How Bayesian Persuasion can Help Reduce Illegal Parking and Other Socially Undesirable Behavior

Penlope Hernndez and Zvika Neeman

June 27, 2018

Abstract

We consider the question of how best to allocate enforcement resources across different locations with the goal of deterring unwanted behaviour. We rely on "Bayesian persuasion" to improve deterrence. For simplicity, we focus on the problem of how to allocate resources in order to reduce the extent of illegal parking. However, the same model can also be applied to many other types of socially undesirable behaviour. Our approach is distinguished by the following five features: (1) we consider a problem in which the principal has to allocate resources and then send messages (persuade) rather than just persuade. (2) Messages are received by drivers in n different neighborhoods, so persuasion is with respect to multiple audiences. (3) The problem is a "constrained convexification" rather than just a convexification problem, where the constraints are due to resource and probability restrictions. This implies that convexification may be partial rather than complete as is usually the case in Bayesian persuasion models, and we provide conditions under which this is the case. (4) Even though the basic problem is not linear, we show that it can be cast as a linear programming problem. Finally, (5) we characterize the number of messages needed in order to obtain the optimal solution, and describe conditions under which it is possible to explicitly solve the problem with only two messages.

1 Introduction

This paper addresses the question of how best to allocate enforcement resources across different locations with the goal of deterring unwanted behaviour. The novelty in our approach is that we employ the techniques of "Bayesian persuasion," namely the use of carefully disseminated communication in order to maximize deterrence. To fix ideas and simplify the presentation, we focus on the problem of how to allocate resources in order to reduce the extent of illegal parking. However, the same model can also be applied to many other types of socially undesirable behaviour.

Suppose that a principal observes the realized amount of enforcement resources available and decides how to allocate them across $n \geq 1$ different neighborhoods or locations. It can send messages about the amount of realized resources and their allocation. These messages can be displayed on the city's website, or on electronic street signs. Drivers in each one of the n neighborhoods observe these messages and decide whether or not to park illegally.

A small literature, that started with Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) (see also Sobel (2013), Bergemann and Morris (2017), and the references therein), studies how a sender of information can affect a receiver's beliefs and thereby induce it to act in a way that benefits the sender. The problem analyzed here has a number of special features that distinguish it from the literature on Bayesian persuasion.

First, we consider a problem in which the principal has to allocate resources and then send messages (persuade) rather than just persuade. Second, messages are received by drivers in n different neighborhoods, so persuasion is with respect to multiple audiences. Third, the problem is a "constrained convexification" rather than just a convexification problem, where the constraints are due to resource and probability restrictions. In an alternative representation of the problem, these constraints become deterrence constraints. These constraints imply that in our setting the "space of signals" can induce only a strict subset of the set of distributions of posterior expectations of resources that preserves the mean of the distribution of resources.[1] Moreover, the fact that convexification is constrained implies that it may be partial rather than complete as is usually the case in Bayesian persuasion models, and we provide conditions under which this is the case. Fourth, even though the basic problem is not linear, we show that it can be cast as a linear programming problem. Finally, fifth, we characterize the number of messages needed in order to obtain the optimal solution, and describe conditions under which it is possible to explicitly solve the problem with only two messages: "high" and "low" that indicate that the amount of expected resources is high and low, respectively. The message "low" may be interpreted as a moratorium on parking enforcement in some clearly defined situations. Our results indicate that such a moratorium can be an important part of an optimal enforcement policy. Intuitively, such a moratorium improves overall deterrence because it achieves stronger deterrence when it is not applied. Indeed, casual empiricism suggests that local governments occasionally experiment with such moratoriums. For example, it is supposedly well known and certainly widely believed among residents of Tel Aviv that the city does not enforce parking violations from Friday to Saturday evenings as well as from the evening before to the evening of state holidays.

The question of how to allocate resources in order to achieve deterrence is typically analyzed in the context of what is known as a "security game." A security game is a two-player, possibly zero-sum, simultanuous-move game in which an attacker has to decide where to strike while a defender has to decide where to allocate its limited defense resources.[2] Analysis of such games has been applied by political scientists to the question of how to defend against terrorist attacks (Powell, 2007), and by computer scientists to a host of related issues (see Tambe, 2011, and the references therein). Security games are closely related to Colonel Blotto games (Borel, 1921; Roberson, 2006; Hart, 2008). These are zero-sum simultanuous-move two-player games in which players allocate a given number of divisions to n different battlefields. Each battlefield is won by the player who allocated a larger number of divisions there, and the player who

[1] Gentzkow and Kamenica (2016) and Kolotilin (2017) characterize feasible distributions of posterior expectations or beliefs in somewhat different settings. Le Treust and Tomala (2017) analyze a different problem of constrained convexification.

[2] The fact that in our formulation, the attacker responds only after observing the defender's signal turns our game into a sequential rather than a simultanuous move game.

wins a larger number of battlefields wins the game. As explained above, we consider a security game in which there is uncertainty about the amount of resources available to the defender, with an added stage in which the defender can send a message about the state of the world.[3]

The question addressed here of how to allocate a given amount of law enforcement resources is different from, and complementary to, the questions famously posed by Becker (1968) about how much resources should be allocated to law enforcement and how to divide these resources between enforcement effort that increases the probability that the offender is caught and the penalty imposed on the offender if caught. Polinsky and Shavell (2000) provide a survey of the theoretical literature on the optimal form of enforcement, and Chalfin and McCrary (forthcoming) provide a survey of the relevant empirical literature.

Within the law and economics literature, the two papers that are most closely related to our work are by Lando and Shavell (2004) and Eeckhout et al. (2010) who both consider the question of how to allocate enforcement resources. Both papers show that it may be beneficial to concentrate enforcement on a subset of the population. The following example illustrates their idea. Suppose that deterrence of the entire population requires 10 units of resources, but only 5 units are available. In this case, allocation of the 5 units of resources across the entire population fails to achieve deterrence, but concentration of the 5 units on half of the population (say, on those with lightly colored eyes) successfully deters this half. Our paper is more general in that it considers any number of neighborhoods and it adds uncertainty, and in that we consider the question of how to further improve deterrence through Bayesian persuasion, or communication.

The paper proceeds as follows. The model is presented in Section 2. Section 3 describes the Optimal Ratio Rule and its implications. Section 4 introduces two lemmas that generalize a famous lemma of Aumann and Maschler (1995, p. 25) that are useful for subsequent analysis. Section 5 considers the case of "monotone" problems. In Section 6 we explain the sense in which the problem is a constrained convexification problem. Finally, in Section 7, we briefly address the issue of dynamics, or deterrence over time.

2 Model

Consider a city with $n \geq 1$ different neighborhoods. The set of neighborhoods or locations is denoted $\mathbf{N} = \{1, \ldots, n\}$. Illegal parking is a problem in all of these neighborhoods. The city determines the amount of resources devoted to enforcement in each neighborhood out of the total amount of available resources, denoted r. The amount of available resources is uncertain. We assume that $r = r_k$, $k \in \{1, \ldots, K\}$, with probability π_k, respectively, where $0 \leq r_1 < \cdots < r_K$ and $\sum_{k=1}^{K} \pi_k = 1$. We treat the distribution of resources as exogenously given, but it may obviously depend on the city's decisions, and provides another dimension on which to optimize the allocation of resources. We discuss two ways of endogenizing the distribution of resources in Section 6 below.

We refer to k as the state of the world. The city knows the realization of the state of the world k and hence also the realization r_k, but drivers only know the distribution

[3] Rabinovich et al. (2015) and Xu et al. (2016) have also studied a security game with messages, but in a very different setting.

$\pi = (\pi_1, \ldots, \pi_K)$.

