THE PINCHAS SAPIR CENTER

FOR DEVELOPMENT

Tel Aviv University

המרכז לפיתוח ע"ש פנחס ספיר
ליד אוניברסיטת תל אביב
"עמותה רשומה"
580010221

THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

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

Pen´elope Hern´andez and Zvika Neeman

Discussion Paper No. 8-2021

June 2021

אוניברסיטת תל-אביב, תל אביב יפו 6997801 | טל': 03-6409202 | פקס: 03-6405815
Tel Aviv University, Tel Aviv Yafo 6997801 Israel | Tel. 972-3-6409202 | Fax 972-3-6405815 | email: sapircen@tauex.tau.ac.il

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

Penélope Hernández and Zvika Neeman

January 10, 2021

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. We focus on the case where agents care only about the expected amount of enforcement resources given messages received. Optimization in the space of induced mean posterior beliefs involves a partial convexification of the objective function. We describe interpretable conditions under which it is possible to explicitly solve the problem with only two messages: "high enforcement" and "enforcement as usual." We also provide a tight upper bound on the total number of messages needed to achieve the optimal solution in the general case, as well as a general example that attains this bound.

JEL CODES: D82, D83, K42.

KEYWORDS: Bayesian persuasion, deterrence, constrained convexification.

---

*We thank Johannes Hörner, Alex Smolin, Eilon Solan, Bernhard von Stengel, Sarit Weisburd, Andriy (Andy) Zapechelnyuk, Ro'i Zultan and participants at numerous seminars for helpful comments. Hernández acknowledges the support of the Spanish Ministry of Science, Innovation and Universities and the Spanish Agency of Research, co-funded with FEDER funds, grant ECO2017-87245-R, and by Consellería d'Innovació, Universitats, Ciència i Societat Digital, Generalitat Valenciana, grant AICO/2019/053. Neeman is grateful for support from the Israeli Science Foundation (grant #1465/18) and from the Sapir Center for Economic Development.

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 other types of socially undesirable behaviour such as speeding, free-riding, tax evasion, etc.

The basic idea is the following. Suppose that after a frustrating search for parking, a driver has found an illegal parking spot. She considers whether to park there or not. Suppose that the driver would park illegally if she estimates the probability she would be sanctioned to be less than one-third. Suppose also that the number of parking inspectors employed is such that the probability of a sanction is just one-quarter. This probability is not sufficiently high to deter the driver, so in the absence of any other intervention, she would park illegally.

Suppose that the city monitors the locations of its parking inspectors, and so can inform the driver whether or not she would be sanctioned if she parked illegally in the specific location and time she is considering.[1] Quick reflection reveals that, perhaps counter-intuitively, sharing this location- and time-specific information with the driver would improve deterrence, because with probability one-quarter there is indeed a parking inspector nearby, and in this case the driver would be deterred.

The city can do even better by using the following policy. When a parking inspector is nearby, the city would convey the message that "the likelihood of a sanction is high," and when a parking inspector is not nearby the city would randomize: with probability two-thirds it would still convey the message that "the likelihood of a sanction is high," and with probability one-third, it would convey the message that "the likelihood of a sanction is 'as usual.' " A driver who is informed that the likelihood of a sanction is 'as usual' would of course park illegally because she understands that, given the city's policy, the probability she would be sanctioned is in fact equal to zero. But a driver who is informed that the likelihood of a fine is 'high' would realize that the probability of a sanction conditional on the 'high' message is exactly one-third, and would be deterred from illegal parking. It can be shown that this is the optimal "persuasion policy" for the city in this example; it increases the probability of deterrence from zero with no communication, to three-quarters ($\frac{1}{4} + \frac{3}{4} \cdot \frac{2}{3} = \frac{3}{4}$).

This example relies on two important implicit assumptions. First, it is assumed that the city can commit to its messaging strategy and that the driver is aware of it. We believe that this is not an unreasonable imposition in this case because if it is ever discovered that the city deviated from its policy by sending the high message also when it was not supposed

[1] For example, the city can display such location- and time-specific information on electronic street signs, on its website, or in a dedicated mobile app.

to, the city would lose its credibility and with it the ability to deter future drivers, which it would not want to do.[2]

Second, it is assumed that the city faces a single driver in any place and time, and that it can condition its policy on the driver's threshold probability of deterrence. If instead the city faced a continuum of drivers whose threshold probabilities for deterrence was commonly known to be distributed according to some distribution function F, then not informing drivers would deter a fraction $F(\mu)$ of the drivers, where μ denotes the expected probability that an inspector is nearby ($\frac{1}{4}$ in the example above), because only drivers with thresholds less than μ are sufficiently deterred from illegal parking.[3]

An important insight of Kamenica and Gentzkow (2011) is that by using persuasion, the city can induce any two posterior beliefs that an inspector is nearby, r_L and r_H, with probabilities p_L and $p_H = 1 - p_L$, respectively, provided that $r_L \leq \mu \leq r_H$ and that the expected posterior belief is equal to μ, or such that $p_L \cdot r_L + p_H \cdot r_H = \mu$ (such beliefs are said to be Bayes plausible). This allows the city to deter an expected fraction of $p_L \cdot F(r_L) + p_H \cdot F(r_H)$ of the population of drivers. Kamenica and Gentzkow's second insight is that the policy that maximizes expected deterrence consists of finding the optimal pair of induced posterior beliefs. And, that, for every value of the parameter μ, this optimal policy achieves a level of expected deterrence that is equal to the one achieved by the smallest concave function that lies above F.[4] It therefore follows that the optimal choice of induced posterior beliefs achieves a level of expected deterrence that is no worse than $F(\mu)$, and if the function F is not concave, such a policy is strictly better for the city for at least some values of the parameter μ.

How likely is the function F to not be concave? Concavity of F is equivalent to a decreasing marginal return for enforcement effort. Not much evidence exists on the return to enforcement effort. In a famous experiment that was conducted in Kansas City in 1974 (Kelling et al., 1974) a doubling of police patrols was shown to have virtually no statistically significant effect on street crime.[5] Sherman and Weisbrud (1995) famously criticized the Kansas City experiment by claiming that Kansas City is too large a unit of analysis for a doubling of patrols to produce an effect, or for a true reduction in crime to be statistically significant. Sherman and Weisbrud repeated the Kansas City experiment in Minneapolis two decades after the Kansas City experiment, but restricted it to crime "hot spots, " which can

[2] See Best and Quigley (2020) for a model of persuasion where commitment is justified through a concern for future credibility.

[3] The model lends itself to two mathematically equivalent interpretations. Under the first interpretation, the city faces a single driver whose deterrence threshold (type) is distributed according to F; Under the second, the city faces a continuum of measure one of drivers, whose thresholds are distributed according to F. The analysis and all the results are identical under these two interpretations.

[4] These two insights were adapted from the work of Aumann and Maschler (1995) who developed them in the context of repeated games with incomplete information.

[5] This finding had a big effect on the thinking on deterrence. It convinced both academics and the police itself that "police presence does not deter" (Sherman and Weisbrud, 1995).

be as small as a street corner or a city block. They found that a doubling of police patrols in crime hot spots produced reductions in total crime that ranged from 6 percent to 13 percent (however, "observed disorder" decreased by one-half). Their findings are consistent with the prevailing view that "large increases in dosage may be essential if any effect on crime is to be observed" (Sherman and Weisbrud, 1995). This suggests that assuming that the distribution function F is S-shaped or even convex may certainly be a reasonable assumption in many cases. For simplicity, in the analysis below we take this logic a step further and assume that the function F is given by a threshold function. Namely, we assume that in each time and place a certain, commonly known, amount of enforcement effort is needed to achieve full deterrence. This assumption simplifies the mathematical derivation but similar qualitative results would obtain for any function F that is S-shaped or convex.

