The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

Punitive Police? Agency Costs, Law Enforcement, and Criminal Procedure

Dhammika Dharmapala
Nuno Garoupa
Richard H. McAdams

CESIFO WORKING PAPER NO. 5310
CATEGORY 1: PUBLIC FINANCE
APRIL 2015

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5310*

Punitive Police? Agency Costs, Law Enforcement, and Criminal Procedure

Abstract

Criminal law enforcement depends on the actions of public agents such as police officers, but the resulting agency problems have been neglected in the law and economics literature (especially outside the specific context of corruption). We develop an agency model of police behavior that emphasizes intrinsic motivation and self-selection. Drawing on experimental evidence on punishment preferences, in which subjects reveal a heterogeneous preference for punishing wrongdoers, our model identifies circumstances in which "punitive" individuals (with stronger-than-average punishment preferences) will self-select into law enforcement jobs that offer the opportunity to punish (or facilitate the punishment of) wrongdoers. Such "punitive" agents will accept a lower salary, but create agency costs associated with their excessive zeal (relative to the public's preferences) in searching, seizing, and punishing suspects. In our framework, the public chooses (under reasonable assumptions) to hire punitive police agents, while providing suspects with strong criminal procedure protections, thereby empowering other agents (such as the judiciary) with average punishment preferences to limit the agency costs of excessive zeal. We thus argue that intrinsic motivation and self-selection provide a possible explanation for the bifurcated structure of criminal law enforcement in which courts constrain police with pro-defendant rules of criminal procedure. We also explore various other implications of this framework.

JEL-Code: K420.

Keywords: law enforcement, intrinsic motivation, punishment, criminal procedure.

Dhammika Dharmapala
University of Chicago Law School
USA - 60637 Chicago IL
dharmap@uchicago.edu

Nuno Garoupa
Texas A&M University Law School
Fort Worth / Texas / USA
ngaroupa@ffms.pt

Richard H. McAdams
University of Chicago Law School
USA - 60637 Chicago IL
rmcadams@uchicago.edu

Revised Version - April 2015
We thank the Editor (Omri Ben-Shahar), an anonymous referee, Alon Harel, William Hubbard, Aziz Huq, Andrew Leipold, Jonathan Masur, Eric Posner, the late Larry Ribstein, Jacqueline Ross, Naomi Schoenbaum, seminar participants at Yale University, Stanford University, the University of Chicago, and the University of Illinois, and conference participants at the 2011 American Law and Economics Association annual meetings and the 2013 Law and Society Association annual meetings for helpful comments. Any remaining errors or omissions, are, of course, our own.

"Don't you love catching someone trying to get away with something? I love it. But that's why I could never be a cop because I would be too happy. I would catch somebody speeding and go 'I got you, I got you . . .' Really obnoxious"

Jerry Seinfeld, *Seinlanguage*, 1993, p. 91.

"Quis custodiet ipsos custodes?"

Juvenal, 1st/2nd century

1) Introduction

In a classic account of corporate law, Easterbrook and Fischel (1991, p. 14) state that "To understand corporate law you must understand how the balance of advantage among devices for controlling agency costs differs across firms and shifts from time to time." Broadly speaking, the same point applies to criminal law. Criminal law enforcement depends on the actions of police officers, prosecutors, and others who act as agents of governments and ultimately as agents of the public. As principal-agent theory predicts, law enforcement agents may have objectives that differ from those of their principals. Yet, economic analysis of criminal law and procedure has paid relatively little attention to the principal-agent problem in law enforcement. Since Becker (1968), the dominant issue in the economics of crime has not been agency costs but optimal probability and severity of punishment in order to achieve efficient deterrence. Recent reviews of the economics literature on public law enforcement (e.g., Polinsky and Shavell, 2009) reveal a literature that provides an exhaustive treatment of optimal deterrence, but scant attention to the principal-agent problem. Friedman (1999) identifies this neglect of the principal-agent problem in criminal law, noting that "[t]he orthodox theory of optimal punishment . . . treats criminals . . . as rational self-interested actors," but "treats the enforcement apparatus – police, courts, prosecutors, and legislature – as a philosopher-king, with imperfect information but only the best of motives."

Agency costs are mentioned only in the limited context where the potential criminal is a collective entity, such as a corporation, that consists of principals and agents. This literature thus considers the agency problem for the *targets* of criminal punishment, but not the agency problem for the *creators* and *enforcers* of criminal law – legislators, police, prosecutors, chief executives, and others. There is a robust and important literature on corruption, which includes corruption of public enforcers. However, we argue below that corruption is only one problem of agency, and that it is quite distinct from the issues of intrinsic motivation and self-selection that are highlighted in the model developed in this paper. On the more general agency problem in criminal law, there are only a handful of economic articles (e.g. Friedman, 1999; Hylton and Khanna 2007); these stress issues of rent seeking that are distinct from the concerns of this paper.[1]

This paper seeks to characterize the fundamental contours of the agency relationship between key law enforcement agents – police officers – and the general public, and to derive some implications for understanding the role of criminal procedure protections. We take as a premise that this particular agency problem cannot be solved through performance-based contracting. As discussed in the previous literature (e.g. McAdams, 2012), a fundamental reason is the danger of fabrication – that if law enforcers were paid by the arrest or conviction, they would "frame" individuals to collect their fee (an example of the classical gaming effect by which the agent manipulates the performance measure set by the principal). Another related problem would be an additional effort on the activities targeted by performance measures (for example, arrest or conviction) at the expense of the activities excluded from those performance measures (for example, verification of guilt).

Our approach is more consistent with existing practices in compensating police. In particular, we draw on experimental evidence on "altruistic punishment," in which some (but not all) subjects willingly incur a cost to punish wrongdoers. We posit heterogeneity in the "punishment preferences" of potential agents and argue that these differences cause agents with unrepresentative preferences to self-select into law enforcement jobs that offer the opportunity to punish wrongdoers or at least to facilitate their punishment by prison authorities. This tends to create a divergence of interest between principal and agent: specifically, punitive agents will operate with a lower threshold of doubt for punishing suspects than would a citizen with average preferences for punishment. Thus, in contrast to the problem of shirking – as extensively studied in the context of corporate law and governance – the agency problem here is likely to include the problem of *excessive zeal*.

We develop a simple sequential game-theoretic model to formalize and clarify this idea. In this model, a representative citizen (the principal) offers a fixed wage contract to police. There

[1] There is a growing literature on prosecutors (e.g., Garoupa 2009; Gordon and Huber, 2009; Rasmusen, Raghav and Ramseyer 2009; Ribstein 2011), but aside from the work on corruption there is a particular neglect of our focus, police officers. Note also that while the law and economics literature has tended to ignore agency costs of criminal law enforcement, scholarship on criminal law has sometimes used agency concepts with regard to prosecution and policing (e.g. Bibas, 2009; Richman 2003; Stuntz 2001, pp. 549-550, Stuntz 2008, p. 1974). However, this literature has not developed formal principal-agent models of law enforcement.

are two types of potential applicants – "punitive" individuals (with stronger-than-average punishment preferences), and individuals with punishment preferences similar to those of the principal; each of these types decides whether to apply to join the police, in response to the wage contract that is offered. The (endogenously-selected) police then decide which suspects to arrest, based on their observed probability of guilt. We first present a model in which suspects have only weak criminal procedure protections, and in which all those arrested are punished. There are two equilibrium outcomes, depending on the underlying values of various parameters (although the equilibrium outcome is unique for a given set of parameter values). One outcome involves self-selection - only punitive types apply, the wage is relatively low, and there is a divergence in preferences between the citizen and the police as to the probability threshold above which suspects should be arrested and punished. The other outcome involves a higher wage, which attracts both punitive and nonpunitive types to apply. The agency costs of excessive zeal are mitigated, but the citizen incurs a higher wage cost.

We then modify this framework to introduce strong criminal procedure protections for suspects. In particular, a court (which for various reasons discussed below is assumed to share the principal's preferences) determines which of the arrested suspects will be punished. The equilibrium outcome here involves self-selection, where only punitive types apply. The wage is higher than in the prior self-selection equilibrium because strong criminal procedure protections limit punishment opportunities, but the job remains more attractive to punitive than to nonpunitive types as long as punishment opportunities continue to exist. Now, however, the court limits the agency costs of excessive zeal, thereby reducing the extent to which the principal's payoff is reduced by the excessive punishment of suspects.

Finally, we characterize the conditions under which the citizen is better off with strong rather than weak criminal procedure protections for suspects. Our conclusion is that the citizen-principal may benefit by seeking to control punitive police by using other agents – mostly, judges and juries – who have preferences closer to those of the public. Thus, criminal procedure protections of suspects can reduce agency costs between the general public and police; they enable the public to benefit to some degree from the lower cost of employing intrinsically-motivated police, while avoiding the worst excesses of over-enforcement by these agents.[2]

[2] A possible objection to this approach is that criminal procedure protections for suspects tend to be unpopular with elements of the general public. While this is true, our argument requires only that on average police officers tend to

Our framework potentially derives the basic structure of the criminal justice system – the separation of judicial and executive enforcement powers, and the judiciary's pro-defendant rules of criminal procedure – from the agency problem in law enforcement. We also make some progress in explaining aspects of the compensation of law enforcement agents: society relies on low powered incentives (wages and salary) rather than high powered incentives (bounties) not only because of the risk of fabrication (and possible verifiability constraints), but also because society manages to attract into policing those who are intrinsically motivated to perform the job, thus rendering external incentives less necessary. The presence of internal incentives makes the trade-off between low and high powered incentives more likely to favor low powered incentives. While our model does not explicitly allow for a choice of effort, we argue that a straightforward extension would show that intrinsically motivated police would be less likely to shirk and to take bribes from guilty suspects because they cannot satisfy their punishment preferences without working and arresting the guilty. In a related paper (McAdams, Dharmapala, and Garoupa, 2015), we develop some implications of our framework for understanding doctrinal distinctions between police and nonpolice government actors that have been drawn in US Fourth Amendment jurisprudence, and for the "community caretaking" doctrine (which draws a distinction between the threshold of suspicion required for a search when police are acting in a law enforcement capacity and when they are engaged in community caretaking).

Our paper proceeds as follows. Section 2 reviews the literature. Section 3 develops and solves the formal model. Section 4 discusses the implications, and Section 5 concludes.

2) Literature Review

The principal-agent problem is the subject of a vast literature in economics and related disciplines (such as law and accounting). One branch of this literature that is particularly relevant considers situations in which a principal delegates considerable decisionmaking authority to an agent who may have different preferences. In such circumstances, the principal must decide how much to intervene ex post to supervise the agent (a choice analogous to the strength of criminal procedure protections in our model). An influential contribution in this literature is Aghion and Tirole (1997). They distinguish between "formal" and "real" authority within organizations, and

be *more* hostile to criminal procedure protections than does the average citizen. Ethnographic accounts of policing (e.g. Skolnick, 1966; Manning and van Maanen, 1978) provide considerable support for this notion.

argue that the latter – the effective ability to choose projects or make decisions – is determined by the degree of asymmetric information between the principal and agent. In their model, providing agents with greater authority can induce them to exert greater effort and to acquire and communicate more information, thereby potentially benefiting the principal even in the presence of divergent preferences. Our model falls within this broad tradition. However, it abstracts from the choice of effort by the agent and from the agent's choice of how much information to acquire, in order to focus on issues of intrinsic motivation and self-selection. Thus, our paper is also related to the literature that introduces intrinsically–motivated agents into the principal-agent framework (e.g. Besley and Ghatak, 2005; Prendergast, 2007). However, this literature has not previously focused on police officers or on understanding the implications of intrinsic motivation for criminal procedure.

