Predictive Policing: An Information Design Approach

Shota Ichihashi*

September 14, 2022

Preliminary Draft

Abstract

Citizens may engage in crime, depending on the probability of being searched and their types such as legal earning opportunities. Police observes information about the types and allocates search efforts to catch citizens who commit crimes. I show that the police who has full information about citizens may fail to deter any crimes. I then characterize an information structure that minimizes crime rate. The information structure is partially informative: It provides the police with information about citizens who face low returns to crime without identifying citizens who face high returns to crime. The resulting information structure may render policing unfair.

---

*Queen's University, Department of Economics. Email: s.ichihashi@queensu.ca.

1 Introduction

Law enforcement agencies increasingly adopt predictive policing algorithms, which aim to forecast crimes by using data on people and places, such as crime records, social media posts, financial records, and local environments (Perry, 2013). For example, the "strategic subjects list," which the Chicago Police Department used between 2012 and 2020, listed people who are deemed likely to be involved in gun violence. Along with concerns about fairness, privacy, and transparency, the potential of data and algorithms to transform policing—for good or bad—has been actively discussed.

The use of data and algorithms for law enforcement raises the following question: How does the information available to law enforcement agencies affect the effectiveness of policing? To tackle the question, I introduce an information design approach to a stylized model of policing. The model consists of a police and a unit mass of citizens. The police chooses the probability to search each citizen, and simultaneously, citizens decide whether to commit crime. Citizens differ in their legal earning opportunities ("types"). A citizen's net return to crime is decreasing in their type and the probability of being searched. The police searches citizens to maximize the probability of uncovering criminals subject to a search capacity constraint.[1]

Before allocating search effort, the police observes information about citizens' types. Information is modeled as a segmentation. For example, a segmentation may partition the set of types into intervals and reveal which interval each citizen's type belongs to. Another example is a segmentation the reveals types below some cutoff and pools all types above the cutoff. The police will then learn the exact types of citizens whose net returns to crime exceed the cutoff. I consider arbitrary segmentations and study how the equilibrium depends on the information available to the police.

To begin with, I consider the police who knows the type of every citizen. The fully informed police may not attain a low crime rate, because the police's objective is to catch criminals. Indeed, unless the police's search capacity is high enough, the fully informed police fails to deter any crimes. Intuitively, the fully informed police will only search citizens who

---

[1]Section 2.3 motivates the assumption that the police maximizes the chance of uncovering crimes.

are most likely to commit crimes. Under a limited search capacity, such a policing strategy turns out to maximize crime rate: Each citizen is not searched at all and thus commits crime, or the citizen is searched but still commits a crime because the returns to crime are sufficiently high relative to the search intensity. As a result, unrestricted data collection may defeat its purpose of deterring crimes.

I then characterize an information structure that minimizes the equilibrium crime rate. I call it a crime minimizing segmentation. The segmentation consists of a continuum of segments. Each segment contains a non-convex set of types, so the police may not distinguish citizens who face very high and low returns to crime. At the same time, the segmentation provides nontrivial information: Each segment differs in the highest possible type it contains; in equilibrium, the police searches each segment with a probability that just deters the crime of citizens who have the lowest incentive to commit crime in that segment. Overall, the crime minimizing segmentation illustrates the benefits and costs of informing the police.

The characterization result has two implications. First, the police may need to restrict the use of information when the environment is prone to crime. Specifically, the crime minimizing segmentation is less informative when the police has a lower search capacity or the citizens face higher returns to crime. Second, the equilibrium outcome reveals more information about citizens who face low returns to crime than about citizens who face high returns to crime. This property contrasts with the kind of predictive policing discussed in the public debate, whose purpose is to identify "high-risk" individuals or places (Perry, 2013). In an example, I show that the equilibrium crime rate may increase if the police has detailed information about citizens who face high returns to crime.

Finally, I study how information affects the fairness of policing. I show that the crime minimizing segmentation is "unfair." Roughly, it means that if citizens in one group tend to face lower incomes than citizens in the other group, the lower-income group is exposed to higher average search intensities. The result points to the tension between fairness and effectiveness of policing. In an extension, I characterize a segmentation that minimizes crime rate subject to the constraint that policing is fair, i.e., different groups of citizens are exposed to the same search intensity on average.

A predictive algorithm in practice may perform poorly for two reasons: (i) the algorithm

is fed with data that suffer from econometric endogeneity problem, and (ii) the algorithm does not respect the incentives of the final decision maker, a human (Ludwig and Mullainathan, 2021). In this paper, I focus on Concern (ii) but not on (i). In particular, the model abstracts away from an often criticized feature of predictive policing—that data are generated by past policing decisions, which may exacerbate existing bias in policing (Ensign et al., 2018). Therefore this paper is best viewed as an exercise to examine the potential and limitation of data-driven policing that overcomes Concern (i) but are subject to Concern (ii).

The contribution of the paper is two-fold. First, the paper provides an intuition for how law enforcement agencies should restrict the use of information on individuals and places. In particular, for maximum crime deterrence, law enforcement agencies may need to do the opposite of what predictive policing algorithms deployed in the past aimed for—i.e., agencies should design information to prevent the police from targeting resources to individuals and places that are prone to crime. Second, the paper contributes to the discussion of fairness and policing by highlighting the role of information design to improve or worsen the fairness of policing.

Related work. First, the paper relates to the economic literature on crime and policing, which starts from Becker (1968). The paper is closely related to Persico (2002), who studies whether requiring the police to behave more fairly reduces crime. The key assumption of my model—that the police allocates limited search resources to catch criminals—follows his paper. My analysis uses information design and offers several new insights. First, the information structure that minimizes crime is qualitatively different from the policing strategies in Persico (2002) or the kind of predictive policing discussed in the public debate. Thus my paper offers a new intuition about how information affects the effectiveness of policing. Second, among other things, Persico (2002) finds that there may be no trade-off between fairness and effectiveness of policing. I complement this observation by showing that the trade-off generally exists once we focus on the information structure that is most effective at deterring crimes. The question of what information about citizens should (not) be used for policing is often discussed in the context of racial profiling (Knowles et al., 2001; Persico and Todd, 2005; Bjerk, 2007; Persico, 2009). Law enforcement agencies will likely be collecting

data that go beyond citizens' coarse characteristics, which renders it important to study the general relation between information and policing. To focus on the role of information in policing, the model abstracts away from other important considerations, such as the design of judicial systems, richer responses by potential criminals, and the behavior of potential victims (e.g., Curry and Klumpp 2009; Cotton and Li 2015; Vasquez 2022).

Second, the paper relates to the literature on Bayesian persuasion and information design (see, e.g., Kamenica (2019) Bergemann and Morris (2019) for surveys). Papers such as Lazear (2006), Rabinovich et al. (2015), and Hernández and Neeman (2022) use Bayesian persuasion (or related tools) to study what information to provide to agents who may take socially undesirable actions. In contrast, I study what information about agents' preferences to disclose to the police, who monitors agents. The information design literature provides conditions under which an optimal signal takes a tractable form, such as monotone partitional signals, censorship policies, and nested intervals (e.g., Guo and Shmaya 2019; Dworczak and Martini 2019; Kolotilin et al. 2022). The crime minimizing information structure does not belong to these classes of signals, and to the best of my knowledge, the standard techniques to solve Bayesian persuasion do not apply to my model. The paper also relates to the computer science and economic literature on algorithmic fairness. Jung et al. (2020) relate algorithmic fairness to policing, and Liang et al. (2022) study the general trade-off between fairness and accuracy for the design of classification algorithms.

2 A Model of Predictive Policing

2.1 Setup

There is a police and a unit mass of citizens. Citizens decide whether to commit a crime, where the net returns to crime depend on their heterogeneous types. The police observes information about citizens' types and then allocates a limited search capacity across citizens to catch criminals.

The formal description is as follows. Each citizen knows his legal earning opportunity $x \in [0, 1]$, which is distributed across the population according to a distribution function

F_0 with a positive density f_0. The value x is a citizen's *type*. Let $\mathcal{F}$ denote the set of all cumulative distribution functions on $[0,1]$.[2] For any $F \in \mathcal{F}$ and event A such as $x \geq x^*$, let $F(\cdot|A)$ and $\mathbb{E}_F[\cdot|A]$ respectively denote the distribution function and the expectation operator conditional on that type x satisfies event A. Finally, let $F^{-1}(c) \triangleq \inf\{x \in [0,1] : c \leq F(x)\}$ denote the quantile function of $F \in \mathcal{F}$.

