System,[1] which incentivizes suspects to plead guilty and receive discounted sentences and immunity from certain charges in return for waiving their right to a trial.[2]

In 2017, a study encompassing 90 countries around the globe (Fair Trials 2017) has revealed that the existence of trial waiver systems worldwide has increased by nearly 300% since 1990.[3] In some jurisdictions, trial waivers almost entirely replace trials.[4] In other jurisdictions, later introduced to the system, the adoption of trial waiver systems has taken place rapidly over the course of just a few years.[5] What could explain such a consistent growth, one that stops—judging by the pioneering example of the US—only when the trials are on the verge of extinction? While one might think that this is a natural result—both sides understand this is preferred to the alternative—we show that the result holds even if most defendants are risk seeking (or simply too optimistic about their chances at court) and trial waivers cannot be tailor-made to fit the idiosyncratic acceptance thresholds of different defendants.

[1] The term Trial Waiver refers to various forms of guilty plea bargaining, abbreviated trials and cooperating-witness procedures (Fair Trials 2017).

[2] The right of defendants to a trial includes the right to file a motion to exclude wrongfully obtained evidence from being used against them at trial, to have evidence displayed in court, to hear and cross-examine witnesses, to challenge evidence introduced by the prosecutor, to appeal following a conviction and more.

[3] Before 1990, only 19 of the 90 jurisdictions studied in the research featured trial-waiver systems in law. By the end of 2015, the number has grown to 66, reaching all six major continents and changing practice across a variety of different legal systems and traditions.

[4] For example, trial waivers conclude 97% of federal cases in the US (https://www.uscourts.gov/, accessed on 30/10/2019; see also Figure 2) and 90% of criminal convictions in the higher courts of England & Wales (UK Ministry of Justice 2014).

[5] For example, in Georgia, 12.7% of cases were resolved through Trial Waiver System in 2005 but that figure had soared to 87.8% of cases by 2012. In Russia, deployment of Trial Waivers shot up from 37% in 2008 to 64% only six years later (Fair Trials 2017).

To this end, we develop a simple model that shows that trial-waiver systems generate a positive feedback cycle (reinforcing dynamic). According to the model, the trial-waiver mechanism guarantees a constant increase in the prosecutor's bargaining power, because the prosecutor can use the resources saved in trial waivers to build a stronger case against other defendants. This in turn increases the expected sanction in trial and, as a result, increases the demand for trial waivers, and so on. This suggests that once trial waivers are introduced, they are bound to spread in the criminal justice system. In the model, a risk-neutral (and forward-looking) prosecutor offers a trial waiver (TW) to a group of defendants whose criminal cases share similar characteristics. If some defendants accept the offer of a TW, next-period defendants update their expectations about the expected sentence in trial, which makes them more prone to take the TW.[6] Foreseeing this, the prosecutor will choose to offer the most severe TW (in terms of its punishment) that can trigger the process. This TW is the one that initially convinces only the most risk averse defendant to take it, but then guarantees to instigate the feedback-cycle dynamics that will eventually make all other defendants—including the risk seekers—take it as well (in the steady state). The optimal solution of the prosecutor, who maximizes total sanctioning (like, e.g., in Landes 1971 and in Rhodes 1976), is shown to imply that the welfare of (at least) the non-risk-averse defendants is reduced, and in particular is smaller than their welfare if no trial waivers are introduced. Thus, although the set of alternatives of the defendants expands, an expansion that should supposedly increase their welfare, it is only the prosecutor whose welfare is surely increased.[7]

The mainstream in Law and Economics scholarship holds that trial waivers are executed "in the shadow of trial" (see e.g. Landes 1971, Mnookin and Kornhauser 1979, Easterbrook 1983). If this is the case, then the sanction offered to defendants does not deviate by much from the expected sentence in court, hence at least defendants who are sufficiently risk tolerant should prefer to have their day in court. Thus, this can hardly explain the rapid

[6]This means that, unlike the prosecutor, the defendants are adaptive. This has no effect on the existence of the equilibrium analyzed—it exists also when defendants are forward looking—but we think this is more realistic and better portrays the observations mentioned above. The result is shown to be robust to other fictitious-play dynamics. See further discussion in Sections 3 and 4.

[7]This message resonates with the conclusions of others, e.g. Bar-Gill and Ben-Shahar [2009], but opposite claims are also prevalent (see e.g. Church 1979).

and widespread expansion of trial waivers that we aim to explain. Another explanation often mentioned for this expansion relates it to the "trial-penalty" that is inflicted on defendants who refuse to waive their right for a trial (see Kim 2015 for a survey). Proponents of this explanation claim that plea bargains—the most common form of trial waivers in the US—offer substantial discounts on the sanction that would otherwise be imposed upon conviction in court. This discount arguably goes beyond merely accounting for the defendant's odds of acquittal, hence is very attractive for the defendant. This implicit trial penalty is sometimes quite explicit, when defendants are actually threatened to face a very severe punishment in case they refuse the bargain, making the latter an offer "they cannot refuse". The extent to which this accurately describes the American criminal justice system—where, in case of conviction in trial, judges' hands are tied by severe legislation—is debated. But one clear weakness of this explanation is that it cannot be convincingly applied to a multitude of other criminal justice systems around the world.[8] More generally, the room given to prosecutorial adjudication (Ayal and Riza 2009) varies between countries, and the global aspect of the vanishing-trial phenomenon (see evidence in Section 2) suggests that the trial penalty, whether implicit or explicit, is unlikely to be the sole explanation to this phenomenon. Similarly, in the U.S. prosecutors often run for election, hence attempts of the prosecution to maximize conviction rate—or total number of convictions—rather than total sanctioning (Rasmusen et al. 2009) could partly explain the increased use of guilty pleas, where conviction is guaranteed by definition. However, it is debated to what extent this is true, even in the U.S., and to what extent prosecutors act instead like "benevolent social planners" whose objective function is to maximize total sanctioning. Our model assumes the latter, which is also empirically supported (e.g. by Boylan 2005), and is still able to demonstrate how trial waivers are taking over the system.[9]

As illustrated earlier, in the current paper we highlight an endogenous mechanism

[8] For example, the case of *Bordenkircher v. Hayes*, where the defendant Hayes had refused to secure a five-year imprisonment in a plea bargain and ended up being convicted in trial and getting a mandatory life sentence (under Kentucky's three strikes law), is a famous demonstration of the trial penalty notion but is clearly specific to the American legal system, just like the application of severe punishments as a mandatory minimum sanction for various offenses.

[9] The dynamic conversion to 100% plea bargains that our model produces will only be faster if we let the prosecutor's objective function include also a component of maximizing conviction rate—see Section 4.

that inherently leads to a sustained increase in the use of trial waivers. This mechanism, by which the resources saved by the prosecution in one criminal case are reinvested to increase its effectiveness in other cases, is not entirely novel to this paper. Landes [1971] was probably the first to describe the externalities generated when a trial is waived and, as a result, resources for other cases are freed up, but he acknowledges that these "secondary effects" are largely ignored in his analysis (see footnote 5 of his paper). Posner [2003] makes the same argument and adds that these saved resources will be used by the prosecution "to build a stronger case when bargaining fails" and, as a result, "average sentences will probably be heavier rather than lighter" (Posner 2003, p. 578), but he does not formally model this statement. In the current paper we model this process exactly and examine its outcomes and implications for the criminal justice system as a whole.

The paper that is probably the closest to ours is that of Bar-Gill and Ben-Shahar [2009], who show that the ability to free up resources by using trial waivers allows the prosecutor to make a credible threat in other criminal cases. Specifically, Bar-Gill and Ben-Shahar [2009] present a model in which the prosecutor can take only one out of N defendants to trial, but since no one wants to be that one defendant, all of the N defendants end up signing a plea bargain. While their model results in 100% plea bargains, Bar-Gill and Ben-Shahar do not treat this result as significant. If anything, they try to convince the reader to see it as an artifact, writing "[I]n particular, in...our model the plea rate is 100%, which is clearly unrealistic." The main goal of the current paper is to show that this result is not only realistic but in fact is almost inevitable once trial waivers are introduced to a criminal justice system. One might even view this paper as demonstrating a mechanism that best be called "trial in the shadow of bargaining", because the expected sanction in court is derived from the dominance and expansion of plea bargaining rather than the other way around.

