# Reasonable Doubt and Beyond

Liqun Liu*

November 20, 2018

**Abstract**

I study a static inspection game in which an inspector makes binary verdict based on evidence contaminated by an inspectee's concealment effort. I characterize the existence of Nash Equilibrium under mild assumptions, showing that any equilibrium involves inspector's use of a cutoff conviction strategy on the basis of evidence maturity. I show that perverse equilibrium effects may arise from inspector's trade-off between the classical two types of errors: as the inspector prioritizes on convicting the guilty rather than acquitting the innocent, *in equilibrium* she/he may relax conviction threshold to acquit more inspectees from both groups. The results shed light on the optimal judicial mechanism design, and instrumentally justify the value of respecting liberty in authority's terrorism prevention policy.

**Keywords:** Inspection Game, Asymmetric Information, Beyond a Reasonable Doubt, Monotone Comparative Statics, Monotone Likelihood Ratio Property

*Ph.D. student, University of Chicago Harris School. Email: liqunliu@uchicago.edu.

## 1 Introduction

*Better that ten guilty persons escape than that one innocent suffer*
*—William Blackstone, Commentaries 358*

You are innocent. You know that. You took a cab to the airport. You forgot to empty your pocket. You went through the security check in a hurry. The screen machine beeped. You underwent security check again. You missed the flight.

It is not uncommon in everyday life that screen machines beep at innocent persons, even if they have put significant efforts respecting security rules. Coincidence may occur. Test machines may go bad. Inspectors may make mistakes. The fundamental cause is the adverse selection problem in the jargon of economics: inspectors have to make a binary verdict, but they do not know the true type of inspectees. Since their decisions are based on evidence gathered by the time of inspection, more often than not inspectors make mistakes of convicting the innocent (type one error) and acquitting the guilty (type two error). As such, inspectors' decision problem trades off two errors.

Such strategic interactions can be mapped to political life at large. For instance, a bureaucrat determines the eligibility of applicants for good or service (Ting, 2017). Authority goes after "evidence" to decide whether certain social group is plotting terrorist attacks (Crenshaw, 1994)[1]. Courts trigger conviction by evaluating whether the standard of evidence goes **beyond a reasonable doubt (BARD)**.

Common to all situations are three ingredients. First, an inspector (authority, jury) has to map evidence/signal of the inspectees (applicants, suspects) to inspector's binary actions, in which a cutoff strategy such as BARD is prevalent. Second, the inspectees may put efforts impacting evidence-generation prior to the investigation stage. The moral hazard problem complicates the adverse selection by large. Third, the inspectors do not have to *commit* to their strategies, as they often have discretion implementing their ideals. It would be a miracle for a terrorist group to survive, when its leaders naively believe in authority's promise of a lenient conviction threshold and thus take less measure in conducting terrorist acts.

In this paper, I study the equilibrium existence property and comparative statics of a static inspection game in the shadow of moral hazard and adverse selection. Importantly, under mild assumptions, I prove that any equilibrium features inspector's cutoff conviction strategy on the basis of evidence maturity. Hence, restricting attention to the BARD equilibria is without loss of generality. This result justifies the prevalent "BARD" strategy we observe in daily life. I also analyze how the inspector would change her/his equilibrium conviction strategy as her/his relative value of type one and type two errors, or the **Blackstone ratio**[2], evolves.

[1]For instance, terrorist attacks in May 1878 leads to immediate European-wide investigation of terrorist plots. Even evidence did not support conviction of Dr. Nobiling, the incident helped Chancellor Bismarck's passes anti-socialist bills that was filibustered, which turned Germany into white terror (Crenshaw, 1994, p.39-41)

[2] In the common law tradition, one of the most influential glossary accounting for such balance is the **Blackstone ratio** which encapsulates Sir William Blackstone's famous quote: "*All presumptive evidence of*

## Preview of Results

To understand why any equilibrium conviction strategy has to feature a cutoff rule on evidence, I stress it crucial to understand the evidence generation process. By assuming the innocent people as possessing superior evidence generating technology, in any equilibrium we are supposed to see the innocent people "appearing" more innocent than the criminals. If ever the inspector convicts at a nondegenrate, *more suspicious* interval of evidence realization but acquits in another, then she should have swapped the conviction strategy in these two intervals. In doing so, she would improve the accuracy of conviction.

Generically, BARD equilibria exhibit different strategic natures, thus begetting equilibria taxonomy. In a civil society, the inspector can either pick a "reasonable" threshold to balance two errors, or retreat into inaction and let go all suspects. In the former case, equilibria could be further dichotomized into a "regular" one, in which the criminal appears suspicious while the innocent stays safe; and a "benign one", in which both are safe. Such classification substantively maps into "evidence-abundant" and "evidence-scarce" investigation stages. More importantly, this classification helps understand the non-obvious strategic effects between inspector and inspectee: their actions in the benign equilibrium are complement, whereas in regular equilibrium they become substitutes.

On the basis of equilibria taxonomy, I derive a sequence of comparative statics, and discover the possibility of an counterintuitive equilibrium effect. Notably, as Blackstone ratio increases (or the society places more emphasis on convicting the guilty vis-a-vis acquitting the innocent), the equilibrium conviction threshold should decrease (or getting stricter) in *any* regular equilibrium, and increase in the *extreme* benign equilibrium. The intuition is as follows: in regular equilibrium, an increase in Blackstone ratio induces co-movement of threshold decrease and criminal's effort decrease. The innocent increases efforts to appear "innocent", but this effect is dominated by threshold increase. In the new regular equilibrium, a stricter conviction threshold thus forces higher equilibrium conviction probability for both groups, but it has a further impact on the criminals since they are disincentivized from putting effort. As such, a higher Blackstone ratio is met with a stricter conviction threshold.

A similar mechanism works for the extreme benign equilibrium. In this class of equilibrium, the inspector wants to coordinate with the inspectees on largest/smallest conviction probability. When the conviction threshold increases (more lenient) in response to an Blackstone ratio increase, it further disincentivizes innocents' effort, as they are further away in the tail in terms of conviction probability and thus more sensitive to threshold changes. As such, the innocents get relatively more suspicious but still stay above the benign threshold, and Blackstone ratio is met at a higher level in equilibrium.

My model makes the following contributions. First, along with Siegel and Strulovici (2018) and Silva (2018), it justifies BARD as the optimal conviction strategy from a different angle, namely the evidence-generating stage. This result expands the scope of the economic analysis

*felony should be admitted cautiously; for the law holds it better that ten guilty persons escape, than that one innocent party suffer*" (Blackstone, 1962).It is a succinct way of measuring the social attitude towards acquitting a guilty vis-a-vis convicting an innocent.

of law, and derives testable implications on how the optimal conviction threshold should evolve as a response to external shocks at different evidence-collecting stages. Second, it enriches the political economy models of conflict and terrorism prevention. Applying the results to terrorism prevention, it justifies liberty and human rights from an instrumental viewpoint—even a self-interest authority should not impose guidelines that infringe basic rights, as its implementation might lead to perverse equilibrium effects.

### Relationship to Law and Economics

Existing literature in law and economics often approaches the adverse selection problem with a mechanism design approach (Mookherjee and Png, 1994; Kaplow, 2011, 2017). They inherit the analytic framework pioneered by Becker (1968), Stigler (1970), which emphasizes the deterrence effect of law and the cost of law enforcement. This strand of literature endogenizes conviction threshold in order for optimal deterrence. Mookherjee and Png (1994) focuses on screening different types of criminals, and analyze optimal deterrence in terms of its marginal cost-benefit. Kaplow (2011, 2017) emphasize the deterrence effect in the design of threshold on both single and multiple verdicts.

Common in the mechanism design approach are the (implicitly) prefect commitment assumptions: the designers have to commit to the preset mechanism in the conviction process. In some information environment, the mechanism design approach may lose its appeal. For instance, legal doctrines often entail vagueness on purpose due to court's hierarchic structure (Lax, 2012). In this case, the inspector (jury, judge) has discretion on conviction threshold, making it hard for the inspectees to coordinate on the unique Nash Equilibrium.

This paper contributes to the literature of optimal law making by highlighting classification of equilibrium and its corresponding non-obviously strategic behaviors. In particular, depending on the equilibrium in which they coordinate, inspectees' effort and inspector's conviction threshold may move in the same or different direction(s). To the best of my knowledge, few papers in law and economics approach the screening problem as such. Standard in the law-as-deterrence models is the practice of assuming a constant conviction rate and "black box" evidence-generating process. As evident in my analysis, strategic responses could yield surprising equilibrium effects.