The police has some information about citizens' types. The police's information is described by a *segmentation*, which is a $\mu \in \Delta(\mathcal{F})$ such that

$$\int_{\mathcal{F}} F(x)\mu(dF) = F_0(x),$$

for all $x \in [0,1]$. A *segment* refer to any distribution $F \in \operatorname{supp}\mu$, where $\operatorname{supp}\mu \subset \mathcal{F}$ is the support of μ. Where it does not cause confusion, we write segment $[a,b]$ instead of segment $F_0(\cdot|x \in [a,b])$. A segmentation is a primitive of the game. Each citizen knows his type and the segment he belongs to. The police does not observe types but knows which segment each citizen belongs to.

One example of a segmentation is *no segmentation*, i.e., μ places probability 1 on prior distribution F_0. No segmentation captures the police who has no information beyond the prior distribution of citizen types. *Perfect segmentation,* which informs the police with the exact type of every citizen, is the segmentation μ whose support consists of $\{\delta_x\}_{x \in [0,1]}$, where δ_x places probability 1 on $\{x\}$. Alternatively, the police may have partial information, such as whether a citizen's type is above some cutoff x^*. The corresponding segmentation consists of distributions $F_0(\cdot|x \geq x^*)$ and $F_0(\cdot|x < x^*)$.

Given a segmentation, the police allocates search efforts; simultaneously, each citizen decides whether to commit a crime. Formally, the police chooses a *search strategy* $\sigma :$ $\operatorname{supp}\mu \to [0,1]$, where the police searches citizens in segment $F \in \operatorname{supp}\mu$ with probability $\sigma(F)$. The police has a measure $\overline{S} \in (0,1)$ of searches to allocate; thus the police can choose a search strategy σ if and only if

$$\int_{\mathcal{F}} \sigma(F)\mu(dF) \leq \overline{S}.$$

---

[2] $\mathcal{F}$ is endowed with the weak* topology and the induced Borel σ-algebra.

Each citizen then chooses whether to commit a crime without observing the search strategy. A citizen's payoff of not committing crime equals his type x. If a citizen commits crime but is not searched, his payoff is 1. If a citizen commits crime and is searched, his payoff is $-L < 0$.[3] Thus the expected payoff of any citizen who belongs to segment F and commits a crime is $1 - \sigma(F) - L\sigma(F)$, and a citizen with type x prefers to commit crime if $1 - \sigma(F) - L\sigma(F) \geq x$.[4]

The police wants to maximize the number of successful searches: Given a strategy profile, let $c(x, F)$ denote the probability with which a citizen with type x in segment F commits a crime. The payoff to the police is $\int_{\mathcal{F}} \left(\int_0^1 c(x, F) F(dx) \sigma(F) \right) \mu(dF)$, i.e., the mass of citizens who commit a crime and are searched. The police's payoff is different from the *crime rate*, $c \triangleq \int_{\mathcal{F}} \left(\int_0^1 c(x, F) F(dx) \right) \mu(dF)$.

We study how the information available to the police affects the Bayesian Nash equilibrium of the game. In particular, we characterize a *crime minimizing segmentation*, i.e., a segmentation that permits an equilibrium that has a lower crime rate than any equilibrium under any segmentation. Such an equilibrium is called a *crime minimizing equilibrium.*

2.2 Example

The following example illustrates the basic incentives of the police and citizens.

Example 1. Suppose that $\overline{S} = 0.3$, $L = 0.5$, and types are uniformly distributed on $[0, 1]$, i.e., $F_0(x) = x$. Without segmentation, the police can search citizens randomly with probability $\overline{S} = 0.3$. A citizen will commit crime if and only if $(1 - \overline{S}) - \overline{S}L = 0.55 \geq x$, so the equilibrium crime rate is $F_0(0.55) = 0.55$.

Suppose now that the police faces a binary segmentation that reveals whether a citizen's type is $x < 0.5$ (i.e., high-risk group) or $x \geq 0.5$ (i.e., low-risk group). Let σ_0 and σ_1 be the search probabilities for segment $[0, 0.5)$ and segment $[0.5, 1]$, respectively. The equilibrium search strategy is determined by two conditions: First, the crime rates in both segments must be equal; for example, if more citizens in segment $[0, 0.5)$ were to commit crime than

---

[3]Proposition 1 and Theorem 1 extend to the case in which citizens are heterogeneous not only in the payoffs from not committing crime but also in the benefits of successful crime and the costs of being caught.

[4]Because $0 \leq x \leq 1$, all citizens commit crime if $\sigma = 0$ and no citizen commits crime if $\sigma = 1$. Introducing citizens whose behaviors do not depend on a search strategy—e.g., citizens with $x < 0$ or $x > 1$—complicates the exposition but does not change the main result.

citizens in $[0.5, 1]$ in equilibrium, then the police, who cares about catching criminals, would deviate by decreasing σ_1 and increasing σ_0. The police's indifference condition is written as $F_0(1 - 1.5\sigma_0 | x < 0.5) = F_0(1 - 1.5\sigma_1 | x \geq 0.5)$, which is written as $\sigma_0 - \sigma_1 = \frac{1}{3}$. Second, the police's search strategy must satisfy the capacity constraint, i.e., $0.5\sigma_0 + 0.5\sigma_1 = 0.3$. By solving these two equations, we obtain $\sigma_0 = \frac{7}{15}$ and $\sigma_1 = \frac{2}{15}$. The resulting equilibrium crime rate is 0.6, which is greater than the equilibrium crime rate under no segmentation.

In this example, the crime rate is higher when the police has more information. This is not because the information is coarse: On the contrary, if the police has full information then every citizen will commit crime (Proposition 1). At the same time, it is not the case that any information increases crime. As we show in Section 4.2, the crime minimizing segmentation in Theorem 1 will attain crime rate 0.1.

2.3 Discussion of the Model

Before proceeding to the analysis, I motivate and interpret the model.

Interpretation of a segmentation. A segmentation has two interpretations. First, it may capture some statistical signals of citizens' types, such as their demographic characteristics. For example, each citizen is described by a covariate v that is correlated with his type. If the police tailor her search effort to v then it is equivalent to the police facing a segmentation such that citizens belong to the same segment if and only if they share the same value of v. Second, a segmentation may represent a predictive policing algorithm that recommends an action (i.e., search probabilities) to the police.[5] The assumption that a predictive policing algorithm has to respect the police's incentive—instead of directly controlling the police's behavior—reflects reality: Several papers document that police officers and their commanders have discretions about whether to follow an algorithm's recommendation (Brayne, 2017; Kapustin et al., 2022).[6] For example, Brayne (2017), who conducts the case study of the

---

[5]This interpretation uses the standard argument of Bayesian persuasion that we can without loss of generality replace messages with action recommendations (Kamenica and Gentzkow, 2011).

[6]Kapustin et al. (2022) focus on the Chicago police, which adopted HunchLab's predictive policing tool. They document "substantial variation across districts in the degree to which officers spend more time in HunchLab-flagged boxes." This observation suggests that an algorithm merely provides information and does not automate the deployment of police resources.

Los Angeles Police Department, documents that how long police officers stay in the area designated as high-risk by the algorithm remains within their discretion.

The police's incentive.   We assume that the police maximizes the chance of uncovering crimes instead of minimizing a crime rate. Empirically, Stashko (2020) supports this assumption in the context of the sale of illegal drugs in the US. Moreover, theory and empirical papers adopt the same assumption (Knowles et al., 2001; Persico, 2002; Hernández-Murillo and Knowles, 2004; Persico and Todd, 2005; Antonovics and Knight, 2009) so it would be a natural starting point. Theoretically, the assumption holds even if the police cares about both crime rate and catching criminals so long as the officer has a negligible impact on aggregate crime or cannot commit upfront to a policing strategy. In such a case, the police takes a crime rate as exogenous and acts as if she only cares about catching criminals. Lastly, even though my model does not have an information designer, we may view our exercise as an information design problem in which the designer (e.g., the police department or the society) wants to reduce crime and provides information to police officers who conduct day-to-day patrols to uncover crimes.

Interpretation as place-based predictive policing.   According to Lau (2020), place-based predictive policing aims to identify places that have a high risk of crime, and person-based predictive policing aims to identify individuals who are likely to commit crime. I describe the model as a model of person-based predictive policing; however, the model also fits place-based predictive policing with the following interpretation: There is a unit mass of locations, $[0, 1]$. Each location $\ell \in [0, 1]$ has a random crime opportunity that determines the payoff v_ℓ from successful crime (that is not caught by the police) at location ℓ. There is a representative citizen who decides whether to commit crime in each location, and the citizen's payoffs are additively separable across his payoffs at different locations. The police learns information about $(v_\ell)_{\ell \in [0,1]}$ and decides the probability to search locations. This model captures place-based predictive policing and is equivalent to my model.

3 The Fully Informed Police

I begin with the benchmark in which the police knows the exact propensity of each citizen to commit a crime.