Our model further enables us to provide insights on the parallel increase in conviction rates in trials (see Section 2) and on the expected sanctioning in trial waivers compared to the expected sentence in court. In particular, our model provides a different, novel explanation to the allegedly large discount given in plea bargains as embodied by the aforementioned notion of trial penalty. We show that, despite being indeed more lenient than the expected sentence in court *in the steady state* (i.e., after the system has converged to relying almost solely on trial waivers), the sanctioning in trial waivers is in fact *more* severe than the expected sentence in court were there no trial waivers. This is so because a harsh punishment in a trial waiver, which initially attracts only the most risk averse defendant, becomes over time attractive to all defendants as the sanction in court becomes increasingly more severe, and ends up looking “too soft” to outside observers.

2 Motivating evidence

In this section we present empirical support for the two main observations that our paper aims to explain. Figure 1 below shows the growth in the use of plea bargains in various countries around the world (SOURCE: Fair Trials International 2017). This growth constitutes our **Observation 1**. The most common form of trial waiver is a *plea bargain* (sometimes called a guilty plea). Figure 2 complements Figure 1 by presenting a table that documents (in the second row) an increase in the use of plea bargains using data from US Distric Courts.[10]

[10] SOURCE: U.S. District Courts—Criminal Defendants Terminated, by Type of Disposition; in https://www.uscourts.gov.

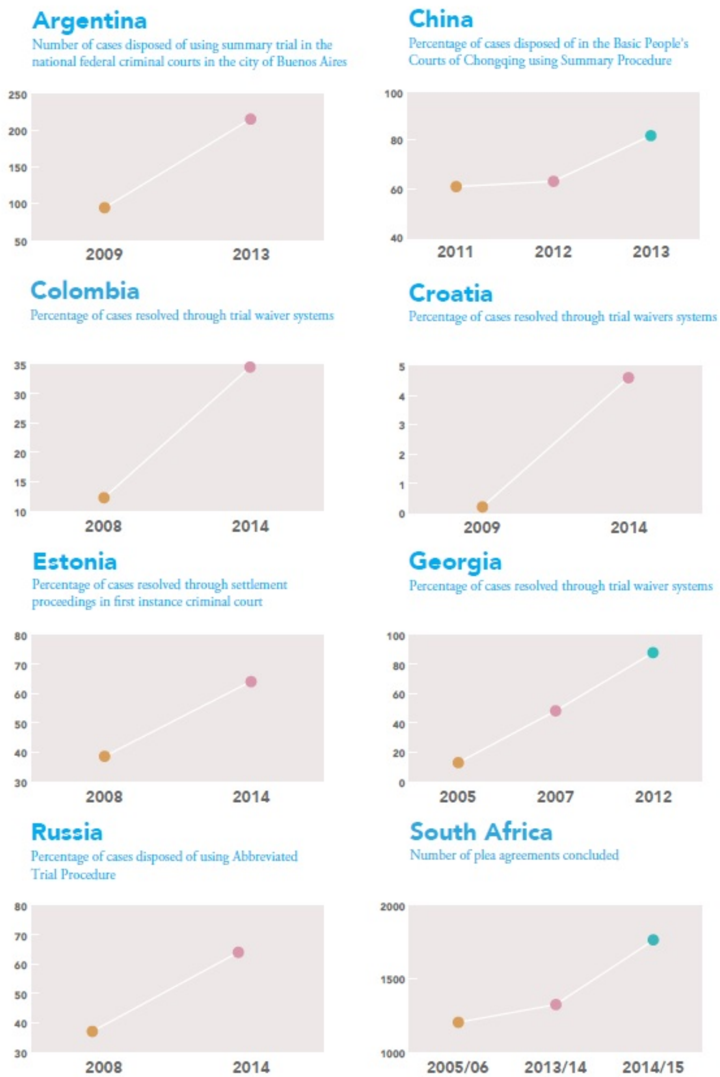


FIGURE 1.— Growth in use of plea bargains in various countries around the world

Year	1981	1991	2001	2011	2019
Conviction Rate	95.6%	97.4%	98.9%	99.5%	99.6%
Guilty Pleas	79.1%	86.3%	93.4%	96.9%	97.4%
Conviction by Jury Trial	78.7%	81.0%	83.3%	84.8%	84.6%

FIGURE 2.— Conviction rates in US Federal Courts. The second row of the table documents the gradual increase in guilty pleas (the most common form of trial waiver). The third row documents the parallel increase in the conviction rate for those going to trial. Overall, both trends contribute to the increase of the overall conviction rate, as depicted in row 1 (For example, in 1981 the calculation is 95.6%=79.1%+78.7%×(100%-79.1%)).

One might naively expect those who refuse to sign a guilty plea to more likely be innocent hence acquitted in court, which in turn should lead to a gradual decrease in the conviction rate in trials as the use of plea bargains increases. However, as can be viewed in the third row of the table in Figure 2, the conviction rate in trials in fact *increases*. This *increase in conviction rate* constitutes our **Observation 2**.

3 The model

Each period $t = 0, 1, 2...$, a continuum of defendants with a unit measure are up to trial for a given crime. They are accused of the same type of crime and face the same strength of evidence hence face the same sanction s if convicted in court. With some abuse of notation, $x \in [0, 1]$ stands for both the defendant and her judicial case, where the absence of an index t reflects the notion that the distribution of cases is stationary. One risk-neutral prosecutor (the state) is in charge of handling their cases. The prosecutor has a fixed per-period budget B that can be allocated as she wishes, but cannot be saved for future periods.[11]

In the benchmark case, plea bargains are not possible. The probability of conviction in court in case x is an increasing function of $i(x)$, which is the investment of the prosecutor in this case. We denote this probability by $p(i(x))$.

The prosecutor has to decide of an allocation of her fixed periodic budget B, with the goal of maximizing the (discounted) total sanctioning subject to the budget constraint:

$$(1) \qquad max \sum_{t=0}^{\infty} \delta^t \left\{ \int sp\left(i_t(x)\right) dx \right\} \; s.t. \; \int i_t(x)\, dx = B \; \forall t,$$

where $\delta \in (0, 1)$ is a discount factor.

First note that the inability to save money for future periods and the stationary distribution of defendants imply that the solution of the within-period optimization problem, $max \left\{ \int sp\left(i(x)\right) dx \right\} \; s.t. \; \int i(x)\, dx = B$, is the same in all periods. Let $i^*(x)$ denote this

[11]One might think of the state prosecution as composed of a large team of specialized prosecutors that each of whom is in charge of a narrow chunk of cases with very similar attributes. In the model we study the dynamics of cases that are handled by one such specialized prosecutor. The simplifying assumption of having only one type of crime can be found also, for example, in Rasmusen et al. [2009], who explain in footnote 8 of their paper why this assumption is innocuous. In their paper, the actual strength of the cases varies, while we vary instead the *perceived* strength of the cases—see below.

solution.[12] Since the cases have identical features and the prosecutor cannot distinguish between defendants (see more below), the probability of conviction—for all defendants—is $E_x[p(i^*(x))]$. In the benchmark case with no plea bargains, the "budget per case" equals $B/1 = B$. Hence, we can simply denote the probability of conviction of defendants in this benchmark case by $P(B)$. That is, $P(B) = E_x[p(i^*(x))]$ where $\int i^*(x)\,dx = B$, and the expected sanction in trial is given by $S(B) \equiv sP(B)$. More generally, we will denote by $P(I)$ the probability of conviction when I is the (average) prosecutor's budget per case. Anticipating the analysis of a criminal justice system with trial waivers, it is worth noting that an increase in the average budget per case for the chunk of cases that are not settled in a trial waiver implies an increase in the probability of these cases ending in conviction in trial, i.e. $P'(I) > 0$.[13]

We proceed now to analyzing the effect of the introduction of trial waivers, and in particular plea bargains (PB), on the proportion of defendants choosing the PB option and on the total sanctioning. A plea bargain here is a proposal made by the prosecutor to a defendant, where the latter is offered a sure sanction b in return for a guilty plea. We assume that b is fixed (across defendants and across time) and costs the prosecutor a fixed cost c, which is smaller than the average investment per case in the benchmark model ($c < B$), reflecting the common notion that plea bargains save litigation resources. In Section 4 we discuss what happens if b can change between periods.