As explained above, we assume that the city may disseminate information about its enforcement effort. We model this possibility by assuming that the city may send a message $m \in \{1, \ldots, M\}$ about the state of the world k.[4] The probability that the city sends message m in state k is denoted by $p_k(m) = \Pr(m|k)$. It follows that

$$p_k(m) \geq 0 \text{ for every } k \text{ and } m, \text{ and } \sum_{m=1}^{M} p_k(m) = 1 \text{ for every } k. \tag{1}$$

The posterior belief that drivers have over the state of the world k upon receiving the message m is denoted

$$\Pr(k|m) = \frac{p_k(m)\,\pi(k)}{\sum_{k'=1}^{K} p_{k'}(m)\,\pi(k')}. \tag{2}$$

Denote the amount of resources allocated to enforcement in neighborhood i in state k when the city sends the message m by $a_k^i(m)$.[5] If message m is sent with probability zero in state k, then $a_k^i(m) \equiv 0$ for every location i.

The city chooses the amounts $a_k^i(m)$ subject to its resource constraint. In every state $k \in \{1, \ldots, K\}$,

$$\sum_{i=1}^{n} a_k^i(m) \leq r_k \tag{3}$$

for every message $m \in \{1, \ldots, M\}$.[6]

The objective of the city is to allocate the amounts of enforcement resources $\left\{a_k^i(m)\right\}$ and send the messages $m \in \{1, ..., M\}$ with probabilities $\{p_k(m)\}$ so as to minimize the extent of illegal parking. The measure of illegal parking in each neighborhood i is given by a function $q^i(a^i(m))$ that is decreasing in the expected amount of enforcement resources $a^i(m) \equiv \sum_{k=1}^{K} a_k^i(m) \Pr(k|m)$ in that neighborhood given message m. For simplicity, we focus on the special case where each q^i is given by a threshold function. Namely, there exists some threshold τ^i such that

$$q^i(a^i(m)) = \begin{cases} 1 & \text{if } a^i(m) < \tau^i \\ 0 & \text{if } \tau^i \leq a^i(m) \end{cases}.$$

Hence, the city's objective is to allocate the amounts of enforcement resources $\left\{a_k^i(m)\right\}$ and send messages with probabilities $\{p_k(m)\}$ so as to minimize the expected social cost of illegal parking as given by

$$\min_{\left\{a_k^i(m)\right\},\{p_k(m)\}} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{i=1}^{n} q^i(a^i(m)) s^i p_k(m)\,\pi_k \tag{4}$$

[4] "No signal" is also a signal.

[5] We show below that conditioning the level of enforcement on the signal on top of just the state of the world may contribute to deterrence.

[6] Observe that there is no need to also sum over the messages in the resource constraint because the constraint only requires that resources add up to no more than what is available given a state of the world and the fact that a specific given message has been sent.

For example, if there are just two locations, just two messages m and m', and r_k units are available in state k, then we need to require that $a_k^1(m) + a_k^2(m) \leq r_k$ and $a_k^1(m') + a_k^2(m') \leq r_k$ rather than the weaker requirement that $p_k(m)\left(a_k^1(m) + a_k^2(m)\right) + p_k(m')\left(a_k^1(m') + a_k^2(m')\right) \leq r_k$ because the city may allocate the entire amount of available resources r_k upon sending any message m.

where s^i, $i \in \{1, \ldots, n\}$, denotes the social disutility generated by illegal parking in neighborhood i, subject to the resource constraint (2) and the constaints imposed by the fact that the $p_k\,(m)$'s are probabilities (1).

Observe that the constraints (1) and (2) are linear in resources $\left\{a_k^i\,(m)\right\}$ and probabilities $\{p_k\,(m)\}$, but the objective function (3) is non-linear both because $q^i(a^i\,(m))$ is a non-linear function of $a^i\,(m)$ and because $a^i\,(m)$ itself is a non-linear function of the probabilities $\{p_k\,(m)\}$.

Alternatively, it is also useful to consider the city's problem as how to allocate the amounts of enforcement resources $\left\{a_k^i\,(m)\right\}$ and send messages with probabilities $\{p_k\,(m)\}$ so as to maximize expected weighted deterrence as given by

$$\max_{\left\{a_k^i(m)\right\},\{p_k(m)\}} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{i=1}^{n} d^i(a^i\,(m)) s^i p_k\,(m)\,\pi_k \tag{5}$$

where the function $d^i(a^i\,(m)) = 1 - q^i(a^i\,(m))$ describes the strength of deterrence and s^i is interpreted as the benefit of deterrence in neighborhood i (which is equal to the decrease in social distutility). Again, the constraints (1) and (2) are linear in $\left\{a_k^i\,(m)\right\}$ and $\{p_k\,(m)\}$, but the objective function (4) is not.

It is helpful to represent the allocation of resources in matrix form, as shown in the next example. Suppose that there are three locations and three states of the world. The allocation of resources is given by:

π_1	$a_1^1(m)$	$a_1^2(m)$	$a_1^3(m)$	r_1
π_2	$a_2^1(m)$	$a_2^2(m)$	$a_2^3(m)$	r_2
π_3	$a_3^1(m)$	$a_3^2(m)$	$a_3^3(m)$	r_3
	τ^1	τ^2	τ^3	

If no messages are sent, then we may denote $m =$; if the message sent reveals the state of the world, then we may denote $m = m_j$ in row j of the matrix.

The case where two messages m_1 and m_2 are sent is represented as follows:

π_1	$a_1^1(m_1)$	$a_1^2(m_1)$	$a_1^3(m_1)$	r_1
π_2	$a_2^1(m_1)$	$a_2^2(m_1)$	$a_2^3(m_1)$	r_2
	$a_2^1(m_2)$	$a_2^2(m_2)$	$a_2^3(m_2)$	
π_3	$a_3^1(m_2)$	$a_3^2(m_2)$	$a_3^3(m_2)$	r_3
	τ^1	τ^2	τ^3	

Message m_1 is sent in states 1 and 2, and message m_2 is sent in states 2 and 3. This example clarifies the reason that not allowing the allocation to depend on the message sent involves a loss of generality: it does not allow the city to sometimes deter only in neighborhoods 1 and 2 in state 2 (when it sends the message m_1), and sometimes deter in neighborhoods $1, 2, 3$ (when it sends the message m_2). This is something that the city may benefit from if the amount of resources available in state 3 permits deterrence in neighborhoods $1, 2, 3$ ($r_3 > \tau_1 + \tau_2 + \tau_3$) but the amount available in states 1 and 2 only permits deterrence in neighborhoods 1 and 2.

The next example, which is similar to an example in Kamenica and Gentzkow (2011), shows that the city may be able to decrease the extent of illegal parking by disseminating information about the realizations of the amount of enforcement effort $\left\{a_k^i\,(m)\right\}$. For simplicity, the amounts of enforcement efforts in this example are independent of the messages, so the index m is ommited, and they are denoted by $\left\{a_k^i\right\}$.

Example 1. Consider a city with one neighborhood. Suppose that drivers park illegally if they perceive the expected amount of enforcement to be smaller than $\tau_1 = 2/5$. Suppose that resources are given by $(r_1, r_2) = (0, 1)$ with probabilities $(\pi_1, \pi_2) = (\frac{2}{3}, \frac{1}{3})$, respectively, and that the social cost of illegal parking is $s_1 = 1$. The fact that there is only one neighborhood greatly simplifies the problem of how to allocate the amount of enforcement efforts $\{a_k^i\}$. The city cannot do better than simply allocate its entire enforcement resources in every state of the world to this single neighborhood, so that $a_1^1 = 0$ and $a_2^1 = 1$. All this information is represented in matrix form as follows:

$$\begin{array}{c|c|c} \frac{2}{3} & 0 & 0 \\ \hline \frac{1}{3} & 1 & 1 \\ \hline & \frac{2}{5} & \end{array}$$

If the city disseminates no information about the state of the world, then drivers park illegally because the expected amount of enforcement is only

$$\frac{2}{3} \cdot a_1^1 + \frac{1}{3} \cdot a_2^1 = \frac{1}{3},$$

which is smaller than the critical threshold $\tau_1 = 2/5$. The expected social cost of illegal parking in this case is 1.