Proposition 1. *If $\overline{S} < \frac{1 - \mathbb{E}_{F_0}[x]}{1 + L}$, the perfect segmentation attains crime rate 1. If $\overline{S} > \frac{1 - \mathbb{E}_{F_0}[x]}{1 + L}$, the perfect segmentation attains crime rate 0.*

The first part states that when the search capacity is low relative to the returns to crime, the fully informed police fails to deter any crimes. Proposition 1 relies on the following lemma, which we will use throughout the analysis.

Lemma 1. *For any segmentation μ, in any equilibrium, almost all segments have the same crime rate.*

The intuition is simple: The police, who aims to catch criminals, will not search a segment that has a lower crime rate than other segments. However, if the police does not search, the segment would have crime rate 1. As a result, no equilibrium has a segment that has a lower crime rate than other segments, i.e., all segments must have the same crime rate.

We can interpret Lemma 1 as follows: Once the police tries to use some variable for predicting citizens' behavior, the variable loses its predictive power.[7] To see this, suppose that each citizen i has some observable characteristic $v_i \in V$. In general, if a citizen's likelihood of committing a crime depends on his type x and the type is correlated with his characteristic v_i, the police can use v_i to predict the probability that citizen i will commit a crime. However, this argument fails in equilibrium: Consider the segmentation generated by the characteristics, i.e., two citizens i and j belong to the same segment if and only if $v_i = v_j$. Lemma 1 then implies that if the police correctly predicts the average behavior of citizens conditional on each $v \in V$ and acts optimally (and citizens best respond), then the likelihood of crime must be independent of $v \in V$. Thus the police cannot use a covariate for meaningfully predicting the behavior of citizens while using the prediction for policing.

The discussion points out that in this model, the value of predictive policing comes from the prediction of citizens' preferences, not their behavior. At the same time, Proposition 1

---

[7]The lemma itself is not new—e.g., it is a generalization of an equilibrium condition in Persico (2002).

says that perfectly predicting every citizen's type may backfire, as we prove now.

Proof. Type x does not commit crime if $x \geq 1 - \sigma - \sigma L$, or equivalently, if $\sigma \geq \sigma(x) \triangleq \frac{1-x}{1+L}$. Inequality $\overline{S} < \frac{1 - \mathbb{E}_{F_0}[x]}{1+L}$ implies that searching type x with probability at least $\sigma(x)$ will violate the capacity constraint. Thus in any equilibrium, some positive measure set of types will commit crime. Under the perfect segmentation, each segment contains a single type. Lemma 1 then implies that almost all types commit crime with probability 1. An equilibrium with crime rate 1 exists: Take a $z \in [0,1]$ such that $\overline{S} = \int_0^z \sigma(x) dF$. It is an equilibrium that the police searches citizens $x \leq z$ with probability $\sigma(x)$ and citizens $x > z$ with probability 0, and all citizens commit crime with probability 1.

If $\overline{S} > \int_0^1 \sigma(x) dF = \frac{1 - \mathbb{E}_{F_0}[x]}{1+L}$, the police can search every type x with at least probability $\sigma(x)$ to deter all crimes. To show any equilibrium has crime rate 0, suppose to the contrary that some equilibrium has a positive crime rate, $c > 0$. Lemma 1 implies that every citizen commits crime with probability c. But inequality $\overline{S} > \int_{[0,1]} \sigma(x) dF$ implies that the police will search some citizens with probability strictly greater than $\sigma(x)$, which means that these citizens prefer to not commit crime. Thus we obtain a contradiction. $\square$

Proposition 1 shows that unrestricted data collection may defeat its purpose of deterring crimes: When the search capacity is limited relative to returns to crime, some citizens will commit crime regardless of the police's search strategy. The fully informed police knows exactly which citizens will commit crime and thus allocates all search efforts to criminals. At the same time, other citizens, whom the police does not search, also find it profitable to commit crime. As a result, the only equilibrium outcome is that all citizens commit crime: A citizen is not searched and thus commits crime or is searched with a positive probability but still commits crime with probability 1, being indifferent between committing and not committing a crime.

The result implies that for policing to lower crime, the police may need to have noisy information about citizens. However, as in Example 1 some information may increase crime relative to no information. Thus we turn to the question of what information the police should have for maximum crime deterrence.

10

4 A Crime Minimizing Segmentation

To characterize a crime minimizing segmentation, we first define a class of segmentations that contain the crime minimizing segmentation.

Definition 1. For any $z \in [0,1]$, the *z-segmentation* is a segmentation whose support consists of a continuum of segments indexed by $i \in [z,1]$, where for each index i, segment G_i is defined by

$$G_i(x) \triangleq \begin{cases} F_0(x) & \text{if} \quad x \leq z, \\ F_0(z) & \text{if} \quad z < x < i, \\ 1 & \text{if} \quad i \leq x \leq 1. \end{cases}$$

The probability measure μ^z over segments $\{G_i\}_{i \in [z,1]}$ is defined by $\mu^z (\{G_i : i \leq x\}) \triangleq \max\left\{ \frac{F_0(x) - F_0(z)}{1 - F_0(z)}, 0 \right\}$ for all $x \in [z,1]$.

Figure 1 depicts a z-segmentation. The segmentation consists of a continuum of segments, $\{G_i\}_{i \in [z,1]}$. Every segment $G_i \in \operatorname{supp} \mu^z$ contains a mass $F_0(z)$ of citizens whose types are distributed according to $F_0(\cdot | x < z)$. The remaining mass $1 - F_0(z)$ of citizens have the identical type i, which is also the index of the segment. As a result, each G_i has a non-convex support and has neither a density nor a probability mass function. The segments $\{G_i\}_{i \in [z,1]}$ are distributed according to $F_0(\cdot | x \leq z)$, which ensures that any z-segmentation averages to the prior type distribution, F_0.[8]

---

[8]For all $x \leq z$ we have $\int_{\mathcal{F}} G(x)\mu^z(dG) = F_0(x)$, and for all $x > z$ we have $\int_{\mathcal{F}} G(x)\mu^z(dG) = F_0(z) + (1 - F_0(z))\frac{F_0(x) - F_0(z)}{1 - F_0(z)} = F_0(x)$.
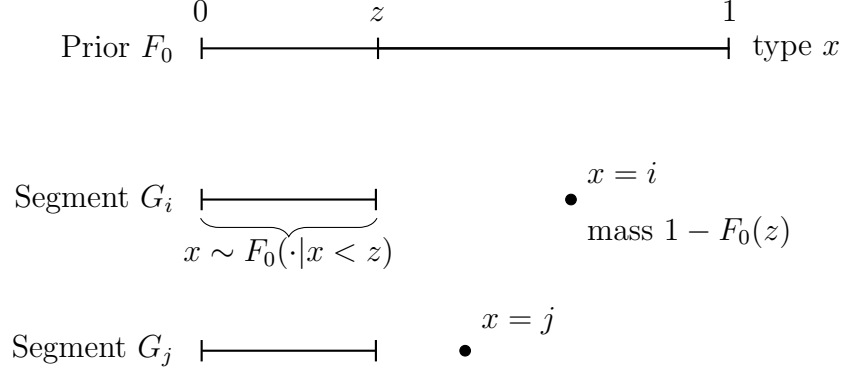
Figure 1: z-segmentation

A z-segmentation is different from the so-called lower censorship policy, which pools all types below some cutoff z and reveals every type above z (e.g., Kolotilin et al. 2022). Under any lower censorship policy, each segment is either δ_x—i.e., a distribution that is degenerate at x—or the truncated distribution $F_0(\cdot|x < z)$. In contrast, any segment in a z-segmentation is a convex combination of δ_x and $F_0(\cdot|x < z)$. The following result characterizes a crime minimizing segmentation (see Appendix B for the proof.)

Theorem 1. *If* $\overline{S} < \frac{1 - \mathbb{E}_{F_0}[x]}{1 + L}$, *the* z^*-*segmentation minimizes crime rate, where* $z^* > 0$ *uniquely solves*

$$\overline{S} = \frac{1 - \mathbb{E}_{F_0}[x|x \geq z^*]}{1 + L}. \tag{1}$$

The minimized equilibrium crime rate is $F_0(z^*)$.

Example 2 (Example 1 continued). Because types are uniformly distributed, we have $\mathbb{E}_{F_0}[x|x \geq z] = \frac{1+z}{2}$. Solving equation (1) we obtain $z^* = 1 - 2\overline{S}(1+L) = 0.1$ (recall $\overline{S} = 0.3$ and $L = 0.5$). Thus the crime minimizing segmentation decreases crime rate from 0.55 to 0.1. The segmentation consists of a continuum of segments, $\{G_i\}_{i \in [0.1,1]}$. Segment G_i consists of mass 0.9 of citizens whose types are i and mass 0.1 of citizens whose types are uniformly distributed on $[0, 0.1]$.