We allow defendants to differ in their certainty equivalence (CE) for facing a trial in which the sanction, in case of conviction, is s. These differences could be due to heterogeneity in risk attitudes or in the perceived strength of evidence, or due to other factors like budget constraints. For simplicity, we will treat these differences mostly as differences in risk attitudes, and make the very plausible assumption that not all defendants are strictly risk seeking.[14] In order to investigate the effect of risk attitudes, we will identify each defendant

[12] Note that the solution may be such that some cases are not taken to court ($i(x) = 0$ for some x's in $[0, 1]$).

[13] To see why this must hold, note that such an increase enables the prosecutor to evenly divide the added budget among all the tried cases so that $i(x) > i^*(x)$ for every tried case, hence $sp(i(x)) > sp(i^*(x))$. The optimization problem of the prosecutor implies that she might be able to do even better than this, but certainly not worse.

[14] Formally, we assume that the mass of defendants who are not strictly risk seeking is strictly positive.

by a unique value z, where $z : [0,1] \to [0,1]$ is a bijection that attaches a value z to each defendant x in a way that generates a (weakly) descending order of risk aversion. That is, $CE(z, P(I))$—the certainty equivalence for defendant z of facing a prosecutor with a budget per case I (hence a probability of conviction $P(I)$)—is monotonously decreasing in z. We assume that $CE(\cdot, \cdot)$ is continuous in both arguments. Our assumption that the most risk averse defendant is not strictly risk loving implies that $CE(0, P(I)) \geq S(I)$ (for any I). The prosecutor knows the distribution of $CE(z, P(I))$ but not the mapping from x to z. In other words, the risk attitude of a defendant is her private information—the attribute z is unobservable by the prosecutor, and by observing x the prosecutor learns nothing meaningful about the defendant.

From the exposition of the decision problem of the defendant, it is clear that her decision whether to accept a PB is dependent on the budget of the prosecutor, which in turn is dependent on the decision of other defendants whether to go to trial or take the plea bargain. Thus, it is basically a coordination problem for the continuum of defendants (within each period). One option to model and solve this decision problem is to have a simultaneous game where all defendants are forward looking and have correct rational expectations about the decisions of other defendants so that they all together simultaneously "jump" to a static equilibrium of this simultaneous game.[15] Beyond its questionable applicability to real-life defendants, such an analysis cannot guide us in selecting an equilibrium when there are multiple ones (as happens here, see more below). While equilibrium refinements that can help us select the equilibrium exist and are commonly used, we believe that their applicability to our case is rather limited. For example, using coalition proofness as a refinement requires that a mass of defendants will be able to successfully coordinate a deviation from a given strategy profile, which is quite implausible, and using instead Pareto optimality as a refinement is not helpful as well given the contradictory motives of the prosecution and the defendants. Therefore, we believe that a more natural approach (with supporting evidence in the data, as was demonstrated in Section 2) is to assume instead that the "judicial game" is played sequentially and that defendants use *adaptive dynamics*. In other words, when choosing

[15] An analysis along these lines can be found in Bar-Gill and Ben-Shahar [2009].

between a trial and a plea bargain, the defendant estimates the expected sanction she faces in trial based on observations from previous periods. The dynamic game then goes as follows.

In period $t = 0$ there are still no plea bargains.[16] At the end of period 0 plea bargains are introduced, when the prosecutor makes a uniform offer b to all future defendants. In each subsequent period, starting from period 1, a new set of defendants with identical properties appears. Each defendant decides whether or not to sign the offered plea bargain. These decisions are taken simultaneously (in each period) and are observed by future defendants, who can then update their belief about the probability of conviction and expected sanctioning. Those who do not sign the PB are taken to court (potentially facing a prosecutor who invests 0 in their specific case, in which case they are dismissed).

The prosecutor is assumed to be forward looking while the defendants, as just explained, are adaptive.[17] In the main text we explore one particular type of adaptive dynamics, sometimes called simply "best-response dynamics" (or "one-recall best-response dynamics"), where the defendants form their belief on the probability of conviction based on the last period, and therefore best reply to the actions of that period. This is a degenerate case of a *fictitious play* (see e.g. Fudenberg and Levine 2009, Young 2015), which is a best reply to a distribution over all (or many) past periods.[18] In Appendix B we show that our results hold also for any recursively-defined fictitious play with a non-zero weight on last-period actions. In general, best response dynamics imply that, given a probability of conviction $P(I)$, defendant z prefers the plea bargain over going to trial if and only if $b \leq CE(z, P(I))$, where the exact version of adaptive dynamics assumed determines how I is computed.

3.1 A static analysis of steady states

We start by analyzing the within-period coordination game of the defendants. For that purpose, assume for now that b is exogenously determined (it will be endogenized later

[16] Reflecting the notion that most (if not all) criminal justice systems started from an initial state of no plea bargaining.

[17] If defendants are also forward looking then the game ends in one period and there are multiple equilibria. These equilibria correspond to the equilibria of a static game with rational players. We do not compute the set of equilibria for this forward-looking version of the model, but the results of such an analysis—and their shortcomings—are discussed in the first paragraph of Section 4.

[18] Young [2015] notes (on p. 365) that when "individuals are not in a position to have much influence on the dynamics...myopic best response behavior serves as a plausible baseline assumption."

in the analysis, when we will also present the updated objective function of the prosecutor and solve for the optimal b). Denote by α_t the proportion of defendants who sign the PB in period t. The (one-recall) best-response dynamics imply that, in period t, defendant z prefers the PB over going to trial if and only if $b \leq CE\left(z, P\left(I_{t-1}\right)\right)$, where $I_{t-1} = \frac{B-\alpha_{t-1}c}{1-\alpha_{t-1}}$.[19] We thus have $\alpha_t = 1 - F_{CE(z,P(I_{t-1}))}\left(b\right)$. Given that $CE\left(z, P\left(I_{t-1}\right)\right)$ decreases in z, it follows that those who sign the PB are the defendants with $z \in [0, \alpha_t]$.

It is easy to see that the two corner cases are steady states. First, if $\alpha_{t-1} = 0$ (no plea bargains at time $t-1$), then any $b > \underline{b} \equiv CE\left(0, P\left(B\right)\right)$ results in $\alpha_t = \alpha_{t-1} = 0$, implying this is a steady state. That is, if there are no plea bargains, hence the prosector's budget for trials is limited and defendants face an expected sentence in trial of only $S\left(B\right)$, any offered PB with sanction b that is not sufficiently small will be rejected by all defendants. However, there does exist a sufficiently small value of b, denoted here by $\underline{b}$, that will make the defendant who is the most reluctant to go to trial (the most risk averse defendant in our case) indifferent between signing PB with sanction $\underline{b}$ and going to trial to face an expected sentence $S\left(B\right)$. Second, if instead $\alpha_{t-1} = 1$ (no trials at time $t-1$), then any $b \leq \bar{b} \equiv CE\left(1, lim_{I\to\infty}P\left(I\right)\right)$ results in $\alpha_t = \alpha_{t-1} = 1$, implying this is also a steady state. That is, if there are no trials hence the prosector's budget for trials is virtually unlimited, any offered PB in which b is not too large will be accepted by all defendants. However, there exists a sufficiently large value of b, denoted here by $\bar{b}$, that will make the defendant who is the most willing to go to trial (the most risk seeking defendant) indifferent between signing a PB with sanction $\bar{b}$ and going to trial to face an expected sentence of $lim_{I\to\infty}sP\left(I\right)$. Logically it makes sense to assume that $\underline{b} < \bar{b}$ (as we will indeed assume), but this is related to the characterization of the inner steady states. To see why, note that an inner steady state (with $\alpha \in (0,1)$) requires that $b = CE\left(\alpha, P\left(\frac{B-\alpha c}{1-\alpha}\right)\right) \equiv b_\alpha$. By substituting $\alpha = 0$ in the expression for b_α we get that $b_0 = \underline{b}$. Similarly, by substituting $\alpha = 1$ we get that $b_1 = \bar{b}$.