Our paper also has parallels with a number of principal-agent analyses that incorporate intrinsic motivation (or related notions such as ideology) in contexts other than criminal procedure. For instance, Gilligan and Krehbiel (1987) explain why the US Congress defers to legislative committees by arguing that ceding decisionmaking authority to committees encourages greater effort by the latter in policymaking; this can make the legislature better off, even when the committee has unrepresentative preferences over policy. Kalt and Zupan (1990) argue that voters can address agency costs by choosing ideologically-motivated political representatives (who share the voters' preferences on the most important issues) who will act in the voters' interests even without monitoring. In a recent contribution, Bubb and Warren (2013) develop a model in which the President appoints ideologically-motivated officials to regulatory agencies in order to mitigate the problem of shirking, while also using centralized regulatory review to constrain these ideological officials. While these models have certain parallels with ours, we address a different setting with some quite distinctive characteristics.

Another branch of the law and economics literature that is related to this paper, but distinct in important respects, addresses agency costs within an organization that is itself the potential offender (as in the context of criminal acts by corporations) – see e.g. Kornhauser (1982), Sykes (1984), Macey (1991), Arlen and Carney (1992), Polinsky and Shavell (1993), Garoupa (2000). More recently, Crocker and Slemrod (2005) analyze the impact of agency costs on illegal tax evasion by corporations, while Desai and Dharmapala (2006) study lawful but aggressive tax avoidance activity by corporations in a principal-agent framework. However, all of these analyses are distinct from ours in that they focus on agency costs as a potential source of crime, and not on agency costs within law enforcement.

There are a few papers that consider principal-agent problems in enforcement, beginning with Becker and Stigler (1974). They emphasize the need to examine the efficiency of enforcement in addition to the efficiency of substantive rules, but their primary focus is on the prevention of corruption by the structure of agent compensation. Easterbrook (1983) describes "criminal procedure as a market system," identifying prosecutors as the public's agents and describing how prosecutors and juries "price" crime, but does not offer a general model of criminal enforcement agents, nor discuss police as agents.

A large economic literature analyzes corruption (e.g. Shleifer and Vishny, 1993; Mookherjee and Png, 1995; Bowles and Garoupa, 1997; Rose-Ackerman, 1999; Polinsky and Shavell, 2001; Echazu and Garoupa, 2010). This literature addresses situations in which an agent departs from some notion of optimal punishment, in the direction of less than optimal punishment, in exchange for a bribe from the offender. While there exist some similarities between this situation and ours in terms of a divergence between the principal's and agent's preferences, there are also important differences between corruption and the types of agency costs that we consider. In particular, corrupt law enforcers are assumed to respond to extrinsic motivations (primarily monetary payments from offenders), whereas we assume intrinsically-motivated agents. The agency costs of excessive zeal that arise as a result of intrinsic motivation and self-selection are quite different in nature from agency costs of corruption. We believe that considering these types of agency costs separately generates interesting new insights, as discussed in Sections 3 and 4 below.

The past contributions to the economic analysis of criminal procedure that are closest to ours are Friedman (1999) and Hylton and Khanna (2007). Friedman (1999) focuses on the problem of rent-seeking, i.e. that criminal enforcers will use their expansive powers to "expropriate other people." Friedman seeks to explain the otherwise economically puzzling choice of prisons as the primary mode of punishment. He argues that the death penalty is far more efficient because it is cheaper to kill someone than to house them in a prison and because, following Becker (1968), the greater severity allows the state to punish fewer individuals, which saves on the costs of detecting crime. Friedman's answer to this puzzle is that inefficient punishments are less susceptible to abuse. First, the weaker the punishment, the less wealth the enforcer can extract by threatening to impose the punishment. Second, the more costly the punishment to the state, the less credible is the threat to inflict the punishment if the bribe is not paid.[3]

Hylton and Khanna (2007) is motivated by a very similar question to ours, namely how to explain the apparent pro-defendant bias of criminal procedure. Like Friedman (1999), they identify the potential dangers of criminal enforcement and use the threat of extortion to explain the rights of criminal defendants. In particular, they argue that criminal procedure protections ameliorate the extraction problem. Procedural protections make it more costly to convict the innocent than to convict the guilty, thus reducing the ability of agents to credibly threaten innocent suspects with false arrest, conviction, and punishment. Hylton and Khanna (2007) specifically offer the bar on double jeopardy and ex post facto laws, the right to a jury trial, and the vagueness doctrine as examples of rules that render it more difficult to extract rents from the innocent.[4]

Friedman (1999) and Hylton and Khanna (2007) address very similar questions to ours and offer important insights. However, they model enforcers as being conventionally self-interested. Our primary objection is thus that this literature addresses one set of puzzles – why do we limit the infliction and intensity of criminal punishments? – by intensifying another puzzle – how do we induce police to work? Although even low-powered incentives within police departments may induce some effort, punishment limitations and procedural protections may cause greater shirking among police by making enforcement against the guilty more difficult. For instance, although procedural protections make it *relatively* more costly to convict the innocent than the guilty, they still make it *absolutely* more costly to convict the guilty. The net result of these two effects is indeterminate and therefore not a clear justification for the procedural rights.[5]

[3] Note, however, that limiting the use of the death penalty may do little to prevent law enforcers from extracting rents from the innocent given that most people would give all of their wealth to avoid a prison term of life or a substantial part of their life. Moreover, the officer who would extract a payment by threatening to frame an innocent person is not the one who bears the expense of incarcerating that individual. Thus, the expense of prison does not deter the individual officer from demanding bribes (unless the individual officer fully internalizes society's costs of imprisonment, in which case there are no agency problems and hence no issue of rent extraction).

[4] A possible objection is that if criminal procedure rights are necessary only to prevent law enforcers from extracting bribes from the innocent, then such rights are necessary only for those who possess resources to extract; thus, the indigent should receive no criminal procedure rights.

[5] Related to shirking is a different type of corruption from the one that Hylton and Khanna (2007) model - the *guilty* offering bribes to the police to abstain from enforcement. It is not clear why self-interested enforcers would ever bother to punish the guilty if they could take bribes instead. Even though procedural protections will decrease the extraction of bribes from the innocent, they will increase the number of the guilty who escape punishment *via*

In general, if enforcers lack intrinsic motivation and external incentives are low-powered, then there is a painful tradeoff: making enforcement cheap imperils the innocent, but making enforcement costly benefits the guilty. We argue in Section 3 below that it is possible to mitigate this tradeoff by making use of the intrinsic motivation of those who choose to join the police.

In constructing a theory of intrinsic motivation among police officers, this paper draws on the experimental literature on punishment. Our model assumes that individuals have preferences for punishing wrongdoers and therefore derive utility directly from facilitating (causally contributing to) the punishment of wrongdoers (without that punishment producing further consequences). This assumption, of course, deviates from the simple and standard assumption that actors care only about ordinary consumption goods. There is already a literature analyzing the consequences of intrinsic motivation in the principal-agent framework (e.g. Besley and Ghatak, 2005; Prendergast, 2007). However, even if one were skeptical about intrinsic motivation generally, our specific assumption about punishment preferences is grounded on what is now an extensive body of experimental evidence that is consistent with the existence of such preferences.

A substantial body of experimental literature has established that humans have a heterogeneous willingness to incur costs to punish those perceived as wrongdoers. The most common design is an iterated public goods or voluntary contribution game. In the standard version of the game (without punishment), the established result is that some individuals make contributions in the early rounds, but that contributions quickly decline to zero over a few rounds. Fehr and Gächter (2000) introduced the novel feature of giving the players a punishment option, where they could incur costs to impose punishment on others. Even in the repeated game, backward induction from the last round implies that the rational and selfish player will not contribute and not punish, but the results reveal that some (but not all) individuals punish free-riders and that punishment can sustain contributions.

Particularly noteworthy for our purposes is heterogeneity in the willingness of subjects, in all these experiments, to lower their own payoff in order to reduce the payoffs of free riders (e.g., Anderson and Putterman 2006, pp. 13-15; Carpenter 2007, pp. 532-33; Henrich et al., 2006

bribery. First, procedural protections involve giving other enforcement agents veto power over punishment, which creates new bribery opportunities. Second, procedural protections decrease the probability of being convicted for the crime *of bribery*, thus increasing the productivity of that activity. With procedural protections, the guilty will be more willing to offer a bribe to enforcers and the enforcers will be more willing to take the bribe.

(Tables 1 and 2)). Subsequent studies have extended this finding in a number of directions. It also extends to the punishment of wrongs done to others rather than to oneself (Fehr and Fischbacher, 2004). Researchers (Henrich et al., 2006) find the intrinsic willingness to punish exists across a wide variety of cultures. Because punishment of free-riders helps to sustain cooperation, a number of evolutionary models explain how the preference could arise (see Henrich et al. (2006, notes 3-10) for citations to this literature).

The willingness of some subjects to inflict costly punishment also persists when there is uncertainty about whether the putative wrongdoer committed a wrong. The basic social dilemma experiments with punishment opportunities described above all involve certainty about subjects' behavior. Thus, punishment choices are made in an environment where potential punishers are sure of their targets' wrongdoing. These results thus cannot distinguish between a taste for punishment of the guilty and a taste for "justice" (i.e. for both punishing the guilty *and* exonerating or refraining from punishing the innocent). Grechenig, Nicklisch and Thöni (2010) modify past experimental designs by providing potential punishers with only a noisy signal regarding whether other subjects defected or cooperated in the past round. If the taste for punishment revealed in previous experiments were in reality a taste for justice, then we would expect that there would be a significant decline in punishment when this uncertainty is introduced (in order to avoid the punishment of possibly innocent subjects). However, Grechenig, Nicklisch and Thöni (2010) find that there is no decline in punishment as a result of uncertainty. This suggests that the taste for punishing wrongdoers is not (at least fully) offset by a corresponding taste for the nonpunishment of the innocent.

The most powerful and directly relevant evidence for our purposes come from two very recent studies that use police officers and police applicants as subjects. Dickinson, Masclet and Villeval (2014) use a subject pool that includes 87 French police commissioners (or individuals who had recently passed the competitive national exam and were on their way to becoming commissioners) as well as nonpolice (primarily student) subjects. Police and nonpolice subjects participated in standard experimental games for testing preferences to punish socially bad behavior (where punishment is costly and without strategic benefit). The study finds that the police subjects are willing to incur greater costs to impose punishment. These results could be attributable either to self-selection into policing or to the effect of police training. Friebel, Kosfeld and Thielmann (2014) address this concern by using as their "police" subjects high school students in the German *länder* of Hesse and Rheinland-Pfalz who have applied to join the police forces, but have experienced no police training. Compared to nonapplicant high school students, the police applicants are willing to incur greater costs to punish, suggesting that self-selection plays a major role in the punitive preferences of police.