This paper also joins an emerging literature on the judicial mechanism design. Siegel and Strulovici (2018) justifies courts' use of BARD as the optimal conviction strategy, focusing on information-gathering during the inspection period in the optimal design of judicial system. Silva (2018) studies the optimal plea bargaining mechanism that elicits (interdependent) information prior to the conviction stage. My model helps justify BARD as the optimal conviction strategy taking account of evidence-generation prior to the conviction stage, and highlights the incentive effects it induces. Our analyses differ mainly in terms of the timing and incentive of information acquisition, and the commitment assumptions.

### Relationship to Terrorism Prevention

My paper has a substantive bearing on terrorism prevention literature, in particular Dragu (2011, 2017). Dragu (2011) concludes that security and liberty may not necessarily conflict with each other: when the authority implements stricter regulations, more terrorism attacks may emerge *in equilibrium*, as the inspectors reduce efforts combating terrorist attacks.

While Dragu's argument grounds on the non-contractiblity of inspectors' action, I show that the perverse equilibrium effect identified in his paper is general and robust. As such, my paper lends support to the normative implication of Dragu (2011) that a civil society should respect human liberty even instrumentally: at the evidence-scarce stage, requiring stricter Blackstone ratio in the hope for more conviction of the criminals is not only morally undesirable, but also often technologically infeasible. Even the most obedient inspectors have to trade off the two-type errors in implementing the policy goal. Often they are *technologically* rather than *intentionally* incapable of adjusting threshold downward to fulfill that policy goal.

In terms of the imperfect attribution problems, my paper resembles that of Baliga et al. (2018). Our approaches differ in the following sense: I analyze a setting in which inspectors and inspectees have to coordinate in a static environment. It makes sense when the inspector does not 1) commit to her/his conviction strategy, or 2) release the conviction threshold. The analysis of Baliga et al. (2018) better describes a dynamic world, where attack-defense happens sequentially.

The rest of the paper is organized as follows: Section 2 describes and intepretes the model of an inspection game. Section 3 justifies BARD as inspector's optimal conviction strategy, and establishes the existence property of Nash Equilibria in the game. Section 4 derives comparative statics. Section 5 discusses the robustness and limitation of the model. Section 6 draws empirical and policy implications. Section 7 concludes.

## 2 Model

### Setup

Two players, an inspectee (hereafter "he") and a Police (hereafter "she"), act simultaneously in an inspection game. The inspectee has a "type" $\theta$, either Innocent ($I$) or Criminal ($C$). The Innocent type imposes no threat to the society, while the Criminal type always plots terrorist attack to the society. Formally, $\theta \in \Theta := \{I, C\}$. Order $\Theta$ as $I \succ C$, meaning that the innocent is the "higher type".

The Police wishes to reduce potential social harm done by the Criminal, so she needs to screen actors of both categories, and punishes suspect by an amount $M > 0$. Unfortunately, Police cannot perfectly observe types. She deduces types from evidence $s \in S$ generated by the inspectee, and picks an action $a \in A := \{\text{Convict, Acquit}\}$ accordingly. In other words, Police's conviction strategy $\sigma : S \to A$ maps from evidence to binary action. Since the state of the world $\Theta$ is binary, I can without loss of generality order $s \in S = \mathbb{R} \cup \{\pm\infty\}$ with higher

$s$ being more "suspicious" of being a Criminal[3]. In the context of BARD, $s$ can stand for how "reasonable" a doubt is. Let $G$ denote the (measurable) set of all evidence realization $s$ at which conviction would be convicted. Due to imperfect observability, she faces the classical type one and type two error.

Inspectees act only according to their types, but they can choose to act carefully with an effort. Their strategy $e : \Theta \to \mathbb{R}_+$ specifies how "favorable" the level of evidence $e$ one would like to induce before inspection happens. In the case of airport security check, $e$ measures how "ready" an inspectee is when she/he stands in line waiting for machine screening. Naturally, one may interpret $e$ as the consequence of effort: should one put more effort, would one obtain evidence more favorable to her/him. Since I order $s$ so that higher $s$ corresponds to more suspicious evidence, favorable (higher) outcome $e$ should correspond to lower evidence level $s$ in order to have an interesting problem. With slightly abuse of notation, I write $e_\theta$ as both strategy and the outcome generated by type $\theta$.

While inspectees of both types may deliver the same outcome $e_\theta = e$ and thus appear identical before the screen machine, they differ in the type-$\theta$-specific cost function $C(\theta, e) = C_\theta(e)$ associated with the outcome $e$. It is reasonable to think of an innocent person as possessing superior evidence generation technology relative to the criminal. After all, an innocent person needs only to "behave himself", while a criminal has to mimic the good person by trying hard.

Furthermore, random noise $\epsilon$ independent of action or type, interpreted as the noise of the test machine, also influences the evidence. As such, I assume that evidence $s_\theta$ generated by type $\theta$ respects a linear technology

$$s_\theta = -e_\theta + \epsilon$$

which implies that one appears more innocent after implementing more favorable (larger) outcome $e_\theta$. I assume that $\epsilon$ is distributed according to standard Gaussian distribution. Needless to say, the PDF $\phi$ is atomless, single-peaked, symmetric around 0 with unbounded support. Furthermore, $\phi$ has the **monotone likelihood ratio property (MLRP)**: the likelihood ratio $\mathcal{L}(s, e_I, e_c) = \frac{\phi(s+e_I)}{\phi(s+e_c)}$ is strictly **decreasing** in $s$ for all $e_I > e_c$.

From now on, I call $e$ "effort" instead of "effort-induced outcome" to make terminology more natural.

## Preferences

All players in the game are vNM utility maximizers. Turn to inspectees first. $C$ wants to do harm to the society. If he can successfully escape the inspection process, he gains by $b$ the amount of harm he does to the society. Otherwise he suffers from punishment $M > 0$. To shield himself from inspection, he makes an efforts $e_c$ at a cost $C_c(e_c)$. His utility function after putting effort $e_c$ is

$$u_c(\sigma, e_I, e_c) = Prob\{\text{Convict}\}[-M] + (1 - Prob\{Convict\})b - C_c(e_c)$$

[3]This is feasible. See Karlin (1968) or Siegel and Strulovici (2018).

$I$'s preference is similar to $C$ but she cares about escaping from convictions only. Her cost of effort is $C_I(e_I)$. She wants to maximize

$$u_I(\sigma, e_I, e_c) = Prob\{\text{Convict}\}[-M] - C_I(e_I)$$

As will be evident later, it is instructive to write $M_\theta = \begin{cases} M, \theta = I \\ M+b, \theta = c \end{cases}$ and define $H_\theta(e) = \frac{C_\theta(e)}{M_\theta}$ as the "effective cost" function for type-$\theta$ inspectees. It measures how difficult type-$\theta$ implements effort $e$ relative to her/his benefit. In doing so we can summarize inspectee of type $\theta's$ objective as maximizing

$$\begin{aligned} u_\theta(\sigma, e_\theta) =& Prob\{\text{Acquit}\} - H_\theta(e_\theta) \\ =& Prob\{s_\theta \in G^c\} - H_\theta(e_\theta) \quad \forall \theta \in \{I, C\} \end{aligned}$$

The inspector $P$ cares about minimizing type one and type two errors, and she trades off two errors at the **Blackstone ratio** $\alpha$. Her utility is

$$\begin{aligned} u_p(\sigma, e_I, e_c) &= Prob\{\text{Acquit } I\} + \alpha Prob\{\text{Convict } C\} \\ &= Prob\{\text{Acquit } I\} - \alpha Prob\{\text{Acquit } C\} + \alpha \end{aligned}$$

This utility function exactly captures $P$'s motivation: she wants to trade off two errors at the Blackstone ratio $\alpha$, a reduced-form measure of society's attitude towards the relative harm of convicting innocent over acquitting guilty.

Such functional form can also be micro-founded in another intuitive way. Suppose society consists of $\lambda$ share of innocent person and $1-\lambda$ criminals. Further assume that the marginal cost of acquitting criminal versus convicting an innocent person is $H$. Then by defining $\alpha = \frac{(1-\lambda)H}{\lambda}$, we may interpret $P's$ objective as minimizing social harm. Of course, $\alpha$ could be influenced by exogenous shock such as terrorist attacks: when citizens prioritize on convicting the guilty than acquitting the innocent ($H \uparrow$) or the perceived population share of criminal increases ($1-\lambda \uparrow$), then $\alpha$ should increase.

## Assumptions

I impose the following assumptions on the parameters values and strategic spaces: $e_i \in [0, \infty); M, b \in (0, \infty)$. Standard in practice, it says the inspectees exert nonnegative efforts $e$; punishment $M$ and the return to harm for the criminal $b$ have to be positive. I also make the following parameter restrictions:

**Assumption 1.** $\alpha \in (0, 1)$

This assumption says that the Blackstone ratio has to sit in $(0, 1)$. It reflects the normative justification that civil society views the mistake of convicting the innocent as more severe compared to acquitting a guilty person.