How does $b_\alpha = CE\left(\alpha, P\left(\frac{B-\alpha c}{1-\alpha}\right)\right)$ change as α increases? It decreases through the first

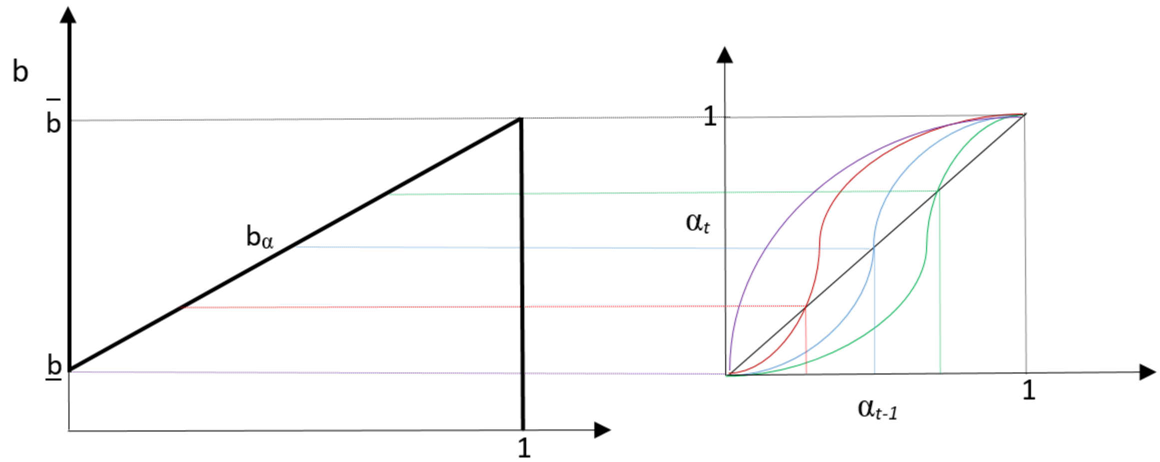
[19] Note that $P\left(I_{t-1}\right)$, which denotes the probability of conviction in period $t-1$, is the ratio between the mass of defendants who are convicted in trial and the mass of defendants who do not sign the PB in that time period $(1-\alpha_{t-1})$, where the latter could potentially include defendants in whose cases the prosecutor does not invest resources (hence are practically not prosecuted).

argument (risk-attitude effect), but increases through the second argument (budget effect). Which effect is stronger? This is basically an empirical question. As will become clear soon, the empirical evidence presented in Section 2 supports the predictions of the model when the budget effect is stronger (so that b_α increases in α and as a result, by continuity, $\underline{b} \leq \bar{b}$). Furthermore, the purpose of this paper is to demonstrate a specific mechanism: by choosing to sign the plea bargain, defendants increase the per-case budget of the prosecutor, which in turn makes other defendants more prone to sign the plea bargain as well. In other words, this mechanism highlights the budget effect. For these reasons, we will assume that the budget effect is stronger than the risk-attitude effect. This is captured by the following assumption.

ASSUMPTION 1 b_α *weakly increases in* α *in the range* $(0, 1)$.

When is this assumption likely to hold? Broadly speaking, b_α increases in α if plea bargains are effective in saving expenses for the prosecution (i.e. c/B is small, making the budget effect large) and/or the distribution of defendants' risk attitudes is sufficiently "dense" (making the risk-attitude effect small). Note that a dense distribution of risk attitudes is quite plausible because, in reality, the intervention of defense attorneys is likely to reduce the variability in risk assessments among the defendants. In Appendix C we demonstrate what exactly a sufficiently "dense" distribution entails, using an example in which the Odds Ratio of conviction is linear in the budget per case for trials ($\frac{p(I)}{1-p(I)} = \frac{I}{B}$) and the disutility of a defendant from conviction is a power function (CRRA utility); and in Section 4 we discuss the case where b_α decreases rather than increases in α. For now, suppose that Assumption 1 holds, so that b_α increases in α. The left panel of Figure 3 depicts the plea-bargain offers that constitute a steady state for the whole range of α.

FIGURE 3.— Steady-state PB as a function of % of defendants who sign the PB



Left panel: The "zig-zag" shape of the graph follows from the existence of a lower bound $\underline{b}$ when $\alpha = 0$, an upper bound $\bar{b}$ when $\alpha = 1$, and the increase of b_α in α in between. b_α is not necessarily linear in α in the range $\alpha \in (0, 1)$ but is monotonically increasing by Assumption 1. Right panel: For any $b \in \left(\underline{b}, \bar{b}\right)$ there is a corresponding phase diagram with three steady states, two of them are stable (at the extremes, where $\alpha_t = \alpha_{t-1} \in \{0, 1\}$) and the third is an unstable inner steady state. The unstable inner steady state is a tipping point: once α_t passes this point there will be convergence to $\alpha_\infty = 1$.

As can be seen on the right panel of Figure 3, the "zigzag" shape of the steady-states $b(\alpha)$ has implications for the number and characteristics of the steady states. When $b > \bar{b}$ the unique steady state is at $\alpha = 0$ and when $b < \underline{b}$ the unique steady state is at $\alpha = 1$. In between, for any $b \in \left(\underline{b}, \bar{b}\right)$, three steady states exist: two stable steady states with $\alpha = \{0, 1\}$ and one unstable inner steady state with $\alpha \in (0, 1)$.[20]

3.2 The dynamic game

We now go back to the dynamic game and present our main results.

PROPOSITION 1 *If (and only if) a non-zero mass of defendants signs the offered PB in period* 1*, then* α *will increase each period until it reaches* 1 *and there are no more trials:* $\alpha_1 > 0 \Rightarrow lim_{t\to\infty}\alpha_t = 1$.

PROOF: See Appendix A. *Q.E.D.*

[20]This multiplicity of equilibria demonstrates why a simultaneous game with perfectly forward-looking players having rational expectations is not warranted here: it will not allow us to say anything meaningful about the state of the criminal justice system if the prosecutor offers a PB of $b \in \left(\underline{b}, \bar{b}\right)$.

Proposition 1 states that once there are defendants who sign the PB, plea bargaining is bound to take over the whole criminal justice system. The intuition is straightforward. If, at the end of period 0, the prosecutor offers $b > \underline{b}$, no defendant will sign it in any subsequent period because they all believe to face an expected sentence of $S(B)$, which does not suffice to convince any of them to take the offered PB. However, if the prosecutor offers $b < \underline{b}$, the defendants who are most reluctant to go to trial will sign the PB and this will start a positive feedback cycle: as b stays fixed, α will keep increasing until a steady state is reached, which in the case of $b < \underline{b}$ is unique at $\alpha = 1$.[21]

We turn now to consider the prosecutor's objective function in order to endogenize b. Modifying (1) to include plea bargains, the prosecutor chooses b at the end of period 0 by solving

$$max_b \sum_{t=1}^{\infty} \delta^t \left\{ \alpha_t(b)\, b + (1-\alpha_t(b))\, S\left(\frac{B - \alpha_t(b)\, c}{1-\alpha_t(b)} \right) \right\}$$

s.t. $\alpha_t(b) = 1 - F_{CE(z,P(I_{t-1}(b)))}(b)$, where $I_{t-1}(b) = \frac{B-\alpha_{t-1}(b)c}{1-\alpha_{t-1}(b)}$.[22] Like in the benchmark case with no plea bargains, the fact that money cannot be saved for future periods and that the prosecutor cannot meaningfully differentiate between defendants implies that the probability of conviction $P(I)$ is a sufficient statistic for the analysis of the dynamic game.[23] The following proposition complements Proposition 1 by showing that the feedback cycle will indeed be triggered in equilibrium, hence the taking over by plea bargains is inevitable.