It is important to bear in mind that the experiments reviewed above are not specifically designed to replicate the context of criminal law enforcement. They do not involve serious criminal wrongdoing or punishment. Nonetheless, it is possible to derive from them some general principles of human motivation and behavior with respect to the punishment of perceived wrongdoers. In particular, the experimental evidence is consistent with the existence of a taste for the punishment of wrongdoers that varies considerably across individuals. The taste for punishment is not confined to those who harm the punisher directly, but extends to wrongdoers in general. Given that punishment occurs in these experiments even when the design entails that punishment is subject to a free rider problem, the utility from punishment seems to result from causing or facilitating the punishment, and not merely from the knowledge that punishment occurs. Finally, the taste for punishing wrongdoers appears not to be fully offset by a corresponding taste for the nonpunishment of the innocent (i.e. the taste for punishment cannot be interpreted simply as a taste for "justice").

We acknowledge that the idea that punishment behavior in these experiments reflects a taste for punishment is only one possible interpretation of this body of results. It is possible that individuals have uniform tastes for punishment, while the willingness to incur costs of punishment varies across the population. However, our interpretation of the evidence as indicating heterogeneous preferences over punishment has some support in this literature. For instance, some experimental studies vary the cost of punishment, including scenarios in which punishment is costless. Anderson and Putterman (2006, p. 8) report an experiment that generated 288 observations in which subjects were faced with costless punishment opportunities. Positive amounts of punishment were inflicted in only 161 of these observations, suggesting heterogeneity in punitive tastes rather than simply in the willingness to incur costs. De Quervain et al. (2004) study neural images of subjects undergoing a punishment experiment and find that the punishment of norm violators activates a reward center in the brain. They set up various treatments in which punishment is costless and others in which it is costly to the punisher. When punishment is costless, they find variation in the degree of brain activation among punishers.

Moreover, those subjects who experience the greatest activation when punishment is costless are also those who are willing to incur the highest costs to punish when punishment is costly. While not necessarily conclusive, this and other evidence suggests some degree of heterogeneity in "deep" preferences over punishment. Moreover, ethnographic accounts of policing (e.g. Skolnick, 1966; Manning and van Maanen, 1978) document the hostility of many police officers to procedural rules imposed on them by the public, and strongly suggest that there is a divergence in preferences (rather than merely in the willingness to incur costs) between the police and the public.

3) The Model

We present our model in four steps. We first show how different punishment preferences imply different preferences for the threshold of doubt, meaning the probability of guilt necessary for an individual to prefer a suspect's search, arrest, and punishment. Second, we specify and solve a simple sequential game of the police hiring and law enforcement process with weak criminal procedure protections (CPP) for suspects. We then modify this model to incorporate strong CPP and solve this model. Finally, we characterize the circumstances in which citizens will prefer strong to weak CPP.

3.1) The Characterization of Preferences over Punishment

We begin by showing that, if law enforcement agents have punishment preferences that differ from the average citizen's punishment preferences, the agents will favor a different threshold of doubt for search, seizure and punishment than would a typical citizen. Let p be the probability that a given suspect is guilty, and let u_C be the principal's utility. We refer to the principal as the representative "citizen." Her preferences (also shown in Table 1) can be represented as follows:

$u_C = 0$ if the suspect is guilty and punished, or if the suspect is innocent and not punished;

$u_C = -L$ if the suspect is guilty and not punished

$u_C = -\beta L$ if the suspect is innocent and punished.

The citizen's utility is thus normalized to zero when the punishment decision is correct. Relative to this baseline of zero, utility is lower when a guilty suspect is not punished ($-L$) and when an innocent suspect is punished ($-\beta L$). This is a very general characterization of the principal's preferences, requiring in essence only that the citizen has a preference that punishment and nonpunishment be directed towards the appropriate targets. Consequently, these "truth-seeking" preferences are widely used in the scholarly literature (e.g. Dharmapala and McAdams, 2003). The parameter β represents the relative cost of punishment errors that involve punishing the innocent, relative to errors that involve nonpunishment of the guilty.[6]

Of course, the decision to punish will generally have to be made under conditions of uncertainty about guilt. Thus, the citizen will wish to punish whenever the *expected* utility from punishment exceeds the *expected* utility from nonpunishment. Given a suspect with probability *p* of guilt, the expected utility from punishment is $-(1 - p)\beta L$ while the expected utility from nonpunishment is $-pL$. The former will exceed the latter when p is sufficiently large. Specifically, the citizen wishes to punish suspects iff:

$$p > \frac{\beta}{\beta+1} \equiv p^* \tag{1}$$

The relative magnitude of the parameter β determines the threshold p^* – for instance, if punishing the innocent is very costly compared to not punishing the guilty (i.e. β is substantially larger than 1), p^* will be close to 1. As β approaches infinity, p^* approaches 1. If $\beta = 1$ (so that wrongful punishment and wrongful nonpunishment are equally costly), then $p^* = ½$. If β approaches zero, p^* approaches 0.

The most natural interpretation of the experimental evidence reviewed in Section 3 above is that there exists a significant element of the population that has a taste for the punishment of wrongdoers – i.e. that derives extra utility *M* (relative to the average preferences characterized in Table 1) from causing or facilitating the punishment of the guilty. To simplify the subsequent algebra, and without loss of generality, we define $M = mL$, where $m > 0$ is a parameter that represents the intensity of the taste for punishment of the guilty (relative to the loss L from erroneous punishment). While these "punitive" preferences (represented in Table 2) seem to capture the experimental evidence most simply and parsimoniously, a more general characterization of the preferences of potential agents would introduce a second variable $n > 0$ to capture extra utility (relative to the average preferences in Table 1) from causing or facilitating the nonpunishment of the innocent, as would reflect "exoneration" preferences. An agent with

[6] The citizen's preferences might conform to Blackstone's (1765-69) famous dictum that ". . . it is better that ten guilty persons escape, than that one innocent suffer" - i.e. that β would substantially exceed 1. However, our results do not require this assumption, and so we do not impose it here. See Volokh (1997).

both $m > 0$ and $n > 0$ could be described as having more intense preferences for justice (both the punishment of the guilty and the nonpunishment of the innocent) than does the average citizen; the combination may be termed "justice" preferences.

If potential agents have pro-justice preferences and $m = n$, then the divergence characterized in Remark 1 below between the preferred thresholds of doubt for punishment of these agents and the average citizen will not exist. Indeed, if $m < n$, it is possible that the agent will prefer a higher threshold of doubt than does the citizen. While conceding that these are theoretical possibilities, we impose the assumption that $m > n$ (given this assumption, setting $n = 0$ as in Table 2 represents an innocuous normalization of m). We do so based on both experimental evidence and some general conceptual considerations. First, the experimental data supports the existence of punitive preferences, but we have not discovered similar evidence of exonerative preferences. To the contrary, the available evidence, though scant, suggests that pro-justice preferences are not particularly common or intense. As discussed in Section 2 above, Grechenig, Nicklisch and Thöni (2010) find that introducing uncertainty about whether individuals are guilty does not reduce the propensity of other experimental subjects to punish them. If preferences were pro-justice, rather than merely punitive, we would expect a significant decline in punishment in order to avoid possible punishment of the innocent. Perhaps more important, even if strong pro-exoneration or pro-justice preferences exist, there is no evidence of their heterogeneity, which is demonstrated for punishment preferences. Heterogeneity is crucial here because it generates the potential for divergence between the citizen's preferences and those of agents and drives self-selection.

Second, the punishment utility that we posit arises from causing or facilitating the punishment outcome. Given this, causing the punishment of the guilty and causing the nonpunishment of the innocent are unlikely to be symmetric in the law enforcement context. In particular, it seems untenable that refraining from punishing innocent individuals can produce utility. The problem is that, except for the one or few individuals guilty of a crime, *the entire population* is innocent. And most of these innocent individuals face no risk of being wrongfully punished, though a law enforcement agent with the power to punish can *cause* their nonpunishment by choosing to refrain from taking action against them. Thus, while we regard as plausible a function in which individuals lose utility from causing the punishment of the innocent, it seems implausible to posit that individuals gain utility from not punishing any innocent person. That a police officer would gain utility from not arresting an individual who appears perfectly blameless strikes us as similar to positing that a person gains utility each time he does *not* strike any individual in his presence (or, for that matter, that he does not hit himself in the head).

We do, however, recognize that *n* can be positive in a more limited case: not whenever an innocent person is not punished but whenever an innocent suspect is *exonerated*. We define an exoneration as occurring where, at time 1, the perceived probability p of a suspect's guilt is high enough to make his punishment seem plausible and then, at time 2, someone produces evidence that proves the probability to be far lower, eliminating the chance of punishment. However, by its very nature, the job of law enforcement offers many more opportunities for causing punishment than causing exoneration. A law enforcement officer frequently encounters the opportunity to make an arrest, to gather evidence that might lead others to make an arrest or to convict, or to inflict informal punishment through the process of stopping, searching, and arresting (including the use of unlawful excessive force). Punishment facilitation is a standard part of the job. By contrast, the officer has only a relatively rare occasion to liberate or prevent the detention of an individual who at an earlier point appears guilty but who, after investigation, is apparently not. One reason for this rarity is that, outside of detective fiction, the person who first appears to be guilty usually is. So once an individual's p reaches a point where he is a serious suspect, further investigation usually fails to find evidence that exonerates the suspect. More generally, most crimes are unsolved. An unsolved crime provides no exoneration opportunities, but does offer the opportunity for detecting and apprehending the perpetrator.

Another reason for doubting the importance of pro-justice or pro-exoneration preferences within the police arises from the presence of promotion incentives. While these incentives are low-powered, it is nonetheless likely to be true that officers with better records at closing cases are more likely to receive promotions. Those who spend a lot of time seeking to exonerate suspects will end up with worse-looking records and be less likely to receive promotion. This factor may make policing less attractive to pro-justice types, or make them less likely to persist in a police career.[7]

Given the assumptions in Table 2, a punitive agent wishes to punish suspects iff:

$$p > \frac{\beta}{m+\beta+1} \equiv p^m \qquad (2)$$

[7] We are indebted to an anonymous referee for suggesting this point.

It follows from Equations (1) and (2) that:

Remark 1: i) $m > 0$ implies that $p^m < p^*$

ii) p^m is decreasing in m

Thus, whenever agents are punitive ($m > 0$) there will exist a divergence between the preferences of agents and those of the average citizen. Moreover, this divergence will be greater the more punitive are the agents (i.e. the larger is m). In the analysis that follows, we focus for concreteness on this divergence in preferences (i.e. that punitive agents wish to punish suspects with a lower threshold of doubt) as a source of agency costs in law enforcement. However, a richer framework could (without fundamentally changing our results) incorporate additional forms of divergence in preference – for instance, punitive agents may have a greater taste for authoritarianism or violence.

