University of Chicago Law School

Chicago Unbound

Coase-Sandor Working Paper Series in Law and Economics Coase-Sandor Institute for Law and Economics

2015

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

Richard H. McAdams
dangelolawlib+richardmcadams@gmail.com

Dhammika Dharmapala
dangelolawlib+dhammikadharmapala@gmail.com

Nuno Garoupa
Nuno.Garoupa@chicagounbound.edu

Follow this and additional works at: https://chicagounbound.uchicago.edu/law_and_economics

Part of the [Law Commons](#)

Recommended Citation

Richard H. McAdams, Dhammika Dharmapala & Nuno Garoupa, "Punitive Police? Agency Costs, Law Enforcement, and Criminal Procedure" (Coase-Sandor Institute for Law & Economics Working Paper No. 644, 2015).

This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

**Illinois Public Law and Legal Theory
Research Papers Series No. 13-47**

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

Dhammika Dharmapala,* Nuno M. Garoupa,
and Richard H. McAdams*****

***Professor, University of Illinois College of Law**

****Professor, H. Ross & Helen Workman Research Scholar, and Co-Director for the
Illinois Program on Law, Behavior and Social Science, University of Illinois**

*****Bernard D. Meltzer Professor of Law and Aaron Director Research Scholar,
University of Chicago Law School**

This paper can be downloaded without charge from the Social Science Research Network
Electronic Paper Collection:

<http://papers.ssrn.com/abstract=2278597>

CHICAGO

COASE-SANDOR INSTITUTE FOR LAW AND ECONOMICS WORKING PAPER NO. 644
(2D SERIES)

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

Dhammika Dharmapala, Nuno Groupa, Richard H. McAdams

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

June 2013

This paper can be downloaded without charge at:
The University of Chicago, Institute for Law and Economics Working Paper Series Index:
<http://www.law.uchicago.edu/Lawecon/index.html>
and at the Social Science Research Network Electronic Paper Collection.

Punitive Police?

Agency Costs, Law Enforcement, and Criminal Procedure

Dhammika Dharmapala

dharmap@illinois.edu

University of Illinois College of Law

Nuno Garoupa

ngaroupa@illinois.edu

University of Illinois College of Law

Richard H. McAdams

rmcadams@uchicago.edu

University of Chicago Law School

June 2013

Abstract

Criminal law enforcement depends on the actions of public agents such as police officers, but there is no standard economic model of police as public agents. We seek to remedy this deficiency by offering an agency model of police behavior. We begin by explaining why the standard contracting solutions are unlikely to work. Instead, we follow recent literature exploring intrinsic motivation and posit heterogeneity in the preferences of potential agents. Drawing on experimental evidence on punishment preferences (so-called “altruistic punishment”), in which subjects reveal a 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 and be less likely to shirk, but create agency costs associated with their excessive zeal (relative to the public’s preferences) in searching, seizing, and punishing suspects. Under plausible 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, two kinds of agents are better than one. We explore various implications for police shirking, corruption, and the content of the criminal procedure rights that the judiciary enforces.

Acknowledgments: We thank Alon Harel, William Hubbard, Aziz Huq, Andrew Leipold, Jonathan Masur, Eric Posner, 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*, Bantam Books, 1993, p. 91.

"Quis custodiet ipsos custodes?"

Juvenal, 1st/2nd century

1) Introduction

In a classic account of corporate law, Easterbrook & Fischel (1991:14) 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. As a result, to understand criminal law enforcement, "one must understand how the balance of advantage among devices for controlling agency costs differs across" governmental entities. To further paraphrase Easterbrook & Fischel: "Without answering difficult questions about the effectiveness of different devices for controlling agency costs, one cannot determine the appropriate allocation" of governmental powers to create and enforce criminal law, nor the individual rights to limit that power.

This point is rather obvious and it has not gone entirely unnoticed. Yet economic scholars of criminal law and procedure pay relatively little attention to the principal-agent problem. Since Becker (1968), the dominant issue in the economics of crime has not been agency costs but optimal probability and severity of punishment. Recent reviews of the economics literature on public law enforcement (e.g., Polinsky & Shavell 2009) reveal a literature that provides an exhaustive treatment of optimal deterrence, but scant attention to the principal-agent problem. 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 (e.g., Echazu & Garoupa 2010; Polinsky & Shavell 2001; Rose-Ackerman 1999), which includes corruption of public enforcers, but corruption is only one problem of agency. On the more general agency problem in criminal law, there is only a handful of economics articles (e.g. Friedman 1999 and Hylton & Khanna 2007). There is a growing literature on prosecutors (e.g.,

Bibas 2009; Boylan 2005; Garoupa 2009; Gordon & Huber 2009; Rasmusen, Raghav & Ramseyer 2009; Ribstein 2011), but aside from the work on corruption there is a particular neglect of our focus, police officers. Overall, there is no sustained or comprehensive analysis of the agency problem and no conventional model for agents of the criminal justice system.

Notwithstanding this neglect, we show below that there is nothing obvious about how society motivates or controls its police force or its other criminal law enforcement agents. Indeed, the agency issue is fundamental because the public achieves optimal punishment only through the actions of agents. Moreover, the problem of agency transcends the various alternative scholarly approaches to crime and criminal law – the prescriptions of retributivist and other moral theories of criminal punishment, no less than the recommendations of economic theories of optimal deterrence, are implemented not by philosopher-kings, but by police and other agents. There is limited practical value to theoretical refinements in optimal deterrence theory (or to philosophical refinements of retributivist theory) if the criminal law enforcer has no incentive or ability to implement the refinement.¹ Rather, what we really need to know is how to minimize agency costs and how to determine punishment given the remaining agency costs.

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. The conventional wisdom is that one cannot solve this particular agency problem through contract. Police are *not* paid for performance; they are not given bounties or paid a piece-rate for each legitimate search, arrest, or conviction, or the level of crime, but paid a wage or salary. The standard explanation for using only weak incentives 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 state some reasons for believing this concern to be serious.

Thus, we turn to a different approach, one more consistent with existing practices in compensating police. Following recent literature on the intrinsic motivation of agents, especially Prendergast (2007), we draw on experimental evidence on “altruistic punishment,” in which many subjects willingly incur a cost to punish wrongdoers. Section 4 reviews this literature. We posit heterogeneity in the “punishment preferences” of potential agents and show how those differences cause agents with unrepresentative preferences to self-select into enforcement jobs. Our specific claim is that individuals with relatively intense intrinsic motivations for punishment will self-select into the job of policing. We show that law enforcement will be

¹ Consider, for example, Becker’s famous prescription that optimality requires *inversely* calibrating the severity of sanctions with the probability of detection. His specific recommendation was to increase severity and decrease the certainty of punishment. Most economists appear not to have noticed that the incentives motivating police and prosecutors usually cause them to spend *greater* effort to investigate and prosecute crimes carrying *heavier* penalties (Boylan, 2005). If the legislature raises the sanction for drunk driving or using a gun in a robbery, that will usually cause the probability of arrest and prosecution for those crimes to rise rather than fall.

dominated by “punitive” agents who are drawn into law enforcement jobs that offer the opportunity to punish wrongdoers or at least to facilitate their punishment by prison authorities.²

Assuming that punishment preferences are not observable and a competitive wage, the public will find that its police are more punitive than it is. The citizen-principal may regard this outcome as desirable for two reasons. First, when agents are compensated in part by the utility from doing their job, here by satisfying punitive preferences, they will work for a lower wage. Second, intrinsically motivated police are less likely to shirk and to take bribes from guilty suspects because they cannot satisfy their punishment preferences without working and arresting the guilty.

On the other hand, there are serious costs. The fact that the agents have a different preference than the average member of the public creates a divergence of interest between principal and agent. Specifically, punitive agents will operate with a lower threshold of doubt for convicting suspects and a lower standard of suspicion for searching or seizing persons or property than would a citizen with average preferences for punishment. Punitive police will also favor harsher punishments and be more eager to use force against suspects than would the average citizen. 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*.

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. In our account, criminal procedure protections offset excessive enforcement zeal by giving other enforcement agents, with preferences closer to the public, the power to determine guilt (a jury), the power to authorize searches and seizures (the judiciary), and the power to determine the sentence (a judge, jury, or sentencing commission). Of course, criminal procedure protections reduce the attractiveness of policing to those with punitive preferences; however, the tendency for self-selection of punitive individuals will not be eliminated as long as punishment opportunities continue to exist. *Thus, two agents (or two types of agents) are better than one.* Instituting criminal procedure protections enables 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.³

² We expect this framework will also prove useful in examining agency problems for other criminal law actors, such as legislators, prosecutors, chief executives.

³ 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 be *more* hostile to criminal procedure protections than does the average

Our model thus derives the basic structure of the criminal justice system – the separation of judicial and executive enforcement powers – from the agency problem in law enforcement. We also make some progress in explaining an economic puzzle about 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, 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.

Our paper proceeds as follows. Section 2 identifies the scope and severity of the principal-agent problem in criminal law. Section 3 summarizes the relevant literature on the principal-agent problem in criminal law. In Section 4, we summarize the behavioral literature finding that many individuals have a preference for punishing wrongdoers. Section 5 sets out the model. This section shows how the public can gain by paying police a lower wage and using criminal procedure rights to control investigation and punishment of suspects. It also extends the model to consider the problems of shirking and corruption. Section 6 discusses how the model explains some otherwise puzzling facts about police behavior and some otherwise puzzling features of the law of criminal procedure. Section 7 concludes.

2) The Magnitude of the Principal-Agent Problem in Criminal Law Enforcement

The neglect of agency within the law and economics of crime is puzzling. To be clear, others have noted the basic problem and we review the few relevant papers below. But there is no sustained attention to agency costs in criminal enforcement, no standard theory of criminal justice agents, much less anything approaching the sophistication of the literature on optimal punishment. That literature grows increasingly refined while generally ignoring the agency problems of implementing its recommendations.

We pause here to consider the scope and intensity of the principal-agent problem in criminal enforcement. Even though our paper primarily addresses only two such agents – the police and the courts – we believe a full statement of the problem will demonstrate the urgency of our efforts to address some part of it. So we identify the problem in its entirety, considering all the agents involved. We take as the basis for comparison the agency problem in corporate law. The agency problem in criminal law is possibly greater, though we shall be satisfied if we merely convince the reader that the problems are equal in magnitude.⁴

citizen. Ethnographic accounts of policing (e.g. Skolnick, 1966; Manning and van Maanen, 1978) provide considerable support for this notion.

⁴ This section borrows from McAdams (2012).

2.1 Inattentive Principals

Consider first the principal – the public. Like shareholders in a firm, the public is inattentive. Citizens have a collective action problem in monitoring and motivating their governmental agents. First, no one has a strong incentive to be informed, given that an individual's ability to influence the agents is so small. For elected agents (legislators, mayors, district attorneys, sheriffs, and some judges), a citizen can influence the agent by threatening to vote against her or him in the next election. For unelected agents (police officers and some judges), the citizen can vote against the incumbent who could have fired or disciplined those agents. However, one vote will almost never affect the outcome. An individual might also incur costs to influence the votes of others, but this solution presents the collective action problem that each individual prefers to let someone else become informed and then organize voters to punish ineffective agents. Thus, with the exception of information needed to minimize one's chances for being the victim of crime, it is rational for most citizens to be ignorant about crime rates, criminal law, and criminal processes.

Indeed, in several respects, it is costlier for the public to monitor criminal law agents than it is for shareholders to monitor the corporation. One reason is that there is better information available about corporations than criminal enforcers. Frequently, shareholders care only about movements in share price because their only concern with corporate behavior is financial. Under standard theory, the best prediction of future share price is the present share price, which (potential) shareholders can easily verify. Even if one rejects the efficient markets hypothesis and thinks it is possible to "beat the market," the law mandates a supply of useful information for making judgments about future share price. Securities regulations require various disclosures and create civil and criminal liability for misrepresentation.

For criminal law enforcement, not only is there less available information, there is also ambiguity about the central criminal justice metric – the crime rate. The FBI's Uniform Crime Reports certainly understate crime rates, while the National Crime Victimization Survey might overstate or understate it, depending on how well people remember and report victimization. City governments and police seem to think that the public relies on crime reports for judging the crime rate because there is evidence of such agents manipulating the numbers to create the appearance of improvement (Rashbaum, 2010). Unlike the manipulation of information about publicly traded stocks, the law creates no civil or criminal liability for criminal justice agents who "massage" crime data.