In the crime minimizing equilibrium, the police searches segment $G_i \in \operatorname{supp} \mu^z$ with probability $\sigma(i) = \frac{1-i}{1+L}$. The police's search deters fraction $1 - F_0(z^*)$ of citizens, who has type i, from crime, because type i is indifferent between committing and not committing a crime given search probability $\sigma(i)$. The remaining fraction $F_0(z^*)$ of citizens, whose types

are below i, will commit crime. As a result, every segment G_i has crime rate $F_0(z^*)$. The police's search strategy is optimal, because every segment G_i has the same crime rate $F_0(z^*)$ and the police exhausts search capacity $\overline{S}$ because of equation (1). Thus the z^*-segmentation has an equilibrium with crime rate $F_0(z^*)$. When we have $\overline{S} - \frac{1-\mathbb{E}_{F_0}[x]}{1+L} < 0$, the equilibrium crime rate is positive.

Intuitively, the z^*-segmentation minimizes crime by (i) adding noise to incentivize the police to search citizens who are responsive to policing and (ii) revealing information that enables the police to avoid the over-policing of citizens. To minimize crime rate, the police should allocate search effort to citizens who have high types x, because they face low returns to crime and will choose not to commit crime at a relatively low search intensity. However, the police is motivated to catch criminals. To incentivize such a police to search citizens with high x, each segment G_i pools citizens who respond to policing (i.e., type i) with citizens who will commit crime (i.e., types below z^*). At the same time, segments $\{G_i\}_{i \in [z^*, 1]}$ differ in the highest type i, and the police will search segment G_i with a probability that is just enough to deter type i from committing crime. Thus the z^*-segmentation provides the police with information that prevents over-policing.[9]

The model still allows the case in which the fully informed police attains the minimum crime rate (i,e., when $\overline{S} \geq \frac{1-\mathbb{E}_{F_0}[x]}{1+L}$). However, the benefit of keeping the police uninformed becomes more salient when the police can adjust search capacity at cost. In Section 6, we study an extension in which the police incurs an increasing convex cost $C(\overline{S})$ to choose $\overline{S}$. In this case, the condition $\overline{S} < \frac{1-\mathbb{E}_{F_0}[x]}{1+L}$ endogenously arises, and the crime minimizing segmentation is partially informative as described in Theorem 1.

Informativeness of the crime minimizing segmentation. One implication of Theorem 1 is that the police should have less information when the environment is prone to crime. To formalize the idea, we define the following notion.

Definition 2. Fix the prior type distribution F_0, and take parameters $(\overline{S}_0, L_0)$ and $(\overline{S}_1, L_1)$. We say that *parameter $(\overline{S}_1, L_1)$ is more prone to crime than parameter $(\overline{S}_0, L_0)$* whenever

---

[9]Formally, there is no over-policing in the crime minimizing equilibrium in the sense that no citizen will be searched with a probability strictly greater than the probability that makes the citizen indifferent between committing and not committing crime.

$1 - \overline{S}_1 - \overline{S}_1 L_1 > 1 - \overline{S}_0 - \overline{S}_0 L_0$, or equivalently, when the equilibrium crime rate under no segmentation is higher at $(\overline{S}_1, L_1)$ than at $(\overline{S}_0, L_0)$.

A parameter becomes more prone to crime if search capacity $\overline{S}$ and loss L from being caught decrease. However, Definition 2 is more general because it allows parameter changes such that one of $\overline{S}$ and L decreases whereas the other increases.

Corollary 1. *If parameter $(\overline{S}_1, L_1)$ is more prone to crime than parameter $(\overline{S}_0, L_0)$ then:*

1. *the equilibrium crime rate under the crime minimizing segmentation is greater under $(\overline{S}_1, L_1)$ than $(\overline{S}_0, L_0)$; and*

2. *the crime minimizing segmentation is less informative in the sense of Blackwell (1951, 1953) under $(\overline{S}_1, L_1)$ than $(\overline{S}_0, L_0)$.*

Part 1 states that any change of parameter $(\overline{S}, L)$ that increases the baseline crime rate (under no segmentation) will also increase the minimized crime rate. In general (say) a higher $\overline{S}$ and a lower L will push the equilibrium crime rate to the opposite directions; Part 1 enables us to examine the impact of such a change without solving the crime minimization problem for each parameter.

We might think that the police benefits more from information when the police's resources are limited or the citizens face high crime rate—e.g., a police chief quoted in Pearsall (2010) states that "predictive policing is the perfect tool to help departments become more efficient as budgets continue to be reduced." Part 2 offers a cautionary tale: If the environment is prone to crime, the police should be endowed with less information. A key assumption is that the police wants to catch criminals. When the environment is prone to crime, the police is inclined to spend her resource on uncovering crimes than deterring crimes, because the latter involves monitoring citizens who will not commit crime on the equilibrium path. The crime minimizing segmentation prevents such a distortion by adding more noise.

4.1   Information and the Allocation of Search Capacity

We study the crime minimizing equilibrium in terms of how the police allocates search efforts across the population. Given a search strategy, let $\sigma(x)$ be the probability with which

the police searches citizen type $x \in [0,1]$. Assuming that $\overline{S} < \frac{1-\mathbb{E}_{F_0}[x]}{1+L}$, we depict search probabilities for three outcomes. In Figure 2, the black solid line refers to the equilibrium search probabilities under no segmentation. The police with no information can only search each citizen randomly with probability $\overline{S}$. As a result, citizen type x^N that satisfies $x^N = 1 - \overline{S} - \overline{S}L$ becomes indifferent, and thus any citizen type $x < x^N$ will commit crime.

The red thick line refers to the crime minimizing equilibrium. Compared to no segmentation, the police now searches types in $[z^*, x^N]$ with higher probabilities and types in $[x^N, 1]$ with lower probabilities. The search probability for types below z^* remains the same, so the resulting search probability is non-monotone in types. The resulting search strategy is

$$
\sigma^C(x) = \begin{cases} \overline{S} & \text{if} \quad x < z^* \\ \frac{1-x}{1+L} & \text{if} \quad x \geq z^*. \end{cases} \tag{2}
$$

By tailoring search probabilities to each citizen type, the police can save search efforts and expand the set of citizens who do not commit crime.

Finally, the blue dashed line refers to the commitment strategy—i.e., the search strategy the police would choose if the police could use full information and commit to any search strategy before citizens make decisions. In order to maximize the mass of innocents, the commitment strategy never searches citizens whose types are below cutoff x^F, because deterring them from crime is costly. The commitment strategy implies that the police only searches citizens who do not commit crime, so it cannot be part of equilibrium.
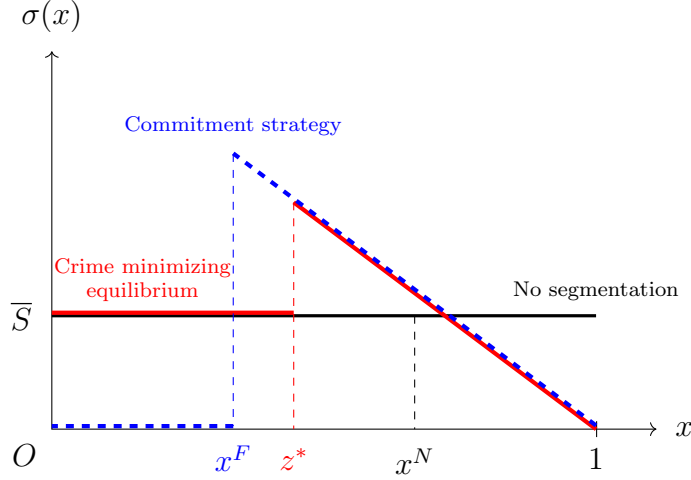
Figure 2: The average search probabilities

4.2 Comparison to the Predictive Policing "in the Wild"

The crime minimizing segmentation has a starkly different informational property from the kind of predictive policing described in the public debate. I illustrate the difference in an example. We use the same parameters as Example 1, i.e., the type distribution is uniform, and we have $\overline{S} = 0.3$ and $L = 0.5$.