It is important to emphasize that the punishment utility *m* experienced by punitive agents as assumed in Table 2 is distinct in significant respects from the truth-seeking preferences (u_C) of the citizen and agents. The latter may be termed "utility from enforcement accuracy." This reflects the preferences in Table 1, and is experienced by all individuals as a public good. This type of utility arises simply from knowing that suspects with *p* above some threshold are punished (and that suspects with *p* below the threshold are not punished). It does not depend on personally causing or facilitating this punishment. On the other hand, the punishment utility *m* represents additional utility derived by a punitive type from personally causing or facilitating the punishment of suspects. This is assumed to be a private good that is contingent on serving in a law enforcement capacity. A punitive police officer would experience punishment utility even if (counterfactually) any other police officer in the same position would also have arrested the same suspect in the same circumstances.[8] It follows (when we model the decision to apply to the police force below) that the incremental gain in utility from joining the police is only the punishment utility – the utility from enforcement accuracy is experienced whether or not an individual joins the police. Clearly, both utility from enforcement accuracy and punishment utility – while differing in many respects – are nonpecuniary in nature. In the model below, both the citizen and potential police applicants must trade off these forms of nonpecuniary utility against the wages that are paid or received. We thus assume that the parameter $\gamma > 0$ captures

[8] This assumption can be grounded in a growing body of experimental evidence that suggests that people feel more responsible for actions than for inactions and tend to attribute responsibility to the person who makes the relevant affirmative decision, even if others would have acted similarly (see e.g. Arlen and Tontrup (2014)).

the tradeoff between monetary payoffs on the one hand and nonmonetary payoffs (both utility from enforcement accuracy and punishment utility) on the other.

To characterize more precisely the citizen's utility from enforcement accuracy, suppose that the police follow a rule of punishing suspects iff $p > p*$; then, the citizen's utility from enforcement accuracy is:

$$-\int_0^{p*} pLdp \ -\int_{p*}^{1}(1-p)\beta Ldp = -\frac{p^2}{2}|_0^{p*} - \beta L(p-\frac{p^2}{2})|_{p*}^{1} \quad (3)$$

$$= -\frac{\beta L}{2}\left(1-\frac{\beta}{\beta+1}\right) \equiv u_c^{p*} \quad (4)$$

Thus, u_c^{p*} is the utility from enforcement accuracy that the citizen derives in a scenario where police follow the citizen's preferences. If the police are punitive and are unconstrained by strong criminal procedure protections (CPP), the police will punish suspects whenever $p > p^m$; the citizen's utility from enforcement accuracy is then:

$$-\int_0^{p^m} pLdp \ -\int_{p^m}^{1}(1-p)\beta Ldp = -\frac{p^2}{2}|_0^{p^m} - \beta L(p-\frac{p^2}{2})|_{p^m}^{1} \quad (5)$$

$$= -\frac{\beta L}{2}\left(1-\frac{\lambda\beta}{\beta+1}\right) \equiv u_c^{p^m} \quad (6)$$

where:

$$\lambda = \frac{m(\beta+1)}{(m+\beta+1)^2} < 1 \quad (7)$$

Thus, $u_c^{p^m}$ is the utility from enforcement accuracy that the citizen derives in a scenario where punitive police follow their own preferences rather than those of the citizen. It follows from Equations (5)-(7) that:

Remark 2: $m > 0$ implies that $0 < \lambda < 1$ and hence that $u_c^{p^m} < u_c^{p*}$

The citizen's utility from enforcement accuracy is thus lower – i.e. more negative - when the agent is punitive (and there exist agency costs of excessive zeal).

To characterize the punishment utility received by a punitive police office, consider first a scenario in which a punitive police officer is unconstrained (i.e. can punish suspects whenever $p > p^m$). The utility from enforcement accuracy and from punishment – in combination – can be expressed as:

$$-\int_0^{p^m} pLdp + \int_{p^m}^{1}[pm-(1-p)\beta]Ldp = u_c^{p^m} + \int_{p^m}^{1} pmLdp \quad (8)$$

$$= u_c^{p^m} + \frac{mL}{2}[1-\frac{\beta^2}{(m+\beta+1)^2}] \quad (9)$$

The first term in Equation (9) is the utility $u_c^{p^m}$ from enforcement accuracy. As highlighted previously, this utility would be experienced by this individual even if she were not part of the police (as long as the counterfactual involves another punitive type being in the police). Thus, the incremental utility from serving in the police (ignoring the wage received) is the punishment utility:

$$\frac{mL}{2}[1 - \frac{\beta^2}{(m+\beta+1)^2}] \equiv v_m^{p^m} \qquad (10)$$

Similarly, if a punitive police officer is constrained (e.g. by strong CPP) to only punish suspects when $p > p*$, then her punishment utility is:

$$\int_{p*}^{1} pmLdp = \frac{mL}{2}[1 - \frac{\beta^2}{(\beta+1)^2}] \equiv v_m^{p*} \qquad (11)$$

It follows from Equations (10) and (11) that:

Remark 3: i) If $m = 0$, then $v_m^{p^m} = v_m^{p*} = 0$

ii) $m > 0$ implies that $v_m^{p^m} > v_m^{p*}$

iii) $v_m^{p^m}$ and v_m^{p*} are each increasing in m

Thus, nonpunitive or "neutral" agents (who share the citizen's punishment preferences) experience no punishment utility. This is because their preferences are defined in relation to those of the citizen, and does not necessarily entail that neutral agents have no punishment preferences at all.[9] When $m > 0$, punishment utility is lower when agency costs of excessive zeal are lower. The more punitive a punitive type is, the larger the punishment utility.

3.2) Sequential Game with Weak Criminal Procedure Protections

This subsection describes and solves a simple sequential game that captures elements of the processes of hiring law enforcement agents and the punishment of suspects. In the model, a representative citizen (the principal) offers a contract to police officers. Potential applicants (with punitive and nonpunitive preferences) then decide whether to apply to join the police. Once hired, the police then encounter suspects (with varying probabilities of guilt p) whom they choose to arrest or not. A court verifies each suspect's p, but as CPP is weak the court has no power (apart from this verification function) to alter police decisions about punishment.

[9] Thus, "neutral" means neutral compared to the citizen/principal. If the citizen is punishment-preferring to some degree, as is likely, we will still refer to the agent as punishment-neutral if he is punishment-preferring to exactly the same degree (and the punitive type, for example, will then consist of those who are *more* punishment-preferring than the citizen).

In the first stage of this game, it is assumed that the contract offered by the citizen consists simply of a wage *w*, and is not conditioned on performance or other outcome measures. As the standard solution for an agency problem is a contract that creates incentives that align the interests of the agent with the interests of the principal, the assumption that this is impossible or impractical in the context of policing requires some explanation. While theoretically optimal contracts are frequently more complex than the contracts we actually observe in practice, they usually bear some structural similarity. By contrast, the contracts we observe for police are nowhere close to what would be necessary to solve the agency problem. Police are *not* paid for performance; they are not given bounties or paid a piece-rate for each legitimate search, arrest, or conviction. Their compensation is not tied to the crime rate in their area. Instead, police are paid a wage or salary, creating only very low powered incentives. Even those low powered incentives assume that an officer might be terminated for poor performance, but the evidence suggests that police officers, many of whom are unionized, face only a weak threat of being fired for poor performance (except during their initial probationary period). While police bureaucracies may reward good performance with coveted assignments and promotions, this incentive is sufficiently noisy as to create only weak incentives as well.

The standard explanation for using only weak incentives for police is the danger of fabrication – that if law enforcers were paid by the arrest or conviction, they would "frame" individuals to collect their fee. We interpret Juvenal's famous query, "who will watch the watchers?," as reflecting, among other things, the difficulty of preventing such fabrication, given the control that the watchers have over the relevant information.[10] There is also the separate, standard concern (Holmstrom and Milgrom, 1991) that high powered incentives tied to one index of job performance (e.g., arrests) will inefficiently cause agents to ignore other less observable areas of job performance (e.g., public safety).[11]

[10] Even if this problem were not inevitable, the fact that American jurisdictions have not paid police in this manner for many decades has allowed legislatures to enact extremely broad laws on the assumption that the police will use discretion to enforce the law only against a few violators when other factors justify enforcement. Given the current breadth of criminal statutes, including what many commentators refer to as "overcriminalization," e.g., Dillonns, 2012; Podgor, 2005, the use of high powered incentives would cause havoc by motivating police to enforce minor offenses to the maximum, not just the literal enforcement of widely violated traffic regulations, but also crimes involving the regulation of noise, littering, copying of copyrighted works, trivial thefts (e.g., a pen taken home from work), or minor assaults justified by social norms (e.g., a gentle tap in response to rude behavior).

[11] Holmstrom and Milgrom (1991) show that in multitask principal-agent settings, it may well be optimal for the principal to use low-powered incentive structures in order to avoid a substitution of effort from measurable to less

Alternatively, the police might affect the crime rate by successful deployments that prevent crime rather than only by arrests after crime occurs. For that reason, one might want the contract to pay police by the amount of crime in their vicinity. There are a variety of problems with this approach, but we will discuss only two. Most obvious is the loose correlation between what an officer or a precinct does and the local crime rate, given the other variables affecting crime: economic and demographic fluctuations, cultural and technological change, and the decisions of government actors in other domains, such as education, housing, and the economy. Even where enforcement is the key explanatory variable, the local police share responsibility with state and federal enforcement agents, as well as each government's legislative decisions over funding, the federal and state judiciary's criminal law and procedure decisions, and the federal and local prosecutor. Crime control is a complicated type of "team production," where the decisions of other agents may swamp the effects of good or bad policing, thus muting the effect of high powered incentives.

The second problem is police manipulation of crime rate data. Where paying police by arrest encourages police to overstate crime so they can make more arrests, paying police by the crime rate encourages them to understate crime, so they appear to being doing better. Even with salaried police, there are media reports (e.g. Bernstein and Isackson, 2014; Rashbaum, 2010) of this kind of manipulation, where police discourage citizens from reporting crimes or recharacterize serious crimes as being less serious (as by understating the value of the object stolen). Perhaps the public could use victim surveys instead of reports, but victim surveys don't work for many important crimes: murder, corruption, illegal sale of drugs or weapons, etc.

Perhaps none of these points fully justifies the failure to compensate police in a way that creates high powered incentives for them to act in the principal's interest. But if so, it is a puzzle that we only observe low powered incentives. A strength of our approach is that it helps to resolve the puzzle. We claim that society manages to attract into policing those who are intrinsically motivated to perform the job, thus rendering external incentives less necessary. The presence of internal incentives makes the trade-off between low and high powered incentives more likely to favor low powered incentives.

measurable tasks. One of their examples focuses on schoolteachers, among whom a compensation structure that focuses only on student test scores may induce a neglect of teaching creativity and other nonmeasurable skills.