In any event, the evaluation of criminal law enforcers requires estimating the effect of their behavior on *future* crime rates. Even if everyone were perfectly aware of past and present crime rates, there is no reason to believe that the best estimate of the future crime rate is the present crime rate. Moreover, the law does very little to supply citizens with information on which to base their predictive estimates. The best analogue to the firm's underlying financial

data might be the way police and prosecutors presently allocate their enforcement resources. There is no legal duty, though, to disclose such information and there is some reason to keep it secret. If the police reveal the existing number of wiretaps for particular offenses or prosecutors publicize their policy not to prosecute certain offenses, they may aid criminals in structuring their crimes to avoid detection or prosecution. Such matters are typically exempted from state Freedom of Information Act laws. Yet even if citizens had pristine data on crime rates *and* law enforcement strategies, the principal's job would remain exceedingly complex because most principals will not care about just one aggregated crime rate. Individuals have different concerns about different categories of crime and different geographical areas where crime occurs. Most individuals would agree to weigh violent crime more heavily than nonviolent crime, but there is disagreement over how to rank the priorities of deterring white-collar fraud, immigration offenses, and illegal drug sales. Typically, individuals care more about their neighborhood's crime rate than their city's, and more about their city's crime rate than the national rate, but they still care somewhat about the latter (given that individuals travel outside their city and care about people who live elsewhere). Thus, instead of there being some simple crime index to observe, such as the S&P 500, the citizen-principals ideally need to create and monitor their own personalized crime index or to monitor multiple disaggregated dimensions of different criminal justice outputs.

This kind of heterogeneity matters in a more fundamental way. However it is aggregated, the crime rate is not the only output by which many citizens want to measure the performance of criminal law agents. Citizens also care about the costs of crime control: taxes used for criminal law enforcement; the rate of wrongful convictions (and not just as it may bear on the crime rate by its weakening deterrence); negative externalities of criminal investigations such as privacy intrusions, wrongful arrest, or excessive force; and inequalities in victimization, investigation externalities, or punishment. Probably some citizens care only about the crime rate and the tax costs of minimizing it, but some care about all of these other effects. For them, there is no observable index like the S&P 500 that measures all of these cost variables. Instead, the principal needs to monitor all the dimensions of concern. This kind of citizen is like a shareholder who cares about his or her firm's environmental and labor policies, in addition to future share price. This is the problem of *multiplicity of objectives*.

In sum, the citizen-principals in criminal law – members of the public – are not only rationally ignorant, but also have multiple objectives, creating the need to monitor different metrics of agency action rather than only one.

2.2 Independent Agents

The problem gets worse when we consider the *multiplicity of independent agents* that the principal needs to monitor. Many nations divide governmental powers into legislative, executive, and judicial functions, but that means that the divided powers are nonetheless, in

most cases, “bundled” over all policy domains (Besley & Coate, 2003; 2008; Gersen, 2010; Gersen & Berry, 2008). The legislative branch, for example, has only legislative powers, but it has those powers over economic policy, education policy, crime policy, and everything else. An alternative political design is to “unbundle” power by policy area and aggregate across function so that a single elected officer or policy board would possess all the governmental power – legislative, executive, and judicial – but only for a specific policy domain, such as energy or consumer safety. Gersen (2010) and Gersen & Berry (2008) intensively explore this trade-off for public law, analogizing the question of optimal government structure to the firm’s choice of dividing itself into divisions according to function (e.g., design, manufacturing, distribution) or product (e.g., Chevrolet and Oldsmobile).

If we applied this unbundling idea to criminal law, the result would be radical and worrisome: we would allocate to a single official or board – a “czar” – all the power now held by different branches to create, adjudicate, and enforce criminal law. However the lower-level agents (police, lawyers, prison guards, etc.) were structured, they would then be subordinate to the plenary power of the czar, who would have the power to reward and punish the subordinates, for example by termination. Many Americans would find such an aggregation of power terrifying (literally permitting one person to be judge, jury, and executioner), but it would have one advantage: the public could exert its influence by monitoring a single agent – the czar – who would control and be accountable for all aspects of criminal enforcement. In this respect, the crime policy czar would be like a corporate board of directors because the shareholder can monitor only the board (while the board indirectly monitors all the other agents by monitoring those at the top of the managerial hierarchy who monitor those below).

For good reasons, our system instead divides governmental power by function, but this choice forces the citizen-principals to monitor more than one agent in the domain of criminal law. Indeed, our system appears to divide power more finely in the domain of criminal policy than any other. Part of the reason is that, in criminal law, we introduce governing institutions – the grand jury and a trial jury with the power to “nullify” the law (because criminal acquittals, unlike civil verdicts, are final) – that we use nowhere else. At many levels of government, we also split executive power, making police and prosecutors independent of each other. At the state level, the governor may control the state police, but not the attorney general, who is separately elected. At the local level, the mayor (or a legislative body such as the city council) may control the municipal police, but not the district attorney, who is separately elected. We also split the executive by state and local government. For example, within a state, only a governor can pardon an individual and he or she can do so independent of the wishes of the local police, prosecutor, and jury who arrested, charged, and convicted the individual. As a result, instead of a hierarchy with a czar or board of directors at the top, the criminal system has many horizontally aligned, independent agents, each exercising some discretion not subject to review by the others.

As one way to understand this multiplicity of independent agents, consider how many American agents have the power to prevent criminal punishment of an individual. In most cases, seven sets of independent agents must assent:

- (1) Prior to the individual's act for which he or she is to be punished, the legislature must have defined the act to be criminal.
- (2) In most cases, the police must make an arrest. (Although the prosecutor can legally initiate prosecution without police assistance, in most cases she or he lacks the information to begin investigating a crime and the physical power to bring the individual into court.)
- (3) The prosecutor must bring charges against the individual.
- (4) In most jurisdictions, the grand jury must indict the individual. Admittedly, grand juries tend to do what prosecutors ask and the refusal of one grand jury to indict does not prevent the prosecutor from seeking an indictment from another grand jury. Still, the grand jury's assent is (in most jurisdictions) legally necessary and in rare cases its refusal blocks prosecution.
- (5) The judiciary must allow the trial against the individual to proceed and uphold any verdict of guilt. The trial judge has the formal power to terminate a case by entering a judgment of acquittal, which is not subject to prosecutorial appeal because of the Double Jeopardy Clause. Moreover, the trial court can render decisions on evidence and jury instructions that make conviction impossible as a practical matter. Appellate courts can reverse a conviction and allow no further trial.
- (6) The petty jury must convict the individual. The trial jury has the absolute power to acquit, even to the point of "nullifying" the law by acquitting against the evidence.
- (7) In most jurisdictions, the chief executive (the president or a governor) must not exercise the power to pardon the individual or fully commute his or her sentence.

Thus, there are seven independent veto points over the criminal punishment of an individual. Admittedly, we can usually ignore the vetoes that are rarely exercised, such as the refusal by grand juries to indict and the granting of pardons by governors. But if we drop the formal description in favor of an empirical account of the independence of agents, we would need to add witnesses and private police. If victims and other witnesses refuse to report crimes,

then the police will never apprehend the offender. In some settings, such as firms and universities, private police (or their employers) exercise sufficient control over an area that, as a practical matter, the public police and prosecutor will never get involved for many crimes unless asked. Frequently, the law fails to impose any obligation on these citizens to report crime and cooperate with police, but even where the law does impose such requirements, these agents remain functionally independent.

Punishment veto points are only one way in which criminal law agents operate independently of one other. Consider the public's need for information about how its agents are operating. To a significant degree, each agent has private information about itself that it is free not to share. The judiciary operates in a substantially public fashion, but the rest of the agents do not. Legislation is public, but the lobbying and logrolling behind the legislation is private. Much police work is secret. Prosecutorial bargaining is private. Grand jury proceedings are protected by requirements of secrecy, as are the deliberations of trial juries. Chief executives usually state some reasons for their pardon decisions, but, again, the lobbying, logrolling, and deliberations may remain less public. All of this means that, when it comes to the public's access to information about how its criminal agents perform, the public relies on those agents, such as the legislature, which can pry loose information by conducting hearings and enacting Freedom of Information Act laws. The public relies on prosecutors and grand juries to investigate criminal malfeasance by police and governors. In the end, however, the public's need for information is frequently met, if at all, by yet another agent, though a nongovernmental one: the media. The public relies on the media to collect, synthesize, and disseminate information about crime and the performance of other criminal law agents. Nonetheless, the media too is not part of a hierarchy; it is protected from government control by the First Amendment.⁵

We share the conventional view that the governmental power of criminal law must be divided, but the agency problem does identify a cost to the decision to divide powers into so many independent agents. We do not offer a comprehensive normative analysis of this basic decision of political organization, but a strength of our approach is that it helps to explain how

⁵ Note that defense lawyers are omitted from the list. That might seem appropriate because, in the adversary system, defense lawyers are supposed to be the agents of the criminal defendant, not the public. However, the omission is not appropriate if we step back and ask why the public chooses to have an adversary process rather than one of the alternatives, such as an inquisitorial process. Presumably, if the adversary system is good for society, it is because it is more accurate and that accuracy depends on having defense counsel challenge the government's case. The theory is that defense counsel prevent the wrongful convictions that bureaucratic laziness or group think would otherwise produce. Therefore, the public has an interest in having defense counsel properly perform its role. In this view, defense lawyers are the public's agents (just as the Vatican's Devil's advocate in beatification procedures was an agent of the Vatican). Defense counsel may then fail as agents, by shirking, which fails to prevent false convictions, or by some types of excessive zeal, which produce false acquittals. Nonetheless, as with the rest of the agents, defense lawyers are quite independent from other agents.

the costs of multiple agents are constrained. Using two criminal law agents – police and courts – we show how the citizen-principal benefits by using one agent, with preferences close to the principal, to monitor another agent, with preferences divergent from the principal. If we can show why two types of agents are better than one, then we may have identified a crucial factor for justifying the multiplicity of criminal agent types that we actually observe.

2.3 The Absence of a Contracting Solution

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. While the 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 incentive assume that an officer might be terminated for poor performance, but the evidence suggests that police officers 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. 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 many 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, 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). 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).⁶

Alternatively, the contract might 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 be doing better. Even with salaried police, there are media reports (e.g. 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.

In sum, the agency problem in criminal law is severe; it is possibly more severe than it is in corporate law. We do not propose a comprehensive model. Instead, we focus on only two agents: police and courts. Yet those agents are an obvious and important starting point. Moreover, our focus on the mechanism of selection by intrinsic preferences points the way for future work discussing other agents.

⁶ 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 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.

3) The Economics Literature on Criminal Enforcement Agents

The law and economics literature has largely ignored these concerns. There are a few papers that consider principal-agent problems, beginning with Becker & Stigler (1974), who emphasize the need to examine the efficiency of enforcement in addition to the efficiency of substantive rules. They offer a number of interesting points for constructing a general theory of enforcement agents, but only model 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. Because these discussions remained isolated, Friedman (1999) identifies what we above called the 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.”

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.

Hylton & Khanna (2007) offer a similar public choice analysis, identifying the potential dangers of criminal enforcement, and using the threat of extortion to explain the rights of criminal defendants. They identify two concerns: (1) “simple corruption,” where the law enforcement agent receives a payment from an individual in return for not arresting or prosecuting him; and (2) “inter-group wealth extraction,” where one group pays enforcement agents to arrest or prosecute members of a different group. Both types of expropriation are costly because they undermine deterrence and cause individuals to take costly precautions to avoid being extorted. Hylton & Khanna then argue that punishment limitations and criminal procedure protections ameliorate the extraction problem. Limitations on punishment work in the way Friedman (1999) identified. 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 & Khanna 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.

Because Friedman (1999) and Hylton & Khanna (2007) come closest to the present paper, we pause to note some important weaknesses in their account. Against Friedman (1999), we note that limiting the use of the death penalty does very 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, Friedman ignores another agency problem: 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. So 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).