We consider a general model in which a principal observes the realized amount of enforcement resources available and decides how to allocate them across $N \geq 1$ different locations. We refer to the realized amount of resources as the state of the world. The principal can commit to a policy of sending public messages about the amount of realized resources and their allocation. Drivers in each one of the N locations observe these messages and decide whether or not to park illegally.[6]

For simplicity, we assume that drivers in each location care only about the expected amount of enforcement resources allocated to their location given the message they heard. This implies that the principal also cares only about the mean of the posterior beliefs that are induced by its messages.

We generalize the example to any number of locations and any distribution of available resources and consider in addition the question of how best to deploy the available enforcement resources. The principal's problem is written as a problem of the minimization of social cost subject to a set of constraints that combine both the distribution of resources and the probabilities with which different messages are sent. It is possible to write the problem, equivalently, as a problem of the maximization of expected deterrence subject to the same constraints.

We describe two complementary approaches for addressing this problem. Under both approaches, it is useful to identify messages with the subsets of locations on which they achieve deterrence. This allows us to easily show that the principal cannot benefit from generating endogenous uncertainty to facilitate persuasion. A second benefit of this identification that is useful for the first approach is that it allows us to replace constraints over the distribution of resources with deterrence constraints that require that messages indeed achieve deterrence on the subset of locations where they are supposed to do so. We show that no loss of optimality is implied by restricting attention to allocations that satisfy a so called "Optimal Ratio Rule." This rule implies that enforcement resources should be al-

[6] We assume that the principal cannot send private driver- or location-dependent messages. This assumption simplifies the discussion and is plausible given the applications considered. Because drivers decisions' are independent and impose no externalities on each other, this assumption has no effect on implementable outcomes.

located proportionally to the deterrence thresholds in those locations where deterrence is achieved, conditional on any message and state of the world. The Optimal Ratio Rule implies that the principal's problem, although nonlinear as stated, can nevertheless be recast as a linear programming problem where social cost is minimized subject to the usual probability constraints and deterrence constraints.

The second approach is based on the idea that the problem can be divided into two sub-problems: first, allocate resources optimally for a certain amount of expected resources; and second, choose what message to send in what state of the world and with what probability. Each message induces a posterior belief about the expected amount of resources, and optimization requires that the optimal distribution of messages be chosen.

The first sub-problem is a knapsack problem, and the second sub-problem is a Bayesian persuasion problem.[7] As mentioned above, Kamenica and Gentzkow (2011) showed that Bayesian persuasion can be viewed as picking the optimal distribution of Bayes plausible posterior beliefs. When the principal cares only about the mean of posterior beliefs, as is the case here, Bayes plausibility is equivalent to the requirements that: (1) the expectation of mean posterior beliefs is equal to the mean of the distribution of resources, and (2) the prior distribution of resources is a mean preserving spread of the distribution of posterior means.[8] As explained above, optimization in the space of induced posterior beliefs, induces a convexification of the payoff to the principal under different posterior beliefs.[9] However, in the space of mean posterior beliefs, the requirement that mean posterior beliefs are second-order-stochastically-dominated by prior beliefs implies that convexification *in this space* might be "partial" rather than "full."

We describe interpretable conditions under which it is possible to explicitly solve the problem with only two messages: "high enforcement" and "enforcement as usual" that indicate that the amount of expected resources is high and low, respectively. The message "enforcement as usual" 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 is possible to achieve stronger deterrence than would be achieved otherwise, when the moratorium is not applied.[10] We also provide a tight upper bound on the

[7] In the knapsack problem there is a collection of items, each with a given weight and a given benefit. The objective is to select a subset of the items that maximizes the sum of benefits subject to a constraint on the total weight allowed.

[8] The second requirement is due to the fact that the principal's messages induce a garbling of the receiver's mean posterior beliefs relative to the true state (see Blackwell, 1953; Gentzkow and Kamenica, 2016; and Kolotilin, 2018).

[9] More precisely, optimization induces a convexification of payoffs if the problem is to minimize social cost. It induces a concavification of payoffs if the problem is the maximization of deterrence, as described in the example above.

[10] 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

total number of messages needed to achieve the optimal solution in the general case, as well as a general example that attains this bound.

Related Literature

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.[11] 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 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.

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 (2019) 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. Our paper is more general in that we consider any number of locations, we add uncertainty, and we consider the question of how to further improve deterrence through Bayesian persuasion, or communication.

Finally, there is a rich literature that started with Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) that studies how a sender with commitment ability can affect a receiver's beliefs and thereby induce it to act in a way that benefits the sender.[12] Within this literature, Kolotilin (2018) has observed that Bayesian persuasion problems may be rep-

---

before to the evening of state holidays.

[11] 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.

[12] For a recent survey of the more general literature on "information design" of which this literature is an important part see Bergemann and Morris (2019).

resented as (infinite dimensional) linear programming problems, and characterized the optimal solution through the dual problem. Kolotilin et al. (2017) and Alonso and Câmara (2016) have considered Bayesian persuasion problems with many receivers. In the model of Kolotilin et al. (2017) receivers' actions are independent and impose no externalities on each other so the sender may optimally apply the same persuasion scheme to all of the receivers. It follows that the fact that there are many receivers makes no difference, as is the case here. In the model of Alonso and Câmara (2016) receivers' actions do impose externalities on each other and the optimal persuasion scheme involves the cultivation of special coalitions of receivers. As mentioned above, the fact that when persuasion is projected into the space of posterior means, full convexification of the underlying objective function in this space may be impossible and the distribution of posterior means needs to be second-order-stochatically-dominated by the prior distribution has been observed by Kamenica and Gentzkow (2011).[13] Methods for solving the problem in this case have been developed by Gentzkow and Kamenica (2016), Kolotilin et al. (2017), Dworczak and Martini (2019), and Kleiner, Moldovanu, and Strack (2020).

The rest of 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 the 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. In Section 7, we briefly address the issue of dynamics, or deterrence over time. Finally, Section 8 concludes with a discussion of the practicability of our approach.

2 Model

Consider a city with $N \geq 1$ different locations. Illegal parking is a problem in all of these locations. The city determines the amount of resources devoted to enforcement in each location out of the total amount of available resources, denoted r. The amount of available resources is uncertain.[14] It is given by r_k, $k \in \{1, \ldots, K\}$, with probability π_k, respectively, where $0 \leq r_1 < \cdots < r_K$ and $\sum_{k=1}^{K} \pi_k = 1$. The expected amount of resources is denoted by $E[r]$. 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 7 below.

We refer to k as the state of the world. The city knows the realization of the state

[13] See footnote 8.

[14] The state of the world may be uncertain because the availability of enforcement resources in any given moment may be subject to random noise. Moreover, as we show below, such noise is necessary for persuasion to be effective and so may be purposefully introduced by the principal to improve deterrence.

of the world k and hence also the realization r_k, but drivers only know the distribution $\pi = (\pi_1, \ldots, \pi_K)$.