Predictive policing is often described as a way to identify individuals or places that are prone to crime (Perry, 2013; Lau, 2020). To capture this idea, let us consider a segmentation that provides the police with detailed information about citizens who face high returns to crime. The resulting segmentation is an upper-censorship segmentation, which reveals the types below a cutoff and pools the types above the cutoff. As an example, suppose that a segmentation reveals the types of citizens if $x < 0.2$ and pools all citizens with $x \geq 0.2$. In equilibrium, every citizen with $x < 0.2$ commits a crime with the same probability $c^U \in (0, 1)$, which implies that citizens with type x must face search probability $\frac{1-x}{1+L}$.[10] As a result, the search capacity allocated to the set of types $[0, 0.2]$ is $\int_0^{0.2} \frac{1-x}{1+L} = 0.12$, and the search capacity for segment $[0.2, 1]$ is $\overline{S} - 0.12 = 0.18$. The mass of segment $[0.2, 1]$ is 0.8, so each citizen in the segment is searched with probability $0.18/0.8 = 0.225$. The resulting crime rate in

---

[10]Crime rate c^U cannot be 0 because the minimum crime rate is 0.1 as shown in Example 2. Crime rate c^U cannot be 1 because the police will search segment $[0, 2, 1]$ with a positive probability, in which case types close to 1 will not commit crime.

segment $[0.2, 1]$ is $c^U = F_0(1 - 0.225(1 + L)|x \geq 0.2) = 37/64 \approx 0.58$. In equilibrium all segments have the same crime rate; thus the equilibrium crime rate under the upper-censorship segmentation is approximately 0.58, which is greater than the equilibrium crime rate under no segmentation (which is 0.55).

Enabling the police to identify individuals who are prone to crime may render policing less effective, because the information distorts the allocation of search efforts and increases the equilibrium crime rate. This idea echoes the concern in, e.g., Brayne (2017):

> Directing resources toward people and places statistically more likely to be associated with criminal activity increases the probability that such people (and people in such places) will be caught if they do something wrong, while reducing the probability of discovering and prosecuting wrongdoing by other people in other locations—the ones from whom the algorithms distract police attention (p. 109).

The crime minimizing segmentation—which attains crime rate 0.1 in this example—differs from the upper-censorship segmentation in two aspects. First, the crime minimizing segmentation does not allow the police to identity citizens with high returns to crime. Specifically, every segment contains types below 0.1 and type above 0.1. By pooling citizens with high and low types, the segmentation prevents the over-policing of citizens who have low types. Second, the crime minimizing segmentation lets the police to tailor search effort to the types above a cutoff, which is 0.1 in the example. This property contrasts with the upper-censorship segmentation, which lets the police to tailor search probabilities to types below cutoff 0.1.

The analysis of the crime minimizing segmentation has the following implication. First, a predictive policing algorithm that aims to identify "at-risk" individuals or places might backfire by distorting the allocation of policing resource. The current public debate makes a similar claim based on the idea that algorithms may use data that reflect existing bias in policing. The analysis of this paper shows that the same concern may apply even if an algorithm is implemented without the issues of bad data. Second, from the perspective of crime deterrence, an effective predictive policing algorithm may do the exact opposite by (i)

concealing information about individuals or places that are highly prone to crime and (ii) providing information about those that are moderately responsive to policing. These two objectives may look contradictory, but the crime minimizing segmentation offers a way to strike the balance.

5   (Un)fairness of Crime Minimizing Segmentation

In this section, I show that the crime minimizing segmentation is unfair in the sense that it exposes different groups to different search intensities. Each citizen is now endowed with his *group identity* that takes values r or b. For each group $g \in \{r, b\}$, let $p_g \in (0, 1)$ denote the fraction of the overall population that belongs to group g, and let F_g denote the type distribution across the citizens in group g, so that we have $p_b F_b + p_r F_r = F_0$. Distribution function F_g has a positive density, f_g. Unless otherwise stated, we assume that the police does not observe the group identity of citizens. Throughout this section, we assume the following:

Assumption 1. Distribution F_b is greater than F_r in the hazard rate order, i.e., $\frac{f_r(x)}{1-F_r(x)} \geq \frac{f_b(x)}{1-F_b(x)}$ for all $x \in [0, 1)$. Also, distribution F_g satisfies $\overline{S} < \frac{1-\mathbb{E}_{F_g}[x]}{1+L}$ for each $g \in \{b, r\}$.

If distribution F_b is greater than F_r in the hazard rate order, then F_b is greater than F_r in the first-order stochastic dominance. We need a condition stronger than the usual stochastic dominance because we will compare the conditional distributions of F_b and F_r in terms of the first-order stochastic dominance. Interpreting type x as one's legal earning opportunity, the first part means that group b's income distribution is stochastically greater than group r. The second part simplifies the analysis.

The crime minimizing segmentation in Theorem 1 does not depend on group identity. Thus the probability $\sigma^C(x)$ that the police searches type x in the crime minimizing equilibrium depends only on x. However, because group identity is correlated with types, the two groups will be searched with different probabilities on average.

Proposition 2. *In the crime minimizing equilibrium, the following holds.*

1. *Group r is exposed to more policing than group b on average:*

$$\int_0^1 \sigma^C(x)dF_r(x) \geq \int_0^1 \sigma^C(x)dF_b(x). \tag{3}$$

2. *Across the citizens who do not commit crime, group r is exposed to more policing than group b on average:*

$$\int_0^1 \sigma^C(x)dF_r(x|x \geq z^*) \geq \int_0^1 \sigma^C(x)dF_b(x|x \geq z^*),$$

where z^* is the highest type that commits crime.

Part 1 uses the notion of fairness in Persico (2002), and Part 2 uses the notion of "fairness in the treatment of innocents" in Durlauf (2006). As we can see in Figure 2, the counterpart of Part 2 for "criminals" does not hold: Within the set of citizens who commit crime, the crime minimizing equilibrium is fair because all such citizens face the same search probability $\overline{S}$ regardless of their type or group identity.

Trade-off between fairness and effectiveness. Persico (2002) finds that an equilibrium crime rate may be lower when the police is required to search all citizens with equal probability than when the police bases search probabilities on group identity. In other words, there may be no trade-off between fairness and effectiveness of policing. Proposition 2 states that once we consider the set of all information structures, the trade-off generally exists, because the most effective policing strategy is unfair. The crime minimizing segmentation enables the police to shift search probabilities from high types to moderate types (see the red line in Figure 2). Even though the police's search probability for low types do not change, if one group tends to face worse legal earning opportunities, the crime minimizing segmentation exposes the low-income group to more intensive policing. This observation echoes the concern that predictive policing may render policing more unfair.

Introducing a fairness constraint. In the model, a predictive policing algorithm may render policing unfair partly because the objective of the algorithm is to minimize crime rate.

A natural remedy, which speaks to "ethical algorithm design," is to introduce an additional constraint that captures the fairness of policing (e.g., Kearns and Roth (2019)). In Appendix D, we characterize a segmentation that minimizes crime rate subject to the constraint that in equilibrium, the police searches groups b and r with the same probability on average (i.e., equation (3) holds with equality). Two insights emerge. First, the fairness constraint may not prohibit the police from using information: Proposition 3 shows that under a certain condition the crime minimizing "fair" segmentation is different from no information. Second, in some case, the fairness constraint necessitates information disclosure: Proposition 4 shows that if the police directly observes the group identity of citizens, we may restore the fairness of policing only by revealing additional information about types.

6   Extension: Endogenous Search Capacity

The crime minimizing segmentation is partially revealing if and only if $\overline{S} < \frac{1-\mathbb{E}_{F_0}[x]}{1+L}$ (see Proposition 1 and Theorem 1). This result relies on the assumption that search capacity $\overline{S}$ is inelastic: If the police can adjust $\overline{S}$ at costs, the crime minimizing segmentation is partially revealing for all parameters. To show this, we modify the game as follows: The police now chooses a search capacity $\overline{S}$ at cost $C(\overline{S})$, which is strictly increasing, strictly convex, smooth, and satisfies $C(0) = 0$. The police chooses a search capacity and a search strategy at once, and simultaneously, citizens choose whether to commit crime. The police's payoff is the mass of criminals she catches minus cost $C(\overline{S})$. In this model, the condition $\overline{S} < \frac{1-\mathbb{E}_{F_0}[x]}{1+L}$ arises endogenously and we obtain the following result.

Claim 1. *If the police can increase $\overline{S}$ at cost, the crime minimizing segmentation is a z-segmentation with $z \in (0,1)$. The equilibrium crime rate is positive, and the crime minimizing segmentation is neither the perfect segmentation nor no segmentation.*

The proof follows from the supply and demand interpretation of the model, where we view search capacity $\overline{S}$ as quantity and crime rate c as price. If the equilibrium crime rate is c, the police chooses search capacity $\overline{S}$ that solves $c = C'(\overline{S})$. The supply of search capacity $\overline{S}(c)$ is strictly increasing. If the equilibrium search capacity is $\overline{S}$, the crime rate $c(\overline{S})$ is determined

as the equilibrium crime rate of the original model with (exogenous) search capacity $\overline{S}$.[11] Note that $c(\overline{S})$ implicitly depends on the segmentation.