The city can do better by fully revealing the state of the world to the drivers. In this case, when the state of the world is $k = 1$, drivers would realize that there is no enforcement because $a_1^1 = 0$ and would park illegally, but when the state of the world is $k = 2$, drivers would be deterred from parking illegally because $a_2^1 = 1$, which implies that the expected social cost of illegal parking in this case is

$$\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{2}{3}.$$

The city can do even better by providing partial information about the state of the world as follows: when $k = 2$ it sends the message H, and when $k = 1$, it sends messages H and L with probability $1/2$ each. When drivers receive the message L they know that $k = 1$ and so the amount of enforcement is $a_1^1 = 0$ and so they park illegally. However, when they receive the message H, their posterior belief about the amount of enforcement is

$$\begin{aligned} E\left[a^1 \left| m = H\right.\right] &= \frac{\pi(H|1)\pi_1}{\pi(H|1)\pi_1 + \pi(H|2)\pi_2} \cdot a_1^1 + \frac{\pi(H|2)\pi_2}{\pi(H|1)\pi_1 + \pi(H|2)\pi_2} \cdot a_2^1 \\ &= \frac{\frac{1}{2}\cdot\frac{2}{3}}{\frac{1}{2}\cdot\frac{2}{3} + 1\cdot\frac{1}{3}} \cdot a_1^1 + \frac{1\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{2}{3} + 1\cdot\frac{1}{3}} \cdot a_2^1 \\ &= \frac{1}{2} \cdot a_1^1 + \frac{1}{2} \cdot a_2^1 \\ &= \frac{1}{2}. \end{aligned}$$

The fact that this posterior belief is larger than the critical threshold $\tau_1 = 2/5$ implies that drivers don't park illegally. This signaling strategy further decreases the expected social cost of illegal parking from $\frac{2}{3}$ to the probability that the city sends the signal L, or to[7]

$$\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 = \frac{1}{3}.$$

[7] The city can decrease the expected social cost of illegal parking even further to $\frac{1}{6}$ by sending the signals L and H with probabilities $\frac{1}{4}$ and $\frac{3}{4}$, respectively, when $k = 1$ and just the signal H when $k = 2$. This is the lowest possible value of the expected social cost in this example.

■

It is also possible to illustrate by example that the optimal allocation of enforcement resources depends on whether the city is able to disseminate information or not: a city that can disseminate information about its enforcement allocates its resources differently than a city that does not. The reason that this is so is clarified in the general analysis below, so we do not provide a specific example for this.

3 The Optimal Ratio Rule

For any probabilities and allocations $p_k(m)$ and $\left\{a_k^i(m)\right\}$, each message m achieves deterrence on some set of locations $S(m) \subseteq \{1, \ldots, n\}$. We may thus identify each message m with the set $S(m)$ on which it deters provided we add the following deterrence constraint:

$$a^i(m) \equiv \sum_{k=1}^{K} a_k^i(m) \Pr(k \,|m) \geq \tau^i \tag{6}$$

for every location $i \in S(m)$, and for every message $m \in M \equiv 2^{\{1,\ldots,n\}}$ that is sent with a positive probability. The set of messages includes a message that achieves no deterrence (or that achieves deterrence on the empty set, $\varnothing \in M$). And no loss of generality is implied by the assumption that exactly one message deters on any given set of locations. This is because if two messages m and m' deter on the same set of locations then they can be merged into one message $m \cup m'$.

The identification of messages with the set of locations on which they achieve deterrence clarifies that persuasion, or the sending of messages, can only be useful if there is some underlying uncertainty.

Proposition 1. *Persuasion is ineffective without true underlying uncertainty. If there is only one state of the world, then there exists an optimal solution that does not involve (non-trivial) persuasion.*

Proof. Suppose that there is only one state of the world. Optimality requires that in this state a message m_1 that is such that $S(m_1)$ maximizes the value of deterrence is sent with probability one. Sending another message m_2 that induces the same or less deterrence is either unnecessary or strictly dominated.

π_1	$a_2^1(m_1)$	$a_2^2(m_1)$	$a_2^3(m_1)$	r_1
	$a_2^1(m_2)$	$a_2^2(m_2)$	$a_2^3(m_2)$	
	τ^1	τ^2	τ^3	

■

The next result shows that no loss of generality is implied by restricting attention to a specific class of allocations of resources.

Proposition 2 (the "Optimal Ratio Rule"). *Given probabilities $\{p_k(m)\}$ and an allocation $\left\{a_k^i(m)\right\}$, the same probabilities together with the allocation $\left\{a_k^{i*}(m)\right\}$ such that:*

For every location $i \in S(m)$, for every state k, and for every message m that is sent with a positive probability at k,

$$a_k^{i*}(m) = \frac{\tau^i r_k}{\sum_{j \in S(m)} \tau^j};$$

and for every location $i \notin S(m)$, or messages m that are sent with probability zero,

$$a_k^{i*}(m) = 0;$$

achieves equal or better deterrence than $\left\{a_k^i(m)\right\}$.

Proof. Fix probabilities $\{p_k(m)\}$ and an allocation $\left\{a_k^i(m)\right\}$. For every location $i \in S(m)$ that is deterred by message m,

$$\sum_{k=1}^{K} \Pr(k \,|m) \, a_k^i(m) \geq \tau^i.$$

Summing over $i \in S(m)$ and changing the order of summation yields

$$\begin{aligned} \textstyle\sum_{i \in S(m)} \tau^i &\leq \textstyle\sum_{i \in S(m)} \sum_{k=1}^{K} \Pr(k \,|m) \, a_k^i(m) \\ &\leq \textstyle\sum_{k=1}^{K} \Pr(k \,|m) \sum_{i \in S(m)} a_k^i(m) \\ &\leq \textstyle\sum_{k=1}^{K} \Pr(k \,|m) \, r_k \end{aligned}$$

where the last inequality follows from feasibility (1).

It therefore follows that

$$\tau^i \leq \sum_{k=1}^{K} \Pr(k \,|m) \frac{\tau^i r_k}{\sum_{j \in S(m)} \tau^j}$$

and so the allocation $a_k^{i*}(m) = \frac{\tau^i r_k}{\sum_{j \in S(m)} \tau^j}$ for every $i \in S(m)$, state k, and message m, and $a_k^{i*}(m) = 0$ for every $i \in \mathbf{N} \setminus S(m)$, state k, and message m, also achieves deterrence of the set $S(m)$. ■

The next example illustrates the intuition for this result.

Example 2. Consider the case in which the city has three neighborhoods with the corresponding thresholds $\tau^1 = 2$, $\tau^2 = 3$ and $\tau^3 = 4$. There are three equally likely states, with the resources $r_1 = 1$, $r_2 = 8$ and $r_3 = 14$, respectively. The city allocates its resources and sends two messages m_1 and m_2 as depicted in the following matrix:

$\frac{1}{3}$	$1-p$	1	–	–	1
	p	1			
$\frac{1}{3}$		2	3	5	10
$\frac{1}{3}$		3	6	5	14
		2	3	4	

Message m_1 is sent in state 1 with probability $1-p$, and message m_2 is sent in state 1 with probability p, and in states 2 and 3.