In the second stage of the game, individuals observe the contract offered in stage 1, and decide whether to apply to join the police. We normalize the size of the police force to 1 (thus, the wage w may be interpreted as both the wage received by the police officer and as the total wage cost incurred by the principal). We assume that there exists a continuum of punitive types (with m > 0) and nonpunitive types (with m = 0) in the population. The assumption that there is a continuum of each type of individual entails that (when either or both types of individual chooses to apply) each applicant is "small" in relation to the applicant pool. In particular, the decision by an individual to apply or not does not significantly affect the composition (in terms of punitiveness) of the applicant pool. The fraction of punitive types in the population is assumed to be ρ, and the tiebreaking assumption that individuals apply when indifferent is imposed. The reservation wage is assumed to be equal for punitive and nonpunitive types, and is denoted by w^R.[12] The citizen is assumed to be unable to observe whether an applicant is punitive or not, and chooses the police office at random from among those who choose to apply.

The citizen is assumed to have nonpunitive preferences (i.e. m = 0), with the following objective function:

$$\gamma u_C - w \tag{12}$$

where u_C is the citizen's utility from enforcement accuracy, w is the wage paid to the police officer, and $\gamma > 0$ is a parameter that represents the tradeoff between monetary and nonmonetary payoffs. In choosing whether or not to apply to join the police, potential applicants seek to maximize:

$$\gamma v_m + w \tag{13}$$

Here, v_m is the utility from punishment (which is zero for nonpunitive potential applicants).

It is important to emphasize that while potential applicants experience utility from enforcement accuracy (along with the citizen), this does not play any role in their decision to apply or not. As noted above, utility from enforcement accuracy is a public good, the utility from which does not depend on service as a police officer. Moreover, the assumption above that each applicant is "small" in relation to the applicant pool entails that an individual's application decision does not change expected police behavior. For instance, in an equilibrium in which only

[12] This assumption is made for simplicity and because there is no compelling reason to believe that one or the other type would have a higher reservation wage. It is possible to generalize this assumption by allowing for different reservation ages for each type, but this would not fundamentally alter the results, while making the notation more cumbersome.

punitive types apply, the counterfactual scenario for any individual punitive applicant is that if he were not to apply then some other punitive individual would be the police officer; in an equilibrium in which both types apply, the counterfactual scenario for any individual punitive applicant is that if he were not to apply, the probability of the police officer being punitive remains ρ. Note also that the police officer's effort in finding suspects is assumed to be exogenous, in order to focus attention on agency costs of excessive zeal. However, if we were to introduce a choice of effort, it is reasonable to expect that punitive officers would exert more effort and this would merely reinforce their value from the perspective of the citizen.

The sequence of decisions can be summarized as follows (see also Figure 1):

1) The citizen chooses a contract consisting of a wage *w* to offer to the police officer; in so doing, the citizen maximizes Equation (12)

2) Individuals observe the contract offered in stage 1, and decide whether to apply to join the police; in so doing, they maximize Equation (13). The police officer is chosen randomly from among those who choose to apply.

3) The police officer hired in stage 2 encounters a continuum of suspects with varying probability of guilt *p*. The police officer decides which suspects to arrest.

4) The court observes *p* for arrested suspects. It has no power to overturn the arrest decision in stage 3; suspects arrested in stage 3 are punished.

Note that the assumption that the court can observe and verify *p* entails that police cannot credibly fabricate evidence. Under our assumptions, a given suspect's *p* is simply exogenous. If fabrication were possible, then a reasonable presumption would be that the court (even under weak CPP) would have the power to overturn the arrest decision in stage 3 and free the suspect. We do not allow fabrication here because that would distract from the central issues that the model is intended to address. The possibility of fabrication would arguably be more relevant in a setting with performance-based compensation for police, a type of contract that we rule out here for the various conceptual and practical reasons discussed earlier. In reality, fabrication may remain a possibility even in the absence of performance-based compensation. However, some degree of verifiability by courts of police allegations against suspects is required for criminal procedure provisions to have any impact at all. Moreover, the possibility of fabrication may generally be expected to make intrinsically motivated punitive types more (rather than less) attractive as police officers, as they would be more intrinsically motivated to identify the guilty rather than to fabricate evidence against the innocent.

The exogenous nature of p also abstracts from the effort involved in finding evidence against suspects. In this respect, we depart from the framework of Aghion and Tirole (1997), which models asymmetric information between the principal and agent and analyzes the agent's investment of effort in acquiring information. This simplification enables us to focus on our paper's central issues, which relate to the choice of punishment given a known probability of guilt and to the role of intrinsic motivation. However, if we were to introduce a choice of effort by the officer in learning p,[13] then it would be reasonable to expect that punitive officers (who have an intrinsic reason to identify the guilty) would exert more effort along this dimension than would nonpunitive officers. This would make the employment of punitive officers even more attractive than under our assumption of an exogenous p, and in that sense would reinforce the basic conclusion that the citizen will in many circumstances wish to employ punitive officers notwithstanding the divergence in preferences over punishment.

The game described above can be solved using backwards induction. The final stage is quite trivial, as the court has no significant power by assumption. Note, however, that its verification of p ensures that the police cannot fabricate evidence. In the third stage, a punitive officer will arrest iff $p > p^m$ (as derived earlier – see Equation (2)), while a nonpunitive officer will arrest iff $p > p*$ (as derived earlier – see Equation (1)). In the second stage, potential applicants will compare the (weighted) sum of the wage and their punishment utility (if any) with their reservation wage w^R in deciding whether to apply. A punitive type will apply iff $w \geq w^R - \gamma v_m^{p^m}$, while a nonpunitive type will apply iff $w \geq w^R$.

In the first stage, the citizen has one of two undominated strategies, depending on the values of the relevant parameters (as specified below). The first (strategy I) is to offer a wage $w = w^R - \gamma v_m^{p^m}$. This induces only punitive types to apply. The police officer who is randomly chosen is therefore necessarily a punitive type, and uses a $p > p^m$ cutoff for punishing suspects. This leads to a payoff for the citizen of:

[13] Note that in addition to the police officer's effort in learning p, the court may also expend effort and cost at the trial stage in determining the probability of guilt. In doing so, the court may have a cost advantage in that it must expend this effort only for those cases where there is an arrest (and no plea bargain), whereas the police must do so for all potential suspects to determine which of them to arrest (by analogy with Shavell's (2013) framework with respect to the enforcement differences between regulation and the negligence rule).

$$- w^R + \gamma(v_m^{p^m} + u_c^{p^m}) \quad (14)$$

The second strategy (Strategy II) is to offer a wage $w = w^R$. This induces both punitive and nonpunitive types to apply. Thus, the randomly chosen police officer is punitive with probability ρ (and uses a $p > p^m$ cutoff for punishing suspects) and nonpunitive with probability $1 - \rho$ (and uses a $p > p*$ cutoff for punishing suspects). This leads to an expected payoff for the citizen of:

$$- w^R + \gamma[(1-\rho)u_c^{p*} + \rho u_c^{p^m}] \quad (15)$$

It is important to note that in Strategy II, the same wage (w^R) is paid to the police officer, regardless of whether the randomly-chosen officer who is hired happens to punitive or nonpunitive. Of course, while the officer's type is unobservable ex ante, it can eventually be inferred ex post from her arrest behavior. However, we assume that the wage offer made in stage 1 is credible and binding on the citizen, and is not revised ex post. It follows that if a punitive officer is chosen under Strategy II, she derives rents in the sense that the wage plus her punishment utility strictly exceeds her reservation wage. This might be viewed as an informational rent in that it arises because punitiveness is unobservable to the citizen.

Whether Strategy I is better than Strategy II for the citizen depends on the following condition:

Condition 1: $(1-\rho) < \frac{v_m^{p^m}}{u_c^{p*} - u_c^{p^m}}$

Intuitively, the left-hand-side represents the fraction of nonpunitive types in the applicant pool. The right-hand-side represents the ratio of the punitive type's punishment utility (when unconstrained by strong CPP) to the loss of utility from enforcement accuracy caused by (unconstrained) punitive police. Note that Condition 1 is more likely to be satisfied when there is a larger fraction of punitive types in the population. This is evident when Condition 1 is expressed purely in terms of primitive parameters, as follows:

$$(m + \beta + 1)^2 \left(m - \frac{(1-\rho)\beta^2}{\beta+1}\right) - \rho m \beta^2 > 0 \quad (16)$$

The derivative of the LHS of Equation (16) with respect to ρ can be expressed as:

$$\frac{\partial(LHS)}{\partial\rho} = \frac{m^2}{\beta+1} + m + \beta + 1 > 0 \quad (17)$$

Thus, an increase in the fraction of punitive types in the population unambiguously makes Condition 1 more likely to hold. However, the impact of other parameters on whether Condition 1 is satisfied is ambiguous. For example, a higher m increases the wage savings that the citizen can achieve by hiring a punitive officer, but it also increases the policy distortion (i.e. the divergence of preferences between the citizen and the officer) and so reduces the citizen's utility from enforcement accuracy.

When Condition 1 is satisfied, the citizen will choose Strategy I; otherwise, the citizen will choose Strategy II. The equilibrium outcomes of the game can thus be characterized as follows:

Proposition 1: If Condition 1 holds, then the equilibrium outcome is:

1) The citizen sets $w = w^R$ - $\gamma v_m^{p^m}$

2) Only punitive types apply to join the police

3) The police officer arrests suspects iff $p > p^m$

4) The court verifies p; suspects with $p > p^m$ are punished

Proof: Suppose that Condition 1 holds, and consider any wage $w > w^R$ - $\gamma v_m^{p^m}$ and $w < w^R$. For any wage in this range, it is clear that nonpunitive types will not apply (as their payoff is w and so their reservation wage w^R exceeds the payoff from joining the police). Thus, the police officer will be punitive and offering the higher wage merely creates rents for the punitive officer, with no change in police behavior. Now consider $w = w^R$. This induces both punitive and nonpunitive types to apply, but the citizen is worse off with this outcome than with Strategy I. Rearranging Condition 1 yields:

$$v_m^{p^m} + u_c^{p^m} > (1-\rho)u_c^{p^*} + \rho u_c^{p^m}$$

By comparing this expression to Equations (14) and (15), it is clear that the citizen is worse off with this outcome than with Strategy I. Any wage $w > w^R$ is clearly dominated. Thus, it follows that under Condition 1 the citizen's optimal strategy is to set $w = w^R$ - $\gamma v_m^{p^m}$. The other results follow straightforwardly from the backwards induction argument in the text.

Proposition 2: If Condition 1 does not hold, then the equilibrium outcome is:

1) The citizen sets $w = w^R$

2) Both punitive and nonpunitive types apply to join the police

3) If the randomly chosen police officer is punitive, she arrests suspects iff $p > p^m$; if the randomly chosen police officer is nonpunitive, she arrests suspects iff $p > p^*$

4) The court verifies p. Suspects with $p > p^m$ are punished if the randomly chosen police officer is punitive; Suspects with $p > p^*$ are punished if the randomly chosen police officer is nonpunitive.

Proof: Suppose that Condition 1 does not hold, and that the citizen were to reduce the wage slightly below w^R. Then, the nonpunitive types will not apply, and the police officer will be punitive. Given that Condition 1 does not hold, it follows from Equations (14)-(17) that the citizen is worse off than by setting $w = w^R$. Any wage $w > w^R$ is clearly dominated. Thus, it follows that when Condition 1 does not hold, the citizen's optimal strategy is to set $w = w^R$. The other results follow straightforwardly from the backwards induction argument in the text.