One objection to Hylton & Khanna (2007) is an odd implication they fail to explore: 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. Law enforcers cannot obtain bribes from people who own nothing of value so, on this theory, the indigent should receive no criminal procedure rights, which does not conform to the rights they attempt to explain (especially the right to a defense lawyer paid by the state).

Our primary objection, however, is that Hylton & Khanna deal with one set of puzzles – why do we limit the infliction and intensity of criminal punishments? – by ignoring another puzzle – how do we induce police to work? Indeed, their analysis intensifies the latter puzzle because, while focusing on expropriation of the innocent, their model ignores the adverse effect that procedural protections and punishment limitations have on enforcement against the guilty. Though the 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 – decreasing the relative costs, while increasing the absolute costs of convicting the guilty – is indeterminate and therefore not a clear justification for the procedural rights. Shirking makes this problem worse. Friedman (1999) and Hylton & Khanna (2007) model enforcers as conventionally self-interested, without intrinsic motivation, but do not explain why self-interested enforcers would ever bother to put forth effort to convict the guilty. Even if low-powered external incentives induce enforcers to put forth some effort, anything that increases the costs of convicting the guilty may greatly reduce the effort to convict the guilty.

Related to shirking is a different type of corruption from the one that Hylton & Khanna model. “Simple corruption” not only involves enforcers demanding bribes from the innocent in order to abstain from enforcement; it also involves *the guilty* offering bribes to the police to abstain from enforcement. Hylton & Khanna do not address this form of corruption and do not explain why self-interested enforcers would ever bother to punish the guilty if they could take

bribes instead.⁷ Even though procedural protections will, as Hylton & Khanna claim, decrease the extraction of bribes from the innocent, they will increase the number of the guilty who escape punishment *via* bribery. First, procedural protections involve giving other enforcement agents veto power over punishment, which creates new bribery opportunities. The jury trial right, for example, makes it productive to bribe the jurors. The right to judicial invalidation of vague laws or retroactive punishments makes it worthwhile to bribe the judge. 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.

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 instead follow Prendergast (2007), who emphasizes that bureaucrats sort themselves by intrinsic preferences and then perform the job in part because the work gives them utility.

4) The Existence of Punishment Preferences

We begin by assuming 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 deviates from the simple and standard assumption that actors care only about ordinary consumption goods. We believe that Prendergast (2007) has already vindicated the assumption of intrinsic motivation by the explanatory power that this kind of agency model can provide. Yet even if one were skeptical about intrinsic motivation generally, we believe our specific assumption about punishment preferences is justified, not only from the explanatory power we derive from our model and the absence of a competing framework, but also by the empirical literature supporting the existence of such preferences.

A notable field study finding that individuals did not free ride as much as expected on the punishment of others is Ostrom, Walker and Gardner (1992). The experimental evidence is obviously better at ruling out strategic incentives for punishment. The first suggestion that individuals derive utility from punishing wrongdoers comes from the standard result of ultimatum games. As Fehr & Fischbacher (2005:8) summarize: “The ultimatum game (UG) (Guth et al. 1982) nicely illustrates that a sizeable number of people from a wide variety of

⁷ Admittedly, the bribes themselves are a form of punishment, but the structure of enforcement discretion encourages non-deterrent bribes. First, the bribe recipient never wants to deter the crime because future violations are a source of future bribes. Second, different enforcers – police, prosecutors, juries, judges, chief executives (with pardon power) – each have a veto on punishment and may compete against each other to sell their veto, lowering the bribe far below the deterrent level.

cultures (Henrich et al. 2001; Roth et al. 1991) facing high monetary stakes (Cameron 1999; Hoffman et al. 1996; Slonim/Roth 1998) are willing to hurt others to . . . punish unfair behaviour.” Across many UG experiments, respondents routinely reject proposers’ offers of less than 25% of the monetary allocation, even though the one-shot and anonymous design ensures that the rejection cannot create a future benefit and is therefore monetarily costly to the respondent. A reasonable interpretation is that the respondent gains utility by punishing the proposer for an unfair offer and this utility exceeds the monetary loss.

Beginning with Fehr & Gächter (2000; 2002), many experiments have validated this interpretation. The most common design is an iterated public goods or voluntary contribution game. See also Bouchet, Page & Putterman (2006); Masclet et al., (2003). 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 & Gächter (2000) introduced the novelty 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 show that individuals do punish free-riders and that punishment can sustain contributions.⁸ Along similar lines, Fudenberg & Pathak (2010) gave players the option of punishing but created common knowledge that the punishment decisions would not be revealed until the end of all play. That structure clearly eliminates the strategic incentive to punish, but subjects punished nonetheless.

Falk, Fehr & Fischbacher (2005) studied the motivations of punishers. In one treatment of a 3-person one-shot prisoners’ dilemma with punishment, they find that 42% of subjects cooperate *and* punish defectors, but they also find that 19% of individuals spitefully defect *and* punish. Defecting punishers distribute punishment almost equally between cooperators and other defectors. Like most of these experiments, the punisher incurred a cost of punishment less than the cost that this punishment imposed on the target. But the researchers ran a second treatment where the punishment reduced the punisher’s wealth by exactly the same amount that it decreased the wealth of the person punished. This change nearly extinguished punishment by defectors, though it caused only a small decrease in the punishment cooperators imposed on defectors. Falk, Fehr & Fischbacher therefore conclude that defectors are motivated by the desire to increase the positive gap between their wealth and the wealth of other players, while cooperators punish out of retaliation for behavior they perceive as unfair. Casari & Luini

⁸ Abbink, Irlenbusch & Renner (2000) find similar results in a game where one party can take money from the second, who can respond by engaging in costly punishment of the first (but not get the money back). Anderson & Putterman (2003) and Carpenter (2003a) confirm that punishment preferences follow the law of demand, as the quantity demanded falls as the price of punishment rises. Kosfeld, Okada & Riedl (2009) move from decentralized to centralized enforcement, showing experimentally that players will frequently vote to create sanctioning institutions in equilibrium.

(2006) also conclude that punishers act out of an emotional need for revenge. Finally, de Quervain et al. (2004) studied neural images of subjects undergoing a punishment experiment and find that the effective punishment of norm violators activated a reward center in the brain, supporting the claim that people gain directly from “altruistic punishment.”

One question left open by many of these experiments is whether individuals have preferences to punish those only who wrong *them* or to punish all wrongdoers, even those who wrong others. Newer experiments offered individuals the opportunity to punish wrongdoers when the potential punisher was not himself a victim of the wrongdoing. For example, Fehr & Fischbacher (2004) designed an experiment in which a third party observed two individuals play a Dictator game, where the third party had the option of incurring costs to punish the Dictator. Again, the game is one-shot and anonymity precludes any reputational effects. Yet nearly 60% of third parties expended money to punish Dictators who gave the second player less than half of the monetary allocation the experimenter provided. Fehr & Fischbacher (2004) obtained similar results when the third party anonymously observed two parties play a one-shot prisoners’ dilemma game (the third-party punishment game or 3PP). The observer was willing to expend resources to punish a player who defected when the other player cooperated. In both cases, individuals spent less on punishment than they did in Ultimatum Games where they were themselves the victim of the unfair behavior, but they still seem to gain utility from punishing wrongdoers regardless of who the victim is. These results are quite robust. Henrich et al. (2006) reports on punishment studies, using the UG and 3PP games, across 14 diverse, non-American cultures with adult as well as student populations. The universal result was that some significant number of people engaged in costly punishment of norm violators even though there was no strategic reason for doing so, but there was great variation in the amount of altruistic punishment across populations.

The social dilemma experiments with punishment opportunities described above all involve certainty about subjects’ behavior (in particular, whether they defected or cooperated in the past round). 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). Thus, a particularly relevant extension of this literature is the experiment reported by Grechenig, Nicklisch and Thöni (2010). They 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, for our purposes, that the taste for punishing wrongdoers is not (at least fully) offset by a corresponding taste for the nonpunishment of the innocent.

It is important to bear in mind that the experiments reviewed above are not specifically designed to replicate the context of 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 establishes the existence of a taste for the punishment of wrongdoers. This preference 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”).

5) A Model of Intrinsically Motivated Police

We present our model in three 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 show how self-selection leads police to have punishment preferences that diverge from the citizen-principal. Third, we show how criminal procedure rights, enforced by punishment-neutral courts, can benefit the citizen-principal while still leaving enough intrinsic benefit to police to attract punitive types into the job.

5.1 Punishment Preferences and the Threshold of Doubt

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 be the principal’s (or society’s) utility. We refer to the principal as the representative “citizen.” Her preferences can be represented as follows:

- $u = 0$ if the suspect is guilty and punished, or if the suspect is innocent and not punished;
- $u = -L$ if the suspect is guilty and not punished
- $u = -\beta L$ if the suspect is innocent and punished.

Alternatively, these preferences can be represented as in Table 1:

Table 1: The Citizen's Preferences

Suspect is:	Probability	Utility	
		Punish	Not Punish
Guilty	p	0	$-L$
Not	$1 - p$	$-\beta L$	0

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.⁹ The parameter β represents the relative cost of punishment errors that involve punishing the innocent, relative to errors that involve nonpunishment of the guilty. The citizen's preferences might typically be expected to 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 (see also Volokh (1997)). However, our results do not require this assumption, and so we do not impose it here.

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 that it exceeds the critical threshold p^* defined as:

$$p^* = \beta / (1 + \beta) \quad (1)$$

The threshold p^* indicates whether a given suspect should be punished, from the standpoint of the citizen's preferences. Above the critical level, the individual should be punished; below the critical level, the individual should not be punished. The relative magnitude of the costs, β , determines the threshold – 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. More precisely, as β approaches infinity, p^* approaches 1. If $\beta = 1$ (so that wrongful punishment and wrongful nonpunishment are equally costly), then $p^* = 1/2$. If β approaches zero, p^* approaches 0.

⁹ For this reason, preferences of this form, often referred to as "truth-seeking" preferences, are widely used in the scholarly literature (see e.g. Dharmapala and McAdams, 2003).

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 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 seem to capture the experimental evidence most simply and parsimoniously, we can characterize the preferences of potential agents more completely by introducing a second variable, n , to capture utility from causing the exoneration of the innocent, as summarized in Table 2:

Table 2: The Preferences of Potential Agents

Suspect is:	Probability	Utility	
		Punish	Not Punish
Guilty	p	mL	$-L$
Not	$1 - p$	$-\beta L$	nL

The parameters m and n could both be positive in that an individual could gain utility from punishing the guilty ($m > 0$) or from *not* punishing the innocent ($n > 0$). Parameters m and n could also both be negative in that an individual could lose utility from *not* punishing the guilty ($m < 0$) or from punishing the innocent ($n < 0$). For m , the positive and negative possibilities are plausibly symmetric: punishing the guilty could produce utility, while its exact opposite – not punishing the guilty – could produce disutility.¹⁰

But this plausible symmetry does not hold for n . Though punishing any innocent individuals can produce disutility, it is untenable that its precise opposite – not punishing any 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

¹⁰ A distinction is sometimes drawn between crimes characterized by uncertainty about identity (where there is no doubt that a crime occurred, but the perpetrator is unknown) and those characterized by uncertainty about actions (where the potential offender’s identity may be known, but there is uncertainty about whether her actions were criminal) – see e.g. Lando (2006). For the former type, punishing the innocent may typically entail exonerating the guilty (as investigations will end after the wrongful conviction). Thus, there may be an added disutility associated with wrongful punishment that may be particularly intense for those who care most about punishing the guilty. This possibility can be accommodated through an appropriate redefinition of m and n . Moreover, a potential law enforcement agent can, in general, reasonably expect to encounter both types of crimes.

from causing the punishment of the innocent, it seems unreal 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.

Now consider the relevant types of individuals these preferences make possible. To begin, we give only brief consideration to individuals for whom m or n is (a net) negative because their intrinsic motivation will cause them to select away from the job of policing and we are interested in explaining who selects into the job. If $m < 0$, the individual loses utility from failing to punish the guilty. Because the police often fail to identify or arrest a perpetrator, such a utility function generates costs that the individual would subtract from the nominal wages for the job. Such individuals find law enforcement work less attractive than other work paying the same wage but not putting the individual in the position to cause or *fail to cause* punishment of the guilty. Similarly, if $n < 0$, then the individual loses utility from punishing the innocent. Because there is some probability $(1 - p)$ that each punished individual is innocent, such an individual suffers an expected utility loss for each suspect whom he arrests or whose arrest or conviction he facilitates. Again, such a person wants to avoid the job of policing.