Proposition 2 *For any $\delta \in (0,1)$, the prosecutor offers $b < \underline{b}$ in period 0 and, once the new steady state is reached, all defendants sign the PB.*

Proof: See Appendix A. *Q.E.D.*

[21]If b equals exactly $\underline{b}$ then there is another steady state at $\alpha = 0$ but this steady state is not stable. For ease of exposition we ignore this knife-edge case in the analysis.

[22]Since b is a choice variable now, we denote by $\alpha_t(b)$ the share of defendants who sign the plea bargain in period t when the offered plea-bargain sanction is b. The prosecutor in fact solves the optimization problem in two steps: first, for any possible b, she chooses the budget allocation among the defendants who do not sign the bargain (in each period t), and then she maximizes over b.

[23]I.e., there is no need to know how the prosecutor divides I among the defendants who do not sign the plea bargain.

Proposition 2 states that the introduction of plea bargains into a criminal justice system will ultimately terminate the use of trials in that system (a.k.a. "the vanishing trial"). To see why it is always (i.e. for any $\delta \in (0,1)$) beneficial for the prosecutor to drive the system toward 100% plea bargains, even at the cost of lowering the sanction to be below $\underline{b}$, note first that any offer $b > \underline{b}$ implies that, in all periods, the total sanctioning equals $S(B)$ in each period, because the shift from trials to plea bargaining described in Proposition 1 is not triggered. From the point of view of the prosecutor, this can be improved upon by offering a plea bargain that equals the CE of a risk neutral defendant in the state with no plea bargains (i.e. offering $b = CE(\tilde{\alpha}, P(B)) = S(B)$, where $z = \tilde{\alpha}$ is a risk neutral defendant).[24] Since, by assumption, at least some defendants are not risk seeking, the dynamic process will be triggered. Furthermore, in any period $t \geq 1$, some defendants sign the PB hence get a sanction $b = S(B)$, while others go to trial, where their sanction is larger than $S(B)$ because they face a stronger prosecutor compared to the benchmark case with no plea bargains. Overall, this is for sure better for the prosecutor compared to any offer $b > \underline{b}$ that does not trigger plea bargaining. Hence, even if this is not the prosecutor's optimal strategy, the optimal one must be to offer some $b < \underline{b}$.

Proposition 2 thus confirms **Observation 1** (continuous growth in plea bargaining), while **Observation 2** (regarding the parallel increase in conviction rates in trials) holds as well: as the share of defendants signing the PB increases during the dynamic process ($\alpha_t \uparrow$), the budget of the prosecutor for handling the cases that go to trial increases as well ($I_t = \frac{B-\alpha_t c}{1-\alpha_t} \uparrow$), and so does the conviction rate in trial ($P(I_t) \uparrow$).

Let $b^*(\delta)$ denote the optimal plea-bargain offer given a discount factor δ. It is easy to see that any defendant z for whom $CE(z, P(B)) < b^*(\delta)$ is worse off once plea bargains are introduced, because instead of having a disutility of $CE(z, P(B))$ (her loss in the benchmark case), she either signs the PB and gets a sanction $b^*(\delta)$ that is greater than $CE(z, P(B))$, or faces a prosecutor with average per-case budget of $I > B$ implying an expected loss of $CE(z, P(I)) > CE(z, P(B))$. At the same time, the other defendants, for whom $CE(z, P(B)) > b^*(\delta)$, are better-off when signing the PB compared to their

[24]If all defendants are risk averse then consider instead the offer $b = S(B) < CE(1, P(B))$.

benchmark sanction $S(B)$. However, the share of defendants who are better off when plea bargains are introduced might be very small, and in fact completely vanishes if the prosecutor is sufficiently patient, as stated in the next proposition.

PROPOSITION 3 *If $\delta \to 1$ then (i) $b \to \underline{b}$ and (ii) in the steady state, all defendants are worse off than with no PB.*[25]

PROOF: See Appendix A. *Q.E.D.*

Since $\delta \to 1$ (the prosecutor is very patient), the prosecutor cares mostly about the steady state, hence wants b to be as close as possible to $\underline{b}$ (from below). Thus, of all values below $\underline{b}$, which by Proposition 2 guarantee conversion to 100% plea bargains, she will choose the highest possible.[26] Furthermore, in all periods $t \geq 1$, *all defendants* are (weakly) worse off than with no plea bargains, because all those who sign the offer get a sanction $\underline{b} = CE(0, P(B)) \geq CE(z, P(B))\ \forall z \geq 0$, and, as explained earlier, those who choose instead to go to trial (at any point in time) face a stronger prosecutor ($I > B$). By offering $b \to \underline{b}$, the prosecutor basically exploits the vulnerability of the defendant who is the "weakest link" in the chain. This might be the defendant who is the most risk averse as modeled here, but may also be simply a poor defendant, who is unable to bear the cost of the long litigation in court. By "breaking" this weakest "link", the prosecutor gets her way into "breaking" stronger and stronger defendants, until all defendants surrender and sign the PB.

4 Discussion

Beyond the predicted expansion of plea bargaining, an important implication of the model is that—at least if some defendants are risk averse—the introduction of PB is likely to increase total sanctioning, both in the long term and in the short term. In the long term, sanctioning increases because in the new steady state it is heavier than before plea bargains are introduced (given that $\underline{b} = CE(0, P(B)) > \mathrm{S}(B)$). In the short term, some risk averse defendants are pushed to replace moderate chances of conviction in trial by a harsher plea

[25]By $b \to \underline{b}$ we formally mean that for any $\varepsilon > 0$, $\exists \delta \in (0, 1)$ s.t. $b > \underline{b} - \varepsilon$.

[26]The proof shows that for any two offers b_1 and b_2, such that $b_1 < b_2 < \underline{b}$, the prosecutor prefers b_2 over b_1.

bargain, and those who forgo the plea-bargaining option face increasingly higher probability of conviction. This state of affairs is supported by the data presented in Figures 4 and 5 below.