**Assumption 2.** *$H_\theta(e)$ satisfies the following condition:*

1. *For $\theta \in \{I, C\}, H_\theta(e)$ is smooth and increasing. Furthermore, $H_\theta(0) = H'_\theta(0) = 0, H''_\theta > 0, H'''_\theta \geq 0$*

2. *For* $\theta \in \{I, C\}$, $H_\theta(e)$ *satisfies the strict Spence-Mirrlees Property (SMP). That is,* $\frac{\partial H_c(e)}{\partial e} > \frac{\partial H_I(e)}{\partial e}$ *for all* $e \in (0, \infty)$

Part 1 of this assumption characterizes the shape of the cost function. It basically says that putting efforts is relatively easy at the zero ground, but prohibitively difficult after some input. Substantively, inspectees often need to put minimal efforts substantiating their alibis. But supplying information beyond this level would induce enormous cost of efforts.

Part 2 is a standard practice in economic theory to "compare" technology. In this context, it says that the innocent is technologically more advantaged relative to the criminal. It guarantees that in equilibrium, the innocent people "appear" more innocent.

**Assumption 3.** *Assumption 3 holds if any one of the following is true:*

1. $H''_\theta > \int_{\mathbb{R}} |\phi''(x)|dx$. *In the case of standard Gaussian,* $\int_{\mathbb{R}} |\phi''(x)|dx = 1 + \frac{1}{2\sqrt{2}}$
2. *Inspectees select on the largest or smallest* $e^*_\theta$ *from the best response correspondence* $\{e^*_\theta(\sigma)\}$

This assumption handles the uniqueness part of inspectees' best-response correspondence. Condition 1 simply adds an extra condition to the shape of cost function, which guarantees the concavity of inspectees' maximization program. Under Condition 1, Condition 2 is satisfied trivially. Hence, it is the stronger of the two.

Condition 2 is a common practice in the literature of monotone methods. It roughly says that inspectees of both types use the same decision rule in picking elements from the set of maximizer $\{e^*_\theta(\sigma)\}$. The assumption is trivially true if inspectees' best response is unique. But when inspectees of both types have multiple optimal choices, this assumption rules out "inconsistent decision rules" in the following sense: we allow inspectees of both types to select on the largest (smallest) $e_\theta$ from the set of maximizers, but not that one type always selects on the largest $e_\theta$ while the other selects on the smallest one.

All parameters but types are common knowledge to all players.

## Comments on the Model Setup

Before analyzing the model, I make the following comments on the setup:

First, I do not consider the deterrence effect of conviction on the act $b$. In other words, $C$ cannot alternate the *nature* of his action; his choice variable (effort) can only scale down his probability of being convicted. If $C$ could ever successfully escape the inspection, then he will implement his ideal level of social harm, which confers a private benefit $b$.

Second, $H_\theta(0) = 0$ implies that the innocent person does not prevail the criminal in terms of technology if neither provides any effort. This assumption captures the idea that it is not the type, but the effort induced by the type, that makes a difference in evidence generation. Imagine a murder investigation. There is a blood stain on the wall. Both innocent and guilty person summoned in the police station need to perform blood test. If neither puts any effort taking the lab test, then they are indistinguishable. "Type" therefore does not grant the innocent an innate advantage. However, the innocent can easily separate from the criminal by simply agreeing to perform the test and mimic whatever the criminal does. As such, it is reasonable to assume that type impacts evidence generation through effort choice.

# 3 Analysis

Since the inspection game is static, the solution concept would be Bayes Nash Equilibrium. An inspection equilibrium is a strategy tuple $\tau = (\sigma^*, e^*)$, where $e^* = (e_I^*, e_c^*)$ in which no one has any incentive to deviate.

Generally speaking, it is hard to characterize inspector $P's$ optimal strategy. To simplify the problem, I prove that it is without loss of generality to restrict attention to a class of "BARD" strategy. In this strategy, $P$ convicts with probability one whenever $s \geq s^*$.

I start with a crucial observation.

**Proposition 1.** *Given Assumptions 2 and 3 and an arbitrary conviction strategy $\sigma$, it must be that the $e_\theta^*(\sigma)$ is uniquely selected. Furthermore, $e_\theta^*(\sigma)$ is isotone in $\theta$.*

This proposition guarantees that in any equilibrium, the innocent people would appear more innocent than the criminals.

**Theorem** (BARD)**.** *Assume Assumptions 2 and 3. Then any equilibrium involves the BARD strategy.*

I suppress the subscript of $s_\theta$ in the proof below. The idea of proof is that higher evidence is a better indicator of a criminal type. Hence, any equilibrium involving conviction at lower evidence and acquitting at higher evidence could be improved upon by swapping the conviction decision.

*Proof.* Suppose not. Then there must be two disjoint intervals $I, I'$ with $s < s'$ for all $s \in I, s' \in I'$ such that $P$ convicts at $s$ but acquit at $s'$ . Since $s = -e + \epsilon$ and $e_I^* > e_c^*$, by Proposition 4 in Milgrom (1981) $s'$ is a more "favorable" signal in the sense of being the criminal type. Let $G$ denote the posterior distribution of type $\theta = C$ after $P$ observes evidence, then the "favorableness" implies that $G(\cdot|s')$ dominates $G(\cdot|s)$ in the sense of strict first-order stochastic dominance (FOSD).

Now consider the new conviction strategy: fixing action after observing $s \notin I \cup I'$, but convict at $s'$ and acquit at $s$. Under the new strategy, an additional $[G(C|s' \in I) - G(C|s \in I)]$ fraction of bad people are convicted correctly (and vise versa, the same fraction of good people are acquitted correctly). By FOSD, this amount is strictly positive[4], thus strictly improving $P's$ utility. □

[4] To see it, $G(\cdot|s \in I)$ is the integral of an indicator function on the interval $I$ against the posterior CDF of $\theta$.

This theorem establishes that we can without loss of generality restrict attention to equilibrium involving BARD strategy. From now on, I abuse notation and write $P's$ strategy as $s^*$, since she just needs to figure out the best threshold at which to trigger the auto-machine of conviction. Since any equilibrium involves the BARD strategy, I rewrite an inspection equilibrium as a strategy tuple $\tau = (s^*, e^*)$. It is an *nontrivial equilibrium* if each vector of $\tau$ is bounded. If in equilibrium $s = \infty$, it is an *inaction equilibrium*, as the police let go all suspects. We are interested the existence of a nontrivial equilibrium.

Note that since any equilibrium admits the BARD structure, I can also rewrite the payoffs in terms of the normal CDF $\Phi$:

$$\begin{aligned} u_\theta(s, e_\theta) &= \Phi(s + e_\theta) - H_\theta(e_\theta) \quad \forall \theta \in \{I, C\} \\ u_p(s, e_I, e_c) &= \Phi(s + e_I) - \alpha\Phi(s + e_c) \end{aligned}$$

This is because the inspectees are acquitted with probability $Prob(s_\theta \leq s^*) = Prob(\epsilon \leq s^* + e_\theta) = \Phi(s^* + e_\theta)$.

I further divide nontrivial equilibria into two categories. Call a type-$\theta$ inspectee "safe" if in equilibrium $e^*_\theta + s^* \geq 0$, and "suspicious" if $e^*_\theta + s^* \leq 0$. Hence, in equilibrium safe type will be convicted for less than half chance, while the suspicious will be convicted more than half chance. A nontrivial equilibrium is *regular* if the innocent is safe and the criminal is suspicious. Instead, an equilibrium is *benign* if all inspectees are safe. As will be evident in Lemma 2, the restriction $\alpha \in (0, 1)$ ensures that the classification is exhaustive.

Both concepts are useful mapping the model to the real world. Benign equilibrium is particular useful to understand the conviction strategy when the evidence has not yet matured at the time of investigation. It implies that the inspector cannot accurately categorize inspectees into "safe" and "suspicious". For instance, the investigation of on-line hostile speech may admit a benign equilibrium, as often these actions arise from emotional eruption which hardly can be converted into harmful social consequences. By contrast, at a relative evidence-abundant stage of investigation, it would be relatively easy to screen out the innocent from a group of suspects. At this stage, the guilty persons are indeed more suspicious no matter how carefully they implement it.

I start equilibrium characterization by looking at the first order conditions for each player, and analyze the feasibility of an interior solution. It is easily checked that the FOC of three players are given as follows:

$$\text{Police} \qquad \frac{\phi(s + e_I)}{\phi(s + e_c)} = \alpha \tag{1}$$

$$\text{Inspectee} \qquad \phi(s + e_\theta) = H'_\theta(e_\theta) \quad \theta \in \{I, C\} \tag{2}$$

Let us revisit $P's$ problem. It is *a priori* not clear whether her optimization problem is strictly globally concave. As such, the first order condition may not pin down her unique best response. In what follows, I show that FOC is actually necessary and sufficient thanks to the MLRP.