We now list the individuals who have either a reason to select into policing or not to select away from it. All potential agents are assumed to share the citizen's preferences with respect to erroneous punishment (i.e. to have the same β and L as the citizen).¹¹

(1) "*Punitive*" agents: these agents have the following preferences, $u = mL > 0$ if the suspect is guilty and punished, $u = 0$ if the suspect is innocent and not punished, otherwise the same preferences as the citizen. For punitive agents, $m > 0$ and $n = 0$. Intuitively, these agents derive extra utility from causing or facilitating the punishment of the guilty.

(2) "*Pro-exoneration*" agents: these agents have the following preferences, $u = nL > 0$ if the suspect is innocent and suspected but exonerated, $u = 0$ if the suspect is guilty and

¹¹ It is possible to imagine that potential agents may differ from the citizen along these dimensions as well. Generalizing the agents' preferences in this way, however, complicates the analysis, and adds relatively little, for example to the types of variation in preferred thresholds for punishment that are characterized below.

punished, otherwise the same preferences as the citizen. For pro-exoneration agents, $m = 0$ and $n > 0$. Intuitively, these agents derive extra utility from causing or facilitating the exoneration of innocent suspects who at one point appear guilty.

(3) “*Pro-justice*” agents: these agents have the following preferences, $u = mL > 0$ if the suspect is guilty and punished, $u = nL > 0$ if the suspect is innocent and suspected but exonerated, otherwise the same preferences as the citizen. For pro-justice agents, $m > 0$ and $n > 0$. Intuitively, these agents have a more intense preference for justice than the citizen – they derive utility *both* from causing or facilitating the punishment of the guilty and from causing or facilitating the exoneration of the suspected innocent.

(4) “*Punishment-neutral*” agents: these agents have the same preferences as the citizen, and are hence “neutral” from the perspective of the citizen’s preferences.¹² In this case, $m = n = 0$.

Given the preferences described, an agent will want to punish a suspect with guilt of probability p if and only if p exceeds that agent’s critical threshold p' :

$$p > p' = (n + \beta) / (1 + n + m + \beta) \quad (2)$$

The relationship between the threshold desired by a particular individual and the one chosen by the citizen depends on how n relates to $m\beta$. Let us start with the case where $n = m\beta$. In this case $p' = p^*$. Thus, a “pro-justice” agent with $n = m\beta$ will have a threshold identical to that of the citizen; even though this agent’s preferences (for both punishment of the guilty and exoneration of the suspected innocent) are more intense than the citizens, the intensities cancel out.

The important result is that punishment preferences can change an individual’s threshold for conviction (and other law enforcement decisions that depend on the probability of guilt). Consider the case of a punitive agent, where $n < m\beta$. In this case, $p' < p^*$. A punitive agent wants a threshold lower than the citizen’s. Intuitively, increasing the benefits of punishment increases the disutility of wrongful acquittals relative to wrongful convictions, which pushes down the threshold probability necessary to favor punishment. Conversely, for a pro-exoneration agent, $n > m\beta$ and consequently $p' > p^*$. A pro-exoneration individual wants a threshold above the one favored by the citizen.

¹² Thus, “neutral” means neutral compared to the citizen/principal. If the citizen is punishment-preferring to some degree, 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).

Now the significance of preference asymmetry described above is apparent: *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.

Thus, there are two reasons to suppose that the average police officer will be more punitive than the average citizen. First, the experimental data supports the existence of punitive preferences, but we have not discovered similar evidence of exonerative preferences.¹³ Without exonerative preferences, the only self-selection that occurs is of punitive types. Second, even assuming that punishment preferences and exoneration preferences are of equal frequency and intensity, the unequal opportunities in law enforcement will make such jobs more attractive to punitive agents than pro-exoneration agents. In either case, punitive agents will disproportionately select into policing. Pro-justice types, if they exist, will also select into policing, but because they are both punitive and pro-exoneration, their presence does not affect our conclusion that the average police officer is more punitive than the average citizen.¹⁴

Now we turn to how the principal will react to the prospect of this skewed selection.

¹³ To the contrary, the available evidence, though scant, suggests that pro-justice preferences are not particularly common or intense. As discussed 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.

¹⁴ Note also that pro-justice types who mirror the citizen are ideal and "knife-edged" in that perfect correspondence with the citizen's preferences requires that the condition $n = m\beta$ holds precisely. With a continuous distribution across potential agents of m and n , it is unlikely that a significant number of agents would precisely satisfy this condition.

5.2. The Model of Police Self-Selection

Our basic claim is that the agent trades off wages with intrinsic costs or benefits of the job and the principal trades off wages with the degree of divergence between her preferences and the preferences of intrinsically-motivated agents. The principal can therefore pay lower wages to an agent who gains utility intrinsically from the job, but must pay higher wages to an agent who intrinsically loses utility. The principal can benefit more from lower wages than she loses from attracting into the police force the type of agents whose intrinsic motivation creates a divergence from her own preferences.

5.2.1 The General Framework

We present a basic model here.¹⁵ Our model has two actors, the citizen and the enforcement agent. The suspect, who plays no strategic role, has a probability of guilt p drawn from a distribution $G(p)$ (with density $g(p)$) with support in $[0, 1]$. The citizen (principal) has punishment-neutral preferences, as characterized in Table 1 above. The citizen cannot carry out law enforcement activities directly, and so must hire an agent. In doing so, the citizen cares about both the achievement of law enforcement objectives and the wage w paid to the police officer (which is funded by taxation).

The police officer (agent) is assumed to have a general n and m , as characterized in Table 2, although our focus will be on the case where punitive police predominate. Potential police officers also have a reservation wage w^R (the wage in the best available alternative occupation, assumed to be the same for all types).¹⁶ The timing of the model is shown in Figure 1 and summarized below:

Stage (1): Citizen hires a police officer at wage w .

Stage (2): Police officer encounters a suspect with random probability of guilt p .

Stage (3): Police officer decides whether or not to punish the suspect (after observing p).

Note that, at this point, we assume that the police officer has the discretion to decide whether to seize *and punish* the suspect, with the suspect enjoying no criminal procedure protections.

In Stage (3), the officer uses a critical probability (p^C), punishing the suspect if and only if p exceeds p^C . This critical probability may be specified in the officer's employment contract, depending on the contractual setup that is assumed. If so, the contractual probability may be the probability favored by the principal (p^*) or more generally any other p . On the other hand,

¹⁵ A more general formulation is presented in Appendix A; Appendix B extends the model to encompass shirking and corruption among law enforcement agents.

¹⁶ While we assume that the reservation wage is independent of punishment preferences n and m , the results can be easily extended to incorporate the possibility that the reservation wage is type-dependent.

p^c may be left contractually unspecified, in which case it is implicitly the agent's preferred probability (p').

The expected utility gained by the agent includes the salary paid by the principal and the intrinsic net benefit from punishing or not punishing with a given threshold p^c . The principal will set a wage that reflects the intrinsic costs and benefits the agent incurs from performing the job as the agent will perform it. The agent compares the job of enforcement agent to other jobs and demands a reservation wage $-w^R$ – adjusted for these intrinsic effects.

Given this set-up, how will the principal write the contract? First, consider the contract that will arise under the strong assumptions that types are directly observable by the citizen and that the citizen can contract on the agent's performance. If the market for agents is competitive (i.e. the participation constraint is binding), the participation constraint is solved by equalizing the agent's expected payoff to the reservation salary. Thus, a natural solution is to specify a contract in which $p^c = p^*$ and $w \geq w^R$. Since the principal will fire the agent for deviating from the contract, agents comply with p^* .

What about a more realistic setting, where the principal cannot observe the agent's type or choice of threshold and cannot contract on performance? The principal cannot contract directly on performance for informational reasons discussed in Section 2.3. There we noted that modern states pay police an hourly wage rather than a commission for each valid arrest (or other action) because of the fear that police would be induced to fabricate cases. Nonetheless, in standard principal-agent problems, where it is infeasible to contract on effort, it is usually assumed that it is possible to contract on a non-manipulable output measure (such as sale or stock price) that results from the agent's effort. As discussed in Section 2.3, the natural outcome variable for police effort is the crime rate, but this is subject to potential manipulation.

Given the inability to observe agents' types and the inability to contract on p or some reasonable proxy, the only instrument available to the principal is the wage w . The principal has two possible hiring policies. He can set w at a high level, such that all types of agents are available to be hired (pooling equilibrium). That is, a sufficiently high wage will induce those who have little intrinsic motivation to join the pool of potential officers. Of course, the intrinsically motivated will obtain greater benefits from the job at any given wage, and so will remain in the pool. However, the fraction of the pool that is intrinsically motivated will fall as the wage rises (and the pool becomes more representative of the population). Alternatively, the principal can set a lower salary and exclude "pro-exoneration" agents as well as "pro-justice" agents with a combination of n and m such that $p' \geq p^*$ (separating equilibrium). That is, the principal can save on wage costs and rely on the "in-kind" compensation derived by those with punitive motivation in order to attract agents.

Under the separating equilibrium, the salaries are lower but the choice of threshold is distorted (since only "punitive" agents as well as "pro-justice" agents with a combination of n and m such that $p' < p^*$ are hired). In the pooling equilibrium, the salaries are higher but some

“pro-exoneration” agents as well as “pro-justice” agents with a combination of n and m such that $p' \geq p^*$ are hired. As a consequence, in the pooling equilibrium, the distortion in threshold is less significant.

In order to assess the pooling equilibrium, we need to denote by a the proportion of individuals in the pool of potential agents who only work for the principal if the salary is high. When a is low, the separating equilibrium is likely to be optimal. The reasoning is that “punitive” agents as well as “pro-justice” agents with a combination of n and m such that $p' < p^*$ dominate the pool of the agents in both equilibria, and the separating equilibrium allows for savings in enforcement costs. As a increases, the citizen is faced with a significant trade-off. When a approaches one, the pooling equilibrium is optimal since there is no distortion concerning the threshold and there are not enough “punitive” agents to generate savings in enforcement costs.

With either the pooling or separating equilibrium (but especially the latter), police will be more punitive than the citizen for the reasons noted in section 4.1. The agency slack allows punitive or exonerative police to extract more utility from the job by, respectively, punishing or exonerating suspects contrary to the wishes of the principal. But exoneration opportunities are rare, so the effect tilts in favor of punitive types and they will find the job more attractive than other types given their ability to extract intrinsic utility. Moreover, we believe that the experimental data shows that punitive types are quite common, so that the savings to the principal in the separating equilibrium are likely to be substantial. That means that it is likely that instead of the citizen resisting the self-selection by punitive types by seeking a pooling equilibrium, the citizen will encourage that self-selection by seeking a separating equilibrium.

5.2.2 A Simple Example

The conclusions of the model can be simply illustrated by the following numerical example. Suppose that there are 3 types of suspects:

- S_L – low probability of guilt
- S_M – intermediate probability of guilt
- S_H – high probability of guilt

A police officer randomly encounters 100 suspects per year, with probability $1/3$ of encountering each type of suspect in any given encounter. The citizen wishes to punish only S_H , while punitive agents wish to punish both S_H and S_M (assume that even punitive agents do not wish to punish S_L). The reservation wage for a police officer (of any type) is \$100,000. Punitive agents gain utility equivalent (in monetary terms) to \$600 from punishing a suspect perceived to be guilty (i.e. either S_H or S_M). The citizen suffers no disutility from the punishment of S_H , but suffers disutility equivalent in monetary terms to \$720 from the punishment of S_M . This

disutility arises because the citizen's preferred threshold for punishing suspects has been violated.¹⁷ All parties are assumed to be risk-neutral.