As an example, assume that F_0 is uniform, $L = 0.5$, and $C(S) = \frac{\gamma}{2}S^2$ with $\gamma > 0$. Figure 3 depicts the crime rates under no segmentation $c^N(\overline{S}) = 1 - 1.5\overline{S}$ (black downward sloping line), the crime rates under the crime minimizing segmentation $c^*(\overline{S}) = 1 - 3\overline{S}$ (red dashed line), and the inverse supply function of search capacity (upward sloping line). The equilibrium search capacity under the crime minimizing segmentation is a solution to $c^*(\overline{S}) = \gamma \overline{S}$, so we have $\overline{S}^* = \frac{1}{3+\gamma}$. The minimum crime rate is $c^*(\overline{S}^*) = 1 - \frac{3}{3+\gamma} > 0$. Thus the crime minimizing segmentation is the c^*-segmentation, i.e., it is never fully revealing for any $\gamma > 0$.
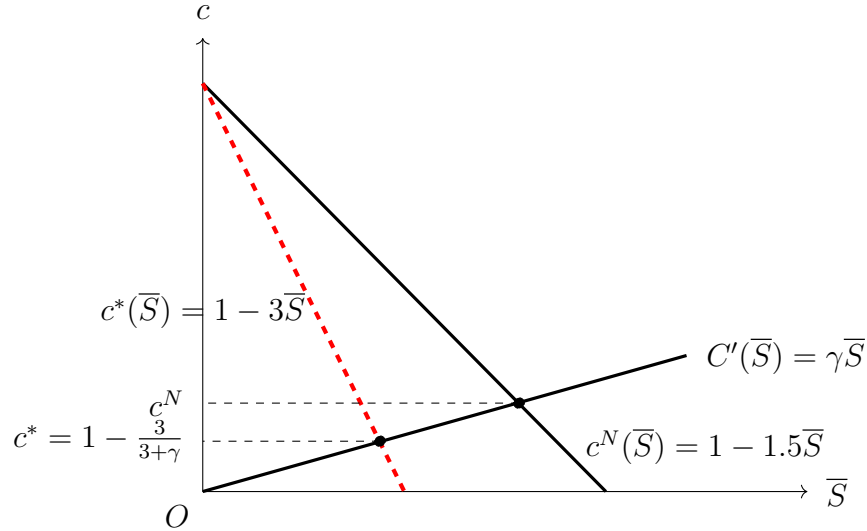

Figure 3: Equilibria with endogenous search capacity: $F_0 \sim U[0,1]$, $L = 0.5$, and $C(S) = \gamma S^2/2$.

Generally, any crossing point of the inverse supply curve $\overline{S}(c)$ and the inverse demand curve $c(\overline{S})$ is an equilibrium of the game with endogenous $\overline{S}$. To find the crime minimizing equilibrium of this game, we can define $c(\overline{S})$ as the minimized crime rate of the original game in which the search capacity is $\overline{S}$. Because $c(\overline{S})$ is strictly decreasing and hits 0 at $\overline{S} = \frac{1 - \mathbb{E}_{F_0}[x]}{1+L}$ whereas $c(\overline{S})$ is strictly increasing, the unique crossing point $(\overline{S}^*, c^*)$ satisfies

---

[11]There may be multiple equilibria given a search capacity $\overline{S}$, in which case $c(\overline{S})$ is a set. However, the multiplicity of equilibria does not affect the argument that follows because we characterize an equilibrium that attains the minimum crime rate, instead of characterizing all equilibria.

$c^* > 0$ and $\overline{S}^* < \frac{1 - \mathbb{E}_{F_0}[x]}{1+L}$. The corresponding segmentation is the z-segmentation where $z = F_0^{-1}(c^*) > 0$. Thus the equilibrium segmentation never fully revealing.

References

Antonovics, Kate and Brian G Knight (2009), "A new look at racial profiling: Evidence from the boston police department." *The Review of Economics and Statistics*, 91, 163–177.

Becker, Gary S (1968), "Crime and punishment: an economic approach." *Journal of Political Economy*, 76, 169–217.

Bergemann, Dirk and Stephen Morris (2019), "Information design: A unified perspective." *Journal of Economic Literature*, 57, 44–95.

Bjerk, David (2007), "Racial profiling, statistical discrimination, and the effect of a colorblind policy on the crime rate." *Journal of Public Economic Theory*, 9, 521–545.

Blackwell, David (1951), "Comparison of experiments." *Proceedings of the second Berkeley symposium on mathematical statistics and probability*.

Blackwell, David (1953), "Equivalent comparisons of experiments." *The annals of mathematical statistics*, 24, 265–272.

Brayne, Sarah (2017), "Big data surveillance: The case of policing." *American sociological review*, 82, 977–1008.

Cotton, Christopher and Cheng Li (2015), "Profiling, screening, and criminal recruitment." *Journal of Public Economic Theory*, 17, 964–985.

Curry, Philip A and Tilman Klumpp (2009), "Crime, punishment, and prejudice." *Journal of Public economics*, 93, 73–84.

Durlauf, Steven N (2006), "Assessing racial profiling." *The Economic Journal*, 116, F402–F426.

Dworczak, Piotr and Giorgio Martini (2019), "The simple economics of optimal persuasion." *Journal of Political Economy*, 127, 1993–2048.

Ensign, Danielle, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian (2018), "Runaway feedback loops in predictive policing." In *Conference on Fairness, Accountability and Transparency*, 160–171, PMLR.

Guo, Yingni and Eran Shmaya (2019), "The interval structure of optimal disclosure." *Econometrica*, 87, 653–675.

Hernández, Penélope and Zvika Neeman (2022), "How bayesian persuasion can help reduce illegal parking and other socially undesirable behavior." *American Economic Journal: Microeconomics*, 14, 186–215.

Hernández-Murillo, Rubén and John Knowles (2004), "Racial profiling or racist policing? bounds tests in aggregate data." *International economic review*, 45, 959–989.

Jung, Christopher, Sampath Kannan, Changhwa Lee, Mallesh Pai, Aaron Roth, and Rakesh Vohra (2020), "Fair prediction with endogenous behavior." In *Proceedings of the 21st ACM Conference on Economics and Computation*, 677–678.

Kamenica, Emir (2019), "Bayesian persuasion and information design." *Annual Review of Economics*, 11, 249–272.

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian persuasion." *American Economic Review*, 101, 2590–2615.

Kapustin, Max, Terrence Neumann, and Jens Ludwig (2022), "Policing and management." *National Bureau of Economic Research*.

Kearns, Michael and Aaron Roth (2019), *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Knowles, John, Nicola Persico, and Petra Todd (2001), "Racial bias in motor vehicle searches: Theory and evidence." *Journal of Political Economy*, 109, 203–229.

Kolotilin, Anton, Timofiy Mylovanov, and Andriy Zapechelnyuk (2022), "Censorship as optimal persuasion." *Theoretical Economics*, 17, 561–585.

Lau, Tim (2020), "Predictive policing explained." https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained (accessed July 31, 2022).

Lazear, Edward P (2006), "Speeding, terrorism, and teaching to the test." *The Quarterly Journal of Economics*, 121, 1029–1061.

Liang, Annie, Jay Lu, and Xiaosheng Mu (2022), "Algorithmic design: Fairness versus accuracy." In *Proceedings of the 23rd ACM Conference on Economics and Computation*, 58–59.

Ludwig, Jens and Sendhil Mullainathan (2021), "Fragile algorithms and fallible decision-makers: lessons from the justice system." *Journal of Economic Perspectives*, 35, 71–96.

Pearsall, Beth (2010), "Predictive policing: The future of law enforcement." *National Institute of Justice Journal*, 266, 16–19.

Perry, Walt L (2013), *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation.

Persico, Nicola (2002), "Racial profiling, fairness, and effectiveness of policing." *American Economic Review*, 92, 1472–1497.

Persico, Nicola (2009), "Racial profiling? detecting bias using statistical evidence." *Annu. Rev. Econ.*, 1, 229–254.

Persico, Nicola and Petra E Todd (2005), "Passenger profiling, imperfect screening, and airport security." *American Economic Review*, 95, 127–131.

Rabinovich, Zinovi, Albert Xin Jiang, Manish Jain, and Haifeng Xu (2015), "Information disclosure as a means to security." In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 645–653, Citeseer.

Shaked, Moshe and J George Shanthikumar (2007), *Stochastic orders*. Springer.

Stashko, Allison (2020), "Do police maximize arrests or minimize crime." *Evidence from Racial Profiling in US Cities. SSRN*, 3132046.

Vasquez, Jorge (2022), "A theory of crime and vigilance." *American Economic Journal: Microeconomics*, 14, 255–303.