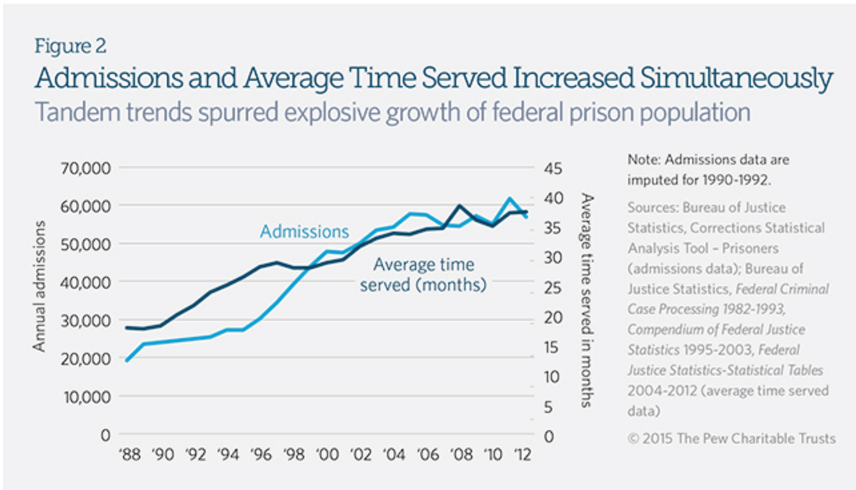


Figure 4.— Average Time Served Increases

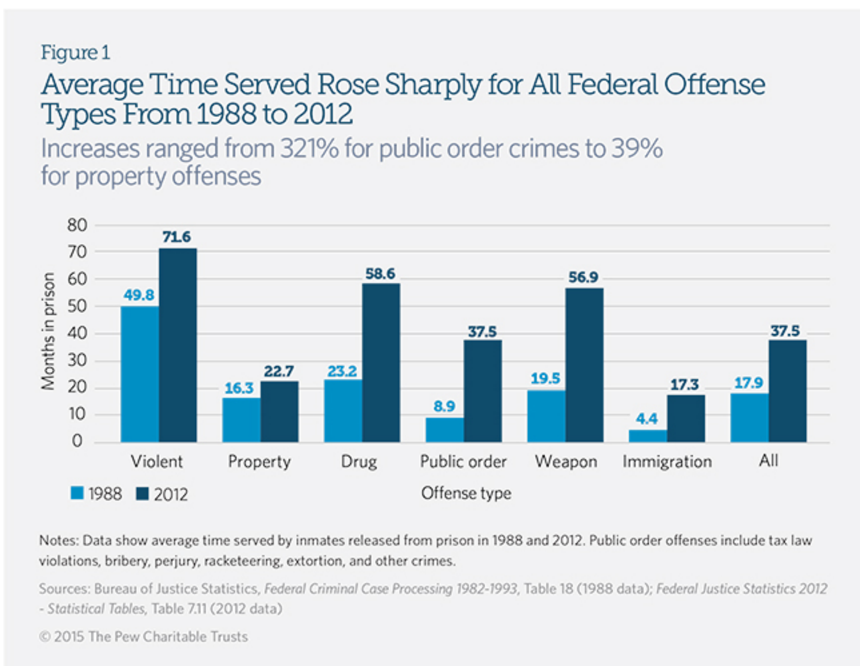


Figure 5.— Average Time Served per Offense

Regarding the robustness of our results, as shown in Appendix B, our results hold much more broadly in terms of the dynamics of the game. In particular, they hold for any fictitious-play dynamics (with non-zero weight on last-period actions). The use of adaptive dynamics is important for the analysis for two reasons. First, it helps us rule out "crazy" solutions that a static model with rational players cannot rule out. For example, with the appropriate system of beliefs, one could construct a static equilibrium where the plea bargain offer is $b_{\hat{\alpha}}$ for some arbitrary $\hat{\alpha} \in (0, 1)$, and the equilibrium proportion of plea bargains equals exactly this $\hat{\alpha}$. This can be an equilibrium if the common belief, shared by the prosecutor and the defendants, is that $\alpha = 0 \; \forall b > \underline{b}$ except for $b = b_{\hat{\alpha}}$ and if $\hat{\alpha}$ is not too small.[27] However, clearly, this is a non-plausible equilibrium supported by non-plausible beliefs. Second, the adaptive dynamics form a sensible characterization of real criminal justice systems, where new defendants constantly appear and react to the situation they observe.

However, the result does depend on Assumption 1, according to which b_α (weakly) increases in α in the range $(0, 1)$. While this assumption is quite plausible, as explained in Section 3, let us consider for a moment the case where b_α decreases rather than increases in α, which could depict a population of defendants with large differences in risk attitudes. In such a case, the middle part of the graph in the left panel of Figure 3 becomes a decreasing line, and therefore there is a unique value of α corresponding to each b. A patient prosecutor would then choose $b = b_{\vec{\alpha}}$ where $\vec{\alpha} \equiv argmax_\alpha \left\{\alpha b_\alpha + (1-\alpha) S\left(\frac{B-\alpha c}{1-\alpha}\right)\right\}$. The solution to this optimization problem cannot be determined without making specific assumptions on $P(I)$, but might potentially result in a mixture of trials and plea bargains in the long-run steady state. Thus, an alternative view of our model is that it gives a sufficient condition for convergence to 100% plea bargains: under Assumption 1, plea bargains are predicted to take over the criminal justice system.

In our model, we assume that the prosecutor tries to maximize total sanctioning. This assumption is standard (see e.g. Landes 1971, Rhodes 1976). But what if the prosecutor cares also about maximizing the conviction rate (or, for that matter, the total number of convictions)? It is easy to show that our main results would still hold. In particular, any

[27] A sufficient (and far from necessary) condition is that $\hat{\alpha} > \tilde{\alpha}$, where $\tilde{\alpha}$ is the risk neutral defendant. This condition implies that $b_{\hat{\alpha}} = CE\left(\hat{\alpha}, P\left(\frac{B-\hat{\alpha}c}{1-\hat{\alpha}}\right)\right) < CE\left(\tilde{\alpha}, P\left(\frac{B-\hat{\alpha}c}{1-\hat{\alpha}}\right)\right) = S\left(\frac{B-\hat{\alpha}c}{1-\hat{\alpha}}\right)$. This in turn ensures that the expected sanction, which is a weighted average of the PB-sanctioning $b_{\hat{\alpha}}$ and of the expected sanction in trial $S\left(\frac{B-\hat{\alpha}c}{1-\hat{\alpha}}\right)$, is greater than $b_{\hat{\alpha}}$, which is itself greater than $\underline{b}$, implying that offering $b_{\hat{\alpha}}$ is preferred by the prosecutor over offering $b \to \underline{b}$ and getting $\alpha = 1$ in the long run.

decrease in the suggested sanction b monotonically increases the proportion of defendants who sign the bargain at time t both directly (by making the bargain more attractive) and indirectly (through making the prosecutor stronger because less defendants are left to be taken to court). And this increase in the proportion of defendants who sign the bargain further implies more convictions at time t, both because any signed bargain is a conviction and because the chances of convicting in trial is larger the less defendants choose the trial option. Thus, Proposition 2 still holds, potentially with a lower value of b (in the limit, when the prosecutor cares *only* about the conviction rate, b will not exceed $CE(1, P(B))$, and all defendants will sign the plea bargain already in $t = 1$). Proposition 3 holds as well, as long as there is a positive weight on the maximization of total sanctioning, because when the prosecutor is sufficiently patient she cares only about the long term effect—when the system has converged to 100% plea bargaining—and the maximal value of b that leads to this convergence is $b \to \underline{b}$, in line with the result stated in Proposition 3.

Finally, our model assumes that the prosecutor knows $\underline{b}$ (because he knows the distribution of defendants' certainty equivalences to a trial). This might seem not realistic. However, strictly speaking this is not necessary. Instead, the prosecutor can simply start off by gradually decreasing b till the first defendant concedes. Proposition 1 then ensures conversion to 100% plea bargains. A natural follow-up question is: why wouldn't the prosecutor also be allowed to change b after first introducing it in $t = 0$? For this our answer is two-fold. First, while during a negotiation with defendants it is a common practice to offer concessions, a signed agreement between a prosecutor and a defendant serves also as a precedent for future plea bargains. Then, since plea bargains are subject to the scrutiny of judges, constantly raising the sanction for the same type of crime (to take advantage of the increased willingness of defendants to sign) is probably not a viable option. Second, from a theoretical point of view, solving the model while letting b depend on the period t is impossible without making further specific assumptions on the model parameters in a way that determines the exact shape of the phase diagram in Figure 3. Still, some things *can* be said without characterizing exactly the optimal periodic sanctioning. In particular, if we let the prosecutor change b (either during the dynamic process or at the end), the prosecutor can impose $b = \bar{b}$ (or close to it) with 100% plea bargains. This is clearly doable once α approaches 1, as can be seen in Figure 3.

5 Conclusions

Despite claims that its expansion is limited outside adversarial-legal systems, evidence are mounting that also in inquisitorial-legal systems the "vanishing trial" phenomenon continues to expand. Once Trial Waiver systems were introduced, most—if not all—of these legal systems have witnessed a steep decline in trials. Our model explains why introducing trial waivers into a legal system is likely to trigger a process whose ultimate result is that all defendants choose to waive their right for a trial. This happens although the prosecutor is assumed to be maximizing total sanctioning rather than conviction rate. The discretion given to prosecutors if and how to incentivize suspects to waive their right for a trial further implies that the introduction of trial waivers enhances prosecutorial power and reduces the welfare of some, if not all, defendants.