Thus, the simple game that we have specified and analyzed in the subsection has two equilibrium outcomes, depending on parameter values (the equilibrium is unique, however, for any given set of parameter values). When Condition 1 holds, there is a self-selection equilibrium, in which the citizen sets a relatively low wage and attracts punitive types into the police. The savings in wage costs outweigh the reduced utility from enforcement accuracy (under Condition 1). When Condition 1 does not hold, the equilibrium involves both types being represented in the police, with higher wage costs for the citizen, but a greater alignment of preferences between the principal and the agent, and lower agency costs of excessive zeal.

3.3) Sequential Game with Strong Criminal Procedure Protections

In this subsection, we modify the game in Section 3.2 by adding strong criminal procedure protections, specifically an expanded role for the court in determining which suspects are to be punished. The first three stages of the game are identical to those in the previous model. In the fourth stage, the court observes p for arrested suspects (as before). Now, however, it decides which suspects will be punished; for example, it can decide whether to convict or acquit the suspect. Admittedly, this is a highly stylized and simplified representation of criminal procedure protections that ignores many nuances and complexities. However, it serves to capture in very simple form the idea of providing enhanced procedural rights for suspects. We solve this modified game (depicted in Figure 2) by backwards induction and characterize the equilibrium below.

The central assumption of this modified game is that the court shares the preferences of the citizen. As this assumption plays a key role, it is worth explaining in detail. As argued in the previous subsection, the citizen in our framework faces a tradeoff between the agency costs of excessive police zeal and the wage costs of hiring less zealous police. This dilemma can be made less sharp if the public can establish institutional structures that entail other agents "policing" the excessive zeal of the police. This requires, of course, that the public has the ability to select agents who have punishment preferences closer to their own than are those of the police. The approach on which we focus is to provide stronger procedural protections for suspects, thereby in effect empowering courts to play a greater role in deciding whether suspects are punished.

In order to achieve this, it is crucial that courts have preferences that are closer to those of the representative citizen than are those of punitive police. This may be arguably true of judges who are directly elected, as judicial elections would be expected to produce judges with punishment preferences identical to those of the public. Judicial appointment might do the same for two reasons. First, the prestige of being a judge is so high, especially in the highest appellate courts that monitor lower courts, that few lawyers turn down the opportunity, which leaves little opportunity for self-selection. Second, those who are appointed tend to be successful lawyers from a variety of fields, not just former prosecutors. All of this is consistent with the common observation that judges are less pro-prosecution than are police, which makes plausible our assumption that judges are closer to being punishment-neutral, i.e., the same as the public. A second institution is the jury. Suppose either that the legal system coerces a broad cross-section of citizens to serve as jurors or that civic virtue dominates as the motivation for jury service. In either case, the jury might be punishment-neutral, that is, its members might on average have the same punishment preferences as the median member of the public.[14, 15]

In the final stage of the game, a court that shares the citizen's punishment preferences will choose to punish suspects iff $p > p^*$ (see Equation (1) above). In the third stage, the decisions of the police officer regarding arrests will now be influenced by the anticipation of the court's choice – i.e. will be made in the shadow of the court's preferences. We assume that the

[14] Moreover, even if courts (i.e. judges and juries) are on average as punitive as the police, the former may still exercise a restraining influence on the latter as long as there is heterogeneity among judges and among juries. On those occasions when the court is less punitive than the police, it will constrain the set of suspects who are punished. On the occasions when the court is more punitive than the police, the court cannot easily cause the police to arrest additional suspects, and so the decisions of the (less punitive) police as to which suspects to arrest will stand. Note that this scenario is outside the scope of our model, in which there is no heterogeneity among judges or among juries.

[15] It is true that few cases go to trial, which may seem to reduce the relevance of the preferences of judges and juries. However, it is reasonable to assume that plea bargains are made in the shadow of the outcomes that would result at trial. Thus, police behavior is likely to reflect the anticipated decisions of the court, even if the probability of any given case going to trial is very small.

police officer faces a small cost of arresting a suspect who is subsequently acquitted. Equivalently, we could assume this cost is zero, but impose a tiebreaking assumption that the officer does not arrest if indifferent.[16] Then, the police office (whether punitive or not) will arrest iff $p > p*$. In the second stage, a punitive type (anticipating that she will be constrained by strong CPP if she joins the police) will apply iff $w \geq w^R$ - γv_m^{p*} (i.e. the punitive type now requires a higher wage in order to apply). As before, a nonpunitive type will apply iff $w \geq w^R$.

In the first stage of the game, the citizen now has only one undominated strategy, which is to offer a wage $w = w^R$ - γv_m^{p*}. Intuitively, there is no longer any value to attracting nonpunitive types into policing by offering a higher wage. Because police are constrained by strong CPP via the court, police behavior is identical under our assumptions whether the police officer is punitive or nonpunitive. The wage $w = w^R$ - γv_m^{p*} induces only punitive types to apply, and leads to a payoff for the citizen of:

$$- w^R + \gamma\left(v_m^{p*} + u_c^{p*}\right) \tag{18}$$

Thus, the equilibrium can be characterized as follows:

Proposition 3: With strong CPP, the equilibrium outcome is:

1) The citizen sets $w = w^R$ - γv_m^{p*}

2) Only punitive types apply to join the police

3) The police officer arrests suspects iff $p\ >\ p^*$

4) The court verifies p; suspects with $p\ >\ p^*$ are punished

Proof: Consider any wage $w > w^R$ - $\gamma v_m^{p^m}$ and $w < w^R$. For any wage in this range, it is clear that nonpunitive types will not apply (as their payoff is w and so their reservation wage w^R exceeds the payoff from joining the police). Thus, the police officer will be punitive and offering the higher wage merely creates rents for the punitive officer, with no change in police behavior. Now consider $w = w^R$. This induces both punitive and nonpunitive types to apply, but even if the police officer ends up being nonpunitive, police behavior is identical as any officer will arrest suspects iff $p\ >\ p^*$ (anticipating the court's choice). Any wage $w > w^R$ is clearly dominated. Thus, it follows that the citizen's optimal strategy is to set $w = w^R$ - γv_m^{p*}. The other results follow straightforwardly from the backwards induction argument in the text.

[16] It is in theory possible that this cost may be negative, in the sense that the officer derives a benefit from arresting a suspect who is expected to be freed, perhaps because the process of arrest itself serves as a form of punishment. This possibility could be incorporated by assuming two different kinds of punishment – informal punishment (including the arrest itself) imposed by the police, and formal punishment imposed by the court. Strong CPP would affect the latter but not necessarily the former (although requiring warrants for arrests and imposing constraints on excessive force may affect the former). This extended model would lead to substantially similar conclusions, although the gains to the citizen from using strong CPP would be somewhat reduced.

Thus, when strong CPP for suspects are added to our simple game, the only possible equilibrium outcome involves self-selection by punitive types into the police force. Intuitively, when procedural protections limit the agency costs of excessive zeal, there is no reason for the citizen to seek to attract nonpunitive police through a higher wage. Of course, this is a simplification – in reality, there would be multiple types of agency costs of excessive zeal, and some such costs – e.g. the excessive use of force against clearly guilty suspects – may not be fully constrained by the court. Nonetheless, strong CPP at least reduces the value of attracting nonpunitive types. The wage in this equilibrium is higher than in the self-selection equilibrium in Section 3.2, because strong criminal procedure protections limit punishment opportunities. However, given the equilibrium wage, the job remains more attractive to punitive than to nonpunitive types as long as punishment opportunities continue to exist. With strong CPP, however, the court limits the agency costs of excessive zeal, thereby reducing the extent to which the principal's payoff is reduced by the excessive punishment of suspects.

3.4) The Choice between Weak and Strong Criminal Procedure Protections

The two previous subsections have characterized the equilibrium outcomes under both weak and strong CPP. The question that arises naturally from this is what type of CPP the citizen would prefer. Suppose that we envisage a scenario in which the citizen chooses the rules of criminal procedure at a prior "constitutional" stage of decisionmaking, anticipating the equilibrium outcomes that would occur in the wake of this choice. Then, holding everything else equal, the citizen will choose strong CPP under the following circumstances:

Proposition 4: If the representative citizen is able to choose the strength of CPP, she will impose strong CPP iff:

i) Condition 1 holds and:

$$v_m^{p*} + u_c^{p*} > v_m^{p^m} + u_c^{p^m}$$

or,

ii) Condition 1 does not hold

Proof: i) If Condition 1 holds, then under weak CPP the citizen will choose Strategy I and obtain the payoff in Equation (14). Under strong CPP, the citizen will obtain the payoff in Equation (18). Thus, strong CPP will be chosen when

$$- w^R + \gamma\left(v_m^{p*} + u_c^{p*}\right) > - w^R + \gamma(v_m^{p^m} + u_c^{p^m})$$

i.e. when $$v_m^{p*} + u_c^{p*} > v_m^{p^m} + u_c^{p^m}$$

ii) If Condition 1 does not hold, then under weak CPP the citizen will choose Strategy II and obtain the payoff in Equation (15). Strong CPP will thus be preferred when:

$$- w^R + \gamma\left(v_m^{p*} + u_c^{p*}\right) > - w^R + \gamma[(1-\rho)u_c^{p*} + \rho u_c^{p^m}]$$

i.e. when $$v_m^{p*} > \rho\left(u_c^{p^m} - u_c^{p*}\right)$$

This is always satisfied as the RHS < 0 and the punishment utility v_m^{p*} is positive by assumption. Thus, strong CPP is always optimal when Condition 1 does not hold.

If Condition 1 does not hold (so that the equilibrium strategy under weak CPP is Strategy II), then strong CPP will always be preferred – the outcome under strong CPP reduces wage costs while also increasing the citizen's utility from enforcement accuracy. If Condition 1 holds (so that the equilibrium strategy under weak CPP is Strategy I), then whether strong CPP is optimal depends on a comparison between a punitive officer's loss of punishment utility from strong CPP (i.e. $v_m^{p^m} - v_m^{p*}$) with the citizen's gain in utility from enforcement accuracy (i.e. $u_c^{p*} - u_c^{p^m}$). This can be viewed as being analogous to the efficiency condition that would be derived from a pure utilitarian social welfare function – when the citizen's gains exceed the punitive officer's loss, the citizen will choose strong CPP.

In the circumstances characterized in Proposition 4, the citizen-principal benefits by seeking to control punitive police by using other agents – mostly, judges and juries – who have preferences closer to those of the public. Strong CPP can thus reduce agency costs between the general public and police; they enable the public to benefit to some degree from the lower cost of employing intrinsically-motivated police, while avoiding the worst excesses of over-enforcement by these agents.

4) Discussion

As discussed previously, our framework potentially derives the basic structure of the criminal justice system – the separation of judicial and executive enforcement powers, and the judiciary’s pro-defendant rules of criminal procedure – from the agency problem in law enforcement. In this section, we identify some further important implications of our model of intrinsically motivated police. The model does not explicitly include a choice of effort (for instance, in finding suspects or in investing in determining the probability of guilt *p*). However, a straightforward extension of the model to encompass this choice would be expected to yield the result that under reasonable conditions the punitive type will exert greater effort, as this generates more punishment opportunities. Thus, our general framework would predict less shirking than does the standard assumption of ordinary consumer preferences, which we contend is more consistent with the qualitative evidence of police behavior.