Appendix

A Proof of Lemma 1

Proof. Suppose to the contrary that for some segmentation and an equilibrium, there are sets $\mathcal{F}_H, \mathcal{F}_L \subset \mathcal{F}$ such that they have positive measures under μ and any segment in $\mathcal{F}_H$ has a strictly higher crime rate than any segment in $\mathcal{F}_L$. If the police searches a citizen with probability $\sigma \in [0,1]$, the net benefit of committing a crime is $1 - \sigma - \sigma L - x$. Thus if $\sigma = 1$, no citizen commits a crime for any type x; if $\sigma = 0$, all citizens with $x < 1$ commit crime. Crime rates in $\mathcal{F}_H$ being positive implies that the average search intensity across the segments in $\mathcal{F}_H$ is strictly less than 1. Crime rates in $\mathcal{F}_L$ being less than 1 implies that the average search intensity across the segments in $\mathcal{F}_L$ is positive. Thus the police can profitably deviate by allocating search intensities from segments in $\mathcal{F}_L$ to segments in $\mathcal{F}_H$, which is a contradiction. $\qquad\square$

B Proof of Theorem 1

We begin with a lemma.

Lemma 2. *If there is a segmentation and an equilibrium that attains a crime rate of $c \in [0,1]$, then we have*

$$\overline{S} \geq \frac{1 - \mathbb{E}_{F_0}[x | x \geq F_0^{-1}(c)]}{1 + L}. \tag{4}$$

Proof. Throughout the proof, we fix a segmentation and an equilibrium. We define probability measure μ^{nc} as follows: For each measurable set $A \subset [0,1] \times \mathcal{F}$, let $\mu^{nc}(A)$ be the fraction

of citizens whose type x and segment F belong to A, among the mass $1 - c$ of all citizens who do not commit a crime in the equilibrium. In particular, $\mu^{nc}(X \times \mathcal{F})$ is the fraction of citizens whose types are in $X \subset [0,1]$ among all citizens who do not commit a crime. Also let $\sigma(F, x)$ denote the equilibrium search probability that type x receives in segment F.

A citizen with type $x \in \text{supp}\, F$ does not commit a crime only if $x \geq 1 - \sigma(x, F)(1 + L)$. We then have

$$x \geq 1 - \sigma(x, F)(1 + L)$$
$$\Rightarrow \int_{[0,1] \times \mathcal{F}} x d\mu^{nc} \geq \int_{[0,1] \times \mathcal{F}} [1 - \sigma(x, F)(1 + L)] \, d\mu^{nc}$$
$$\Rightarrow \int_{[0,1] \times \mathcal{F}} \sigma(x, F) d\mu^{nc} \geq \frac{1 - \int_{[0,1] \times \mathcal{F}} x d\mu^{nc}}{1 + L}. \tag{5}$$

The search intensity allocated to citizens who do not commit a crime is $\overline{S}^{nc} \triangleq (1 - c) \int_{[0,1] \times \mathcal{F}} \sigma(x, F) d\mu^{nc}$. However, the crime rate is c in almost all segments. Thus for every citizen with type $x \in \text{supp}\, F$ who does not commit a crime, there are $\frac{c}{1-c}$ citizens who belong to segment F, commit a crime, and face search probability $\sigma(x, F)$. Thus the total search intensity in equilibrium is

$$(1 - c)\overline{S}^{nc} + \frac{c}{1 - c}\overline{S}^{nc} = \int_{[0,1] \times \mathcal{F}} \sigma(x, F) d\mu^{nc}.$$

As a result, if the equilibrium crime rate is c, the total search capacity $\overline{S}$ must be at least the right-hand side of inequality (5).

Finally, because the mass of citizens who do not commit a crime is $1 - c$, we have

$$\mathbb{E}[x | x \geq F_0^{-1}(c)] \geq \int_{[0,1] \times \mathcal{F}} x d\mu^{nc}.$$

Combining this inequality with inequality (5), we conclude that

$$\overline{S} \geq \int_{[0,1] \times \mathcal{F}} \sigma(x, F) d\mu^{nc} \geq \frac{1 - \mathbb{E}_{F_0}[x | x \geq F_0^{-1}(c)]}{1 + L}.$$

$\square$

26

Proof of Theorem 1. Lemma 2 implies that given search capacity $\overline{S}$, the crime rate is at least c^*, where

$$\overline{S} = \frac{1 - \mathbb{E}_{F_0}[x | x \geq F_0^{-1}(c^*)]}{1 + L} \tag{6}$$

Indeed, if the equilibrium crime rate is $c < c^*$ then the total search capacity must be at least

$$\frac{1 - \mathbb{E}_{F_0}[x | x \geq F_0^{-1}(c)]}{1 + L} > \frac{1 - \mathbb{E}_{F_0}[x | x \geq F_0^{-1}(c^*)]}{1 + L} = \overline{S},$$

which is a contradiction. Equation (6) indeed has a unique solution c^*, because as a function of c, $\frac{1 - \mathbb{E}_{F_0}[x | x \geq F_0^{-1}(c)]}{1+L}$ continuously and strictly decreases from $\frac{1 - \mathbb{E}_{F_0}[x]}{1+L} > \overline{S}$ to $0 < \overline{S}$ as c increases from 0 to 1.

We show that the z^*-segmentation attains the equilibrium crime rate c^*. Recall that search probability $\sigma(x)$ makes citizens with type x indifferent between committing and not committing a crime. If the police allocates search probability $\sigma(x)$ to segment F_i in the z^*-segmentation, citizens with type i prefers not to commit a crime, whereas any other citizens, whose types are below i, will commit a crime. As a result, the crime rate in segment F_i will be $F_0(z^*)$. Under such a search strategy, the total search intensity is

$$\int_{z^*}^1 \sigma(x) d\mu^c = \frac{1 - \mathbb{E}_{F_0}[x | x \geq z^*]}{1 + L} = \overline{S}.$$

The first equality is by the definitions of $\sigma(x)$ and μ^{z^*} (see Definition 1), and the second equality is by the definition of z^* (see equation (1)). Because the police exhausts her search capacity and the crime rates in all segments are equalized, we have an equilibrium. Therefore the z^*-segmentation associated with this equilibrium attains crime rate $F_0(z^*)$. We have $F_0(z^*) = c^*$ by comparing equations (1) and (6). $\qquad\square$

C Proof of Corollary 1

Proof. Take any parameter $(\overline{S}, L)$. In the equilibrium with no segmentation, citizens commit crime if and only if their types exceed cutoff x^N defined by

$$\overline{S} = \frac{1 - x^N}{1 + L}.$$

In the crime minimizing equilibrium, citizens commit crime if and only if their types exceed cutoff z^* defined by

$$\overline{S} = \frac{1 - \mathbb{E}_{F_0}[x|x \geq z^*]}{1 + L}.$$

If one parameter is more prone to crime than the other parameter, the former has a higher x^N. Comparing the two equations, we conclude that the same change of the parameter increases z^* and the minimized crime rate, i.e., Part 1 holds.

For Part 2, it suffices to show that for any z_0 and $z_1 > z_0$, the z_0-segmentation is more informative than the z_1-segmentation. We construct the z_1-segmentation by garbling the z_0-segmentation as follows: First, we pool segments $\{G_i^0\}_{i \in [z_0, z_1]} \subset \operatorname{supp} \mu^{z_0}$ of the z_0-segmentation into a single segment, say G. Second, we view segment G as consisting of a continuum of segments $\{H_i\}_{i \in [z_1, 1]}$ where $H_i = G$ for all $i \in [z_1, 1]$, and index i is distributed according to $F_0(\cdot | x \geq z_1)$. We then pool each H_i with $G_i^1 \in \mu^{z_1}$, which results in segment $G_i^1 \in \operatorname{supp} \mu^{z_1}$.[12] $\qquad \square$

D Omitted Proofs for Section 5

First, we prove Proposition 2.

Proof. We have

$$\sigma^C(x) = \begin{cases} \overline{S} & \text{if} \quad x < z^*, \\ \frac{1-x}{1+L} & \text{if} \quad x \geq z^*. \end{cases}$$

Because F_b dominates F_r in the hazard rate order, $F_b(\cdot | x \geq z^*)$ is greater than $F_r(\cdot | x \geq z^*)$ in the first order stochastic dominance (Shaked and Shanthikumar, 2007). This implies Part 2 of the proposition because $\sigma^C(x)$ is decreasing in $x \geq z^*$. To show Part 1, note that $\overline{S} = \frac{1 - \mathbb{E}_{F_0}(x|x \geq z^*)}{1+L}$ implies

$$\frac{1 - \mathbb{E}_{F_b}(x|x \geq z^*)}{1 + L} < \overline{S} < \frac{1 - \mathbb{E}_{F_r}(x|x \geq z^*)}{1 + L}. \tag{7}$$

---

[12]Equivalently, we can pool each segment in $\{G_i^0\}_{i \in [z_0, z_1]}$ with each segment $\{G_i\}_{i \in [z_1, 1]} \subset \mu^{z_1}$ to create the z_1-segmentation, which follows more closely with the garbling definition of Blackwell (1951, 1953).