It follows that, when the police officer unilaterally controls the decision to punish, the expected value of punishment to a punitive officer is, for each suspect encounter:

$$(1/3)(0) + (1/3)(600) + (1/3)(600) = \$400$$

For 100 encounters, the expected punishment utility is \$40,000, so a punitive officer is willing to work for a \$60,000 wage. That is, a punitive officer is just as happy with a \$60,000 wage in law enforcement as with a \$100,000 wage in a field outside law enforcement with no punishment opportunities. The citizen thus saves \$40,000 by employing a punitive officer, which exceeds her expected disutility of $(1/3)(720) = \$240$ for each of the 100 suspect encounters, for a total of \$24,000, based on the prospect that S_M will be wrongly punished (which, given the definition of S_M , does not necessarily mean that S_M is innocent, but that his guilt is not established with sufficient probability for the principal-citizen to prefer punishment).

5.3 Incorporating Criminal Procedure Rights

If, as argued in the previous subsection, the police are dominated by the punitive type, the citizen 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.

One possibility is the judiciary. Some judges are directly elected and judicial elections might 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

¹⁷ Note that both the punitive officer and the citizen share a preference (defined in our model by β and L) for avoiding wrongful punishment, but differ in their interpretation of whether it is wrongful to punish S_M .

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. In this section, we explore the use of such punishment neutral agents to control the punishment-preferring police.

5.3.1 Trial Rights

Now we consider the possible advantage to the citizen of introducing trials right for criminal suspects, to have his guilt depend on a determination made by agents other than the police. Let us redefine the steps of the game, as shown in Figure 2:

Stage (1): Citizen hires a police officer (and offers a wage w).

Stage (2): Police officer encounters a suspect with random probability of guilt p .

Stage (3): Police officer decides whether or not to arrest the suspect (after observing p).

Stage (4): Trial – a jury (with the same preferences as the representative citizen) decides whether or not to convict the suspect (after observing p).

This version of the game thus introduces trial rights for the suspect – i.e. the suspect now has the right to trial by a jury, whereas in our basic model the conviction decision was left in the hands of the police. The implications can be illustrated by extending the numerical example from Section 4.2. When trial rights are introduced, a punitive officer can arrest both S_H and S_M , but only obtain a conviction against the former (because the jury, which by assumption shares the citizen's preferences, will not convict S_M). The assumptions from before can be extended as follows. Punitive officers gain utility equivalent (in monetary terms) to \$270 from arresting a suspect perceived to be guilty (i.e. arresting either S_H or S_M), and \$330 from the postconviction punishment of such a suspect. The citizen suffers no disutility from the punishment of S_H , but suffers disutility equivalent in monetary terms to \$360 from the arrest of S_M , and disutility of \$360 from the postconviction punishment of S_M . Thus, for each encounter, the expected value of punishment to a punishment-preferring agent is:

$$(1/3)(0) + (1/3)(270) + (1/3)(600) = \$290.$$

For 100 encounters, the punishment utility is \$29,000, so a punishment-preferring agent is willing to work for a \$71,000 wage. Compared to the baseline scenario in Section 4.2, creating trial rights costs the citizen \$11,000 in additional police wages, but avoids the expected disutility of $(1/3)(360) = \$120$ per encounter, or \$12,000 for 100 encounters/year, from the prospect that S_M will be wrongly convicted. Thus, the citizen can reduce agency costs by employing the court system to constrain some of the punitive officer's excessive zeal (by removing the officer's discretion over the conviction decision). However, the citizen must still endure the disutility of

knowing that some (relatively) innocent suspects will be arrested, in particular the expected disutility of $(1/3)(360) = \$120$ per encounter, or \$12,000 for 100 encounters/year, from the prospect that S_M will be wrongly arrested.

5.3.2 *Rights Limiting Police Investigation*

Now we consider the possible advantage to the citizen of introducing investigation rights for criminal suspects. These are rights that limit the power of police to investigate crime, such as the Fourth Amendment rights which make certain searches, arrests, and property seizures illegal. Investigation tools such as arrest and search are important to our analysis because they create external costs on the targets arrested or searched. As Feeley (1979) famously noted, “the process is the punishment.” Punitive police have an incentive to impose those process costs to punish individuals to a greater extent than citizens would want. Punitive police derive utility directly from the costs their arrest or search imposes (which is why in Section 5.3.1, they arrest individuals the judicial agents will not convict). They also derive benefits from arrest or search indirectly, when the investigation produces evidence that increases the probability of conviction and formal punishment. In each case, the police agents’ preferences deviate from the citizen’s in the direction of being more punitive. That is, punitive police will have a lower threshold of suspicion for searching or seizing a suspect than will the typical citizen. Citizens can thus potentially benefit from placing investigative constraints, *via* judicially enforced criminal procedure rights, on police agents.

This point can be illustrated by modifying the example from Section 5.3.1. To keep the example simple, it is highly stylized and limited to arrest, but similar arguments apply to searches and interrogations. We redefine the stages of the model as follows (see also Figure 3):

Stage (1): Citizen hires a police officer (and offers a wage w).

Stage (2): Police officer encounters a suspect with random probability of guilt p .

Stage (3): A judge (assumed to be punishment-neutral, like the representative citizen) issues a warrant for seizure of the suspect (i.e. arrest), after observing p .

Stage (4): If a warrant has been issued in stage (3), the police officer decides whether or not to arrest the suspect.

Stage (5): Trial – a jury (with the same preferences as the representative citizen) decides whether or not to convict the suspect (after observing p).

Here, merely to illustrate, we sacrifice realism in two respects: First, we assume that the same probability p determines both the lawfulness of arrest and the jury’s willingness to convict, where in reality the former is based on probable cause and the latter is based on proof beyond a reasonable doubt, a much higher standard. Second, we assume away the effect that the police officer’s arrest has on the evidence against the suspect, ignoring the possible discovery of new

evidence during an interrogation or search incident to arrest (which raises the level of suspicion from probable cause to beyond a reasonable doubt). We do not believe that relaxing these assumptions affects the qualitative results of the example.¹⁸ Moreover, we use the warrant requirement merely to illustrate. We get the same results if we focus on other Fourth Amendment rules, such as the arrest requirement that (roughly) the police have probable cause to believe the suspect has committed a crime.

With these assumptions, the judge in stage (3) will only issue a warrant for the arrest of S_H (who will then subsequently be convicted by the jury in stage (5)). The police officer will wish to arrest S_M as well, but will be unable to do so given the warrant requirement. Thus, the expected utility of punishment for a punishment-preferring officer is:¹⁹

$$(1/3)(0) + (1/3)(0) + (1/3)(60,000) = \$20,000$$

A punitive officer will hence be willing to work for an \$80,000 wage. Relative to the basic model, creating criminal procedure protections (both trial rights and a warrant requirement) costs the citizen \$20,000 in additional police wages, but avoids the expected disutility of $(1/3)(72,000) = \$24,000$ from the prospect that S_M will be wrongly arrested and convicted. Table 3 summarizes this and the previous numerical examples.

¹⁸ Note that the two effects push in opposite directions. The higher standard of proof for conviction rather than arrest means that the officer will gain *less* from arrest than we suppose because, in the real world, some arrestees are not convicted. Yet the possibility that the officer will discover new evidence against the suspect means that the officer can gain *more* from the search than we suppose because, in the real world, the officer makes formal punishment possible by a successful search.

¹⁹ This discussion assumes that the warrant requirement applies to the arrest (as is true for arrests of people in their homes) and that the warrant requirement is enforced to an extent sufficient to deter the police from violating it. . For the purposes of our argument, it does not matter whether the remedy for violations of the warrant requirement takes the form of an exclusionary rule or of damages paid by the police officer.

Table 3: Criminal Procedure Protections (Punitive Officer)

	Punitive Officer		Citizen	
Criminal Procedure Rights	Expected Punishment Utility	Wage Demanded	Expected Agency Cost	Savings in Wage Costs
None	\$40,000	\$60,000	\$24,000	\$40,000
Trial Rights	\$29,000	\$71,000	\$12,000	\$29,000
Trial Rights + Warrant Requirement	\$20,000	\$80,000	\$0	\$20,000

6) Implications

In this section, we identify some important implications of our model of intrinsically motivated police. First, the model predicts less shirking than does the standard assumption of ordinary consumer preferences, which we contend is more consistent with the qualitative evidence of police behavior. Second, the model implies a greater need for courts to constrain the actions of police than to constrain identical actions by other governmental agents, which explains an otherwise puzzling distinction made in the constitutional case law. Third, the model implies that optimal governmental design (involving the bundling and separation of powers) will often be to use government jobs with narrow policy domains to attract intrinsically motivated agents with preferences that diverge from the public, who are in turn monitored by agents with broad policy domains and preferences closer to the public. We observe this design in judicial review of police and administrative agencies.

6.1: Less Shirking, More Working

In the introduction 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. Now consider a related puzzle: 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 1985).

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 (Levitt 2002, 2004; Vollaard & Hamed 2012), which seems unlikely without effort. Indeed, the public generally rates the police highly at preventing and solving crime, as well as being honest and ethical.²⁰

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).²¹ 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.²² 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

²⁰ See Sourcebook of Criminal Justice Statistics 2003, at 121 (reporting 2000 and 2002 data showing that more than 60% of respondents stated that their local police were "excellent" or "pretty good" at preventing or solving crime); Sourcebook of Criminal Justice Statistics 2006, at Table 2.13.2006 (from 1998 to most recent year reported, 2006, more than half of survey respondents had "a great deal" or "quite a lot" of confidence in police to protect them from violent crime); Sourcebook of Criminal Justice Statistics 2010, at 2.20.2010 (from 1999 to the most recent year reported, 2010, more than half of respondents say police rank "very high" or "high" in being honest and ethical compared to other occupations; approximately 90% rank police as average or above).

²¹ 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).

²² 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).

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.²³

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.

Note also the implications for bribery. The conventional recommendation is to pay more (Becker & Stigler, 1974:6). They specifically recommend an “entrance fee” – that 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. But such bonds are non-existent. This may be due to wealth constraints among potential enforcers. However, once we introduce intrinsic motivation, we can explain not only why bonds are non-existent, but also why this absence may not necessarily be inefficient. 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.²⁴

²³ 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.

²⁴ Police unions, by raising police pay, might attract individuals less intrinsically motivated to punish, who are more willing to take bribes (and otherwise shirk). More specific implications depend on context. Some crimes create more opportunity for bribery than others. A significant variable is the presence or absence of witnesses other than those involved in committing the crime. Vice crimes are rife with potential for bribery because there are often no witnesses other than those involved in the crime. Either the crimes are “victimless” or there are victims, but they lack personal knowledge necessary to provide useful testimony at trial. Another example may be crimes with diffuse and uniformed victims, such as antitrust violations. By contrast, some crimes naturally do have witnesses – such as victims – who may bring about political scrutiny if police accept a bribe rather than arrest the guilty party. Examples are burglary, robbery, extortion, rape, and even murder (though the central victim of murder cannot press police to investigate, there are usually family members or friends who can). Thus, if vice cops or antitrust enforcers are more likely to take bribes, then the conventional analysis says that we should pay these

6.2: The Nature of Criminal Procedure Rights

In the analysis above, we derive the basic structure of criminal procedure rights from our model, specifically the right to trial and the limitations on search and seizure. Now we wish to show that the model explains some doctrinal details in the law of (at least) American criminal procedure, specifically two elements of search and seizure law: (1) the distinction some cases implicitly make between law enforcement agents – police – and other governmental actors and (2) the distinction some cases explicitly draw between police engaging in criminal law enforcement and police engaging in other activities.