**Lemma 1.** *$P's$ utility function $u_p(s, e_I, e_c)$ has a unique maximizer $s^*$ given by Equation (1) if only if $e_I > e_c$.*

*Proof.* ($\Leftarrow$): Under $e_I > e_c$, $\mathcal{L}(s, e_I, e_c) = \frac{\phi(s+e_I)}{\phi(s+e_c)}$ is strictly decreasing in $s$. Furthermore, the ratio $\mathcal{L}$ takes values from $+\infty$ to 0. Hence, there must be a unique $s^*$ that solves Equation (1).

For all $s < s^*$, $\mathcal{L}(s, e_I, e_c) > \alpha$ by MLRP so that $u_p$ is increasing; for all $s > s^*$, $u_p$ is decreasing. Hence, at $s^*$ $P'$ maximizes her utility $u_p$. In other words, the $s^*$ solving first order condition is $P's$ best response to $(e_I, e_c)$.

($\Rightarrow$) Suppose $e_I < e_c$, then $u_p$ admits a global minimum at $s^*$ using the similar argument. If $e_I = e_c$, then $\mathcal{L} = 1 > \alpha$. Hence, no equilibrium is feasible. □

Now let us establish some useful fact for proving existence properties.

**Lemma 2.** *$e^*_\theta(s)$ is decreasing in $s$ if type-$\theta$ inspectee is safe, and increasing if suspicious. Moreover, $s^* + e^*_\theta(s)$ is always increasing in $s$.*

This lemma has already delivered some surprising strategic effect. Conditional on the existence of an equilibrium, this lemma says that a tightened conviction threshold may induce co-movement of inspectee's effort choice! Consider the incentive of an inspectee when $s$ declines. On the one hand, she/he may either increase effort to counteract $s$, or decrease effort *further* to exploit the "economy" of saving effort cost. If he is safe, the first effect dominates, which speaks to our intuition of "compensating" for regulation using more effort. However, when he is suspicious, the inspectee's return to the second option way dominates the first. As such, he wants to economize efforts at a cost of being convicted more often.

Now I investigate the sufficient conditions. Since it is impossible to compute equilibrium analytically, I introduce the **best responding likelihood ratio function** to analyze equilibrium behavior. It is given by

$$LR(s) = \frac{\phi(s + e^*_I(s))}{\phi(s + e^*_c(s))}$$

From now on, write $e^*_I(s), e^*_c(s)$ as $e^*_I, e^*_c$ for short. The $LR$ function differs from $\mathcal{L}$ in that $e^*_\theta$ vary with $s$. Parametrizing equilibrium by Blackstone ratio $\alpha$, the following lemma characterizes interior equilibrium by way of $LR(s)$:

**Proposition 2.** *$(s(\alpha), e_i(s(\alpha)))$ is an interior equilibrium if and only if $LR(s) = \alpha$*

*Proof.* Note that $e^*_I > e^*_c$ by Proposition 1. The result follows from the definition of equilibrium, and sufficiency of FOC established in Lemma 1. □

Basically, the $LR$ function tests for some $s$ whether it may solve the system of equations $(1) - (2)$. It is motivated by the relaxed problem (2): fixing the choice of $s = s^*$, one can back out the unique triple $(s^*, e^*_I, e^*_c)$. As such, the system has only one degree of freedom. It remains to test whether this $s^*$ also solves (1). I vary $s$ from $-\infty$ to $+\infty$, and check if any one of them works. Clearly, $LR(s)$ is continuous, because the best responses $e^*$ as a function of $s$ given in (2) are continuous in $s$.

Now it is crucial to characterize the shape of $LR(s)$. Due to single-peakedness of $\phi$, we can establish the "interior behavior" of $LR$.

**Lemma 3.** *Given Assumption 2-1, there exists $\underline{s} < \bar{s}$ such that $LR(\underline{s}) > 1, LR(\bar{s}) < 1$. Furthermore, in any equilibrium such that $s$ lies between $\underline{s}$ and $\bar{s}$, it is a regular equilibrium; any equilibrium with $s \geq \bar{s}$ corresponds to a benign equilibrium.*

Now turn to the asymptotic behavior of $LR(s)$. We say that the innocent and criminals are **asymptotically indistinguishable** if $LR(s) \to 1$ for $s \uparrow \infty$. The definition is tested against the limiting behavior when the police sets a loose conviction threshold. Substantively, the condition says that when conviction threshold becomes sufficiently loose, inspectees of two groups should appear very similar to each other. In many real-world situations such as online speech regulation, indistinguishability is a reasonable description. With this condition, the Blackstone Ratio $\alpha$ would cross $LR(s)$ twice, one at the "regular regime" and other at the "benign regime".

From a modeling perspective, while *prima facie* a lenient threshold as such would induce little effort of the inspectees alike, it is not necessarily true that the induced likelihood ratio would asymptotically approach 1. In the Appendix, I show that this condition is guaranteed by Assumption 2 under normal distribution. Furthermore, there exists $s_0 \geq \bar{s}$ such that $LR$ attains its minimum.

Here is the main existence theorem of this paper.

**Theorem 1** (Existence)**.** *Given Assumption 1-3, an inspection equilibrium exists. Furthermore,*

*(i) For any $0 < \alpha < LR(s_0)$, only inaction equilibrium exists.*

*(ii) For any $LR(s_0) \leq \alpha < LR(\bar{s})$, only regular equilibrium exists.*

*(iii) For any $LR(\bar{s}) \leq \alpha < 1$, regular and benign equilibria coexist.*

*Proof.* Part (i) is obvious.It suffices to prove Part (iii). Note that Lemma 1 guarantees $e_I^* > e_c^*$ for any choice of $s > 0$. By Lemma 8 in the Appendix, indistinguishability condition is met. By intermediate value theorem (IVT), any $\alpha \in [LR(\underline{s}), 1)$ would crosses $LR(s)$ at least twice. By Lemma 3, one of them occurs at the benign regime and the other at regular. □

In fact, we can say more about the equilibria thus constructed.

**Lemma 4** (Monotonicity in regular equilibrium)**.** *In any regular equilibrium, $LR(s)$ is decreasing in $s$.*

*Proof.* By Lemma 2, $s + e_\theta^*(s)$ is increasing in $s$. By single-peakedness and the definition of a regular equilibrium, $\phi(s + e_I^*)$ decreases and $\phi(s + e_c^*)$ increases. Hence $LR(s)$ is strictly decreasing. □

**Corollary.** *If the inspection game admits a regular equilibrium, then it is the unique regular equilibrium.*

As evidence from analysis, with Assumptions 1–2, benign and regular equilibria coexist. We may select on one of them motivated by the level of evidence maturity. Since regular equilibrium is unique if any, it remains to sharpen the prediction of benign equilibrium.

Unfortunately, there does not exist a similar monotonicity result with respect to benign equilibrium[5]. This forces us to select on the set of benign equilibria.

To sharpen predictions of benign equilibrium, we may select on extreme benign equilibria using normative criteria. Even in a civilized society, the inspector is often ideologically committed. A police may manage to coordinate on the harshest/softest threshold with the inspectees depending on whether he favors a benign or harsh social consequence. It may also reflect the incentive that $P$ feels like shirking/hardworking as long as his equilibrium strategy is rationally justified, and the most extreme threshold is conducive to this goal. Consistent with this motivation, I focus on a set of "extreme benign equilibrium" $s^* \in \{\bar{s}^*, \underline{s}_*\}$ from the set of equilibrium $\{\tau_J\} = (s^*_J, e^*_J)$ such that exactly either $\bar{s}^* \geq s$ or $\underline{s}^* \leq s$ for any $s$ consistent with some $\tau \in \tau_J$. Even better, we can deduce comparative statics for empirical predictions.

# 4 Comparative Statics

## 4.1 Blackstone Ratio $\alpha$

A substantively important topic in terrorism prevention is how the authority should respond to the aftermath of an attack. An exogenous shock such as "9.11" attack immediately raises citizens' concern about national security, which induces them to be more intolerant towards acquitting the guilty. In other words, citizens tend to be more forgiving toward mistakes of convicting the innocent. External shock as such is reflected as an increase in Blackstone ratio $\alpha$. To see the intuition, if $\alpha \uparrow \infty$, it basically indicates that $P$ would like to reduce the probability of acquitting a criminal at all cost.

As such, one may naively expect that the optimal threshold $s$ should decrease responding to any $\alpha$ increase. In this way, the police can curb the rampant social turmoil due to terrorist attack even at the cost of convicting more innocent person. In the following analysis, I show that this intuition is at best incomplete. Since the system of equation (1)-(2) does not admit a closed form solution, I deduce comparative statics indirectly.