Trial waiver systems have considerable implications for human rights and for the rule of law. They transfer power to the hands of prosecutors and, as studies have shown, they might contribute to the coercion of innocent defendants to plead guilty (Langbein 1978, Alschuler 1981, Stuntz 2001, Lundberg 2018). They also play a role in increasing prison populations and racial stratification in prisons (Savitsky 2012). Ultimately, as Stuntz [2001] puts it, "the real boundaries of criminal liability are defined by law enforcers, not by the law". These implications and others are beyond the scope of this paper, but any country that considers embracing a trial waiver system should take into account its deterministic character, which tends to lead to the extinction of trials wherever they are introduced (Lynch 2009).

Acknowledgments

We would like to thank Robert John Aumann, Benjamin Bachi, Oren Bar-Gill, Oren Gazal Ayal, David Lagziel, Ran Smorodinski and Ro'i Zultan for their valuable comments. All remaining errors are ours.

References

Cynthia Alkon. The increased use of "reconciliation" in criminal cases in central asia: A sign of restorative justice, reform or cause for concern? *Pepperdine Dispute Resolution Law Journal*, 8(3):41, 2007.

Albert W Alschuler. The changing plea bargaining debate. *Calif. L. Rev.*, 69:652, 1981.

Oren Gazal Ayal and Limor Riza. *Plea-Bargaining and Prosecution in Criminal Law and Economics*, volume Criminal Law and Economics, chapter 6, pages 145–170. Edward Elgar, 2009. Criminal Law and Economics.

Oren Bar-Gill and Omri Ben-Shahar. The prisoners' (plea bargain) dilemma. *Legal Analysis*, 737(2):37, 2009.

Richard T Boylan. What do prosecutors maximize? evidence from the careers of us attorneys. *American Law and Economics Review*, 7(2):379–402, 2005.

Thomas W. Church. In criminal defense of âbargain justice'. *Law & Society Review*, 1979.

Robert J Conrad Jr and Katy L Clements. The vanishing criminal jury trial: From trial judges to sentencing judges. *Geo. Wash. L. Rev.*, 86:99, 2018.

Frank H Easterbrook. Criminal procedure as a market system. *The Journal of Legal Studies*, 12(2):289–332, 1983.

International Fair Trials. The disappearing trial: towards a rights-based approach to trial waiver systems. Technical report, 2017.

Drew Fudenberg and David K Levine. Learning and equilibrium. *Annu. Rev. Econ.*, 1(1):385–420, 2009.

Marc Galanter. The vanishing trial: An examination of trials and related matters in federal and state courts. *J. EMPIRICAL LEGAL STUD.*, page 112, 2004.

The Japan Times. Japan's first-ever plea bargain. The Japan Times, 2018.

Andrew Chongseh Kim. Underestimating the trial penalty: An empirical analysis of the federal trial penalty and critique of the abrams study. *Mississippi Law Journal*, 84(5), 2015.

William M. Landes. An economic analysis of the courts. *Journal of Law & Economics*, 14(1):47, 1971.

John H Langbein. Torture and plea bargaining. *The University of Chicago Law Review*, 46(1):3–22, 1978.

MÃ¡ximo Langer. From legal transplants to legal translations: the globalization of plea-bargaining and the americanization thesis in criminal procedure. *HARVARD INTERNATIONAL LAW JOURNAL VOLUME*, 45(1):65, 2004.

Alexander Lundberg. When do the innocent plead guilty? *Available at SSRN 3100845*, 2018.

Elizabeth M. Lynch. Maybe a plea, but is it a bargain? an initial study of the use of simplified procedure in china. *China Rights Forum*, 1, 2009.

Robert H Mnookin and Lewis Kornhauser. Bargaining in the shadow of the law: The case of divorce. *The Yale Law Journal*, 88(5):950–997, 1979.

Richard A. Posner. *Economic analysis of law*. Aspen, 6 edition, 2003. ISBN 0735534748.

Eric Rasmusen, Manu Raghav, and Mark Ramseyer. Convictions versus conviction rates: the prosecutor's choice. *American Law and Economics Review*, 11(1):47–78, 2009.

William M Rhodes. The economics of criminal courts: A theoretical and empirical investigation. *The Journal of Legal Studies*, 5(2):311–340, 1976.

Douglas Savitsky. Is plea bargaining a rational choice? plea bargaining as an engine of racial stratification and overcrowding in the united states prison system. *Rationality and society*, 24(2):131–167, 2012.

William J. Stuntz. The pathological politics of criminal law. *Michigan Law Review*, 100(3):505–600, 2001.

H Peyton Young. The evolution of social norms. *economics*, 7(1):359–387, 2015.

A Proofs

A.1 Proof of Proposition 1

Proof: First note that $\alpha_0 = 0 \Rightarrow S(I_0) = S(B)$. The IF direction: If $\alpha_1 > 0$, then it must be that $b \leq CE(0, P(B)) = \underline{b}$ (because a non-zero mass of defendants have chosen to sign the bargain). In this case, Assumption 1 implies that $b < b_\alpha \; \forall \alpha \neq 0$, hence there is a unique steady state, in which $\alpha = 1$.[28] Therefore, there will be gradual conversion to this unique steady state, implying that $lim_{t\to\infty}\alpha_t = 1$. The ONLY IF direction: If $\alpha_1 = 0$, then $S(I_1) = S(I_0) = S(B)$, implying that also $\alpha_2 = 0$, hence $S(I_2) = S(I_1) = S(B)$ and so on, so that $\alpha_t = 0$ for any $t > 0$.[29] *Q.E.D.*

A.2 Proof of Proposition 2

Proof: To see why it is always (i.e. for any $\delta \in (0,1)$) beneficial for the prosecutor to drive the system toward 100% plea bargains, even at the cost of lowering the sanction to be below $\underline{b}$, note first that any offer $b > \underline{b}$ implies that, in all periods, the expected sanctioning equals $S(B)$, because the process of plea bargaining described in Proposition 1 is never triggered. Now consider an offer of $b = CE(\tilde{\alpha}, P(B)) < CE(0, P(B)) = \underline{b}$, where $z = \tilde{\alpha}$ is a risk neutral defendant.[30] Noting that $CE(\tilde{\alpha}, P(B)) = S(B)$, we get that *in any period* $t \geq 1$, some defendants (first those with $z \leq \tilde{\alpha}$, later others too) sign the PB hence get a sanction $S(B)$, while others go to trial, where their sanction is for sure larger than $S(B)$ because they face a prosecutor whose budget is larger than B. Overall, this is for sure better for the prosecutor compared to any offer $b > \underline{b}$ that does not trigger plea bargaining hence results in a sanction of $S(B)$ in all periods. *Q.E.D.*

A.3 Proof of Proposition 3

Proof: (i) Consider two plea offers, b_1 and b_2, such that $b_1 < b_2 < \underline{b}$. Then there exists a time period T such that

(2)

$$\sum_{t=T}^{\infty} \delta^t \left\{ \alpha_t(b_1) b_1 + (1-\alpha_t(b_1)) S\left(\frac{B-\alpha_t(b_1)c}{1-\alpha_t(b_1)}\right) \right\} < \sum_{t=T}^{\infty} \delta^t \left\{ \alpha_t(b_2) b_2 + (1-\alpha_t(b_2)) S\left(\frac{B-\alpha_t(b_2)c}{1-\alpha_t(b_2)}\right) \right\},$$

[28]Note that if $b = \underline{b}$ while $\alpha > 0$, convergence again can be only to $\alpha = 1$.

[29]Note also that $\alpha_1 = 0$ only if $b \geq \underline{b}$, hence we start from a stable steady state (or, in the case of $b = \underline{b}$, we stay in the steady state where $\alpha = 0$ because $\alpha_1 = \alpha_0 = 0$).