Earlier, we noted that intrinsic motivation helps to explain the fact that modern governments usually choose to motivate law enforcers extrinsically with only low powered incentives. Internal incentives make external incentives less necessary, so the trade-off (given the risk that strong external incentives cause police to fabricate evidence and divert police from unmeasured tasks not rewarded with high powered incentives) favors weaker external incentives. A related question is the following: given these low powered incentives, why there isn’t more shirking among police than there is? We answer the question with the same point: that those who select into policing are intrinsically motivated to perform the job.

Because the state uses only low-powered incentives to motivate most of its bureaucrats, Wilson (1989:56) noted that it “is surprising that bureaucrats work at all.” With police however, a variety of factors could reconcile economic theory with observed behavior: (1) some external employment incentives (e.g., the opportunity for promotion and attractive assignments) motivate *some* work; (2) group norm enforcement might motivate work if police precincts have pro-work norms; (3) police officers might work to gather evidence of crime merely to put themselves in a position to receive bribes from the guilty for not arresting them; and (4) some evidence suggests that some police officers actually do shirk quite a bit. (e.g., Mastrofski et al. 1994; Walsh 1986).

While some police are corrupt and some shirk, the evidence suggests that many police officers do not take bribes and work more than they shirk (e.g., Brehm & Gates 1997), making about 14 million arrests per year (FBI, 2006). Some econometric analysis supports the idea that adding officers to a police force decreases crime (e.g. Levitt, 2002; Vollaard and Hamed, 2012), which seems unlikely without effort. Nonetheless, these facts do not necessarily demonstrate that police work beyond the level explained by weak external incentives. It is difficult to say how much effort those incentives predict.

What is puzzling is the fact that police perform their duties even in situations of great personal risk. In 2009, for example, 48 law enforcement officers were feloniously killed in the line of duty, another 47 were accidentally killed, and 57,268 were assaulted (of which, over a quarter sustained injuries).[17] The largest category of those assaulted – about one-third – were responding to a disturbance call, such as a bar fight or domestic quarrel. Almost half of those killed by accident are in a car crash, many resulting from a high speed police chase of a suspect.[18] Even if low power incentives are sufficient to induce law enforcement officers to make millions of stops, searches, and arrests each year, it remains puzzling why officers would endanger their lives via risky driving behavior and the prompt answering of calls about violent disturbances. Neither the low powered incentives nor the possibility of receiving bribes from the guilty seems worth these risks. One might posit that group camaraderie and norms of professionalism explain these effort levels. But those forces merely imply that individual officers from a unit will tend to work or shirk at equal rates. The same forces could cause officers to *increase* shirking, allowing the group to gain the maximum benefit from the job. So we need a more basic reason why policing norms encourage work as much as they do, given only low-powered incentives.[19]

We propose that self-selection produces a police force with many intrinsically motivated individuals. Low monetary wages plus the opportunity to punish wrongdoers ensures that those attracted to policing are among those most strongly motivated to punish. The degree of shirking is far less than it would be in the absence of the self-selection of intrinsically motivated individuals. Thus, we think that our agency model of policing captures some essential features of law enforcement that are otherwise missing from the literature.

Our framework also has implications for bribery. The conventional recommendation is to pay more (Becker and Stigler, 1974, p. 6). They specifically recommend an "entrance fee" – that

[17] See FBI Uniform Crime Reports, Law Enforcement Officers Killed and Assaulted 2009, at http://www.fbi.gov/about-us/cjis/ucr/leoka/2009 (last accessed 13 June 2011).

[18] See NHTSA, Characteristics of Law Enforcement Officers' Fatalities in Motor Vehicle Crashes, January 2011, at http://www-nrd.nhtsa.dot.gov/Pubs/811411.pdf (last access 13 June 2011).

[19] We do not discount the importance of organizational culture. Although it does not ultimately explain the choice between a norm of working and a norm of shirking, informal sanctions can turn a small bias into a large one. If most police are punitive, they may make life hard for police who are of other types. So even a small initial tendency towards self-selection by punitive types can become reinforced to the point of becoming dominant.

upon taking an enforcement job, a person posts a bond that will be paid back when the person leaves the job, if they were not corrupt. While such bonds are rarely posted in this explicit form (perhaps due to wealth constraints among potential enforcers), pensions and other forms of deferred compensation may play a similar role in inducing honesty and effort.[20] However, the existence of intrinsic motivation can complement such mechanisms, and also suggests that in some circumstances the conventional prescription of paying a higher wage may *increase* bribery. Offsetting the conventional point, low pay disproportionately attracts those with intrinsic motivation, who are harder to bribe. The preferences make the person's compensation depend in part on doing his job, even if failing to do the job is not detected.[21]

In a related paper (McAdams, Dharmapala, and Garoupa, 2015), we develop some implications of our framework for understanding doctrinal distinctions between police and nonpolice government actors that have been drawn in US Fourth Amendment jurisprudence. The Supreme Court has sometimes demanded more justification for searches by police officers than for the same searches conducted by government agents other than police. For example, teachers and government employers are sometimes allowed to conduct searches without a warrant when a police officer would require a warrant.[22] This distinction is a puzzle because the text of the Fourth Amendment does not distinguish police and because the social value of police searches aimed at crime control is often higher than the value of non-police searches aimed at other purposes. We explain the puzzle by the fact that police are likely to differ systematically in their preferences, to be more punitive than most other government agents, and therefore to have lower thresholds of doubt for justifying searches, requiring greater judicial scrutiny.

We also explain the related puzzle that the Supreme Court is more permissive of police searches motivated by purposes other than law enforcement, as where the police engage in

[20] This point is analogous in some respects to Lazear's (1981) explanation of rising age-earnings profiles for workers. But as it is difficult to terminate poorly performing police, it is difficult to terminate the pension rights of police. See, e.g., *People ex rel. Madigan v. Burge*, 18 N.E.3d 14 (Illinois 2014) (dismissing on jurisdictional grounds a state attorney general lawsuit seeking to terminate pension of retired police officer convicted of perjury and obstruction of justice for lying about his use of torture on suspects interrogated in police custody).

[21] It might be thought that bribes extracted from the guilty represent a form of informal punishment that would be valued by punitive officers. However, the nature of bribery entails that the bribe must necessarily be less burdensome to the offender than would the formal punishment for the offense. Wealth constraints would make some bribes far lower than the cost the offender incurs from incarceration imposed after conviction. This implies that bribery would be less attractive (relative to arresting suspects) to punitive police than they would be to nonpunitive police.

[22] Although eventually abandoned, the Court also once drew distinctions between police and non-police personnel in the application of the good faith exception to the exclusionary rule and in the warrant requirement for administrative searches of homes. See McAdams, Dharmapala, and Garoupa (2015).

"community caretaking" by entering a home believing someone inside requires emergency medical assistance. Punitive preferences drive police to be overzealous of the punitive parts of the job only; there is no necessary divergence between in the preferences of police and citizens regarding the performance of the non-punitive aspects of the job. Indeed, if the time spent on community caretaking is an opportunity cost to punitive police who might prefer to be apprehending criminals, such police will be more likely to shirk at these parts of the job. Thus, in this domain, there is no need for extra judicial scrutiny and greater costs to such scrutiny. In sum, the theory of punitive preferences helps to explain and justify otherwise puzzling Fourth Amendment doctrine.

5) Conclusion

Agency problems are pervasive in criminal law enforcement, yet there has been little analysis of the principal-agent issues in law enforcement in the law and economics literature. In this paper, we begin the task of filling this gap in the scholarly literature. We examine self-selection into law enforcement jobs by intrinsically motivated agents. In identifying the intrinsic motivations that are most likely to be prevalent in this context, we draw on the lessons of the experimental literature on altruistic punishment. Our model identifies circumstances in which "punitive" individuals (with stronger-than-average punishment preferences) will self-select into police jobs that offer the opportunity to punish (or facilitate the punishment of) wrongdoers. We identify both costs and benefits of this type of intrinsic motivation. "Punitive" agents will accept a lower salary and be less likely to shirk, but create agency costs of excessive zeal in searching, seizing, and punishing suspects. Under a reasonable set of assumptions, the public chooses to hire punitive police agents, while submitting them to monitoring by other agents (such as the judiciary) with average punishment preferences. Thus, our analysis sheds new light on the perennial question: *Quis custodiet ipsos custodes?*

References

Aghion, Phillipe, and Jean Tirole. 1997. Formal and Real Authority in Organizations. *Journal of Political Economy* 105: 1-29.

Anderson, Christopher M., and Louis Putterman. 2006. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior* 54: 1–24.

Arlen, J., and S. W. Tontrup. 2014. Does the Endowment Effect Justify Legal Intervention? The Debiasing Effect of Institutions. NYU Law and Economics Research Paper No. 14-18.

Becker, Gary S. 1968. Crime and Punishment: an Economic Approach. *Journal of Political Economy* 76: 169-217.

Becker, Gary S., and George J. Stigler. 1974. Law Enforcement, Malfeasance, and Compensation in Enforcers. *Journal of Legal Studies* 3:1.

Bernstein, David, and Noah Isackson, The Truth About Chicago's Crime Rates: The city's drop in crime has been nothing short of miraculous. Here's what's behind the unbelievable numbers, *Chicago Magazine* (April 7, 2014).

Besley, Timothy, and Maitreesh Ghatak. 2005. Competition and Incentives with Motivated Agents. *American Economic Review*, 95: 616-636.

Bibas, Stephanos. 2009. Rewarding Prosecutors for Performance, *Ohio State Journal of Criminal Law* 6:441.

Blackstone, William. 1765-69. *Commentaries on the Laws of England.* Oxford.

Bowles, Roger, and Nuno Garoupa. 1997. Casual police corruption and the economics of crime. *International Review of Law & Economics* 17: 75-87.

Brehm, John, and Scott Gates (1997), *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor: Michigan University Press.

Bubb, Ryan, and Patrick L. Warren. 2014. Optimal Agency Bias and Regulatory Review. *Journal of Legal Studies* 43: 95-135.

Carpenter, Jeffrey P. 2007. The Demand for Punishment. *Journal of Economic Behavior & Organization* 62: 522–542.

de Quervain, Dominique J.-F., et al. 2004. The Neural Basis of Altruistic Punishment. *Science* 305: 1254-58.

Dickinson, David L., David Masclet, and Marie Claire Villeval. 2014. Norm Enforcement in Social Dilemmas: An Experiment with Police Commissioners (GATE Working Paper 1416), archived at http://perma.cc/D242-9MGJ.

Dharmapala, Dhammika, and Richard H. McAdams. 2003. The Condorcet Jury Theorem and the Expressive Function of Law: A Theory of Informative Law. *American Law and Economics Review*, 5, 1-31.

Dillonns, Zach. 2012. Foreward to Symposium on Overcriminalization, 102 *Journal of Criminal Law and Criminology* 102: 525.