As an example of the first distinction, consider several cases applying the so-called “special needs” doctrine. In *New Jersey v. TLO*,²⁵ the Supreme Court upheld the search of a teenager’s purse by a public high school principal based on a lower standard of suspicion – “reasonable grounds” – than the probable cause standard needed to justify a police officer’s search of the purse (which in some cases would also require a warrant). In two later cases, the Court upheld high school programs of random drug testing of students under circumstances in which it would not permit the police to administer such tests.²⁶ The Court upheld the Federal Railroad Administration’s (FRA’s) regulations requiring railroads to test their employees for drugs and alcohol after certain accidents and safety incidents.²⁷ Though the Court would ordinarily require individualized suspicion and a warrant for such intrusions, it did not require such a standard for searches required by the FRA regulations. Finally, when government is an employer, as where it operates a government hospital, its searches of employee spaces – e.g., enclosed offices, individual desks and lockers – is governed by the fourth amendment. Nonetheless, the Court has applied a lower standard to justify government searches of its employees’ private spaces than the warrant and probable cause it would ordinarily require.²⁸

It should be obvious that the terminology of “special needs” the Court used to justify its opinions does *not* refer to an especially powerful need. Where the Court allows a principal to search a student’s purse, it is not because the need to enforce school rules is more important than the need to enforce criminal laws. Discovering cigarettes to enforce a no-tobacco rule is obviously less important than finding evidence in a murder or rape case. Similarly, for drug tests of students or employees, the social interest is the same whether the tests are administered by school officials, bureaucratic supervisors, or police. The same is true if the search aims to find stolen property – the interest is the same whether the police or some other government agent

particular enforcers more than their reservation wage, while the self-selection model says we should pay them less (in order to restrict the pool of applicants to intrinsically-motivated individuals).

²⁵ 469 US 325 (1985).

²⁶ *Vernonia School District v. Acton* 515 US 646 (1995) and *Board of Education v. Earls*, 536 US 822 (2002).

²⁷ *Skinner v. Railway Labor Executives’ Assn.*, 489 US 602 (1989).

²⁸ See *O’Connor v. Ortega*, 480 US 709 (1987). See also *Treasury Employees v. Von Raab*, 489 US 656 (1989).

conducts the search. Thus, it is a puzzle that the Court distinguishes the levels of justification required for these searches.

Our answer to the puzzle is that the police are more likely than other governmental actors to be punishment-preferring compared to the public and thus police require more judicial monitoring than other state actors. Police self-select for the job because it allows them the opportunity to punish and facilitate the punishment of wrongdoers. School teachers do not self-select in this way because only a small part of the job is the punishment of wrongdoers (and the small part that is about punishment involves far less significant wrongdoing than what police punish, plausibly generating far less punishment utility).²⁹ A bureaucracy like the Federal Railroad Administration does not offer significant punishment opportunities and so its members would have no reason to be as punishment-preferring as police. Nor is there any reason to suppose that the government employees running a state hospital are attracted by punishment opportunities.

Now consider a second distinction. Fourth Amendment law is more deferential to police when they engage in activities with some primary purpose other than criminal law enforcement. For example, if police want to enter a home without consent to find evidence of a crime, and face an exigent circumstance like the imminent destruction of evidence (which excuses a warrant), they need probable cause to believe the home contains evidence of a crime. But if police want to enter a house because they suspect someone is in immediate need of medical care or rescue from danger, then they only need to meet a lower standard of a “reasonable basis” for the belief.³⁰ The latter activity is known as “community caretaking.” Livingston (1998).

We see this distinction again in the roadblock cases. If the primary purpose of the roadblock is “ordinary law enforcement” (directed at an occupant of the car³¹), such as the detection of narcotics (and there is no exigency such as an escaped prisoner), the Fourth Amendment forbids the roadblock because it inevitably seizes motorists for whom the police have no “individualized” suspicion.³² Yet if the primary purpose is one other than ordinary law enforcement, such as removing drunk drivers from the road, then the Court is willing to engage in a balancing test, under which they have upheld such checkpoints.³³ Removing drunk drivers

²⁹ Note that the central point in our account of *New Jersey v. TLO* is not that the optimal level of suspicion required to search a student’s purse is lower than the level in a typical criminal context (though this may be true), but rather that the divergence of preferences between the public and the school principal regarding this level of suspicion is lower than the divergence of preferences between the public and a punishment-preferring police officer in a typical criminal context.

³⁰ See *Brigham City v. Stuart*, 547 U.S. 398 (2006).

³¹ See *Illinois v. Lidster*, 540 U.S. 419 (2004) (applying a different standard when the roadblock was aimed at acquiring information for ordinary law enforcement directed at someone not in the car stopped).

³² See *Indianapolis v. Edmond*, 531 US 32 (2000).

³³ See *Michigan Dept. of State Police v. Sitz*, 496 U.S. 444 (1990).

from the road may seem like law enforcement, because it terminates an ongoing crime, but the point is that the removal of drunk drivers generates immediate protection for the public independent of (and in the absence of) any arrest or prosecution. As long as the primary – even if not the exclusive – purpose is the benefit that accrues without arrest or prosecution, it is not the “ordinary interest” in crime control. The Court refers to this idea in the cases discussed above – such as drug testing in public schools – where the school demonstrates that its primary purpose is not crime control in part by the fact that it does not turn over positive drug tests to police. When a state hospital adopts a policy that tests pregnant women for cocaine use and threatens to turn over positive results to the police, the Court holds that the primary purpose is ordinary law enforcement and therefore that the policy violates the Fourth Amendment.³⁴ In short, “special needs” are not especially strong needs; they are just goals other than the prosecution and punishment of crime.

The puzzle is why courts would want to draw this line. From an efficiency perspective, it seems inexplicable because the “ordinary” law enforcement needs may be more important than the “special” needs. But an agency perspective does explain the distinction, in two ways. First, police officers engaged in community caretaking activities will not anticipate a significant likelihood of arrest or punishment opportunities. Thus, their threshold probability for entering a house will not be affected (i.e. lowered) by the prospect of inflicting punishment. Thus, there is less likely to be a divergence of preferences between the public and the police (even if they are punishment-preferring) with regard to the threshold probability required to enter a house where they suspect someone is in immediate need of medical care. Indeed, based on anecdotal observation, Livingston (1998, p. 274) states that “the potential for overzealousness is often reduced when police serve community caretaking, as opposed to law enforcement ends.”

Second, up to now, we have emphasized that if police are punishment-preferring, they will not shirk as much as conventional models predict; instead they will be zealous about the apprehension and punishment of individuals they suspect of crime. But now that we have introduced the fact that police have jobs other than law enforcement – such as community caretaking and removing unsafe drivers from the road – we have to revise the point. Punishment-preferring police *will* shirk on aspects of the job that do not involve punishment. Indeed, because non-punishment tasks represent an opportunity cost, diverting time from punishment, punishment-preferring police will shirk *more* than punishment-neutral police. If so, then courts will create better incentives by demanding less justification for the activity that punishment-preferring police will underperform. Imposing a higher standard for punishment

³⁴ *Ferguson v. Charleston*, 532 US 67 (2001). Another example is the “inventory search” exception, which allows police to search the contents of a car they have lawfully impounded. The Court has emphasized that the purpose of such a search is not to find evidence of a crime, but to have a list of valuables left in the car in case a property dispute arises when the car is returned to its owner. See *Florida v. Wells*, 495 U.S. 1 (1990).

activities than for “other” activities will raise the external costs to the police for punishment work and lower the relative costs of non-punishment work, offsetting the tendency of punishment preferring police to do too much punishment work and not enough community care-taking or other non-punishment work.

Aside from Fourth Amendment rights, we derived from our model the need for trial rights. We do not claim that the model illuminates the doctrinal details of the trial right, except in one respect: the right to a jury. As noted above, our model plays on the difference in the intrinsic motivation of police and other agents who checks the power of police. Regarding searches and seizures, the other agents are judges. There are a variety of reasons that judges might be punishment-neutral or closer to being punishment-neutral than the police. Nonetheless, although we do not model it, there is likely to be some significant variance among individual judges and some may be as punitive as the police are. In this case, judges will fail as monitors of searches and seizures, but for trial rights the jury provides a redundant check. If the state coerces jury participation or if individuals participate out of a sense of civic obligation, there will *not* be self-selection of the punitive. We do not want to overstate the probability that a jury of twelve people is actually representative of the public, but the jury can create a redundancy in the monitoring of police. First, the judge (and usually a grand jury) has to be satisfied that there is enough evidence to justify a trial, the course of which the judge can heavily influence by a variety of rulings. Second, the jury has to be satisfied that there is enough evidence to convict. Both actors hold an effective veto over conviction; the odds that both the judge and the jury are as punishment-preferring as the police is much smaller than the odds that either one of them is.

6.3: Bundling and Separation of Powers

As stated in section 2.2, a basic feature of democratic government is the separation of powers, which serves various ends including the founders’ concern with preventing tyranny. Besley & Coate (2003; 2008) observe that there are two fundamental approaches to separating or bundling powers: by (1) functions or (2) policy domain. The American constitution separates functions into legislative, executive, and judicial branches, but some American local governments divide power by subject matter and bundle different functions, as a governing board might exercise legislative and executive power on some narrow subject like water, education, or wildlife management (Gersen, 2010; Gersen & Berry, 2008). The interesting puzzle is that local American jurisdictions that do allocate some power according to policy domain rather than function never do so with criminal law enforcement. To the contrary, criminal law is perhaps the only area in which power is divided both by function *and* policy domain. As noted in section 2.2, the local legislative, executive, and judicial functions are separated and the executive is further divided into separate police and prosecutor agents, who are also distinct from the mayor (or other chief local executive).

Our agency model offers an explanation. The separation or bundling of functions or policy domains affects the self-selection and motivation of agents. When agents exercise power across many policy domains, the job will attract a heterogeneous mix of individuals who on average are more like the median citizen-principal. Thus, the fact that judges, legislatures, and chief executives (presidents, governors, and mayors) are policy generalists means that those who self-select into such positions (or are elected) tend to have overall policy preferences close to the median voter. By contrast, when agents exercise power in a single policy domain, the job will attract individuals who have unusually strong intrinsic preferences for that one policy. When the government agent is primarily empowered to enact policy in one direction (e.g., to facilitate punishment), the agents' preferences will be systematically different from the citizen-principal. The fact that the job of police and prosecutors is mostly about facilitating criminal punishment means that they will be more intensely motivated by such policy and, for reasons stated above, more punishment-preferring than the median voter (though police also facilitate public safety as by rescuing the injured and career-minded lawyers often seek prosecutorial experience to create opportunities in politics and large civil firms that value trial experience). We describe the system as functioning because the motivation of police make them good agents for apprehending criminals as long as they are monitored by judicial agents who are punishment neutral. If we combined the judicial function with the police and prosecutorial function, or if we merely created courts that only had jurisdiction over criminal matters (as we do with "drug courts"), then we would lose the distinction between the motivation of police and judges. Both would be more punishment-preferring than the public, and the judiciary would no longer be a good check on the police.

Gersen & Berry (2008:1418-21) discuss how the structure of power (bundling and separation, by function and subject) will affect the public's selection and the candidate's self-selection for public office. But they focus on the effect of structure on the expertise of office-holders, not the effect on the motivation of the self-selecting agents.³⁵ Bubba and Warren (2013) highlight the interaction of shirking and policy bias among bureaucrats in administrative agencies. However, their focus is on agents shirking in the production of information. Our model of self-selection provides one new and very general factor for deciding how best to divide and bundle government power. *The bundling and separation of powers affect what kind of agent is attracted to the job and therefore whether and how their intrinsic motivation differs from the public.* One option for solving the agency problem is to try to select every government agent to share the public's motivation as much as possible, which we would ordinarily facilitate by giving the agent power across policy domains (which implies that any limitation on power will

³⁵ Gersen (2010:340) comes closest to our idea when he briefly mentions that subject matter limits may encourage more "energetic" government because those attracted to the job are more interested in the particular subject matter that defines it.

be achieved by dividing functions). But another option – which we show is sometimes superior – is to create one agent with a limited policy domain, to intentionally attract individuals with motivations different than the public, and to use a second agent, with power across policy domains, and therefore more like the public, as a monitor of the first. We might generalize the point by saying that *generalist agents are more like the public and therefore are well suited to monitoring specialty agents, who are less like the public*. More simply: generalists should monitor specialists.

Although our focus is criminal enforcement, we note that this more general point gives an agency-based justification to judicial review more broadly, such as the review of specialty administrative agencies. Just as the police may have stronger preferences for punishment than the citizen-principal, an administrative agency in charge of, for example, employee welfare may have stronger preferences for the protection of employees than the citizen-principal. Thus, where Bubb & Warren (2013) imagine that the President selects a generalist administrative agency (OIRA) to monitor specialty agencies (EPA), we identify why generalist courts more commonly serve the function of reviewing the self-selected bureaucrats of specialty agencies.