The city achieves deterrence with message m_2 but not with message m_1. Thus, a larger probability p implies a larger probability of deterrence, but if p is too large, then the city loses deterrence in the third location. The maximum probability p that allows the city to deter in all three locations is $p=\frac{1}{2}$. The overall probability of deterrence (in all three locations) with this probability $p=\frac{1}{2}$ is $\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}+\frac{1}{3}=\frac{5}{6}$.

If however the city allocates its enforcement resources proportionally to the deterrence thresholds in the three locations as implied by the Optimal Ratio Rule, then it can achieve more deterrence. The allocation according to the Optimal Ratio Rule is depicted in the following matrix:

	1			
$\frac{1}{3}$ p	$\frac{2}{9}\times 1$	$\frac{3}{9}\times 1$	$\frac{4}{9}\times 1$	1
$\frac{1}{3}$	$\frac{2}{9}\times 10$	$\frac{3}{9}\times 10$	$\frac{4}{9}\times 10$	10
$\frac{1}{3}$	$\frac{2}{9}\times 14$	$\frac{3}{9}\times 14$	$\frac{4}{9}\times 10$	14
	2	3	4	

With this allocation, the city can set $p=\frac{3}{4}$ and achieve deterrence in all three locations with probability $\frac{1}{3}\cdot\frac{3}{4}+\frac{1}{3}+\frac{1}{3}=\frac{11}{12}$.

The Optimal Ratio Rule implies that the problem can be recast as a problem of choosing the probabilities $\{p_k(m)\}$ so as to maximize deterrence, subject to the probability constraints (1) and the deterrence constraint (5) applied to $\left\{a_k^{i*}(m)\right\}$ as follows:

$$\max_{\{p_k(m)\}}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{i\in S(m)}s^i p_k(m)\,\pi_k \tag{7}$$

subject to the probability constraints (1) and the deterrence constraint:

$$a^{i*}(m)\equiv\sum_{k=1}^{K}a_k^{i*}(m)\Pr(k\,|m)\geq\tau^i \tag{8}$$

for every location $i\in S(m)$, and for every message $m\in M$ that is sent with a positive probability.

The objective function (6) is linear, but the deterrence constraint is not because the conditional probabilities $\Pr(k\,|m)$ are not linear in the probabilities $\{p_k(m)\}$, and because the constraint is only imposed on messages that are sent with a positive probability rather than on all messages. Nevertheless, as shown by the next proposition, the problem can be recast as a linear programming problem.

Corollary. *The problem* min (3) *subject to the probability and resource constraints* (1) *and* (2)*, respectively, can be recast as the linear programming problem:*

$$\max_{\{p_k(m)\}}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{i\in S(m)}s^i p_k(m)\,\pi_k \tag{9}$$

subject to the probability constraints (1) and the deterrence constraints:

$$\sum_{k=1}^{K} p_k(m)\,\pi(k) a_k^{i*}(m) \geq \tau^i \sum_{k=1}^{K} p_k(m)\,\pi(k) \tag{10}$$

for every location $i \in S(m)$ and message $m \in M$.

Proof. The problem max (8) subject to the probability and deterrence constraints (1) and (9) is a linear programming problem. The objective function (8) is obtained from (3) upon substitution of the resources according to the Optimal Ratio Rule. The deterrence constraints (9) are obtained from the deterrence constraints (7) upon multiplication of both the right- and left-hand-sides of the constrain by the denominator of the conditional probability $\Pr(k\,|m) = \frac{p_k(m)\pi(k)}{\sum_{k'=1}^{K} p_{k'}(m)\pi(k')}$. The deterrence constraints can be imposed on all messages because for messages that are not sent with a positive probability in $p_k(m) = 0$, which trivially satisfies the deterrence constraint. ■

The result that the problem can be recast as a linear programming problem is useful because there are several well known algorithms for solving linear programming problems that work very well in practice. We do not think that the type of problem described here is likely to be very large in practice anyway, but another advantage of linear programming problems is that they can be solved in time that is polynomial in the size of the input of the problem. However, here, the size of the input is the product of the number of states and the number of messages, $k \times 2^n$, which is exponential in the number of locations, n. In the next subsection, we show that it is enough to send just two messages in the optimal solution, but because it is impossible to tell which two messages should be sent, the solution is still exponential in the number of locations, n.

4 Persuasion

A set of probabilities $\{p_k(m)\}$ describes the probability that each message m is sent in each state k. The Optimal Ratio Rule specifies how the total available resources should be allocated across the different locations in the set $S(m)$ when message m is sent in state k.

Obviously, the decision of whether to send any message m (that deters on $S(m)$) in state k depends on the total amount of resources available in state k, as well as on the city's persuasion or signaling objectives. In this section we provide two useful results about the beliefs that the city can induce about the total amount of available resources conditional on any message m.

Denote the posterior expected amount of resources conditional on message m by $r\,|m \equiv \sum_{k=1}^{K} p(k\,|m)\, r_k$ and the probability that message m is sent by $\Pr(m) \equiv \sum_{k=1}^{K} p_k(m)\pi_k$. The total expected amount of resources is denoted $E[r]$. Resources must add up, so

$$E[r] \equiv \sum_{k=1}^{K} \pi_k r_k = \sum_{m=1}^{M} \Pr(m) \cdot r\,|m\,.$$

The next lemma is a generalization of a lemma of Aumann and Maschler (1995, p. 25). Denote the posterior total expected amount of resources conditional on two messages, m and m' by

$$r\left|m,m'\right. \equiv \frac{\Pr(m)}{\Pr(m)+\Pr(m')}\cdot r\left|m\right. + \frac{\Pr(m')}{\Pr(m)+\Pr(m')}\cdot r\left|m'\right. .$$

Lemma 1. *Any two messages l and h that are sent with probabilities* $\Pr(l)$ *and* $\Pr(h)$ *and that induce posterior expectations $r\left|l\right. < r\left|h\right.$, can be replaced with two messages l' and h' that induce any two posterior expectations $r\left|l\right. \le r\left|l'\right. \le r\left|h'\right. \le r\left|h\right.$ such that: (1) the overall probability of sending messages l and h is preserved, or*

$$\Pr(l)+\Pr(h)=\Pr(l')+\Pr(h'),$$

and (2) and posterior expectation conditional on the two messages is preserved, or

$$r\left|l,h\right. = r\left|l',h'\right. ,$$

without affecting any of the other messages or the probabilities with which they are sent.

Proof. Sending messages l' and h' instead of messages l and h with any conditional probabilities $\Pr(l'\left|l\right.)=1-\Pr(h'\left|l\right.)$ and $\Pr(l'\left|h\right.)=1-\Pr(h'\left|h\right.)$ preserves the overall probability of sending messages l and h and posterior expectations conditional on the two messages, $\Pr(l)+\Pr(h)=\Pr(l)+\Pr(h')$, and $a\left|l,h\right. = a\left|l',h'\right.$, respectively. Messages l' and h' induce posterior expectations $a\left|l\right. \le a\left|l'\right. < a\left|l,h\right. < a\left|h'\right. \le a\left|h\right.$ if the conditional probabilities $\Pr(l'\left|l\right.)$ and $\Pr(l'\left|h\right.)$ are chosen to satisfy the following two equations:

$$a\left|l'\right. = \frac{\Pr(l)\Pr(l'\left|l\right.)a\left|l\right. + \Pr(h)\Pr(l'\left|h\right.)a\left|h\right.}{\Pr(l)\Pr(l'\left|l\right.)+\Pr(h)\Pr(l'\left|h\right.)}$$

and

$$a\left|h'\right. = \frac{\Pr(l)\Pr(h'\left|l\right.)a\left|l\right. + \Pr(h)\Pr(h'\left|h\right.)a\left|h\right.}{\Pr(l)\Pr(h'\left|l\right.)+\Pr(h)\Pr(h'\left|h\right.)}.$$