As explained above, we assume that the city can commit to a policy of disseminating 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. 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

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

and the amount of expected resources available conditional on message m is denoted $r(m) \equiv \sum_{k=1}^{K} r_k \Pr(k|m)$.

Denote the amount of resources allocated to enforcement in neighborhood i in state k when the city sends the message m by $a_k^i(m)$.[15] 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 $k \in \{1, \ldots, K\}$,

$$\sum_{i=1}^{N} a_k^i(m) \leq r_k \tag{2}$$

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

The objective of the city is to allocate the amounts of enforcement resources $\{a_k^i(m)\}$ 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 location 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 location given message m.

For simplicity, we focus on the special case where the measure of illegal parking in each location is given by a threshold function. Namely, for each location $i \in \{1, \ldots, N\}$ there exists some threshold $\tau^i > 0$ such that

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

In itself, the assumption that an individual driver is deterred from illegal parking if the probability of sanction is above a certain threshold involves no loss of generality if we assume that the payoffs from parking legally, parking illegally and being sanctioned, and

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

parking illegally without being sanctioned are themselves constant for each individual in each location. It follows that a continuum of drivers whose thresholds are distributed according to some continuous cumulative distribution function F would induce a continuous function $q_i = 1 - F$. If the distribution of drivers' thresholds F is bell shaped, then the cumulative distribution of drivers' thresholds would be S-shaped (see the concluding section for further elaboration of this claim). The assumption that the function q^i is a threshold function is an extreme version of this case, when all the drivers in location i happen to employ the same threshold rule.[16]

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

$$\min_{\{a_k^i(m)\},\{p_k(m)\}} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{i=1}^{N} q^i(a^i(m)) s^i p_k(m) \pi_k \tag{3}$$

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

Importantly, we assume that the city can commit to its strategy. That is, it determines the allocation and probabilities $\{a_k^i(m)\}, \{p_k(m)\}$. Then, it observes the state of the world k and draws a message m that is transmitted to drivers using the probabilities $\{p_k(\cdot)\}$. There can be no effective persuasion as described here without an ability to commit. As mentioned above, we believe that in the context of the problem studied here, of a central authority that seeks to deter socially unwanted behavior, the ability to commit is a reasonable assumption. This is because it is reasonable to expect that the central authority would be closely monitored by the media, who would alert the public in case the central authority deviates from its strategy. The short term benefit from deviation is surely smaller than the long term benefit from maintaining deterrence, so a patient central authority has an interest to maintain its ability to commit.

Drivers in different locations care only about what messages imply with respect to enforcement in their own location. The assumption of commitment implies that drivers do not care, nor do they have anything relevant to learn about the state of the world from the level of enforcement in other locations. Of course, if the city cannot commit, then it can send two different locations a message that it will enforce there even though it only has resources available for enforcement in just one location. However, in such a case, drivers would not

[16] The assumption that the q^i's are given by threshold rules greatly simplifies the discussion and description of the solution because it permits an easy identification of the set inflection points of the underlying objective function that is necessary for effective convexification. If the functions q_i are not threshold functions, then it is still possible to solve the problem along the same lines described here, but it would be more difficult to explicitly identify the inflection points necessary for effective convexification. The Optimal Ratio Rule would be a lot more cumbersome and the disutility function $D(r)$ that is described below would not be a step function without this assumption.

necessarily believe the city's message anyway.[17]

Observe that the constraints (1) and (2) are linear in probabilities $\{p_k(m)\}$ and resources $\{a_k^i(m)\}$, respectively, 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 $\{a_k^i(m)\}$ and send messages with probabilities $\{p_k(m)\}$ so as to maximize expected weighted deterrence as given by

$$\max_{\{a_k^i(m)\},\{p_k(m)\}} \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{i=1}^{N} d^i(a^i(m))s^i p_k(m)\,\pi_k \tag{4}$$

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 location i (which is equal to the decrease in social distutility). Again, the constraints (1) and (2) are linear in $\{p_k(m)\}$ and $\{a_k^i(m)\}$, respectively, 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 $m = \varnothing$ in the matrix above; 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 illustrates 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 locations 1 and 2 in

[17] The analysis of equilibrium behavior under limited commitment is outside the scope of this paper. Crawford and Sobel (1982) is the classical reference for this subject. See Lipnowski et al. (2019), and Eilat and Neeman (2020) for a recent discussion of these issues.

state 2 (when it sends message m_1), and sometimes deter in locations $1, 2, 3$ (when it sends message m_2). This is something that the city may benefit from if the amount of resources available in state 3 permits deterrence in locations $1, 2, 3$ ($r_3 > \tau_1 + \tau_2 + \tau_3$) but the amount available in states 1 and 2 only permits deterrence in locations 1 and 2.

3 The Optimal Ratio Rule

Suppose that the probabilities and allocations $p_k(m)$ and $\{a_k^i(m)\}$ satisfy the constraints (1) and (2). Denote the set of locations on which each message m achieves deterrence by $S(m) \subseteq \{1, \ldots, N\}$. It follows that the following deterrence constraint:

$$a^i(m) \equiv \sum_{k=1}^{K} a_k^i(m) \Pr(k \mid m) \geq \tau^i \tag{5}$$

is satisfied for each location $i \in S(m)$, and violated for locations $i \notin S(m)$, for every message $m \in M$ that is sent with a positive probability. It also follows that we may identify messages with the set of locations on which they achieve deterrence. Thus, no loss of generality is implied by the assumption that $M \equiv 2^{\{1,\ldots,N\}}$. 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.[18]

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. That is, the city cannot benefit from endogenously generated uncertainty about enforcement.

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 that is such that $S(m)$ maximizes the value of deterrence subject to the resource constraint is sent with probability one. Sending another message m' that induces the same or less deterrence subject to the resource constraint (on the set $S(m')$) would be either unnecessary or strictly worse. ■

The next result shows that no loss of optimality is implied by restricting attention to a specific class of allocations that spread resources across different locations in a way that is proportional to their deterrence thresholds.

Proposition 2 (the "Optimal Ratio Rule"). *Given probabilities $\{p_k(m)\}$ and an allocation*

[18] If two messages m and m' deter on the same set of locations then they can be merged into one message $m \cup m'$.

$\{a_k^i(m)\}$ that satisfy the probability and resource constraints (1) and (2), the same probabilities together with the allocation $\{(a_k^i)^*(m)\}$ that is given by:

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

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

- *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 $\{a_k^i(m)\}$.

Proof. Fix probabilities $\{p_k(m)\}$ and an allocation $\{a_k^i(m)\}$ that satisfy the constraints (1) and (2). For every location $i \in S(m)$ that is deterred by message m,

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

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 the resource constraint (2)).

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 $i \in \{1, \ldots, N\} \setminus S(m)$, state k, and message m, also achieves deterrence of the set $S(m)$. ■

Intuitively, allocation according to the Optimal Ratio Rule minimizes unnecessary waste of enforcement resources. This is illustrated by the next example.