**Proposition 3.** *Conditional on its existence,*

- *If the inspection game selects on the extreme benign equilibrium, then in equilibrium $s$ is increasing in $\alpha$.*
- *if the inspection game selects on the (unique) regular equilibrium, then in equilibrium $s$ is decreasing in $\alpha$.*

*Proof.* Let $(s, e)$ be the largest benign equilibrium and denote the new equilibrium $(s', e')$ for some $1 > \alpha' > \alpha$. This implies that $LR(s') = \alpha'$ and $LR(s) = \alpha$. Assume towards contradiction that $s' < s$. By IVT, there exists $s'' > s$ such that $LR(s'') = \alpha'$, contradicting that $s'$ is the largest with respect to $\alpha'$. The case for the smallest extreme benign equilibrium is analogus.

The second part follows from the monotonicity of $LR(s)$ in the regular regime. $\square$

I defer the interpretation of this result to Section 6.

[5] We are silent on the number of benign equilibra. To see it, notice that the intermediate value theorem used in proving Theorem 1 is silent on the number of "crossings" with respect to $\alpha$.

### 4.2 Perceived Benefit of Attacks $b$

Another potentially interesting comparative statics lies in the relationship between equilibrium conviction threshold $s$ and criminal's perceived benefit $b$. Substantively, the comparative static results predict ceteris paribus how the authority should respond to an uprising of terrorism activities. The following monotonicity result says that an increase in the perceived benefit $b$ uniformly raises the criminal's input in evidence generation, as long as the nature of equilibrium remains intact.

**Lemma 5.** *$e_c^*(s)$ is increasing in $b$.*

$b\uparrow$ has two effects. First, it generates a new cutoff point $\underline{s}' < \underline{s}$ as defined in Lemma 3 that demarcates regular and benign equilibrium. Second, it raises the $LR$ function for all $s \geq \underline{s}$ while decreases $LR$ for all $s < \underline{s}'$. In these two cases, there is no changes in the *nature* of equilibrium.

**Proposition 4.** *Assume Assumptions 1–3.*

1. *If $1 > \alpha > LR(\underline{s}')$, then in any regular equilibrium, $s^*$ is decreasing in $b$.*
2. *If $1 > \alpha > LR(\underline{s})$, then in the largest extreme benign equilibrium, $s^*$ is decreasing in $b$.*

The theorem speaks to our intuition. In case of an increasing perceived benefit undertaking terrorist activities, the authority should (almost) unambiguously tighten up conviction threshold to punish the criminals, *even at the cost of convicting more innocent person.* Different from Blackstone ratio changes, an increase in $b$ would incentivize a first order criminal's investment in effort. As such, authority's rational response involves tightening up conviction immediately.

An unfortunate consequence to $b\uparrow$ is that the innocent is convicted more while criminals are released more. The fundamental idea is that higher benefit induces criminals to put more efforts mimicking the innocent. The first-order effect is too strong for the police to wash out using stricter conviction threshold.

### 4.3 Preventive Actions and Technological Reversal

At the beginning of this paper, I assume single crossing property of cost functions $H_\theta(e)$. If some common external shock impact inspectees' evidence-generating process so that the inequality is reversed, then technological reversal happens. Such is the case when the political authority takes measures to combat crime, but unfortunately bring about a larger negative impact on the innocents' evidence-generating technology. Fix an arbitrary interior equilibrium. If reversal ever happens, then the primitive model assumptions is violated, and all equilibrium results are gone.

To see it, note that the reversal of technology makes the criminals uniformly more adaptive to conviction technology. As a consequence, the criminals "appear" more innocent than the innocent persons in equilibrium, which forces the police to choose inaction.

The result helps understand preventive actions in the real world. Authorities in many countries adopt on-line sensitive-word detectors in the name of terrorist prevention. However, the detector often imposes a larger impact on innocent's marginal cost of speech, as the criminals can potentially coordinate on a system of coded language for implementing attacks. Once the sensitive-word censor machine gets more accurate, the reversal of technology may happen. Anticipating that they would just convict more innocent, the police would rather wave their hands at any evidence, thus blurring the authority's true policy motive.

# 5 Discussion

## 5.1 The Role of Gaussian Noise

Throughout the paper, I assume that the noise $\epsilon$ is distributed according to standard Gaussian. One may wonder how results are robust to alternative noise distributions. I first discuss the role of Gaussian noise, followed by generalizing main results to a broader class of distribution.

In the proof, Gaussian noise plays three roles.

- The PDF $\phi(x)$ belongs to the Schwartz space[6], ensuring that $\phi', \phi''$ are Lebesgue integrable. This property helps bound inspectees' marginal gains in terms of changing conviction probability, making their optimization problem well-defined. With stronger assumptions on the shape of cost function, inspectees' problem could even be made concave!
- Being in the Schwartz space, $\phi$ decays very fast at infinity. Since inspectee's marginal cost function dominates a linear one, it can be shown that at $s \uparrow \infty$, inspectees of two types put so little effort that they are indistinguishable.
- The PDF $\phi(x)$ has MLRP. Hence first order condition pins down inspector's best response.

Acknowledging this, we can replace the Gaussian noise distribution to the following class without affecting the results.

**Theorem 2.** *Suppose $\epsilon$ admits absolute continuous CDF $F$ with PDF $f$ satisfying the following properties:*

*(i)* $f \in S(\mathbb{R})$*, where* $S(\mathbb{R})$ *is the Schwartz space.*

*(ii)* $f$ *is atomless, single-peaked, symmetric around* $0$ *with unbounded support, and has MLRP.*

*(iii)* $\frac{f'(x)}{f(x)} = P(x)$*, where* $P(x)$ *is a Laurent polynomial (polynomials potentially with negative degrees).*

*Given Assumptions 1-3 with $\epsilon$ distributed as such, then all conclusions hold.*

Seemingly restrictive, these requirements encompass a broad and common class of distribution within the one-dimensional exponential family, including Gaussian, modified (symmetrized) exponential, modified Gamma and so on.

[6]See Folland (2007), page 237 for integration properties of Schwartz class

### 5.2 Assumption 3

Assumption 3 rules out the pathological equilibrium behavior that a criminal might appear more "innonent" than a true one. This assumption endogenously determines that inspectees' **strategies** are *isotone*[7] in types i.e. $e_I^* > e_c^*$. Compare this with Lemma 1, where it requires $e_I > e_c$ as **value**. Clearly, isotone strategies implies the latter. If we drop Assumption 3, there may be equilibrium involving inspectees' non-isotone strategies. It has an impact not on equilibrium existence property, but the optimality of BARD and comparative static analysis.

The remedy to comparative statics is actually simple: we can restrict attention to searching for a class of equilibrium involving inspectees' isotone strategy and do comparative statics therein. As has been proved in Lemma 1, any **selection** of $\{e_\theta^*(\sigma)\}$ belongs to the class of isotone strategy.

It is the justification of the BARD strategy that forces the Assumption 3. Without Assumption 3 there could be non-isotone strategies with $e_I^* < e_c^*$ for some realization of $s \in S \subset \mathbb{R}$ under $\sigma$. Given $P's$ objective, she should convict at all $s \in S$. But $S$ may not be connected in $\mathbb{R}$, thus violating BARD.

### 5.3 Why Static Game

This paper studies a static inspection game to reflect the morale of Schelling's metaphor[8]. As echoed in Schelling's story, the inspector of the game cannot *commit* to its conviction strategy. In the context of terrorism prevention, the terrorists understand that if they naively believe in authority's promise of a lenient conviction threshold and thus take less measure in conducting terrorist acts, more often than not this terrorism group are cracked down rapidly. As such, actors are coordinating on each others' best responses.

The strategic environment considered in this paper is different from the traditional inspection games (see Avenhaus et al., 2002). Standard in literature, the inspector decides whether to investigate at a small but exogenous cost, and the inspectee decides how to respond to potential inspections. Only mixed-strategy equilibrium may exist in this setting, as either player has incentive to deviate when anticipating the opponent's pure strategy. My paper deals with investigation cost/benefit of a complex kind embedding the *coordination of strategy*, thus restoring the possibility of pure-strategy equilibrium.

### 5.4 Why LR Function

I introduce a non-conventional way of proving equilibrium existence. The standard machines fail due to two specialties of this problem, regardless of how one may perturb models specifi-

[7] Isotone is a technical term in the literature of monotone methods. In one-dimension strategy space, it is equivalent to monotone. See McAdams (2005) for a complete treatment.

[8] ... [I]f I go downstairs to investigate a noise at night, with a gun in my hand, and find myself face to face with a burglar who has a gun in his hand, there is a danger of an outcome that neither of us desires. Even if he prefers to leave quietly, and I wish him to, there is a danger that he may think I want to shoot, and shoot first. Worse, there is danger that he may think that I think he wants to shoot. Or he may think that I think he thinks I want to shoot. And so on." (Schelling, 1960, p207)

cations.