The solution to these two linear independent equations in two unknowns is

$$\Pr(l'\left|l\right.) = \frac{a\left|h\right. - a\left|l'\right.}{a\left|h\right. - a\left|l\right.}\cdot\frac{\Pr(l)+\Pr(h)}{\Pr(l)}\cdot\frac{a\left|h'\right. - E\left[a\left|l,h\right.\right]}{a\left|h'\right. - a\left|l'\right.}$$

and

$$\Pr(l'\left|h\right.) = \frac{a\left|l'\right. - a\left|l\right.}{a\left|h\right. - a\left|l\right.}\cdot\frac{\Pr(l)+\Pr(h)}{\Pr(h)}\cdot\frac{a\left|h'\right. - E\left[a\left|l,h\right.\right]}{a\left|h'\right. - a\left|l'\right.}.$$

These two conditional probablities lie between 0 and 1 because $\frac{\Pr(l)+\Pr(h)}{\Pr(l)}\cdot\frac{a|h' - E[a|l,h]}{a|h' - a|l'} = 1$ and $\frac{a|h - a|l'}{a|h - a|l}\cdot\frac{\Pr(l)+\Pr(h)}{\Pr(h)} \le 1$ if and only if $a\left|l'\right. \le a\left|l,h\right.$. ∎

The next lemma provides another useful observation, which is also a generalization of the same lemma of Aumann and Maschler (1995, p. 25), that characterizes the maximal spread that can be obtained between two induces beliefs about the total expected amount of resources.

Lemma 2. *Given a distribution of resources $r_1,\ldots,r_K$, and given any two total expected amounts of resources $R_L < E\left[\ r\ \right] < R_H$, it is possible to send two messages L and H such that*

$$r|L = R_L \qquad r|H = R_H$$

provided that $r_1 \leq R_L$, $R_H \leq r_K$, *and*

$$R_L \geq \frac{\sum_{k=1}^{k'-1} \pi_k r_k + (1-p)\pi_{k'} r_{k'}}{\sum_{k=1}^{k'-1} \pi_k + (1-p)\pi_{k'}}$$

where $k' \in \{1, \cdots, K\}$ *and* $p \in [0, 1)$ *are the unique solution to:*

$$R_H = \frac{\sum_{k=k'+1}^{K} \pi_k r_k + p\pi_{k'} r_{k'}}{\sum_{k=k'+1}^{K} \pi_k + p\pi_{k'}}.$$

Proof. The maximum difference between R_H and R_L is obtained when message H is sent in states $k \in \{k'+1, \cdots, K\}$, message L in states $k \in \{1, \cdots, k'-1\}$, and in state k' messages H and L are sent with probabilities p and $1-p$, respectively, for some state $k' \in \{1, \ldots, K\}$ and probability p. The condition on R_L reflects the lowest possible value of R_L given a set value for R_H under this signaling/persuasion policy. Less extreme messages permit closer values of R_H and R_L. ■

The next example illustrates the restrictions that the distribution of resources imposes on the relationship between the induced posterior expectations about the total amount of resources available $r\,|H$ and $r\,|L$.

Example 3. Consider a case with three states of the world. Resources are given by $(r_1, r_2, r_3) = (0, \frac{1}{2}, 1)$ and the prior is $(\pi_1, \pi_2, \pi_3) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. In this case, $E\,[\,r\,] = \frac{1}{2}$, and

$$R_L \geq \max\left\{\frac{3R_H - 2}{8R_H - 5}, 0\right\}.$$

If $\frac{1}{2} < R_H \leq \frac{2}{3}$ then R_L is unrestricted; the lowest possible value of R_L increases monotonically with $\frac{2}{3} < R_H < 1$; and if $R_H = 1$ then $R_L \geq \frac{1}{3}$.

5 The Monotone Case

We may assume without loss of generality that the locations can be ordered by their importance, or:

$$s^1 \geq s^2 \geq \cdots \geq s^n.$$

In this section, we assume that deterrence thresholds can also be ranked in the same way, or:

$$\tau^1 \leq \tau^2 \leq \cdots \leq \tau^n.$$

We refer to this assumption as the *monotonicity assumption*. Monotonicity allows us to completely solve the problem, but it involves a considerable loss of generality. In particular, it implies that it is also more effective to deploy resources in more important locations, or:

$$\frac{s^1}{\tau^1} \geq \frac{s^2}{\tau^2} \geq \cdots \geq \frac{s^n}{\tau^n}.$$

The monotone case captures a situation where in "more important neighborhoods" as defined by the disutilities $\{s^i\}$, residents are also "better behaved" in the sense of having a lower threshold τ^i for not parking illegally. Indeed, one often hears the complaint that cities care more about law enforcement in "good" compared to "bad" neighborhoods, and it seems that people are generally harder to deter in bad compared to good neighborhoods.

It is straightforward to verify that if it is optimal to deter at location i under some message m, then it is also optimal to deters at location $j < i$. It follows that the number of messages that is needed is only $n + 1$. Namely, in the optimal solution, it is enough to restrict attention only to those messages associated with the sets $\varnothing$, $\{1\}$, $\{1, 2\}$, ..., $\{1, \ldots, N\}$. Moreover, the optimal solution satisfies "nesting." Namely, the sets $S(m)$ can be nested in the sense that $n' < n''$ implies $S(\{1, \ldots, n'\}) \subseteq S(\{1, \ldots, n''\})$.

Monotonicity simplifies the city's allocation problem. If the total expected amount of resources is less than τ^1 then no deterrence is possible. If the total expected amount of resources is more than τ^1 but less than $\tau^1 + \tau^2$ then it is possible to deter only in location 1, and so on. Continuing in the same way we see that devoting all the available resources to deterrence with no messages produces the following non-increasing step-function disutility:

$$D(a) = \begin{cases} \sum_{i=1}^{N} s^i & \text{if } 0 \leq a < \tau^1 \\ \sum_{i=n}^{N} s^i & \text{if } \sum_{i=1}^{n-1} \tau^i \leq a < \sum_{i=1}^{n} \tau^i, \;\; 2 \leq n \leq N \\ 0 & \text{if } \sum_{i=1}^{N} \tau^i \leq a \end{cases}$$

that maps the amount of available expected resources a into disutility. The steps in the function $D(a)$ become longer and lower, as shown in Figure 1 below.