[30]If the defendant $x = 0$ is risk neutral himself, set $\tilde{\alpha} > 0$ to be the largest value corresponding to a risk-neural defendant (where the fact that $\tilde{\alpha} > 0$ follows from the assumption that the mass of defendants who are not strictly risk seeking is strictly positive); if all defendants are risk averse then consider $b = CE(1, P(B))$ instead.

because

$$lim_{t\to\infty}\left\{\alpha_t(b_1)b_1+(1-\alpha_t(b_1))S\left(\frac{B-\alpha_t(b_1)c}{1-\alpha_t(b_1)}\right)\right\}=b_1<b_2=lim_{t\to\infty}\left\{\alpha_t(b_2)b_2+(1-\alpha_t(b_2))S\left(\frac{B-\alpha_t(b_2)c}{1-\alpha_t(b_2)}\right)\right\}.$$

Then, even if the total discounted sanction up to period T is higher under b_1 than it is under b_2 for any $\delta\in(0,1)$, it is still finite, while the sums in equation (2) are infinite for $\delta\to1$. Hence it is guaranteed that, for δ sufficiently close to 1, we will have

$$\sum_{t=1}^{\infty}\delta^t\left\{\alpha_t(b_1)b_1+(1-\alpha_t(b_1))S\left(\frac{B-\alpha_t(b_1)c}{1-\alpha_t(b_1)}\right)\right\}<\sum_{t=1}^{\infty}\delta^t\left\{\alpha_t(b_2)b_2+(1-\alpha_t(b_2))S\left(\frac{B-\alpha_t(b_2)c}{1-\alpha_t(b_2)}\right)\right\},$$

which guarantees that the prosecutor prefers to offer b_2 over offering b_1.

(ii) Since $b\simeq\underline{b}=CE(0,P(B))$, and $CE(z,P(B))\leq CE(0,P(B))\ \ \forall z\in[0,1]$, we get that any defendant who signs the offer is (weakly) worse off. As for the defendants who choose to go to trial, they face a larger expected sanction because, for any $\alpha>0$, we have $I(\alpha)=\frac{B-\alpha c}{1-\alpha}>B$, implying that $S(I)>S(B)$. *Q.E.D.*

B Fictitious-Play dynamics

Suppose that instead of using only the observation of the previous period, defendants use the observations of all past periods. This is called *fictitious play*. In period t defendants best respond to a weighted average of $\alpha_0,...,\alpha_{t-1}$ (with non-zero weight on α_{t-1}). That is, defendant z signs the PB if and only if $b\leq CE(z,P(I(\bar{\alpha}_{t-1})))$, where ω_τ^t is the weight put on α_τ in period t and $\bar{\alpha}_t\equiv\sum_{\tau=0}^{t}\omega_\tau^t\alpha_\tau$. The only constrain we put on $\{\omega_\tau^t\}$ is that they can be recursively defined, i.e. $\bar{\alpha}_t=\sum_{\tau=0}^{t}\omega_\tau^t\alpha_\tau=\omega_t^t\alpha_t+(1-\omega_t^t)\sum_{\tau=0}^{t-1}\omega_\tau^{t-1}\alpha_\tau$. The degenerate case studied in the main text is the case where $\omega_t^t=1$, and the common case of a uniform distribution over all past periods is the case where $\omega_t^t=\frac{1}{t+1}$ (and $\omega_\tau^{t-1}=\frac{1}{t}$ for any τ). We will show that—while convergence is slower—it is still guaranteed to happen, hence Propositions 1 and 2 hold.

Claim 1 If $b<b_0$, then α_t increases till it converges.

Proof: We present a proof by induction. Test for $t=1$: for any $b<b_0$ we have $\alpha_1=1-F_{CE(z,P(B))}(b)>0=\alpha_0$. Next, we will assume that $\alpha_t>\alpha_{t-1}>\alpha_{t-2}>\cdots>\alpha_1$ and prove that also $\alpha_{t+1}>\alpha_t$. Noting that $\bar{\alpha}_{t-1}\equiv\sum_{\tau=0}^{t-1}\omega_\tau\alpha_\tau$, we have $\bar{\alpha}_t=\sum_{\tau=0}^{t}\omega_\tau^t\alpha_\tau=\omega_t^t\alpha_t+(1-\omega_t^t)\sum_{\tau=0}^{t-1}\omega_\tau^{t-1}\alpha_\tau=\omega_t^t\alpha_t+(1-\omega_t^t)\bar{\alpha}_{t-1}>\bar{\alpha}_{t-1}\iff\alpha_t>\bar{\alpha}_{t-1}$, where the latter holds by assumption: $\alpha_t>\alpha_{t-1}>\bar{\alpha}_{t-1}$. Finally, $\bar{\alpha}_t>\bar{\alpha}_{t-1}$ implies that $\alpha_{t+1}=1-F_{CE(z,P(I(\bar{\alpha}_t)))}(b)>1-F_{CE(z,P(I(\bar{\alpha}_{t-1})))}(b)=\alpha_t$. *Q.E.D.*

C Dense risk attitudes

Assumption 1 states that b_α weakly increases in α in the range $(0,1)$. The text then explains that this assumption holds if the distribution of risk attitudes is sufficiently "dense". In this section we demonstrate what this means using an example, in which defendants have CRRA (dis)utilities from the sanction, and the Odds Ratio of conviction is a linear function of the budget per case. Specifically, suppose that the probability is such that $\frac{p}{1-p} = \frac{I}{B}$,[31] and the expected sanction $(S(I))$ is then simply $sp = s\frac{I}{B+I}$. Suppose further that a PB costs the prosecutor no more than one quarter of the cost of handling a trial ($\frac{c}{B} \leq \frac{1}{4}$) and that the disutility from conviction with sanction s is given by $s^{1+y-2yz}$. We will show that a sufficient (and far from necessary) condition for ensuring that $b'_\alpha > 0$ is that $y \leq 0.2$. That is, the disutility can range from slightly convex ($s^{1.2}$, risk averse) to slightly concave ($s^{0.8}$, risk loving).

The CE of defendant z satisfies the following equation: $\omega - CE^{1+y-2yz} = p\left(\omega - s^{1+y-2yz}\right) + (1-p)\,\omega$. Then $CE = sp^{\frac{1}{1+y-2yz}}$, hence $b_\alpha = CE\left(\alpha, P\left(\frac{B-\alpha c}{1-\alpha}\right)\right) = s\left(p(\alpha)\right)^{\frac{1}{1+y-2y\alpha}} = s\left(\frac{I(\alpha)}{B+I(\alpha)}\right)^{\frac{1}{1+y-2y\alpha}}$, and so $b'_\alpha > 0$ if $\psi(\alpha) \equiv \frac{1}{1+y-2y\alpha} ln\left(p(\alpha)\right)$ is an increasing function.

$$\psi'(\alpha) > 0 \Longleftrightarrow \frac{1}{p}\frac{dp}{dI}\frac{dI}{d\alpha}\left(1+y-2y\alpha\right) > -2ylnp$$

$$\Longleftrightarrow \frac{B}{(B+I)^2}\frac{B-c}{(1-\alpha)^2}\left(1+y-2y\alpha\right) > -2yplnp$$

$$\Longleftrightarrow \frac{B(1-\alpha)^2}{(2B-\alpha(B+c))^2}\frac{B-c}{(1-\alpha)^2}\left(1+y-2y\alpha\right) > -2yplnp$$

$$\Longleftrightarrow \frac{B(B-c)(1+y-2y\alpha)}{(2B-\alpha(B+c))^2} > -2yplnp$$

Given that $-plnp \leq e^{-1} < 0.375$, we get that the RHS is smaller than $0.75y$. As for the LHS, both the numerator and the denominator are positive and decreasing in α, hence a lower bound on the numerator is achieved by substituting $\alpha = 1$ and an upper bound on the denominator is achieved by substituting $\alpha = 0$, and we get that the LHS is for sure larger than $\frac{B-c}{4B}(1-y)$. Recalling that $\frac{c}{B} \leq \frac{1}{4}$, we get a sufficient (and far from necessary) condition is $\frac{0.75}{4}(1-y) > 0.75y$. This condition holds whenever $1-y > 4y$, i.e. for any $y \leq 0.2$.

[31] When $\frac{p}{1-p} = \frac{I}{B}$ we get that without plea bargains the probability of conviction is 1/2.