Example 1. Consider the case in which the city has three locations with the corresponding thresholds $\tau^1 = 2$, $\tau^2 = 3$ and $\tau^3 = 4$. There are three equally likely states, with the resources $r_1 = 1$, $r_2 = 10$ 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:

$\frac{1}{3}$	$1-p$	1			1
	p	$\frac{2}{9} \times 1$	$\frac{3}{9} \times 1$	$\frac{4}{9} \times 1$	
$\frac{1}{3}$		$\frac{2}{9} \times 10$	$\frac{3}{9} \times 10$	$\frac{4}{9} \times 10$	10
$\frac{1}{3}$		$\frac{2}{9} \times 14$	$\frac{3}{9} \times 14$	$\frac{4}{9} \times 14$	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 city's problem: minimize expected social cost (3) subject to the probability and resource constraints (1) and (2), can be recast as a problem of choosing the probabilities $\{p_k(m)\}$ so as to minimize the expected social cost (6) below,

$$\min_{\{p_k(m)\}} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{i \in \{1,\ldots,N\} \setminus S(m)} s^i p_k(m) \pi_k \tag{6}$$

(for each message m that is sent with a positive probability, only locations $i \notin S(m)$ contribute to social cost and are therefore included in the sum) subject to the probability constraints (1) and the deterrence constraint (5) applied to $\left\{ (a_k^i)^* (m) \right\}$ as follows:

$$(a^i)^* (m) \equiv \sum_{k=1}^{K} (a_k^i)^* (m) \Pr(k \,|m) \geq \tau^i \tag{7}$$

for every message $m \in M$ that is sent with a positive probability, and for every location $i \in S(m)$.[19]

The objective function (6) is linear in the probabilities, 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: minimize expected social cost (3) subject to the probability and resource constraints (1) and (2), respectively, can be recast as the linear programming problem: minimize expected social cost (6) subject to the probability constraints (1) and the deterrence constraints:*

$$\sum_{k=1}^{K} p_k(m)\,\pi(k)(a_k^i)^*(m) \geq \tau^i \sum_{k=1}^{K} p_k(m)\,\pi(k) \tag{8}$$

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

Proof. The problem: minimize (6) subject to the probability and deterrence constraints (1) and (8) is a linear programming problem. The objective function (6) is obtained from (3) upon substitution of the resources according to the Optimal Ratio Rule. The deterrence constraints (8) 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 $m \in M$ that are not sent with a positive probability $p_k(m) = 0$, 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.

[19] It may be more natural to think of the problem as maximize expected weighted deterrence

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

subject to the same constraints.

4 "Splitting"

As explained in the introduction, the second approach to the problem consists of dividing it into two separate sub-problems. The solution of the second sub-problem uses ideas from the literature on Bayesian persuasion and in particular the idea that it may be possible to "split" a given expected amount of enforcement resources into two or more expected amounts, that would each be induced with a certain probability. In this section, we explain how this can be done.

Each message m induces a belief about the posterior expectation of resources, and so a message policy, in which different messages are sent with different probabilities in different states of the world, induces a distribution of posterior expectations of resources. The expected amount of resources $E[r]$ is thus "split" into different posterior expectations $r(m) \equiv \sum_{k=1}^{K} p(k|m) r_k$, which are each realized with the probability $\Pr(m) \equiv \sum_{k=1}^{K} p_k(m)\pi_k$ with which message m is sent, such that

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

The objective of Bayesian persuasion is to pick the optimal "split," or distribution of messages, from of the set of distributions of posterior expectations of resources that preserve the mean of the distribution of resources and are second-order-stochastically-dominated by the prior distribution as mentioned in the introduction.

In this section we provide two useful results about splitting. The next lemma is a generalization of the famous lemma of Aumann and Maschler (1995). Denote the posterior total expected amount of resources conditional on two messages, m and m' by

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

Lemma 1.[20] *Any two messages L and H that are sent with probabilities $\Pr(L)$ and $\Pr(H)$ and that induce posterior expectations $r(L) < r(H)$, can be replaced with two messages L' and H' that induce any two posterior expectations $r(L) \leq r(L') \leq r(H') \leq r(H)$ 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) the posterior expectation conditional on the two messages is preserved, or

$$r(L, H) = r(L', H'),$$

[20] Letting $k = 2$ and assuming that the city sends only two messages L and H that fully reveal the state of the world (such that $r(L) = r_1$ and $r(H) = r_2$) reproduces Aumann and Maschler's Lemma.

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

The proof of Lemma 1 is similar to the proof given by Aumann and Maschler (1995) and is omitted. The next lemma provides another useful observation, which is also a generalization of the same lemma of Aumann and Maschler. This lemma characterizes the maximal distance that can be achieved between any two induced beliefs about the total expected amount of resources. This maximal distance imposes a constraint on the maximal degree of convexification that can be achieved in our problem as explained in the next two sections.

Lemma 2. *Given a distribution of resources $r_1, \ldots, r_K$, and given any two total expected amounts of resources $r_L < E[r] < r_H$, 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 requirement of second-order-stochastic-dominance imposes on the maximum difference between the high and low posterior expectations $r(H)$ and $r(L)$ when the city sends two messages H and L.

Example 2. Suppose there are three states of the world. Resources are given by $(r_1, r_2, r_3) = (0, \frac{1}{2}, 1)$ and the prior is $(\pi_1, \pi_2, \pi_3) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. The expected amount of resources is $E[r] = \frac{1}{2}$. Suppose the city sends two messages, L and H. As mentioned in the introduction, because sending messages about the state of the world induces a garbling of the drivers' beliefs relative to the true state, it follows from Blackwell (1953) that the distribution of posterior expectations of resources r_H and r_L is second order stochastically dominated by the prior distribution.

In turn, this implies that in this example, the lowest possible posterior expectation r_L depends on the highest induced posterior expectation r_H and is given by

$$r_L = \max\left\{\frac{3r_H - 2}{8r_H - 5}, 0\right\}.$$

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

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 enforcement 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 locations" as defined by the disutilities $\{s^i\}$, drivers 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 monotonicity implies that if it is optimal to deter at neighborhood i under some message m, then it is also optimal to deter at location $j < i$. It follows that the number of messages that is needed is at most $N + 1$. Namely, in the optimal solution, it is enough to restrict attention only to those messages associated with the sets $\varnothing$, $\{1\}, \{1,2\}, \ldots, \{1,\ldots,N\}$.

Monotonicity simplifies the city's allocation problem. If the total expected amount of resources is less than τ^1 then no deterrence is possible. The social cost associated with that is $\sum_{i=1}^N s^i$ because no drivers are deterred from illegal parking. 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 neighborhood 1. The social cost that is associated with that is $\sum_{i=2}^N s^i$ because drivers in neighborhood 1 are deterred from illegal parking, and so on. Continuing in the same way we see that devoting all the available resources to deterrence with no communication produces the following non-increasing step-function social cost:

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

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

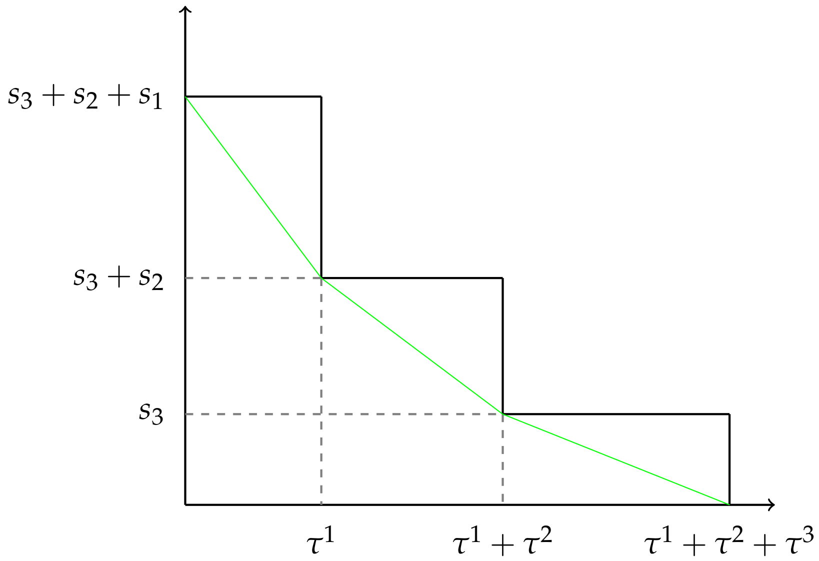

Figure 1: $D(r)$ in the monotone case

Persuasion, or the sending of messages, allows the city to achieve a lower expected social cost than $D(r)$. The value of the city's objective function when it sends messages $1, \ldots, M$ with probabilities $\Pr(1), \ldots, \Pr(M)$ and with induced mean posterior beliefs $r(1), \ldots, r(M)$, respectively, is:

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

The monotone case admits a complete solution of the city's problem of minimize expected social cost (6) subject to the probability and deterrence constraints (1) and (8), respectively, 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* $\sum_{i=1}^{n-1} \tau^i \le E[r] < \sum_{i=1}^{n} \tau^i$ *for some* $2 \le n < N$.[21] *Then, the optimal solution involves the sending of only two messages* L *and* H *such that the posterior expectation* $r(H)$ *is set equal to* $\sum_{i=1}^{n} \tau^i$ *if this is possible given the distribution of resources, and the posterior*

[21] If either $E[r] < \tau^1$ or $\sum_{i=1}^{N} \tau^i \le E[r]$ 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.

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

$$\frac{\Pr(L)D(r(L))}{\Pr(L)+\Pr(H)}+\frac{\Pr(H)D(r(H))}{\Pr(L)+\Pr(H)} \leq \frac{\Pr(L')D(r(L'))}{\Pr(L')+\Pr(H')}+\frac{\Pr(H')D(r(H'))}{\Pr(L')+\Pr(H')}.$$

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

$$\sum_{m\neq L,H}\Pr(m)D(r(m))+\Pr(L)D(r(L))+\Pr(H)D(r(H))$$

to

$$\sum_{m\neq L,H}\Pr(m)D(r(m))+\Pr(L')D(r(L'))+\Pr(H')D(r(H')).$$

If a policy includes two messages L and H that induce posterior expectations $\sum_{i=1}^{n-1}\tau^i \leq r(L)$ and $\sum_{i=1}^{n}\tau^i < r(H)$ then expected disutility can be lowered if the two messages L and H are replaced with messages L' and H' that are such that $r(L') = r(L)$ and $\sum_{i=1}^{n}\tau^i = r(H')$. The straight Line that connects the points $(r(L'), D(r(L')))$ and $(r(H'), D(r(H')))$ still lies strictly below the straight line that connects the points $(r(L), D(r(L)))$ and $(r(H), D(r(H)))$. 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 $r(H) = \sum_{i=1}^{n}\tau^i$ if this is possible given the distribution of resources and $r(L) \geq \sum_{i=1}^{n-1}\tau^i$. The step structure of the function $D(r)$ implies that if the distribution of resources does not allow to set $r(H) = \sum_{i=1}^{n}\tau^i$ then persuasion is unhelpful and no messages should be sent. It also implies that $r(L)$ 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 cost. When $r_H = \sum_{i=1}^{n}\tau^i$ and $r_L =$

$\sum_{i=1}^{n-1} \tau^i$ the expected social cost is obtained on the Green curve at the point $E[r]$. The expected social cost with mean posterior beliefs $r'_H > r_H$ and $r'_L > r_L$ is higher and is obtained on the Black curve at the point $E[r]$. Figure 2 also illustrates the reason that if it is impossible to set $r_H = \sum_{i=1}^{n} \tau^i$ then persuasion is ineffective.

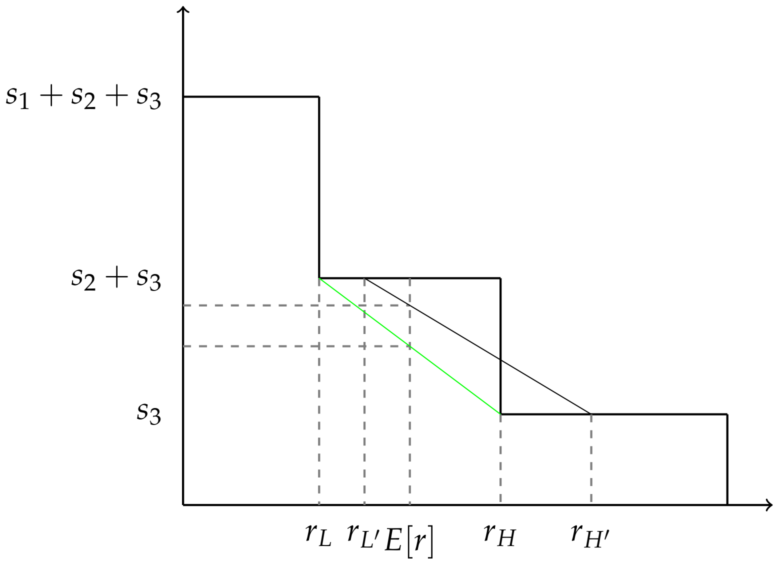

Figure 2: Optimal solution in the monotone case (the Green line generated by optimal messages L and H lies below the Black line generated by messages L' and H' at $E[r]$)

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 $r_L = \sum_{i=1}^{n-1} \tau^i$ when this is impossible given the distribution of resources. The answer to this question is, not surprisingly, negative.[22]

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

[22]Suppose then that $r(L)$ is optimally set at a continuity point of $D(r)$. 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 $(r(L), D(r(L))$ and $(r(H), D(r(H))$ is:

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

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

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

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

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 ($r_H = \sum_{i=0}^{m+1} \tau^i$ and $r_L = \sum_{i=0}^{m} \tau^i$) 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 $r_H = \sum_{i=0}^{m+1} \tau^i$ but $r_L > \sum_{i=0}^{m} \tau^i$, or $r_H < \sum_{i=0}^{m+1} \tau^i$ and 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 convexification problem, and characterize the number of messages needed for the optimal solution. However, we cannot provide an explicit solution of the problem as in the monotone case.

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

$$D_S(r) = \begin{cases} \sum_{i \in \{1,\ldots,N\}} s^i & \text{if } r < \sum_{i \in S} \tau^i \\ \sum_{i \in \{1,\ldots,N\} \setminus S} s^i & \text{if } \sum_{i \in S} \tau^i \leq r \end{cases}$$

that maps the amount of available expected resources r into social cost. If $r < \sum_{i \in S} \tau^i$ then there are not enough resources to deter on the set of locations S. In this case, drivers in all locations park illegally, and the social cost is $\sum_{i \in \{1,\ldots,N\}} s^i$. If on the other hand $r \geq \sum_{i \in S} \tau^i$ then there are enough resources to deter on the set S. In this case, drivers in S are deterred and only drivers outside S park illegally. The social cost is $\sum_{i \in \{1,\ldots,N\} \setminus S} s^i$.

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

$$D(r) = \min_{S \subseteq \{1,\ldots,N\}} D_S(r).$$

Identification of the set of locations S on which the minimum $D_S(r)$ is obtained for any amount of resources r is a knapsack problem. The knapsack problem is a NP hard problem for which there exists a fully polinomial time approximation scheme (FPTAS) (Ibarra and Kim, 1975).