- Fixed point theorems do not work. The reason is that while we can treat inspectees' effort choice as being compact, the inspector has to choose a threshold from a non-compact strategy space. Since the product of strategy space is not compact (let alone being finite), we cannot easily adapt Nash existence theorem to this game.

- Monotone methods do not work. The fundamental reason is that $P$ and $C$ inherit conflicting objectives. As such, the game ex ante cannot possibly be a supermodular one. Tarski theorem does not apply.

Hence, the only tool remaining is to exhaust all possible collections of strategy profiles, which underscores the importance of the $LR$ function.

### 5.5 Extreme Sentence

Earlier work in law and economics such as Becker (1968), Stigler (1970) pay close attention to the deterrence effects of extreme sentences. In Becker's seminal work on crime and punishment, extreme sentence may be socially efficient, as it achieves deterrence while minimizing the social cost of combating crimes. By contrast, Stigler puts more emphasis on the incentive effect when multiple crimes are at concern. He argues that maximal sentence is less than the social optimal, as criminals may substitute away from the lesser of two evils.

Ideally, this paper would analyze the evolution path of equilibrium as sanction level $M$ changes. The technical difficulty persists due to analytic intractability: different from Blackstone ratio/perceived benefit where only one actor is affected, $M \uparrow$ impacts inspectees of both types. Furthermore, extreme sentence will *never* lead to technological reversal. Although the initial technological-advantage of innocent citizen gradually erodes as $M$ increases to infinity, it converges eventually to a state where innocent is still technologically advantaged.[9] Hence, indistinguishability condition and existence conditions both apply, and equilibria of two kinds coexist. As such, the model is silent on the deterrence effect of extreme sentence[10].

## 6 Policy and Empirical Implications

### 6.1 Rationalizing Irrationality: Why "Inaction" May be Observed

In the aftermath a terrorist attack, civil society panics. Citizens blame the authority for their failure in combating the terrorism, making citizens and authority alike prioritize on convicting criminals relative to normal times. A common practice is that the authority increases security level in public area. Witnessing Istanbul airport's attack in 2016, even US airports stepped up security level (Jones, 2016). At the same time, Turkish government exercised severe controls over the Internet (Pelegrin, 2016), which co-evolves with a declining trend in Turkey's freedom

[9] To see it, note that SMP requires $\frac{\partial H_c(e)}{\partial e} > \frac{\partial H_I(e)}{\partial e}$, or $\frac{\partial C_c(e)}{\partial e} > \frac{\partial C_I(e)}{\partial e} \cdot \frac{M+b}{M}$. If reversal happens, it means that $\frac{\partial C_c(e)}{\partial e} \leq \frac{\partial C_I(e)}{\partial e}$ which violate SMP.

[10] The converse of extreme sentence $M \downarrow 0$ is very uninteresting. Solving for the equilibrium, it turns out that $s^* < 0$. The substantive meaning of this equilibrium is that the Police set a strict investigation rule, but the court releases all suspect.

index (FreedomHouse, 2018).

In his elegant paper, Dragu (2011) argues that stricter regulations may induce perverse equilibrium effect: as the authority implements stricter regulations, more terrorism attacks emerge *in equilibrium*. Accordingly, the authority should weigh carefully the strategic effects of regulation rather than evolving into a Leviathan. Counterintuitive equilibrium effect may emerge because regulations and inspector's efforts are substitutes. A tighter regulation disincentivizes inspector's effort of terrorism prevention and simultaneously terrorists' effort in attack implementation, which may lead to more equilibrium attacks if the first effect dominates.

Dragu's surprising finding of perverse equilibrium effect sheds light on a wide range of social phenomena. However, one may wonder how sensitive his results depend on model specifications. Central to the results is the assumption that authority cannot observe or contract on the inspector's effort. It is unclear whether the perverse equilibrium effect would persist if inspector's effort choice is observable, contractible, or centralizable.

By way of the dual moral hazard arguments, my paper justifies Dragu's finding as a robust equilibrium phenomena. It stresses the stage of evidence where terrorism prevention action happens. In a civil society, the inspection agencies may want to select on an extreme benign equilibrium. If by the time of inspections the police has collected relatively scarce evidence and inspectees look alike, then we should expect a loosen conviction threshold, even if the police agrees with society's proposal for prioritizing convicting the guilty. As a consequence, more suspects would flee from the screening process, leading to a larger base for future terrorism attacks.

The rationale for this observation is not due to inspector's irrationality or irresponsibility. Instead, it is due to innocent people's superior technology of evidence-generation. As conviction threshold increases, both inspectees would decrease their efforts so as for the inspector to meet the Blackstone ratio $\alpha$. During normal times when $\alpha$ is low, the inspector can set the threshold to be relatively low (harsher) to induce the innocent people to separate from the criminals. As $\alpha \uparrow 1$, the inspector has to set threshold so high that it almost acts as a rubber stamp.

### 6.2 Social Value

My analysis highlights the importance of holding on to social values even after social catastrophes. Once attacks have successfully incited society's hatred for acquitting the guilty, they may press for higher Blackstone ratio in the conviction process uniformly in all criminal prevention activities. Hence, self-enforcing equilibrium attack increase may happen when the society initially coordinates on the extreme benign equilibrium.

At the extreme, if the Blackstone ratio responds so sensitively to attacks that $\alpha > 1$, terrorists attacks would become detrimental to the normal legal process, as it induces a nontrivial equilibrium in which citizens of all kinds are convicted for more than half chance. Such is the case in Bismarck's Germany, where white terror ensued terrorist attacks.

Even in today's civil society, if the executive branches mandate a harsher Blackstone ratio as the guideline to the police department, society may slip into a "benign" equilibrium in which the inspector fails to detect criminals. This is particularly the case when the police department faces a capacity constraint. Intuitively, police has to adjust $s$ downward which decreases equilibrium attack probability. At the same time, however, the probability of conviction increases. If the police departments eventually reach the cap of convictions, they are forced into either inaction or benign equilibrium, where the criminals and innocent people are indistinguishable.

# 7 Conclusion

This paper studies a static inspection game in the shadow of moral hazard and adverse selection. I prove the existence of a nontrivial equilibrium under a rich set of parameters, justify BARD as the optimal conviction strategy, and deduce comparative statics.

The implications of this simple game is far-reaching. On the positive side, my model provides new insight into the optimal design of legal system and terrorism prevention. The authority should always consider the implementability of policy goals vis-a-vis the Blackstone ratio. When choosing the optimal regulation, authority's delegates (inspectors) have to take account of inspectee' strategic effort choices that impact evidence-generating process. As such, if inspection happens at an evidence-scarce stage where inspectees look similar, authority's guideline of tightening regulation threshold would paradoxically lead to lower conviction probability.

Normatively, my results lend support to Dragu (2011)'s argument that there does not necessarily exist a trade off between security and liberty. The strategic nature of the inspection game prevents any short-sighted executive or legislative response from success. In response to social turmoils, the authority has to take measures analyzing the fundamental causes of terrorist attacks, and put efforts going after crime-specific traits rather than implementing indiscriminate guidelines that infringes citizens' basic rights, in particular the freedom of speech.

# Appendix

# A Proofs

## A.1 Integrability of $\phi^{(n)}$

I state a few basic mathematical facts with respect to the derivative of normal density $\phi$ to simplify the characterization of the inspectees' best response. A function $f$ is Lebesgue integrable if $\int_{\mathbb{R}} |f| < \infty$.

Fact 1: $\phi^{(n)}$ is Lebesgue integrable.

Fact 2: $\int_{\mathbb{R}} |\phi''(x)| dx \leq 1 + \frac{1}{2\sqrt{2}}$

*Proof.*

$$\int_{\mathbb{R}} |\phi''(s)|ds = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |(s^2-1)e^{-s^2}|ds \leq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} s^2 e^{-s^2} ds + 1 = \frac{1}{2\sqrt{2}} + 1 < \infty$$

□

### A.2 Proof of Proposition 1

Before proving the proposition, we need to make sure that inspectees always have a well-defined objective. The following lemma does it.

**Lemma 6.** *Assume Assumptions 2. Inspectees always admit a well-defined best response correspondence.*

In the proof, I suppress all type-subscripts.

*Proof.* Recall that $G$ denote the (measurable) set of all evidence realization $s$ at which conviction would be convicted. For any effort $e$ chosen by the inspectee, he will be convicted if $s = -e + \epsilon \in G$. Hence, the probability that he is convicted $F(e)$ would be

$$\begin{aligned} F(e) &= \int 1\{s \in G\} d\Phi(\epsilon) = \int 1\{-e + \epsilon \in G\} d\Phi(\epsilon) \\ &= \int 1\{\epsilon \in G + e\} d\Phi(\epsilon) = \int_{G+e} \phi(\epsilon) d\epsilon \end{aligned}$$

$G + e$ is the right-shift of the set $G$ by $e$. Rewrite $D = G + e$ and thus $F(e) = \int_D \phi(s)ds$. Now let us characterize the shape of $F$.