7) 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

Abbink, Klaus, Bernd Irlenbusch & Elke Renner (2000), “The moonlighting game: An experimental study on reciprocity and retribution,” 42 J. Econ. Behav. & Org. 265–277

Anderson, Christopher M. & Louis Putterman (2006), "Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism," 54 J. Econ. Behav. & Org. 1-24.

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

Becker, Gary S. & George J. Stigler (1974), "Law Enforcement, Malfeasance, and Compensation in Enforcers," 3 J. Legal Studies 1.

Berry, Christopher R. & Jacob E. Gersen (2008), The Unbundled Executive," 75 U. Chi. L. Rev. 1385

Besley, Timothy (2006), *Principled Agents? The Political Economy of Good Government*, Oxford: Oxford University Press.

Timothy Besley & Stephen Coate, Issue Unbundling via Citizens' Initiatives, 3 Q.J. Pol. Sci. 379 (2008)

Timothy Besley & Stephen Coate, Elected versus Appointed Regulators: Theory and Evidence, 1 J. Eur. Econ. Ass'n 1176 (2003)

Bibas, Stephanos, "Prosecutorial regulation versus prosecutorial accountability," 157 U Pa L Rev 959 (2009).

Blackstone, William (1765-69) *Commentaries on the Laws of England*

Bochet, Olivier, Talbot Page & Louis Putterman, "Communication and punishment in voluntary contribution experiments," 60 J. Econ. Behav. & Org. 11-26

Bowles, Roger & Nuno Garoupa, 1997. "Casual police corruption and the economics of crime," 17 Int'l. Rev. Law & Econ. 75-87

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

- Bubb, Ryan & Patrick L. Warren (2013) "Optimal Agency Bias and Regulatory Review" NYU Law and Economics Research Paper No. 12-47.
- Cameron, L. A. (1999), Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia, in: *Economic Inquiry* 37(1), 47–59
- Carpenter, Jeffrey P. (2007), "The demand for punishment," 62 *J. Econ. Behav. & Org.* 522-542.
- Casari, Marco & Luigi Luini (2006), "Peer Punishment in Teams: Emotional or Strategic Choice?," (unpublished manuscript).
- de Quervain, Dominique J.-F., et al. (2004), "The Neural Basis of Altruistic Punishment," 305 *Science* 1254-58.
- 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.
- Easterbrook, Frank H. (1983), "Criminal Procedure as a Market System," 12 *J. Legal Studies* 289-332.
- Echazu, Luciana & Nuno Garoupa (2010), "Corruption and the Distortion of Law Enforcement Effort," 12 *Amer. Law & Econ. Rev.* 162-80.
- Falk, Armin, Ernst Fehr & Urs Fischbacher (2005), "Driving Forces Behind Informal Sanctions," 73 *Econometrica* 2017-2030
- Federal Bureau of Investigation (2006), "Crime in the United States: 2006," last accessed 6/2/13 at <http://www2.fbi.gov/ucr/cius2006/arrests/>
- Feeley, Malcolm M. (1979), *The Process is the Punishment: Handling Cases in a Lower Criminal Court*, New York, NY: Russell Sage Foundation
- Fehr, Ernst & Urs Fischbacher, "Human Altruism – Proximate Patterns and Evolutionary Origins," 27 *Analyse & Kritik* 6-47 (2005)
- Fehr, Ernst & Urs Fischbacher (2004), "Third-party Punishment and Social Norms," 25 *Evolution & Human Behav.* 63–87
- Fehr, Ernst. & Gächter, Simon (2000), "Cooperation and punishment in public goods experiments," 90 *Amer. Econ. Rev.* 980-994.

Fehr, Ernst, & Simon Gächter (2002), "Altruistic Punishment in Humans," 415 *Nature* 137–140.

Friedman, David D. (1999), "Why not hang them all : the virtues of inefficient punishment", *J. Pol. Econ.* 107: S259-S269

Fudenberg, D. and Pathak, P. (2010), "Unobserved punishment supports cooperation," 94 *J. Pub. Econ.* 78-86.

Gersen, Jacob E. (2010), "Unbundled Powers," 96 *Va. L. Rev.* 301

Gordon, Sanford C. and Gregory A. Huber. 2009. "The Political Economy of Prosecution." 5 *Annual Rev. Law & Soc. Sci.* 8.1-8.22

Garoupa, Nuno (2009), Some reflections on the economics of prosecutors: Mandatory vs. discretionary prosecution, 29 *Int'l Rev L & Econ* 25

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

Guth, W., R. Schmittberger, B. Schwarze (1982), An Experimental Analysis of Ultimatum Bargaining, in: *Journal of Economic Behavior & Organization* 3, 367–388.

Henrich, J., S. Bowles, C. Camerer, E. Fehr et al. (2001), In Search of Homo Economicus: Behavioral Experiments in 15 Small-scale Societies, in: *American Economic Review* 91(2), 73–78

Henrich, Joseph, R., McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, et al. (2006), "Costly Punishment Across Human Societies," 312 *Science* 1767–1770.

Hoffman, E. & V. Smith (1996), On Expectations and the Monetary Stakes in Ultimatum Games, in: *International Journal of Game Theory* 25(3), 289–301

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.

Kosfeld, Michael, Akira Okada, and Arno Riedl. 2009. "Institution Formation in Public Goods Games." *American Economic Review*, 99(4): 1335–55.

Lando, Henrik (2006) "Does Wrongful Conviction Lower Deterrence" 35 *J. Legal Stud.* 327.

Levitt, Steven D. (2004), "Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not," *J Econ Persp* 18, 163

Levitt, Steven D. (2002), "Using Electoral Cycles In Police Hiring To Estimate The Effects Of Police On Crime: Reply," *Am Econ Rev* 92, 1244.

Livingston, Debra (1998) "Police, Community Caretaking, and the Fourth Amendment" 1998 *Univ. of Chicago Legal F.* 261.

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

Masclet, David; Charles Noussair; Steven Tucker; & Marie-Claire Villeval (2003), "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism," 93 *Amer. Econ. Rev.* 366-380.

Mastrofski, Stephen D., R. Richard Ritti & 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," in *The Political Heart of Criminal Procedure: Essays on Themes of William J. Stuntz*, 47-63, Michael Klarman, David Skeel, & Carol Steiker, eds., New York: Cambridge University Press.

Ostrom, E., Walker, J. and Gardner, R. (1992), "Covenants with and without a sword: Self-governance is possible," 86 *Amer. Pol. Sci. Rev.* 404-417.

Polinsky, A. Mitchell (1980), "Private Versus Public Enforcement of Fines," 9 *J. Legal Studies*, 105-127.

Polinsky, A. Mitchell & Steven Shavell (2001), Corruption and Optimal Law Enforcement, 81 *J. Public Econ.* 1-24.

Polinsky, A. Mitchell & Steven Shavell (2009), Public Enforcement of Law, in *Criminal Law and Economics*, Nuno Garoupa, ed., Northampton, MA: Edward Elgar.

Prendergast, Canice (2007), "The Motivation and Bias of Bureaucrats," 97 *Amer. Econ. Rev.* 180-196.

Rasmusen, Eric, Manu Raghav, & Mark Ramseyer (2009); "Convictions versus Conviction Rates: The Prosecutor's Choice," 11 *Amer. Law & Econ. Rev.* 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 Paper No. LBSS11-01.

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

Roth, Alvin, V. Prasnikar, M. Okuno-Fujiwara, & S. Zamir (1991), Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburg and Tokyo: An Experimental Study, in: *American Economic Review* 81(5), 1068–1095

Schulhofer, Stephen, "Criminal Justice Discretion as a Regulatory System," 17 *J Legal Stud* 43, 51-53 (1988).

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

Slonim, R. & A. E. Roth (1998), Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic, in: *Econometrica* 66(3), 569–596

Stuntz, William, "The pathological politics of criminal law," 100 *Mich L Rev* 505 (2001)

Vollaard, Ben & Joseph Hamed (2012), "Why the Police Have an Effect on Violent Crime After All: Evidence from the British Crime Survey," 55 *J. Law & Econ.* 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

Wright, Ronald & Marc Miller (forthcoming), The Worldwide Accountability Gap for Prosecutors," *Wash. & Lee Law Rev.*

APPENDIX A:

A More General Formulation of the Model

In the model in Section 5.2.1, the expected payoff for the agent can be represented formally by:³⁶

$$U = w + \int_0^{p^c} [(1-p)n - p]LdG(p) + \int_{p^c}^1 [pm - (1-p)\beta]LdG(p) \quad (3)$$

The expected utility gained by the agent includes the salary paid by the principal and the intrinsic net benefit from punishing or not punishing with a given threshold p^c . The integral on the left represents the utility gain or loss from not punishing suspects whose probability of guilt is below the critical value specified. The integral on the right represents the utility gain or loss from punishing suspects whose probability of guilt is above the critical value. Intuitively, this last term can be understood as follows. The suspect is punished whenever the suspect's p is between p^c and 1. When the officer punishes the suspect, she incurs a loss of utility because (with probability $(1-p)$) the suspect is in fact innocent. This utility loss reflects the truth-seeking preferences that were defined in Section 4.1 and that are common to the agent and the principal. In addition, the agent experiences a utility gain of mL (again, as defined in Section 4.1) from inflicting punishment, weighted by the probability of guilt p . The agent's net utility from inflicting punishment will reflect both these elements.

The principal will set a wage that reflects the intrinsic costs and benefits the agent incurs from performing the job as the agent will perform it. The agent compares the job of enforcement agent to other jobs and demands a reservation wage – w^R – adjusted for these intrinsic effects, as follows:

$$w = w^R - \int_0^{p^c} [(1-p)n - p]LdG(p) - \int_{p^c}^1 [pm - (1-p)\beta]LdG(p) \quad (4)$$

The results discussed informally in the text follow from this formulation. The rest of this appendix develops a version of the model that allows for more exoneration than the principal desires and for alternative information scenarios.

When $p^c < p^*$, the agent sets p^c rather than p^* as the threshold probability as we have seen in the basic model. However, in this version of the model, when $p^c \geq p^*$, the agent also picks p^c rather

³⁶ As the salary w is in monetary terms while the other terms represent utility gains and losses, we could use some parameter scaling the monetary term relative to the utilities. We normalize this parameter to one for the sake of simplicity (without significant loss of generality, as such a parameter could also be implicitly taken account of in the definition of L).

than p^* as we have imposed in the basic model. The expected payoff for any agent is given by (3).

(A) Principal knows agent's type and observes agent's choice of threshold

The analysis of the basic model applies here. Agents are forced to set the threshold to p^* . As a consequence, the principal simply sets (4). However, in this context, it is unclear which type of agent is cheaper since both "punitive" and "pro-exoneration" have additional gains. The principal will pick the cheapest type of agent. If "pro-exoneration" agents are cheaper then they will be preferred by the principal.

(B) Principal knows agent's type but cannot observe agent's choice of threshold

Types are directly observable by the citizen. However, the citizen cannot contract on the agent's performance since it is unobservable (or observable but unverifiable by a third-party).

In this setting, the principal knows that the agent will choose $p^c = p'$. The agent, once hired, will set a threshold according and determined by her preferences. This is the incentive compatibility constraint in this context.

Agents enjoy additional utility either from punishing more frequently (those with $p' < p^*$) or from exonerating more frequently (those with $p' > p^*$). Therefore, the principal can obtain savings by lowering the salary paid to agents.

In this scenario, the optimizing variables are just the parameters n and m (since the principal is only able to identify the agent's type). One solution is to pick agents that satisfy $n = \theta m$ since for these agents $p^c = p' = p^*$. In this possible solution, there are no distortions in indictment and exoneration, but also no savings in salary.

An alternative solution is to reduce the salary and hire both "pro-exoneration" and "punitive" agents. The principal might achieve a mix of "pro-exoneration" and "punitive" agents such that the average threshold is p^* but with a significant variance. If she does not care about the variance, this solution is strictly better because it saves costs in wages while achieving the principal's desired threshold p^* on average.

(C) Principal observes agent's choice of threshold but does not know agent's type

Types are not directly observable by the citizen. However, the citizen can contract on the agent's performance since it is observable. This situation is equivalent to scenario (A). The threshold is

p^* by contract and therefore the salary is described by (4). The principal will set the lowest possible salary given the existing m and n .