The average search probability for group b is

$$F_b(z^*)\overline{S} + [1 - F_b(z^*)]\frac{1 - \mathbb{E}_{F_b}(x|x \geq z^*)}{1 + L} < \overline{S}. \tag{8}$$

The average search probability for group b is

$$F_r(z^*)\overline{S} + [1 - F_r(z^*)]\frac{1 - \mathbb{E}_{F_r}(x|x \geq z^*)}{1 + L} > \overline{S}. \tag{9}$$

Combining these inequalities, we conclude that group r faces a higher search probability than group b on average. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

We introduce several terminologies and notations. Let $\mathcal{H} \triangleq \Delta([0,1] \times \{g,b\})$ denote the set of all joint distributions of types and group identities, and let $H_0 \in \mathcal{H}$ denote the prior joint distribution induced by (p_b, p_r, F_b, F_r). Define an *extended segmentation* as any element of $\Delta\mathcal{H}$ that averages to H_0, and call $H \in \operatorname{supp}\eta$ an *extended segment*. As before, the police chooses a search strategy $\sigma : \operatorname{supp}\eta \to [0,1]$, which maps each segment to the search probability. Note that a search probability is measurable only with respect to $\operatorname{supp}\eta$, which means that the police does not directly learn about citizens' group identities beyond the information provided by the extended segmentation. Each citizen chooses whether to commit crime, known his (x,g) and the segment he belongs to.

Definition 3. Given an equilibrium, let $\sigma(x,g)$ denote the (expected) probability that a citizen with $(x,g) \in [0,1] \times \{b,r\}$ is searched. The equilibrium is *fair* that two groups face the same search probability, i.e.,

$$\int_0^1 \sigma(x,b)dF_b(x) = \int_0^1 \sigma(x,r)dF_r(x). \tag{10}$$

To state the next result, define z^*, z_b^*, and z_r^* as the solutions to the following equations:

$$\overline{S} = \frac{1 - \mathbb{E}_{F_0}[x|x \geq F_0^{-1}(z^*)]}{1 + L}, \quad \overline{S} = \frac{1 - \mathbb{E}_{F_b}[x|x \geq F_b^{-1}(z_b^*)]}{1 + L}, \quad \text{and} \quad \overline{S} = \frac{1 - \mathbb{E}_{F_r}[x|x \geq F_r^{-1}(z_r^*)]}{1 + L}.$$

$$\tag{11}$$

29

Proposition 3. *In addition to* Assumption 1, *suppose we have*

$$\frac{1 - F_r(z_r^*)}{1 - F_r(z_b^*)} \geq 1 - F_0(z^*). \tag{12}$$

Across all extended segmentations that permit a fair equilibrium, the minimized crime rate is $p_b c_b^* + p_r c_g^*$, which is strictly lower than the crime rate under no segmentation.

Proof. We define $c_g^* \triangleq F_g(z_g^*)$ for each $g \in \{b, r\}$ and $c^* \triangleq F_0(z^*)$. For an equilibrium to be fair, the average search probability for groups b and r must be equal to $\overline{S}$. By the same logic as the proof of Theorem 1, the crime rates within groups b and r are at least c_b^* and c_r^*, respectively. As a result, the crime rate at any fair equilibrium is at least $p_b c_b^* + p_r c_g^*$. Thus to prove the result we construct an extended segmentation that attains the equilibrium crime rate $p_b c_b^* + p_r c_g^*$.

We first partition the overall population $[0, 1] \times \{g, b\}$ into $\{A_1, A_2\}$, where

$$A_r(z) = \{(x, r) : x \geq z\},$$
$$A_b(z) = \{(x, r) : x < z\} \cup \{(x, b) : 0 \leq x \leq 1\}.$$

If we apply the crime minimizing segmentation to group g, citizens in group g commit crime if and only if $x < z_g^*$. Suppose that all citizens behave according to such a group-wise crime minimizing equilibrium. Thus if $z = z_r^*$, the crime rate in the set $A_r(z_r^*)$ equals 0. If $z = 0$, the crime rate in the set $A_r(0)$ equals $c_r^* > p_b c_b^* + p_r c_g^*$. Thus there is a unique value $\hat{z} \in (0, z_r^*)$ such that the crime rates in both $A_r(\hat{z})$ and $A_b(\hat{z})$ equal $p_b c_b^* + p_r c_g^*$. Moreover, inequality (12) implies that $\hat{z} \leq z_b^*$.

Let F_{A_g} denote the type distribution within set $A_g(\hat{z})$. By construction, we have

$$F_{A_r}(\cdot | x \geq z_r^*) = F_r(\cdot | x \geq z_r^*),$$
$$F_{A_b}(\cdot | x \geq z_b^*) = F_b(\cdot | x \geq z_b^*),$$

and $F_{A_r}^{-1}(\hat{z}) = z_r^*$ and $F_{A_b}^{-1}(\hat{z}) = z_b^*$. Thus we have

$$\overline{S} = \frac{1 - \mathbb{E}_{F_{A_b}}[x|x \geq \hat{z}]}{1 + L} \quad \text{and} \quad \overline{S} = \frac{1 - \mathbb{E}_{F_{A_r}}[x|x \geq \hat{z}]}{1 + L}. \tag{13}$$

These equations imply that we can apply the crime minimizing segmentation in Theorem 1 to extended segments $A_r(\hat{z})$ and $A_b(\hat{z})$ to attain crime rate $p_b c_b^* + p_r c_g^*$ in an equilibrium. $\quad\square$

The result highlights the new trade-off between the fairness of policing and fairness of privacy. The segmentation in Proposition 3 ensure that the police searches two groups of citizens at the same frequency. At the same time, two citizens who have the same type x may be allocated to different segments, which implies that the information the police will learn about a citizen's type may depend on their group identity. This observation motivates a question of whether there is a non-trivial segmentation that preserves the fairness of policing and privacy. I leave this question for future research.

Observable group identity. So far, we have assumed that the police does not directly observe group identity. Proposition 3 relies on this assumption: The segmentation pools citizens who have different group identities into the same segment, but the police who observes group identity may search them with different probabilities. Thus if the police observes group identity, the segmentation in Proposition 3 may no longer be fair. Restricting the police's information does not solve the problem. Indeed, if the police has no information about types, the police will base search probabilities only on group identity, which will expose group r to more search under Assumption 1. Thus when group identity is observable, we can restore the fairness of policing only by providing the police with more information about citizens' types. The following result presents the solution to this problem.

Proposition 4. *Suppose that the police observes group identity. If $\overline{S} < \frac{1 - \mathbb{E}_{F_b}[x]}{1 + L}$ and $\overline{S} < \frac{1 - \mathbb{E}_{F_r}[x]}{1 + L}$, the segmentation that minimizes crime subject to fairness attains crime rate $\max(F_r(z_r^*), F_b(z_b^*))$.*

Proof. Let $c_g^* = F_g(z_g^*)$ for each $g \in \{b, r\}$. Without loss of generality, suppose $c_r^* > c_b^*$. First, because the police observes each citizen's group identity g, it is without loss of generality to consider an extended segmentation such that the support of any possible segment does not

31

contain the citizens from both groups. Given that the average search probability to each group $g \in \{r, b\}$ is $\overline{S}$, the crime rate within group r must be at least c_r^*. Then the crime rate within group b must be at least c_r^*. Indeed, if the crime rate within group b is (say) lower than c_r^*, then a positive mass of segments for group b has a lower crime rate than any segment fro group g, which contradicts the equilibrium condition.

We now construct a segmentation for each group such that the crime rate is c_r^*. For group r, we can use the crime minimizing segmentation for F_r. For group b, we consider two cases. Let c_b^N be the equilibrium crime rate under no segmentation given F_b. If $c_b^N \geq c_r^*$, then segmentation $\mu^{c_r^*}$ in Definition 1 applied to F_r will attain crime rate c_r^*. Indeed, as we move z from 1 to 0, the equilibrium crime rate under μ^z segmentation moves from $c_b^N \geq c_r^*$ to $c_b^* < c_r^*$. If $c_b^N < c_r^*$, we consider the following segmentation indexed by z: For each $x < z$, the police fully learns the value of x; if $x \geq z$, the police only learns that $x \geq z$. Because $\overline{S} < \frac{1 - \mathbb{E}_{F_r}[x]}{1 + L}$, as we move z from 0 to 1, the equilibrium crime rate increases from c_b^N to 1, so there is some z such that the segmentation attains crime rate c_r^*. $\qquad\square$