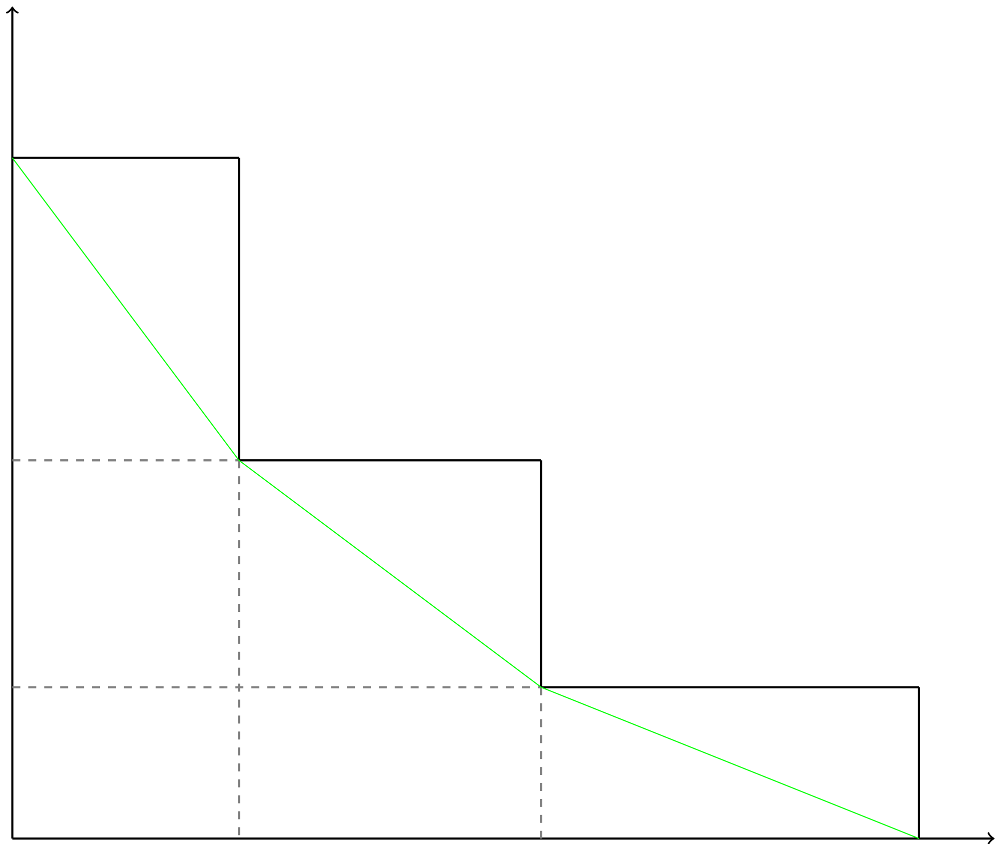

Figure 1: $D(a)$ in the monotone case

The sending of messages, signaling, or persuasion allows the city to achieve a lower disutility than $D(a)$. Recal that $r\,|m$ denotes the posterior expected amount of resources conditional on message m. The value of the city's objective function when it sends messages $1, \ldots, M$ with probabilities $\Pr(1), \ldots, \Pr(M)$, respectively is:

$$\sum_{m=1}^{M} \Pr(m) \cdot D\left(r\,|m\right).$$

The monotone case admits a complete solution of the city's problem with no more than two messages as follows.

Proposition 3. *Suppose that the monotonicity assumption holds. Suppose that the expected amount of resources* $E[r]$ *is such that* $\sum_{i=1}^{n-1} \tau^i \leq E\left[r\right] < \sum_{i=1}^{n} \tau^i$ *for some* $2 \leq n < N$.[8] *Then, the optimal solution involves the sending of only two messages L and H such that the posterior expectation* $a\,|H$ *is set equal to* $\sum_{i=1}^{n} \tau^i$ *if this is possible given the distribution of resources, and the posterior expectation* $a\,|L$ *is set equal to* $\sum_{i=1}^{n-1} \tau^i$ *if this is possible given the distribution of resources, and as low as possible otherwise. If the distribution of resources does not allow to set* $a\,|H = \sum_{i=1}^{n} \tau^i$ *then*

[8] If either $E\left[r\right] < \tau^1$ or $\sum_{i=1}^{N} \tau^i \leq E\left[r\right]$ then the problem is trivial. In the former case, no deterrence is possible, and in the latter case, full deterrence is possible with no messages.

persuasion is unhelpful and no messages (or equivalently just one message) should be sent.

Proof. The proof of Proposition 3 relies on Lemma 1. Suppose that $\sum_{i=1}^{n-1}\tau^i \leq E[r] < \sum_{i=1}^{n}\tau^i$ for some $2 \leq n < N$.

If a policy includes two messages l and h that induce posterior expectations $a\,|l < \sum_{i=1}^{n-1}\tau^i < \sum_{i=1}^{n}\tau^i < a\,|h$, then expected disutility can be lowered if the two messages l and h are replaced with messages l' and h' that are such that $a\,|l' = \sum_{i=1}^{n-1}\tau^i$ and $a\,|h' = \sum_{i=1}^{n}\tau^i$. The step structure of the disutility function $D(a)$ implies that the straight line that connects the points $(a\,|l', D(a\,|l'))$ and $(a\,|h', D(a\,|h'))$ lies strictly below the straight line that connects the points $(a\,|l, D(a\,|l))$ and $(a\,|h, D(a\,|h))$. Therefore, the expected disutility from sending messages l' and h' instead of l and h, which lies on this line at the point $a\,|l,h = a\,|l',h'$, is lower, or

$$\frac{\Pr(l)\,D(a\,|l)}{\Pr(l)+\Pr(h)} + \frac{\Pr(h)\,D(a\,|h)}{\Pr(l)+\Pr(h)} \leq \frac{\Pr(l')\,D(a\,|l')}{\Pr(l')+\Pr(h')} + \frac{\Pr(h')\,D(a\,|h')}{\Pr(l')+\Pr(h')}.$$

It therefore follows that performance of this replacement of messages decreases expected social disutility from

$$\sum_{m\neq l,h}\Pr(m)\,D(a\,|m) + \Pr(l)\,D(a\,|l) + \Pr(h)\,D(a\,|h)$$

to

$$\sum_{m\neq l,h}\Pr(m)\,D(a\,|m) + \Pr\left(l'\right)D\left(a\,\middle|l'\right) + \Pr\left(h'\right)D\left(a\,\middle|h'\right).$$

If a policy includes two messages l and h that induce posterior expectations $\sum_{i=1}^{n-1}\tau^i \leq a\,|l$ and $\sum_{i=1}^{n}\tau^i < a\,|h$ then expected disutility can be lowered if the the two messages l and h are replaced with messages l' and h' that are such that $a\,|l' = a\,|l$ and $\sum_{i=1}^{n}\tau^i = a\,|h'$. The straight line that connects the points $(a\,|l', D(a\,|l'))$ and $(a\,|h', D(a\,|h'))$ still lies strictly below the straight line that connects the points $(a\,|l, D(a\,|l))$ and $(a\,|h, D(a\,|h))$. Therefore, performance of this replacement of messages also decreases expected social disutility as before.

It follows that it is enough to send only two messages L and H in the optimal solution such that $a\,|H = \sum_{i=1}^{n}\tau^i$ if this is possible given the distribution of resources and $a\,|L \geq \sum_{i=1}^{n-1}\tau^i$. The step structure of the function $D(a)$ implies that if the distribution of resources does not allow to set $a\,|H = \sum_{i=1}^{n}\tau^i$ then persuasion is unhelpful and no messages should be sent. It also implies that $a\,|L$ should be set equal to $\sum_{i=1}^{n-1}\tau^i$ if this is possible given the distribution of resources, and as low as possible otherwise. ■

Figure 2 below shows that setting $r\,|H = \sum_{i=1}^{n}\tau^i$ if possible, and setting $r\,|L$ as low as possible but not below $\sum_{i=1}^{n-1}\tau^i$ decreases expected social disutility. It also illustrates the reason that if it is impossible to set $a\,|H = \sum_{i=1}^{n}\tau^i$ then persuasion is ineffective.

– Figure 2 here (Optimal solution in the monotone case; draw two lines to illustrate that the one associated with the optimal solution lies below the other one ... –

The fact that $r\,|L$ should be set as low as possible given the distribution of resources, but not below $\sum_{i=1}^{n-1}\tau^i$, raises the question of whether it may be beneficial to destroy resources in order to set $a\,|L = \sum_{i=1}^{n-1}\tau^i$ when this is impossible given the distribution of resources. The answer to this question is, not surprisingly, negative.[9]

As illustrated by Figure 2 and elaborated further in the next section, when the message L induces a posterior expectation $a\,|L > \sum_{i=1}^{n-1}\tau^i$ the convexification of the function $D(a)$ is partial. The next proposition characterizes the distribution of resources that permit complete convexification of the disutility function $D(a)$ in the monotone case.