**Lemma 7.** *For all $e \in \mathbb{R}_+, |F'(e)| \leq 2, |F''(e)| \leq 1 + \frac{1}{2\sqrt{2}}$*

*Proof.*

$$\begin{aligned} F'(e) &= \lim_{h \to 0} \int_D \frac{\phi(s+h) - \phi(s)}{h} ds \\ &= \int_D \lim_{h \to 0} \frac{\phi(s+h) - \phi(s)}{h} ds \\ &= \int_D \phi'(s) ds \\ &\leq \int_{\mathbb{R}} |\phi'(x)| dx = 2 \int_{-\infty}^{0} \phi'(x) dx = 2\phi(0) \end{aligned}$$

The second line follows from Lebesgue dominant convergence theorem (using $1_D|\phi'(s+\xi)|$ with $\xi \in (0, h)$ as the dominating function. This is feasible due to mean value theorem).

Similarly,

$$\begin{aligned}
F''(e) &= \lim_{h\to 0}\int_D \frac{\phi'(s+h)-\phi'(s)}{h}ds \\
&= \int_D \lim_{h\to 0}\frac{\phi'(s+h)-\phi'(s)}{h}ds \\
&= \int_D \phi''(s)ds \\
&\leq \int_{\mathbb{R}} |\phi''(s)|ds \leq 1+\frac{1}{2\sqrt{2}}
\end{aligned}$$

□

With these results, we know that inspectees' marginal gain in terms of favorable evidence-shift $F'$ is uniformly bounded by $2\phi(0)$. But their marginal cost to effort $H'(e)$ being convex satisfies

$$H'(e) \geq H'(0) + H''(0)e = H''(0)e$$

As such, the effort choice of inspectees is actually restricted to a compact set $[0, \bar{e}]$, with the upper bound $\bar{e}$ given by $2\phi(0)/H''(0)$. This is because outside this compact set, the net marginal gain is always negative[11]. Hence, the inspectees admit well-defined maximization problem.

Furthermore, if we impose Assumption 3-1, then inspectee's problem is **globally concave**, thus admitting a unique maximizer. □

With this observation, we can prove Proposition 1 easily.

*Proof.* Note that

$$\begin{aligned}
u_\theta(\sigma, e_\theta) =& Prob\{s_\theta \in G^c\} \\
=& Prob\{\epsilon \in G^c + e_\theta\} - H_\theta(e_\theta)\ \forall\theta \in \{I, C\}
\end{aligned}$$

$G^c$ is **independent** of $\theta$ due to inspector's imperfect observability of type. By Assumption 2-2, $u_\theta$ has the increasing difference condition defined in Milgrom and Shannon (1994). By theorem 3 in Milgrom and Shannon (1994) it has the single crossing property. Furthermore, being a real-valued function $u_\theta(\sigma, e_\theta)$ is supermodular in $e_\theta$.

With Assumption 3, apply the monotone selection theorem (Theorem 4') in Milgrom and Shannon (1994) to obtain $e_I^* \geq e_c^*$ as an elements of maximizer. By theorem 3 in Edlin and Shannon (1998), $e_I^* \neq e_c^*$. Combine two observations one obtains that $e_I^* > e_c^*$ .

□

[11] Note that net marginal gain is $F'(e) - H'(e) \leq 2\phi(0) - H'(e) \leq 2\phi(0) - H''(0)e \leq 0$ for all $e \geq \bar{e}$. Hence searching for $e^*$ within $[0, \bar{e}]$ is without loss.

### A.3 Proof of Lemma 2

I suppress type in this analysis. Consider inspectees' best response functions

$$\phi(s+e) = H'(e)$$

and SOC

$$\phi_i' < H''$$

where $\phi' = \phi'(s+e_i)$.

One can check the comparative statics of $e$ with respect to the change in $s$.

$$e'(s) = \frac{\phi'}{H'' - \phi'}$$

It is clear that $\phi'(x) < (>)0$ for all $x > (<)0$. Therefore one concludes

$$e'(s) = \begin{cases} -\frac{|\phi'|}{H''+|\phi'|} \in (-1,0) \text{ for } s+e \geq 0 \\ \frac{\phi'}{H''-\phi'} > 0 \text{ for } s+e < 0 \end{cases} \tag{3}$$

Now analyze the threshold change. Fix any triple $(s, e_I, e_c)$, $s + e(s)$ is increasing in $s$ because by (3), $e'(s) + 1 > 0$.

### A.4 Proof of Lemma 3

*Proof.* Consider equation (2). Define $s(\theta) = -(H_\theta')^{-1}(\phi(0))$. At $s(\theta)$, type $\theta$ will best respond by choosing $e_\theta^* = -s(\theta)$, resulting a density $\phi = \phi(0)$ which is maximized by single-peakedness. Since $s(I) < s(c)$, $\bar{s}$ corresponds to $s(c)$ with the induced $LR(\bar{s}) < 1$, and $\underline{s}$ corresponds to $s(I)$ with the induced $LR(\underline{s}) > 1$.

Turn to the second part. By Lemma 2, $s + e^*(s)$ is increasing in $s$. By definition of $\underline{s}, \bar{s}$, $\underline{s} + e_I(\underline{s}) = 0$ and $\bar{s} + e_c(\bar{s}) = 0$. Hence for any $s \in (\underline{s}, \bar{s})$, $s + e_I(s) > 0$ but $s + e_c(s) < 0$. Therefore, any equilibrium with $s \in (\underline{s}, \bar{s})$ is regular. □

### A.5 Indistinguishability Condition

**Lemma 8.** *Assuming 2, then indistinguishability condition is met*[12].

*Proof.* $e_I > e_c$ is immediate by Lemma 1. Note that for $s > \bar{s}$,

$$1 > LR(s) = \frac{\phi(e_I^* + s)}{\phi(e_c^* + s)} \geq \frac{\phi(e_I^* + s)}{\phi(s)} = \frac{\exp[-\frac{(s+e_I^*)^2}{2}]}{\exp[-\frac{s^2}{2}]} = \exp[-e_I^* s]\exp[-(e_I^*)^2/2]$$

Now let $s \to +\infty$. Note that $e_I^* \to 0$ as $s \to +\infty$ from (2). It suffices to show that $e_I^* s \to 0$ as $s \to +\infty$. In other words, $e_I^*$ decays faster than $\frac{1}{s}$.

[12] It also works for $s \downarrow -\infty$ by slightly modifying definition of indistinguishability.

Now bound the size of $se_I^*$. The first order condition for the inspectee is $\phi(s+e) = H'_\theta(e)$. By convexity of $H'_\theta$, $H'_\theta(e) \geq H'_\theta(0) + H''_\theta(0)e$, where $H''(0) > 0$. As such, for $s$ sufficiently large beyond $\bar{s}$

$$e_I^* \leq \frac{1}{H''_\theta(0)}\phi(s+e_I^*) \leq \frac{1}{H''_\theta(0)}\phi(s)$$

Hence

$$\lim_{s\to\infty} e_I^* s \leq \lim_{s\to\infty} s\frac{1}{\sqrt{2\pi}}\frac{1}{H''_\theta(0)}\exp[-\frac{s^2}{2}] = \frac{1}{\sqrt{2\pi}}\frac{1}{H''_\theta(0)}\lim_{s\to\infty}\frac{s}{\exp[s^2/2]} \to 0$$

□

**Lemma 9.** *If indistinguishability condition holds, then $\exists s_0 \geq \bar{s}$ such that $LR(s)$ attains its minimum.*

*Proof.* Pick $\epsilon < 1 - LR(\bar{s})$. By indistinguishability condition, $\exists s^*$ such that $LR(s) > 1-\epsilon$ for all $s \geq s^*$. By construction, $LR(s^*) > LR(\bar{s})$. Moreover, note that $LR$ is decreasing for $s \in (\underline{s}, \bar{s}]$. Hence, $LR(s)$ must attain its minimum $s_0 \in [\bar{s}, s^*]$. □

## A.6 Proof of Lemma 5

*Proof.* Note that $u_c$ is smooth with an interior maximizer, and $\frac{\partial u_c(e,b)}{\partial e}$ is increasing $b$. Now apply theorem 3 in Edlin and Shannon (1998) to get the monotone result. □

## A.7 Proof of Lemma 4

*Proof.* Suppose $b' > b$. Use $e_c^*(s), e_c'^*(s)$ to denote criminal's old/new best response, and $s^*, s'^*$ old/new equilibrium conviction threshold. Focus on the largest benign equilibrium in Case II.