In the monotone case, the steps defined by the social cost function $D(r)$ became longer and lower, but this is not necessarily the case generally. Define the convexification of $D(r)$ from below as

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

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

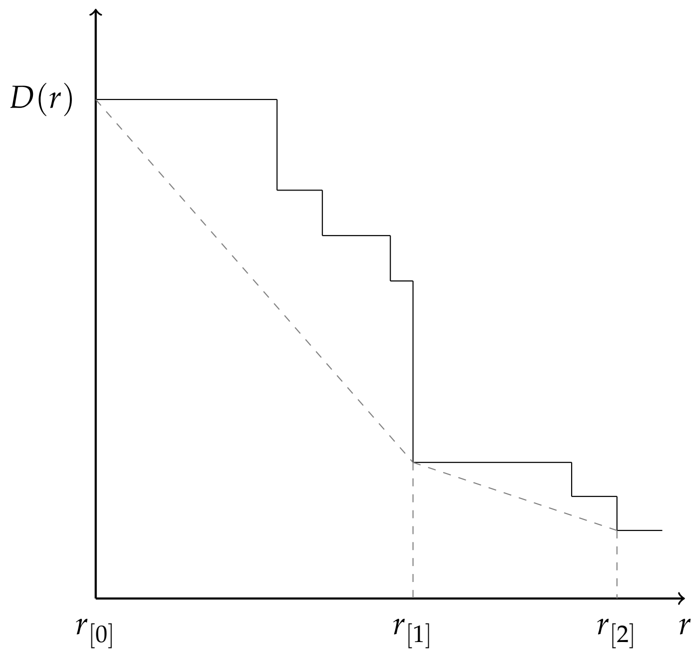

Figure 3: The functions $D(r)$ and conv $D(r)$ in the general case

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

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

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

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

and would have required only two messages. As shown in Figure 4 below, the optimal solution would have involved sending only messages L and H with induced posterior beliefs r_L and r_H that are equal to the consecutive two coincidence points that are such that $r_{[l]} < E[r] < r_{[l+1]}$,[23] with probabilities $\Pr(H)$ and $\Pr(L) = 1 - \Pr(H)$ that are such that $\Pr(L) \cdot r_L + \Pr(H) \cdot r_H = E[r]$.

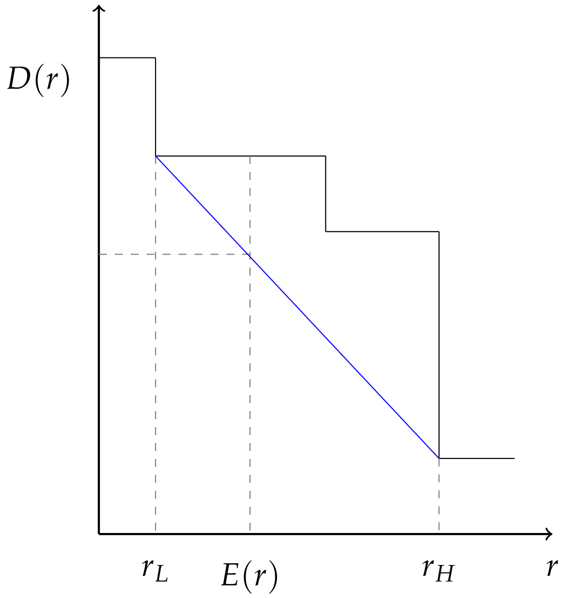

Figure 4: Optimal solution in the unconstrained case involves only two messages

However, as explained in the introduction, the induced distribution of mean posterior belief $\{r(m)\}$ must be second-order-stochastically-dominated by the prior distribution. These implies that the problem is given by the following constrained convexification problem

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

subject to the constraint that the induced distribution of mean posterior belief $\{r(m)\}$ is second-order-stochastically-dominated by the prior distribution.

This additional constraint implies that sometimes three or more messages may generate a lower value of the objective function than just two messages. This is illustrated in the next example.

Example 3. A city has two locations 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 5 below, optimal deterrence with two messages L and H (such that $r(L) < r(H)$) requires that $r(H) = 1$ and $r(L)$ is set as low as possible, which in this case implies $r(L) = \frac{1}{3}$,

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

$\Pr(L) = \frac{3}{4}$ 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 $\frac{1}{4} \cdot \frac{5}{4} + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot \frac{1}{4} = \frac{7}{8} < 1$.

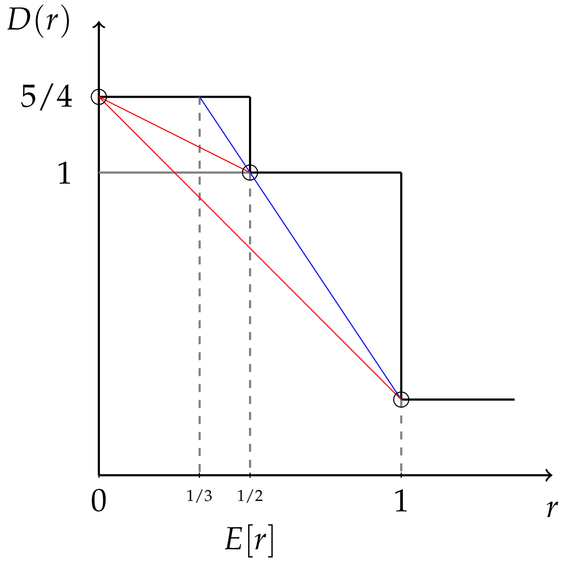

Figure 5: Three messages are better than two

The next proposition 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 $[r_{[l]}, r_{[l+1]}]$. Then, the number of messages needed in order to obtain the optimal solution is no more than the number of steps of the function $D(r)$ in the segment $[r_{[l]}, r_{[l+1]}]$ plus one. If the expected amount of resources coincides with one of the points $r_{[0]}, r_{[1]}, \ldots, r_{[I]}$, 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 $[r_{[l]}, r_{[l+1]}]$. An identical argument to the one used in the proof of Proposition 3 shows that no loss of generality is implied by restricting attention to a set of messages that induce posterior expectations that lie in the interval $[r_{[l]}, r_{[l+1]}]$. This is because any two messages L and H that induce posterior expectations $r(L) < r_{[l]} < r_{[l+1]} < r(H)$, can be replaced by two messages L' and H' that are such that $r(L') = r_{[l]}$ and $r(H') = r_{[l+1]}$ 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 $r_{[l]} \leq r(L)$ and $r_{[l+1]} < r(H)$ can be replaced by two messages L' and H' that are such that $r(L') = r(L)$ and $r(H') = r_{[l+1]}$ 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 $r(L) < a_{[l]}$ and $r(H) \leq r_{[l+1]}$ can be replaced by two messages L' and H' that are such that

$r(L') = r_{[l]}$ and $r(H') = r(H)$ 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 $r(m_i) = r(m_j)$ without affecting any other probabilities or posterior expectations.

Finally, if the expected amount of resources coincides with one of the points $r_{[0]}, r_{[1]}, \ldots, r_{[I]}$, 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 $D(r)$. ■

In recent papers, Doval and Skreta (2018) and Le Treust and Tomala (2019) show that the number of messages needed in order to attain the optimum in a Bayesian persuasion problem is smaller than or equal than the number of states.[24] It follows that the bound on the number of messages identified in Proposition 5 can be tightened and set at the minimum of the number of steps of the function D in the relevant interval and the number of states K. Moreover, the next example attains this bound and thus proves it to be tight.