Proposition 4. *If $r_1 \leq \sum_{i=0}^{m}\tau^i \leq E\,[\,r\,] < \sum_{i=0}^{m+1}\tau^i \leq r_K$ for some $m \leq n-1$ then it is possible to achieve full convexification ($a|H = \sum_{i=0}^{m+1}\tau^i$ and $a|L = \sum_{i=0}^{m}\tau^i$) provided that*

$$\sum_{i=0}^{m}\tau^i \geq \frac{\sum_{k=0}^{k'-1}\pi_k r_k + (1-p)\pi_{k'} r_{k'}}{\sum_{k=0}^{k'-1}\pi_k + (1-p)\pi_{k'}}$$

where $k' \in \{1,\cdots,K\}$ and $p \in [0,1)$ are the unique solution to:

$$\sum_{i=0}^{m+1}\tau^i = \frac{\sum_{k=k'+1}^{K}\pi_k r_k + p\pi_{k'} r_{k'}}{\sum_{k=k'+1}^{K}\pi_k + p\pi_{k'}}.$$

Otherwise, convexification is partial, either $a|H = \sum_{i=0}^{m+1}\tau^i$ but $a|L > \sum_{i=0}^{m}\tau^i$, or persuasion is altogether unhelpful.

Proposition 4 is a corollary of Lemma 2 in the previous section.

6 Constrained Convexification

In this section we extend the analysis performed in the previous section for the monotone case to the general case. We explain the sense in which the problem is a constrained

[9]Suppose then that $a\,|L$ is optimally set at a continuity point of $D(a)$. Decreasing it further necessitates the destruction of resources. We show that such destruction of resources is inefficient.

The equation of the line that connects the points $(a\,|L\,, D(a\,|L))$ and $(a\,|H\,, D(a\,|H\,))$ is:

$$y = \frac{D(a\,|H\,) - D(a\,|L)}{a\,|H\, - a\,|L} \cdot x + D(a\,|L) - \frac{D(a\,|H\,) - D(a\,|L)}{a\,|H\, - a\,|L} \cdot a\,|L\,.$$

If $a\,|L$ is lowered by a small $\varepsilon > 0$, then the expected amount of resources decreases from a to $a - \varepsilon\pi(L)$ and the line of expected distutility connects the two points: $(a\,|L - \varepsilon, D(a\,|L))$ and $(a\,|H\,, D(a\,|H\,))$ is:

$$y = \frac{D(a\,|H\,) - D(a\,|L)}{a\,|H\, - a\,|L + \varepsilon} \cdot x + D(a\,|L) - \frac{D(a\,|H\,) - D(a\,|L)}{a\,|H\, - a\,|L + \varepsilon} \cdot (a\,|L - \varepsilon)\,.$$

Algebraic manipulation shows that the height of the former line at the point where $x = a$ is equal to the height of the second line at the point where $x = a - \varepsilon\pi(L)$. It follows that the destruction of resources does not lower expected disutility.

convexification problem, and characterize the number of messages needed for the optimal solution. However, we cannot provide an explicit solution of the problem like in the monotone case.

Devoting all the available resources to deterrence on the set of neighbourhoods $S \subseteq \{1, \ldots, n\}$ with no messages produces a non-increasing step-function disutility:

$$D_S(a) = \begin{cases} \sum_{i \in \mathbf{N}} s^i & \text{if } a < \sum_{i \in S} \tau^i \\ \sum_{i \in \mathbf{N} \setminus S} s^i & \text{if } \sum_{i \in S} \tau^i \leq a \end{cases}$$

that maps the amount of available expected resources a into disutility.

It follows that the minimal disutility that can be achieved without persuasion, or without sending any messages, is given by the following non-increasing step-function:

$$D(a) = \min_{S \subseteq \mathbf{N}} D_S(a).$$

In the monotone case, the steps defined by the disutility function $D(a)$ became longer and lower, but this is not necessarily the case generally.

Define the convexification of $D(a)$ from below as

$$\text{conv } D(a) \equiv \max \tilde{D}(a)$$

where the maximum is taken over all convex functions $\tilde{D}(a) \leq D(a)$ for all $a \geq 0$. The convexification of $D(a)$ is a piecewise linear, monotone nonincreasing, convex function. The functions conv $D(a)$ and $D(a)$ coincide on points $a \geq \sum_{i=1}^{N} \tau^i$. Denote the points on which conv $D(a)$ and $D(a)$ coincide in the interval $\left[0, \sum_{i=1}^{N} \tau^i\right]$ by $a_0, a_1, \ldots, a_{M_d}$, where $0 = a_0 < a_1 < \cdots < a_{M_d} = \sum_{i=1}^{N} \tau^i$. Each pair of consecutive points a_l, a_{l+1} defines a linear segment of the function conv $D(a)$. There is a finite number of such points because each such point must be a discontinuity point of the function $D(a)$ and there is only a finite number of such discontinuity points. The *number of steps* of the function $D(a)$ in any segment $[a_l, a_{l+1}]$ is given by the number of discontinuity points of $D(a)$ in the segment $[a_l, a_{l+1}]$. See Figure 3 below.

– Figure 3: The functions $D(a)$ and conv $D(a)$ in the general case –

If the distribution of resources imposed no constraints over the distribution of the posterior expectations $\{a\,|m\}$, except of course for the requirement that resources add up, or that

$$\sum_{m=1}^{M} a\,|m \cdot \Pr(m) = E\,[r]$$

then the optimal solution could have been obtained as the solution to the following (unconstrained) convexification problem

$$\min_{\{\Pr(m)\}\{a|m\}} \left\{ \sum_{m=1}^{M} \Pr(m) D\left(\{a\,|m\}\right) : \sum_{m=1}^{M} \Pr(m) = 1, \sum_{m=1}^{M} \Pr(m) \cdot a\,|m = E\,[r] \right\}$$

and would have required only two messages. Specifically, as shown in Figure 4 below, the optimal solution would have involved sending only messages L and H with induced posterior beliefs $a\,|L$ and $a\,|H$ that are equal to the consecutive two coincidence points that are such that $a_l < E\,[r] < a_{l+1}$,[10] with probabilities $\Pr(H)$ and $\Pr(L) = 1 - \Pr(H)$ that are such that $\Pr(L) \cdot a\,|L + \Pr(H) \cdot a\,|H = E\,[r]$.

– Figure 4: Optimal solution in the unconstrained case involves only two messages L and H –

However, the distribution of resources imposes restrictions on the set of posterior expectations $\{a\,|m\}$ that have to be taken into account. These restrictions imply that the problem is given by the following constrained convexification problem

$$\min_{\{\Pr(m)\}\{a|m\}} \left\{ \sum_{m=1}^{M} \Pr(m) D\left(\{a\,|m\}\right) : \sum_{m=1}^{M} \Pr(m) = 1, \sum_{m=1}^{M} \Pr(m) \cdot a\,|m = E\,[r] \right\}$$

subject to the constraint that there exists an assignment of probabilities $\{p_k(m)\}$ that induces the set of posterior expectations $\{a\,|m\}$, or such that

$$a\,|m = \sum_{i=1}^{n} a^i(m) = \sum_{k=1}^{K} a_k^i\,(m) \Pr\left(k\,|m\right)$$

where each conditional probability $\Pr\left(k\,|m\right)$ can be expressed in terms of the probabilities $\{p_k(m)\}$ using Bayes Rule as in (2).

The restrictions that the distribution of resources imposes on the set of posterior expectations $\{a\,|m\}$ implies that sometimes three or more messages may generate a lower value of the objective function than just two messages. The next example describes a situation in which three messages are better than two. Similar examples may be constructed in which four messages are better than three and two, five are better than four, three and two, etc.