Case I: Given condition 1 and monotonicity of $LR$ in regular equilibrium, it must be true that $s^*, s'^* < \underline{s'}$. By Lemma 5, for any $s$ we have $e_c'^*(s) > e_c^*(s)$. Use the definition of regular equilibrium, $s'^* + e'^*(s') < 0$. Hence we have $s'^* + e^*(s') < 0$. Since $\phi(x)$ is increasing in $x$ for $x < 0$, $\frac{\phi(s'^*+e_I^*(s'^*))}{\phi(s'^*+e_c'^*(s'^*))} = \alpha < \frac{\phi(s'^*+e_I^*(s'^*))}{\phi(s'^*+e_c^*(s'^*))}$.

Now suppose towards contradiction that $s'^* > s^*$. Let $s' \downarrow s$, then $\phi(s'^* + e_I^*(s'^*))$ increases but $\phi(s'^* + e_c^*(s'^*))$ decreases. This contradict $s^*$ being the equilibrium threshold under $b$.

Case II: Under $1 > \alpha > LR(\underline{s})$, it must be that $s'^*, s^* > \underline{s}$ in the benign equilibrium. Use the definition of benign equilibrium, $s^* + e^*(s^*) > 0$ and hence $s^* + e'^*(s^*) > 0$. Suppose towards contradiction that $s'^* < s^*$. $\frac{\phi(s^*+e_I^*(s^*))}{\phi(s^*+e_c^*(s^*))} = \alpha < \frac{\phi(s^*+e_I^*(s^*))}{\phi(s^*+e_c'^*(s^*))}$. However, this implies the existence of an $s'' > s^*$ such that $\frac{\phi(s''+e_I^*(s''))}{\phi(s''+e_c'^*(s''))} = \alpha$, contradicting the $s'^*$ being the largest extreme benign equilibrium. □

## A.8 Proof of Theorem 2

*Proof.* With Assumptions 2, Condition 1 implies that inspectee has a well-defined problem. Together with Assumption 3, it is unique. Monotonicity arguments follow, which establishes the optimality of BARD.

Single-peakedness of $f$ ensures Lemma 3. Now check the indistinguishability condition, or whether $LR(s) \to 1$ as $s \uparrow \infty$. Denote $L(x) = \log f(x)$. Since for $s$ sufficiently large under BARD, $e^* \to 0$. Therefore $s + e^* > 0$ and thus $f, L$ are both decreasing when $s \uparrow \infty$. It suffices to show that $L(s + e_I^*) - L(s + e_c^*) \to 0$ as $s \to 0$.

$$0 \geq L(s + e_I^*) - L(s + e_c^*) \geq L(s + e_I^*) - L(s) = L'(s + \xi(s))e_I^*, \quad \xi(s) \in (0, e_I^*)$$

By Equation (2) and proof of Lemma 8, $0 \leq e_I^* \leq \frac{1}{H_I''(0)} f(s)$. Hence,

$$\begin{aligned} L(s + e_I^*) - L(s + e_c^*) &\geq \frac{1}{H_I''(0)} L'(s + \xi) f(s) = \frac{1}{H_I''(0)} f'(s + \xi) \frac{f(s)}{f(s + \xi)} \\ &= \frac{1}{H_I''(0)} [P(s + \xi)] f(s) \end{aligned}$$

First note that we can without loss assume that $P$ is indeed a polynomial. This is because any term of $(s+\xi)$ with negative order is of the form $[\frac{1}{s+\xi}]^n$ which is bounded by some constant for $s$ sufficient large.

Next, recall in the proof of Proposition 1, we know that for any $\bar{e}$ there exists $\tilde{s}$ such that $s \geq \tilde{s}$ implies $e_I^* < \bar{e}$. This observation helps bound $s + \xi$. Now, I show that there exist polynomials $\bar{P}, \underline{P}$ such that $\underline{P}(s) \leq P(s + \xi) \leq \bar{P}(s)$ for all $s > 0$. Write $P(s + \xi) = a_0 + a_1(s + \xi) + ... + a_n(s + \xi)^n$. Note that since $s > 0, s + \xi > 0$, we can define $\bar{P}(s) = |a_0| + |a_1|(s + \bar{e}) + ... + |a_n|(s + \bar{e})^n$ and $\underline{P} = -\bar{P}$. Both are polynomials. Since $f \in S(\mathbb{R}), \lim_{s\to\infty} \bar{P}(s)f(s) = \lim_{s\to\infty} \underline{P}(s)f(s) = 0$, Hence, $\lim_{s\to\infty} P(s + \xi)f(s) = 0$. This proves indistinguishability condition.

Together with Lemma 3, Assumption 1 gives two crossing points of $\alpha$ with respect to $LR$ function. By MLRP, the inspector is best responding at these points.As such, Theorem 1 applies. Comparative static analysis follows. □

## Reference


Rudolf Avenhaus, Bernhard von Stengel, and Shmuel Zamir. Inspection games. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications*, volume 3, chapter 51, pages 1947–1987. Elsevier, 1 edition, 2002.

Sandeep Baliga, Ethan Bueno de Mesquita, and Alexander Wolitzky. Deterrence with imperfect attribution. *mimeo*, 2018. URL http://home.uchicago.edu/~bdm/PDF/deterrence.pdf.

Gary S. Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, 1968. doi: 10.1086/259394. URL https://doi.org/10.1086/259394.

William Blackstone. *Commentaries On the Laws of England.* Boston :Beacon Press, 1962.

Martha Crenshaw. *Terrorism in Context.* Penn State University Press, 1994.

Tiberiu Dragu. Is there a trade-off between security and liberty? executive bias, privacy protections, and terrorism prevention. *The American Political Science Review*, 105(1):64–78, 2011. ISSN 00030554, 15375943. URL http://www.jstor.org/stable/41480827.

Tiberiu Dragu. The moral hazard of terrorism prevention. *The Journal of Politics*, 79(1): 223–236, 2017. URL SSRN:https://ssrn.com/abstract=2983557.

Aaron S. Edlin and Chris Shannon. Strict monotonicity in comparative statics. *Journal of Economic Theory*, 81(1):201 – 219, 1998. ISSN 0022-0531. doi: https://doi.org/10.1006/jeth.1998.2405. URL http://www.sciencedirect.com/science/article/pii/S002205319892405X.

Gerald Folland. *Real Analysis: Modern Techniques and Their Applications.* Wiley; 2nd edition, 2007.

FreedomHouse. World press freedom index 2018, 2018. URL https://rsf.org/en/turkey.

Marcy Jones. Us airports step up security following turkey attack, 2016. URL http://www.fox10phoenix.com/news/arizona-news/us-airports-step-up-security-following-turkey-attack.

Louis Kaplow. On the optimal burden of proof. *Journal of Political Economy*, 119(6):1104–1140, 2011. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/10.1086/664614.

Louis Kaplow. Optimal multistage adjudication. *The Journal of Law, Economics, and Organization*, 33(4):613–652, 2017. doi: 10.1093/jleo/ewx007. URL http://dx.doi.org/10.1093/jleo/ewx007.

Samuel Karlin. *Total positivity, Volume 1.* Stanford University Press, 1968.

Jeffrey R. Lax. Political constraints on legal doctrine: How hierarchy shapes the law. *The Journal of Politics*, 74(3):765–781, 2012. doi: 10.1017/S002238161200045X. URL https://doi.org/10.1017/S002238161200045X.

David McAdams. Isotone equilibrium in games of incomplete information. *Econometrica*, 71 (4):1191–1214, 2005. doi: 10.1111/1468-0262.00443. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00443.

Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(1):157–180, 1994. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2951479.

Paul R. Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, 1981. ISSN 0361915X. URL http://www.jstor.org/stable/3003562.

Dilip Mookherjee and I. P. L. Png. Marginal deterrence in enforcement of law. *Journal of Political Economy*, 102(5):1039–1066, 1994. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/2138656.

Williams Pelegrin. Turkey throttles access to facebook and twitter following airport attacks, 2016. URL https://www.digitaltrends.com/social-media/facebook-twitter-turkey-airport-attack/.

Thomas Schelling. *The Strategy of Conflict*. Harvard University Press, 1960.

Ron Siegel and Bruno Strulovici. Judicial mechanism design. *mimeo*, 2018. URL http://faculty.wcas.northwestern.edu/~bhs675/JMD.pdf.

Francisco Silva. If we confess our sins. *mimeo*, 2018. URL https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxmcmFuY2lzY29zaWx2YTI5MDl8Z3g6NjZmZjBmNmE0MTY5MTc3Mg.

George Stigler. The optimum enforcement of laws. *Journal of Political Economy*, 78(3):526–36, 1970. URL https://EconPapers.repec.org/RePEc:ucp:jpolec:v:78:y:1970:i:3:p:526-36.

Michael M. Ting. Politics and administration. *American Journal of Political Science*, 61(2):305–319, 2017. doi: 10.1111/ajps.12273. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12273.