(D) Principal does not observe agent's choice of threshold and does not know agent's type

Types are not directly observable by the citizen. Furthermore, the citizen cannot contract on the agent's performance since it is not observable. As with the basic model, there is a pooling equilibrium where every potential agent is willing to work for the principal. Alternatively, the principal can lower the salary and achieve a partial separating equilibrium. In this context, it is partial and not full separating as in the basic model because both "pro-exoneration" and "punitive" have utility gains and are therefore willing to work for the principal. As the salary gets lower and lower, only more extreme "pro-exoneration" and "punitive" agents are willing to work (as well as "pro-justice" agents with large n and m).

The choice faced by the principal here is fundamentally the same as in the basic model. However, the trade-off is more complex since in the pooling equilibrium there are still distortions in the threshold (unlike in the basic model). Depending on the distribution of m and n , we can have more or fewer distortions in the pooling equilibrium.

When salaries are reduced, the principal attracts the most extreme "punitive" and "pro-exoneration" types which will generate some distortion on the threshold. As the salary increases, the principal attracts less extreme "punitive" and "pro-exoneration" agents. If these additional individuals reduce the distortion on the threshold, then there is a trade-off to be addressed by the principal much the same way as we have seen in the basic model. If these additional individuals do not reduce the distortion on the threshold, the separating equilibrium with the lowest salary is necessarily optimal.

Appendix B:

Incorporating Shirking and Corruption

Now consider how our model fits into a broader framework that incorporates both shirking and corruption. Shirking lowers the arrest rate of the guilty while a certain kind of corruption raises arrests of the innocent. Following Polinsky & Shavell (2001), we divide corruption into two types: (1) bribery, which is the police taking money from the guilty to refrain from arrest, and (2) extortion, which is the police taking money from the innocent to refrain from manufacturing evidence and making an arrest. (These are not the legal definitions of these crimes). To make credible the threat to arrest the innocent, the practice of extortion may require the police occasionally to frame individuals who resist the extortionate threat. With

these definitions, note that bribery is a form of shirking. The police officer shirks by not working to make arrests the public prefers, whether his motive is to enjoy leisure or to take a bribe. Thus, controlling the agent involves minimizing shirking (including bribery) and extortion.

To keep the model of both factors simple, let us just assume that there is a parameter e for each conviction and for a given L . If positive, it measures an opportunity cost for the agent – forgone leisure or foregone bribery – that cannot be compensated by the principal (because it is unverifiable). Opportunity costs induce shirking, which means arresting fewer of those the public wants arrested. If negative, e measures the opportunity to generate private benefits for the agent through extortion. These private benefits induce extortion, which means arresting more of those the public does not want arrested. As before, the citizen wants p^* as the target for conviction. Given that agents are exposed to shirking or extortion opportunities, the payoffs are now:

Table 4: Potential Agents' Preferences, with Shirking and Corruption

Suspect is:	Probability	Utility	
		Punish	Not Punish
Guilty	p	$mL - eL$	$-L$
Not	$1 - p$	$-\beta L - eL$	nL

We have now a new threshold:

$$p'' = (n + e + \beta) / (1 + n + m + \beta) \quad (5)$$

B.1 Shirking

Let us start by considering shirking (which includes the bribery side of corruption), i.e. $e > 0$. It immediately follows that $p'' > p'$, that is, the threshold is higher under shirking. Fewer suspects are actually punished for a given n and m . In relation to p^* , the critical value is $n = m(\beta - e) - e$. Punishment-neutral agents, with $m = n = 0$, under-punish because of shirking. The effect is enhanced for pro-exoneration types, with $n > 0$ and $m = 0$. However, punitive agents, with $m > 0$ and $n = 0$, might be above or below p^* . In particular, if $e = m(\beta - e)$, they actually choose p^* . The shirking effect goes in the opposite direction to the punishment utility effect. In the limit, there is a combination of e and m such that the threshold is actually p^* . In this case, the citizen would like to hire punitive agents because they are less likely to shirk.

The intuition is that the external incentive to shirk works in the opposite direction to the internal incentive to punish for punitive agents. Thus, the attractiveness of these agents to the citizen is reinforced when we consider the possibility of shirking. Notice that all agents in this specification are more expensive since the salary has to cover the cost eL . From this point of

view, our conclusions from the basic model are unchanged since this cost is borne equally by every type of agent.

B.2 Extortion

Now consider extortion, i.e. $e < 0$. It follows immediately that $p'' < p'$, that is, the threshold is lower under extortion. More suspects are effectively punished for a given n and m . In relation to p^* , the critical value is still $n = m(\beta - e) - e$. Punishment-neutral agents, with $m = n = 0$, over-punish now because of the private benefit. The effect is enhanced for punitive agents, with $m > 0$ and $n = 0$, since they are corrupt and also like punishment. However, for pro-exoneration agents we could have their threshold above or below p^* . The extortion effect goes in the opposite direction to the punishment disutility effect. In the limit, as before, there is a combination of e and n such that the threshold is exactly p^* , namely $n = -e$. In this case, the citizen would like to hire pro-exoneration agents because they are better in achieving the threshold preferred by the citizen.

The intuition is that the external incentive for extortion works in the opposite direction to the internal incentive to avoid punishment for pro-exoneration agents. Thus, in this case, they may be preferred by the citizen. Notice at the same time that all agents in this specification are cheaper since the salary can discount for the private benefit eL .

B.3 Summary

Taken together, opportunities for shirking and extortion create a trade-off. Compared to punishment neutral police, punitive police will engage in less shirking while pro-exoneration police will engage in less extortion.³⁷ For reasons explained above, we predict that the police will be more punishment-preferring than the public. Our positive prediction, therefore, is that the public will face a smaller problem with shirking (including bribery) than conventional models predict.

As a normative matter, the conventional analysis of corruption is that bribery and extortion are both inefficient because they dilute deterrence. For that reason, there is no per se reason to prefer one type of police officer over the other. Of course, because we predict that the police will be punitive, the intrinsic motivation of officers already works against bribery and the law should focus more on extortion. Yet we note an alternative possibility. The argument against extortion is that the punishment of the innocent narrows the gap between the expected punishment of those who commit a crime and the expected punishment of those who don't offend. Yet the irony is that, if extortion is sufficiently pervasive, this point does not hold.

³⁷ Although we have been using "punishment neutral" to refer to police with the average punishment preferences of the public, we can make the same point about rational, selfish police with no punishment preferences at all: they will engage in more shirking than punishment-preferring police and more extortion than punishment-averse police.

Suppose police extort a fixed sum of money from *every* member of the population – the guilty as well as the innocent – in every time period. In this case, the effect is no different than everyone paying a tax to support the police. Because the guilty have to pay the extortion “tax” and then *also* face the threat of additional sanctions for the crime they actually commit, there is no dilution of deterrence. In this scenario, there is a normative reason to prefer extortion to bribery and thus a normative reason to seek punitive police.

Figure 1: Basic Model – Timeline

Figure 2: Timeline of Model with Trial Rights

Figure 3: Timeline of Model with a Warrant Requirement

Readers with comments should address them to:

Professor Richard H. McAdams
University of Chicago Law School
rmcadams@uchicago.edu

Chicago Working Papers in Law and Economics (Second Series)

For a listing of papers 1–600 please go to Working Papers at <http://www.law.uchicago.edu/Lawecon/index.html>

601. David A. Weisbach, Should Environmental Taxes Be Precautionary? June 2012
602. Saul Levmore, Harmonization, Preferences, and the Calculus of Consent in Commercial and Other Law, June 2012
603. David S. Evans, Excessive Litigation by Business Users of Free Platform Services, June 2012
604. Ariel Porat, Mistake under the Common European Sales Law, June 2012
605. Stephen J. Choi, Mitu Gulati, and Eric A. Posner, The Dynamics of Contract Evolution, June 2012
606. Eric A. Posner and David Weisbach, International Paretianism: A Defense, July 2012
607. Eric A. Posner, The Institutional Structure of Immigration Law, July 2012
608. Lior Jacob Strahilevitz, Absolute Preferences *and* Relative Preferences in Property Law, July 2012
609. Eric A. Posner and Alan O. Sykes, International Law and the Limits of Macroeconomic Cooperation, July 2012
610. M. Todd Henderson and Frederick Tung, Reverse Regulatory Arbitrage: An Auction Approach to Regulatory Assignments, August 2012
611. Joseph Isenbergh, Cliff Schmitt, August 2012
612. Tom Ginsburg and James Melton, Does De Jure Judicial Independence Really Matter? A Reevaluation of Explanations for Judicial Independence, August 2012
613. M. Todd Henderson, Voice versus Exit in Health Care Policy, October 2012
614. Gary Becker, François Ewald, and Bernard Harcourt, “Becker on Ewald on Foucault on Becker” American Neoliberalism and Michel Foucault’s 1979 *Birth of Biopolitics* Lectures, October 2012
615. William H. J. Hubbard, Another Look at the Eurobarometer Surveys, October 2012
616. Lee Anne Fennell, Resource Access Costs, October 2012
617. Ariel Porat, Negligence Liability for Non-Negligent Behavior, November 2012
618. William A. Birdthistle and M. Todd Henderson, Becoming the Fifth Branch, November 2012
619. David S. Evans and Elisa V. Mariscal, The Role of Keyword Advertisign in Competition among Rival Brands, November 2012
620. Rosa M. Abrantes-Metz and David S. Evans, Replacing the LIBOR with a Transparent and Reliable Index of interbank Borrowing: Comments on the Wheatley Review of LIBOR Initial Discussion Paper, November 2012
621. Reid Thompson and David Weisbach, Attributes of Ownership, November 2012
622. Eric A. Posner, Balance-of-Powers Arguments and the Structural Constitution, November 2012
623. David S. Evans and Richard Schmalensee, The Antitrust Analysis of Multi-Sided Platform Businesses, December 2012
624. James Melton, Zachary Elkins, Tom Ginsburg, and Kalev Leetaru, On the Interpretability of Law: Lessons from the Decoding of National Constitutions, December 2012
625. Jonathan S. Masur and Eric A. Posner, Unemployment and Regulatory Policy, December 2012
626. David S. Evans, Economics of Vertical Restraints for Multi-Sided Platforms, January 2013
627. David S. Evans, Attention to Rivalry among Online Platforms and Its Implications for Antitrust Analysis, January 2013
628. Omri Ben-Shahar, Arbitration and Access to Justice: Economic Analysis, January 2013
629. M. Todd Henderson, Can Lawyers Stay in the Driver’s Seat?, January 2013
630. Stephen J. Choi, Mitu Gulati, and Eric A. Posner, Altruism Exchanges and the Kidney Shortage, January 2013
631. Randal C. Picker, Access and the Public Domain, February 2013
632. Adam B. Cox and Thomas J. Miles, Policing Immigration, February 2013
633. Anup Malani and Jonathan S. Masur, Raising the Stakes in Patent Cases, February 2013
634. Ariel Porat and Lior Strahilevitz, Personalizing Default Rules and Disclosure with Big Data, February 2013
635. Douglas G. Baird and Anthony J. Casey, Bankruptcy Step Zero, February 2013
636. Oren Bar-Gill and Omri Ben-Shahar, No Contract? March 2013
637. Lior Jacob Strahilevitz, Toward a Positive Theory of Privacy Law, March 2013
638. M. Todd Henderson, Self-Regulation for the Mortgage Industry, March 2013
639. Lisa Bernstein, Merchant Law in a Modern Economy, April 2013
640. Omri Ben-Shahar, Regulation through Boilerplate: An Apologia, April 2013

- 641. Anthony J. Casey and Andres Sawicki, Copyright in Teams, May 2013
- 642. William H. J. Hubbard, An Empirical Study of the Effect of *Shady Grove v. Allstate* on Forum Shopping in the New York Courts, May 2013
- 643. Eric A. Posner and E. Glen Weyl, Quadratic Vote Buying as Efficient Corporate Governance, May 2013
- 644. Dhammika Dharmapala, Nuno Garoupa, and Richard H. McAdams, Punitive Polic? Agency Costs, Law Enforcement, and Criminal Procedure, June 2013