Example 4. A city has two neighborhoods with the thresholds $\tau^1 = \frac{1}{2}$ and $\tau^2 = 1$ and social disutilities $s^1 = \frac{1}{4}$ and $s^2 = 1$. There are three states, with resources $r_1 = 0$, $r_2 = \frac{1}{2}$ and $r_3 = 1$, and probabilities $\pi_1 = \frac{1}{4}$, $\pi_2 = \frac{1}{2}$ and $\pi_3 = \frac{1}{4}$, respectively. Clearly, as shown by Figure 3 below, optimal deterrence with two messages L and H (such that $a\,|L < a\,|H$) requires that $a\,|H = 1$ and $a\,|L$ is set as low as possible, which in this case implies $a\,|L = \frac{1}{3}$, $\Pr(L) = \frac{3}{4}$ and $\Pr(H) = \frac{1}{4}$. The value of the objective function in this case is $\frac{3}{4} \cdot \frac{5}{4} + \frac{1}{4} \cdot \frac{1}{4} = 1$. This is also the value of the objective function with no messages at all or just one message. But with three messages that reveal the state of the world, the expected value of the objective function is $\frac{1}{4} \cdot \frac{5}{4} + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot \frac{1}{4} = \frac{7}{8} < 1$.

– Figure 5: Three messages may be better than two –

[10] If $E\,[r]$ is equal to one of the coincidence points, then the optimal solution requires just one, or no messages at all.

The next theorem bounds the maximum number of messages needed in order to implement the optimal solution.

Proposition 5. *Suppose that the expected amount of resources $E[r]$ is an interior point of the segment $[a_l, a_{l+1}]$. Then, the number of messages needed in order to obtain the optimal solution is no more than the number of steps of the function $D(a)$ in the segment $[a_l, a_{l+1}]$ plus one. If the expected amount of resources coincides with one of the points $a_0, a_1, \ldots, a_m$, then no messages or just one message is needed for the optimal solution.*

Proof. Suppose that the expected amount of resources $E[r]$ is an interior point of some segment $[a_l, a_{l+1}]$. An identical argument to the one used in the proof of Proposition 3 shows that no loss of generality is implied by restricing attention to a set of messages that induce posterior expectations that lie in the interval $[a_l, a_{l+1}]$. This is because any two messages l and h that induce posterior expectations $a\,|l < a_l < a_{l+1} < a\,|h$, can be replaced by two messages l' and h' that are such that $a\,|l' = a_l$ and $a\,|h' = a_{l+1}$ without affecting the probabilities of the other messages or their posterior expectations in a way that decreases expected disutility. And any two messages l and h that induce posterior expectations $a_l \leq a\,|l$ and $a_{l+1} < a\,|h$ can be replaced by two messages l' and h' that are such that $a\,|l' = a\,|l'$ and $a\,|h' = a_{l+1}$ without affecting the probabilities of the other messages or their posterior expectations in a way that decreases expected disutility. A similar argument shows that any two messages l and h that induce posterior expectations $a\,|l < a_l$ and $a\,|h \leq a_{l+1}$ can be replaced by two messages l' and h' that are such that $a\,|l' = a_l$ and $a\,|h' = a\,|h$ without affecting the probabilities of the other messages or their posterior expectations in a way that decreases expected disutility.

There is no need to send two messages that induce the same posterior expectation because any such two messages m_i and m_j can be combined into one message that is sent with probability $\Pr(m_i) + \Pr(m_j)$ and induces the same expected posterior as $a\,|m_i = a\,|m_j$ without affecting any other probabilities or posterior expectations.

Finally, if the expected amount of resources coincides with one of the points $a_0, a_1, \ldots, a_{M_d}$, then no messages or just one message is needed for the optimal solution because as implied by the preceding discussion, it is impossible to obtain a value of the objective function that lies below conv $D(a)$. ■

As in the monotone case, the convexification of the function $D(a)$ may be incomplete in the sense that the optimal solution may lie strictly above the function conv $D(a)$.

7 Endogenous Distribution of Resources & Deterrence over Time

It is possible to endogenize the prior distribution over the amount of available resources in the following way. Suppose that the city employs n inspectors. Each inspector is allocated to a specific day and time, or to several time slots, depending on how many hours it is required to work per day or week. Each inspector shows up to each asigned time slot with probability $1 - \varepsilon$, independently across the different inspectors.

Any assignment of inspectors to time slots generates a prior distribution of resources available in each time slot. It is then possible to optimize over these prior distributions, given that in each time slot, the city allocates the available resources and disseminates information optimally, as described above. The solution of such a problem provides a theory of enforcement operations.

The ideas presented here suggest that it is possible to achieve better overall deterrence through convexification. That is, the city should allocate its insectors randomly so that in some states of the world their number is small and in other states it is large. This as thhis would allow to improve overall deterrence through convexification. Renault et al. (2016) and Ely (2017) provide solutions of related problems. We are hopeful that the methods they developed can be used to solve the dynamic version of the problem presented here as well.

References

[1] Aumann, R. J., and M. Maschler (1995) *Repeated Games with Incomplete Information*, MIT Press.

[2] Becker, G. S. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76, 169-217.

[3] Bergemann, D. and S. Morris (2017) "Information Design: A Unified Perspective," unpublished working paper.

[4] Borel, E. (1921) "La Thorie du Jeu et les quations Intgrales Noyau Symtrique," *Comptes Rendus de l'Acadmie des Sciences* 173, 1304–1308. Translated by Savage LJ, "The theory of play and integral equations with skew symmetric kernels," *Econometrica* 21 (1953) 97–100.

[5] Chalfin, A. and J. McCrary (2017) "Criminal Deterrence: A Review of the Literature." *Journal of Economic Literature*, forthcoming.

[6] Eeckhout, Jan, Nicola Persico, and Petra E. Todd. (2010). "A Theory of Optimal Random Crackdowns." *American Economic Review* 100, 1104-35.

[7] Ely, J. C. (2017) "Beeps," *American Economic Review* 107, pp. 31-53.

[8] Gentzkow, M. and E. Kamenica (2016) "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers & Proceedings* 106(5): 597–601.

[9] Hart, S. (2008) "Discrete Colonel Blotto and General Lotto games," *International Journal of Game Theory* 36: 441–460.

[10] Kamenica, E. and M. Gentzkow (2011) "Bayesian Persuasion," *American Economic Review* 101, 2590-2615.

[11] Kolotilin, A. (2018) "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, 13(2), 607-636.

[12] Lando, H. and S. Shavell (2004) "The advantage of focusing law enforcement effort," *International Review of Law and Economics* 24, 209-218.

[13] Le Treust, M. and T. Tomala (2017) "Persuasion with limited communication capacity," unpublished working paper.

[14] Polinsky M. A. and S. Shavell (2000) "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature* 38, 45-76.

[15] Powell, R. (2007) "Defending against Terrorist Attacks with Limited Resources," *The American Political Science Review* 101, pp. 527-541.

[16] Renault, J., E. Solan, and N. Vieille (2016) "Optimal Dynamic Information Provision," unpublished manuscript.

[17] Roberson, (2006) "The colonel blotto game," *Economic Theory* 29: 1–24.

[18] Rabinovich, Zinovi, Albert Xin Jiang, Manish Jain, Haifeng Xu (2015) "Information Disclosure as a Means to Security," *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*, Bordini, Elkind, Weiss, Yolum (eds.), May 4-8, 2015, Istanbul, Turkey.

[19] Sobel, J. (2013) "Giving and receiving advice." Acemoglu, D., Arellano, M., Dekel, E., eds. *Advances in Economics and Econometrics, Tenth World Congress of the Econometric Society*, Cambridge University Press, 305-341.

[20] Tambe, M. (2011) *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*, Cambridge University Press.

[21] Xu, Haifeng, Rupert Freeman, Vincent Conitzer, Shaddin Dughmi, Milind Tambe (2016) "Signaling in Bayesian Stackelberg Games," *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*, J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9-13, 2016, Singapore.