Easterbrook, Frank H. 1983. Criminal Procedure as a Market System. *Journal of Legal Studies* 12: 289-332.

Easterbrook, Frank H., and Daniel R. Fischel. 1991, *The Economic Structure of Corporate Law* Cambridge, Mass.: Harvard University Press.

Echazu, Luciana, and Nuno Garoupa. 2010. Corruption and the Distortion of Law Enforcement Effort. *American Law & Economics Review* 12: 162-80.

Federal Bureau of Investigation. 2006. Crime in the United States: 2006, last accessed 6/2/13 at http://www2.fbi.gov/ucr/cius2006/arrests/.

Fehr, Ernst, and Urs Fischbacher. 2004. Third-party Punishment and Social Norms. *Evolution & Human Behavior* 25: 63–87.

Fehr, Ernst, and Simon Gächter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90: 980-994.

Friebel, Guido, Michael Kosfeld, and Gerd Thielmann. 2014. Sorting of Motivated Agents: Evidence from Applicants to the German Police, Goethe University Frankfurt mimeo.

Friedman, David D. 1999. Why not hang them all?: The virtues of inefficient punishment. *Journal of Political Economy* 107: S259-S269.

Fudenberg, D., and P. Pathak. 2010. Unobserved punishment supports cooperation. *Journal of Public Economics* 94: 78-86.

Garoupa, Nuno. 2009. Some reflections on the economics of prosecutors: Mandatory vs. discretionary prosecution, *International Review of Law & Economics* 29: 25.

Gordon, Sanford C., and Gregory A. Huber. 2009. The Political Economy of Prosecution. *Annual Review of Law & Social Science* 5:8.1-8.22.

Grechenig, K., A. Nicklisch, and C. Thöni. 2010. Punishment Despite Reasonable Doubt – A Public Goods Experiment with Uncertainty over Contributions. *Journal of Empirical Legal Studies* 7: 847-867.

Henrich, Joseph, Richard McElreath, Abigail Barr, et al. 2006. Costly Punishment Across Human Societies. *Science* 312: 1767–1770.

Holmstrom, B., and P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization* 7: 24-52.

Hylton, Keith N., and Vikramaditya S. Khanna. 2007. Toward a Public Choice Theory of Criminal Procedure. *Supreme Court Economic Review* 15: 61.

Levitt, Steven D. 2004. Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not. *Journal of Economic Perspectives* 18: 163.

Levitt, Steven D. 2002. Using Electoral Cycles In Police Hiring To Estimate The Effects Of Police On Crime: Reply. *American Economic Review* 92: 1244.

Manning, Peter, and John Van Maanen. 1978. *Policing: A View from the Street.* Santa Monica: Goodyear Publishing.

Mastrofski, Stephen D., R. Richard Ritti, and Jeffrey B. Snipes. 1994. Expectancy Theory and Police Productivity in DUI Enforcement. *Law & Society Review* 48: 113-148.

McAdams, Richard H. 2012. Bill Stuntz and the Principal-Agent Problem in American Criminal Law, Pp. 47-63 in *The Political Heart of Criminal Procedure: Essays on Themes of William J. Stuntz*, edited by Michael Klarman, David Skeel, & Carol Steiker. New York: Cambridge University Press.

McAdams, Richard H., Dhammika Dharmapala, and Nuno Garoupa. 2015. The Law of Police, *Chicago Law Review* 82: 137.

Mookherjee, Dilip, and I.P.L. Png. 1995. Corruptible Law Enforcers: How Should They Be Compensated? *Economic Journal* 105: 145.

Podgor, Ellen S., Foreward to Symposium on Overcriminalization: The Politics of Crime. 2005. *American University Law Review* 54: 541 (2005).

Polinsky, A. Mitchell, and Steven Shavell. 1993. Should employees be subject to fines and imprisonment given the existence of corporate liability? *International Review of Law and Economics* 13(3): 239.

Polinsky, A. Mitchell, and Steven Shavell. 2001. Corruption and Optimal Law Enforcement, *Journal of Public Economics.* 81: 1-24.

Polinsky, A. Mitchell, and Steven Shavell. 2009. Public Enforcement of Law. Pp. xx, in *Criminal Law and Economics*, edited by Nuno Garoupa. Northampton, MA: Edward Elgar.

Prendergast, Canice. 2007. The Motivation and Bias of Bureaucrats. *American Economic Review* 97: 180-196.

Rasmusen, Eric, Manu Raghav, and Mark Ramseyer. 2009. Convictions versus Conviction Rates: The Prosecutor’s Choice. *American Law & Economics Review* 11: 47–78.

Rashbaum, William K. 2010. Retired Officers Raise Questions on Crime Data. New York Times, Feb. 6, 2010, last accessed 6/2/13 at http://www.nytimes.com/2010/02/07/nyregion/07crime.html.

Ribstein, L. 2011. Agents Prosecuting Agents. Illinois Program in Law, Behavior and Social Science Working Paper No. LBSS11-01.

Richman, Daniel. 2003. Prosecutors and Their Agents, Agents and Their Prosecutors, *Columbia Law Review*. 103: 749.

Rose-Ackerman, Susan. 1999. *Corruption and Government: Causes, Consequences, and Reform*. Cambridge: Cambridge University Press.

Shavell, S. 2013. A Fundamental Enforcement Cost Advantage of the Negligence Rule over Regulation. *Journal of Legal Studies.* 42: 275-302.

Shleifer, Andrei, and R.W. Vishny. 1993. Corruption. *Quarterly Journal of Economics*. 108(3):599-617.

Skolnick, Jerome H. 1966. *Justice without Trial: Law Enforcement in Democratic Society*. New York: John Wiley and Sons Inc.

Stuntz, William J. 2001. The pathological politics of criminal law. *Michigan Law Review*. 100: 549-550.

Stuntz, William J. 2008. Unequal Justice. *Harvard Law Review* 121: 1969.

Vollaard, Ben, and Joseph Hamed. 2012. Why the Police Have an Effect on Violent Crime After All: Evidence from the British Crime Survey. *Journal of Law & Economics* 55: 901-924.

Volokh, Alexander. 1997. N Guilty Men. *University of Pennsylvania Law Review* 146: 173-216.

Walsh, William F. 1986. Patrol Officer Arrest Rates: A Study of the Social Organization of Police Work. *Justice Quarterly* 2: 271-90.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Table 1: The Citizen's Preferences

Suspect is:	Probability	Utility	
		Punish	Not Punish
Guilty	*P*	0	*-L*
Not	*1 - p*	*-βL*	0

Table 2: The Preferences of Punitive Agents

Suspect is:	Probability	Utility	
		Punish	Not Punish
Guilty	*p*	*mL*	*-L*
Not	*1 – p*	*-βL*	0

Figure 1: Sequential Model with Weak CPP

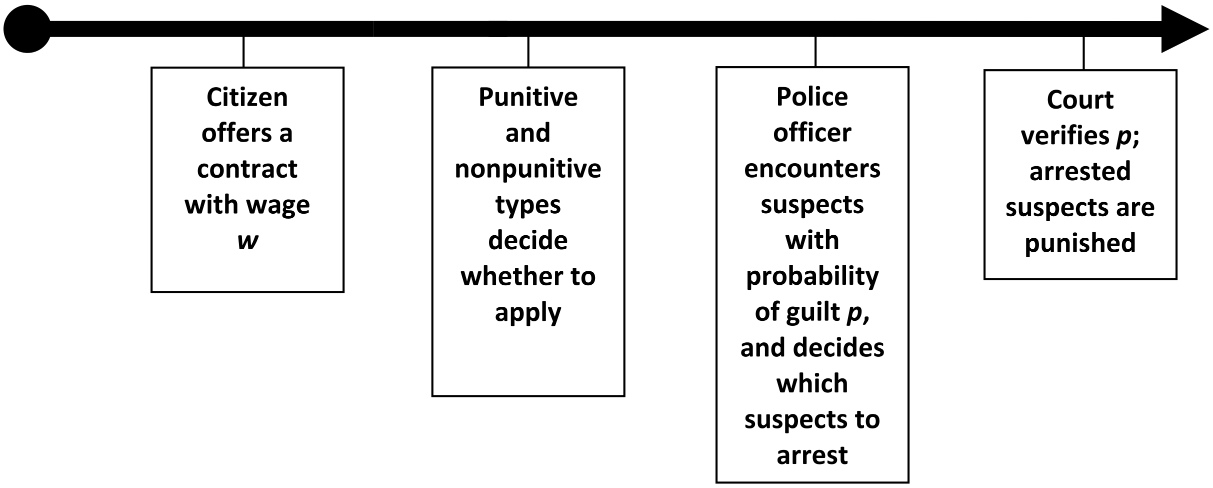
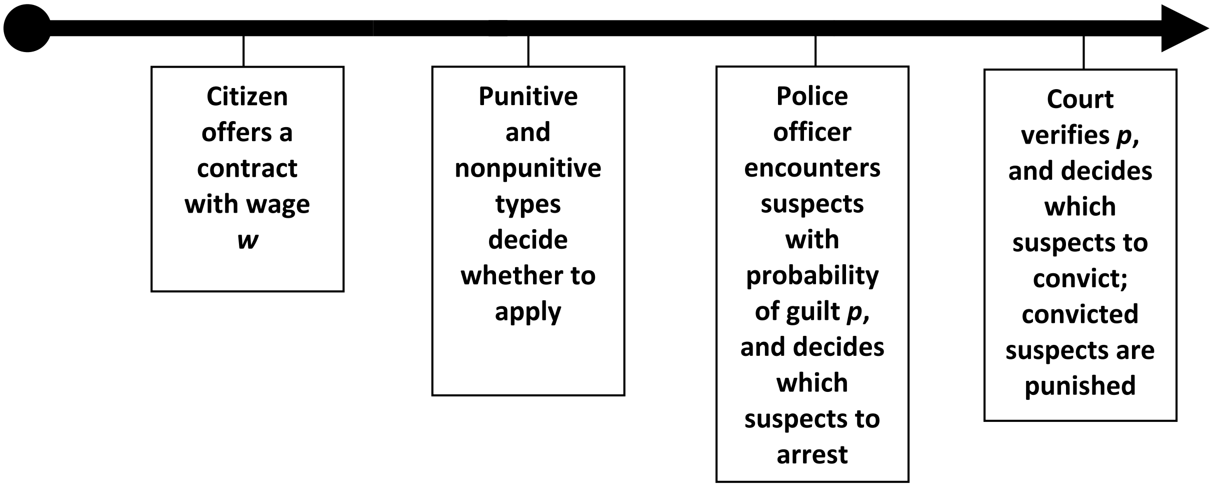

Citizen offers a contract with wage *w*
Punitive and nonpunitive types decide whether to apply
Police officer encounters suspects with probability of guilt *p*, and decides which suspects to arrest
Court verifies *p*; arrested suspects are punished

Figure 2: Sequential Model with Strong CPP


Citizen offers a contract with wage *w*
Punitive and nonpunitive types decide whether to apply
Police officer encounters suspects with probability of guilt *p*, and decides which suspects to arrest
Court verifies *p*, and decides which suspects to convict; convicted suspects are punished