Example 4. Suppose that a city has N locations. The thresholds for deterrence are given by $\tau^1 < \cdots < \tau^N$. The number of states is $K = N$. Suppose that the distribution of resources is given by $r_0 = 0$, $r_1 = \tau^1$, $r_2 = \tau^1 + \tau^2$, $r_3 = \tau^1 + \tau^2 + \tau^3$, and so on. Suppose that the thresholds for deterrence τ^n are increasing in n sufficiently fast so that in state k it is only possible to deter in neighborhhods $1, \ldots, k$. Suppose that the social disutility of illegal parking is increasing in n so that the function D has steps that become taller and longer, as in Figure 6 below.

[24] Their results are established for constrained Bayesian persuasion problems. Here, no additional constraints beyond Bayes plausibility are imposed on the problem so the number of messages needed is smaller than or equal to the number of states.

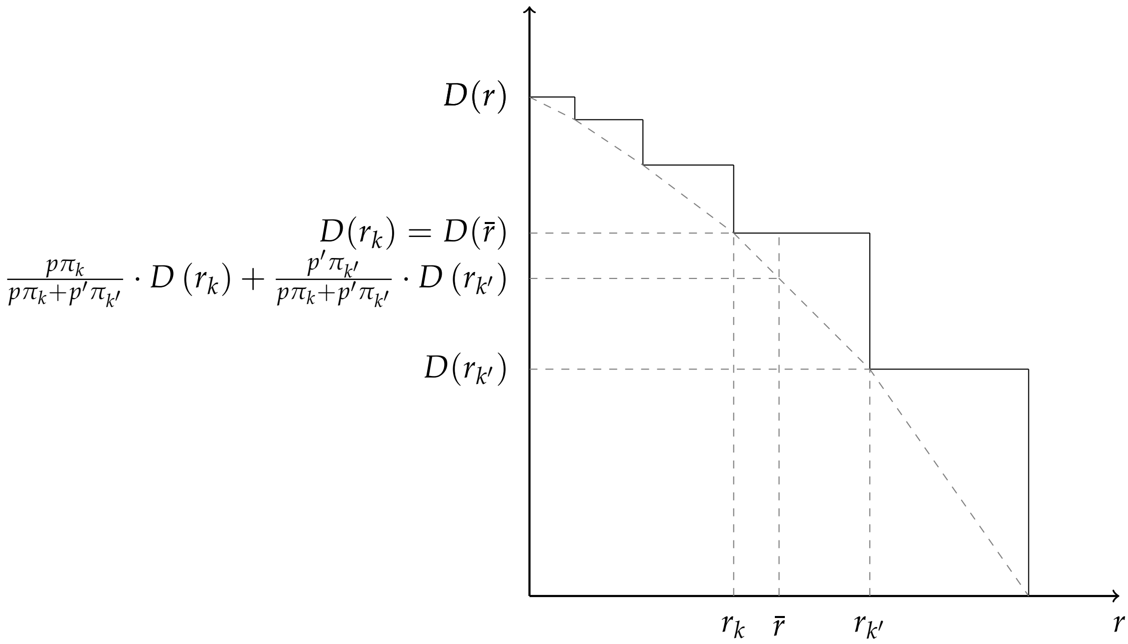

Figure 6: Full revelation of the state

The optimal policy in this example consists of full revelation of the state and in each state k, allocation of resources to locations $1, \ldots, k$. To see that this is optimal note that (1) the allocation rule is optimal because it makes the most efficient use of the available resources and (2) if two different states are pooled together into one message, then it is possible to improve the value of the objective function by simply revealing the state instead. Suppose that message m is sent with probabilities p and p' in states k and k', respectively. This implies that message m generates a payoff of $D(\bar{r})$ where $\bar{r} = \frac{p\pi_k}{p\pi_k+p'\pi_{k'}} \cdot r_k + \frac{p'\pi_{k'}}{p\pi_k+p'\pi_{k'}} \cdot r_{k'}$. If instead the city reveals the state, the expected payoff is $\frac{p\pi_k}{p\pi_k+p'\pi_{k'}} \cdot D(r_k) + \frac{p'\pi_{k'}}{p\pi_k+p'\pi_{k'}} \cdot D(r_{k'})$ which is lower than $D\left(\frac{p\pi_k}{p\pi_k+p'\pi_{k'}} \cdot r_k + \frac{p'\pi_{k'}}{p\pi_k+p'\pi_{k'}} \cdot r_{k'}\right)$ because of the structure of the function D, as can be seen from Figure 6.[25]

It therefore follows that in this example the number of messages in the optimal solution is $N = K$. ■

Finally, we mention that as in the monotone case, the convexification of the function $D(r)$ may be partial in the sense that the optimal solution may lie strictly above the function conv $D(r)$.

[25] Kolotilin (2018) has a similar result. He shows that a policy of full revelation of the state is optimal if and only if for any two states k and k' with resources $r_k < r_{k'}$, respectively, revelation of the state is preferrable to pooling these messages together.

7 Endogenous Distribution of Resources & Deterrence over Time

It is possible to endogenize the prior distribution of the amount of available resources in the following way. Suppose that the city employs K inspectors. Each inspector is allocated to a specific day and time, or to several time slots, depending on how many hours he is required to work per day or week. Each inspector shows up to each assigned 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.

It is also interesting to explore the allocation of enforcement resources over time. Cyclical allocations, where the same distributions are repeated on a daily, weekly or monthly basis can be addressed along the lines described above. Another possibility is where the state of the world evolves according to a Markov process. Ashkenazi-Golan (2021) et al. solve this problem for the case with two states. Interestingly, they show that the optimal strategy is not myopic.

8 Conclusion

Because Bayesian persuasion lowers social cost through convexification of the social cost function $D(r)$, its usefulness obviously depends on whether $D(r)$ is not already convex. The simplifying assumptions we imposed on drivers' behavior imply that $D(r)$ is a step function, and so necessarily nonconvex. However, in practice $D(r)$ may well be a smooth function, and so the question is whether the $D(r)$ function that is likely to arise in practice is nonconvex. This is an empirical question.

To simplify, suppose there is only one neighborhood. Because parking illegally induces a binary bet (an individual who parks illegaly is either sanctioned or not), the drivers' expected utility from such a bet is independent of their risk attitudes. Namely, any two drivers who derive the same utility from parking illegally and being sanctioned and not, derive the same expected utility from the bet, regardless of the curvature of their utility function.[26] Because expected utility functions are invariant to affine transformation, no loss of generality is implied by assuming that drivers' utility from parking illegally and being sanctioned or not are one and zero, respectively.

Thus, whether or not a driver chooses to park illegally or not depends on the utility that the driver derives from parking legally. Each such utility from parking legally induces

[26]Hence, there is no need to invoke the fact that many economists (cf. Rabin, 2000) argue that given people's risk attitudes over large bets, they should be nearly risk neutral with respect to smaller bets.

a threshold probability of sanction, above which the driver would be deterred from parking illegally. It is reasonable to suppose that the distribution of drivers' utilities from parking legally is bell-shaped, which in turn, implies a bell-shaped distribution of the threshold probabilities of sanction above which drivers are deterred from illegall parking. Denote the cumulative distribution of these thresholds by $F(r)$. For any level of sanction or available enforcement resources r, the proportion of drivers who are deterred by the sanction is given by $F(r)$. Social cost, $D(r)$, is proportional to the fraction of drivers who are not deterred by the sanction, or $1 - F(r)$. If the distribution of thresholds is bell-shaped and has an expectation that is strictly between zero and one, then the cumulative distribution $F(r)$ is S-shaped. This implies that $D(r)$ is nonconvex.

Holt and Laury (2002) famously described an experiment where individuals were required to choose between a relatively safe and a relatively risky binary bet, as the probability of failure varies from zero to one. If we interpret the probability of failure as the probability of a sanction, it is possible to interpret the choice of the safe and risky bets as legal and illegal parking, respectively. Holt and Laury (2002) found that the proportion of individuals who chose the safe bet, $F(r)$, is indeed S-shaped. This finding is echoed by Dohmen et al. (2011) who report on a large survey of individuals' risk attitudes with respect to large risks. They find that the distribution of "tolerance for risk" in the population is bell shaped with a small mass of people with extremely low tolerance for risk. This, again, produces an S-shaped cumulative distribution $F(r)$.

The significance of an S-shaped cumulative distribution function $F(r)$ is that it indicates that a "critical level" of risk is needed to achieve a large effect, and that smaller levels produce significantly less deterrence. This type of nonconvexity is necessary to make Bayesian persuasion useful.[27]

Our approach is also related to research in criminology about the effectiveness of what is known as "hot spots policing." It is observed that urban crime is mostly concentrated in a relatively small number of "hot spots" locations, and so vigorous police enforcement in these hot spot is expected to increase overall crime prevention (Braga and Bond, 2008; Braga et al. 2014; Blattman et al. 2019; Mohler et al. 2015). Because police presence attracts a lot of attention, sending a foot patrol or a police cruiser to a hot spot area is akin to revelation of the state of the world in the context of our model. This may indeed be optimal in some cases (cf. example 4) but is not expected to be optimal in general (cf. the example in the introduction). As demonstrated in this paper, persuasion may be more effective if it is possible to "spread enforcement resources over an event that has a higher probability." However, in the context of urban crime, it is not clear how it is possible to indicate that police is nearby without actually sending a patrol, which effectively amounts to a complete revelation of the state.

This suggests that Bayesian persuasion would be more suitable for deterrence of un-

[27]In a paper about tax evasion, Kleven et al. (2011) show that threat of audit letters have a large effect on behavior. They contrast their findings with those of Alm et al. (1992) who found that when penalties and audit probabilities are set at realistic but low levels, their deterrent effect is quite small. This is consistent with the idea that a critical level of enforcement is needed to achieve deterrence.

wanted activity that is less likely to be displaced by display of enforcement to other locations (such as voilent crime compared to opportunistic crime, see Braga et al., 2014) and in those cases where it is not clear whether enforcement is on or off at any given moment, so that it is possible to use communication to shape beliefs, as in the case of illegal parking, speeding to some extent, tax evasion, and free-riding.

Of course, people are probably less than fully Bayesian rational, and certainly, probably not as Bayesian rational as assumed in this paper. However, people in practice definitely respond to messages, even if they don't understand exactly what they mean in terms of implied levels of expected enforcement. A local government who wants to exploit the power of using messages to help regulate behavior would probably not do badly by ensuring that the messages it uses are Bayesian optimal as described in this paper. The use of any other messages risks squandering the government's credibility or not maximizing the potential for deterrence.

References

[1] Alm, J., B. R. Jackson, and M. McKee (1992) "Estimating the determinants of taxpayer compliance with experimental data," *National Tax Journal* 45, 107-114.

[2] Alonso, R. and O. Câmara (2016) "Persuading voters," *American Economic Review* 106, 3590-3605.

[3] Ashkenazi-Golan, G., P. Hernandez, Z. Neeman, and E. Solan (2021) "Dynamic Bayesian Persuasion with Two States", working paper.

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

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

[6] Bergemann, D. and S. Morris (2019) "Information Design: A Unified Perspective," Journal of Economic Literature, forthcoming.

[7] Best, J. and Quigley D. (2020) "Persuasion for the Long Run," unpublished manuscript, Oxford University.

[8] Borel, E. (1921) "La Théorie du Jeu et les Équations Intégrales á Noyau Symétrique," *Comptes Rendus de l'Académie 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.

[9] Braga, A. and B. Bond (2008) "Policing crime and disorder hot spots: a randomized controlled trial," *American Society of Criminology* 46, 577-607.

[10] Braga, A., Papachristos, A., and D. Hureau (2014) "The effects of hot spots policing on crime: an updated systematic review and meta-analysis," *Justice Quarterly* 31, 633-663.

[11] Blattman, C., D. Green, D. Ortega, and S. Tobón (2019) "Place-based interventions at scale: the direct and spillover effects of policing and city services on crime," NBER Working Paper 23941.

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

[13] Crawford, V. and Sobel, J. (1982), "Strategic Information Transmission," *Econometrica* 50(6), 1431-1451.

[14] Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011) "Individual risk attitudes: measurement, determinants, and behavioral consequences," *Journal of the European Economic Association* 9, 522-550.

[15] Doval, L. and V. Skreta (2018) "Constrained information design: Toolkit," unpublished working paper.

[16] Dworczak, P. and G. Martini (2019) "The Simple Economics of Optimal Persuasion," *Journal of Political Economy* 127:5, 1993-2048.

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

[18] Eilat, R. and Z. Neeman (2021) "Communication with Endogenous Deception Costs," working paper.

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

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

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

[22] Holt, C. and S. K. Laury (2002) "Risk aversion and incentive effects," *American Economic Review* 92, 1644-1655.

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

[24] Kelling, G., A. M. Pate, D. Dieckman, and C. Brown (1974) "The Kansas City preventive patrol experiment: Technical report. Washington DC: Police Foundation.

[25] Kleiner, A., B. Moldovanu, and P. Strack (2020) "Extreme Points and Majorization: Economic Applications," unpublished working paper, University of Bonn.

[26] Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011) "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica* 79, 651-692.

[27] Kolotilin, A., T. Mylovanov, M. Li, and A. Zapechelnyuk (2017) "Persuasion of a privately informed receiver," *Econometrica* 85, 1949-1964.

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

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

[30] Le Treust, M. and T. Tomala (2019) "Persuasion with limited communication capacity," *Journal of Economic Theory* 184, 104940.

[31] Lipnowski, E., D. Ravid, and D. Shishkin (2019) "Persuasion via Weak Institutions," working paper.

[32] Mohler, G. O., M. B. Short, S. Malinowski, M. Jhonson, G. E. Tita, A. Bertozzi, and P. J. Brantingham (2015) "Randomized controlled field trials of predictive policing," *Journal of the American Statistical Association* 110, 1399-1411.

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

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

[35] Rabin, M. (2000) "Risk aversion and expected utility theory: A calibration theorem," *Econometrica* 68, 1281-1292.

[36] Renault, J., E. Solan, and N. Vieille (2016) "Optimal Dynamic Information Provision," *Games and Economic Behavior* 104, 329-349.

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

[38] Sherman L. W. and D. Weisbrud (1995) "General deterrent effect of police patrol in crime "hot spots": A randomized controlled trial," *Justice Quarterly* 12:4, 625-648.

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

[40] Weisburd, D. and C. Telep (2014) "Hot spots policing," *Journal of Contemporary Criminal Justice* 30, 200-220.