



BUSINESS SCHOOL
Te Kura Pakihi

ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1406

May 2014

Quantity restrictions with imperfect enforcement in an over-used commons: Permissive regulation to reduce over-use?

Jeong-Yoo Kim and Nathan Berg

Address for correspondence:

Nathan Berg
Department of Economics
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Email: nathan.berg@otago.ac.nz
Telephone: 64 3 479 8648

Quantity restrictions with imperfect enforcement in an over-used commons: Permissive regulation to reduce over-use?

Jeong-Yoo Kim*and Nathan Berg[†]

May 21, 2014

This paper presents a model of quantity regulation aimed at mitigating externalities from over-use of a commons: for example, restrictions on use of automobiles, fisheries, computer networks and electronic stock quotation systems with high-frequency traders. The model provides a counter-intuitive answer to the question of what happens when quantity restrictions are legislated but enforcement is imperfect. If the probability of enforcement depends on both violation rates and enforcement expenditures, then equilibrium congestion can become worse as the quantity restriction becomes more severe. Stricter regulation causes more agents to violate the regulation which consequently reduces the probability of detection. Aggregate payoffs respond nonmonotonically to stricter regulatory rules. We find an interior near-optimal solution which is neither too permissive nor too strict. We show, however, that this near-optimal quantity regulation falls short of achieving socially optimal levels of use. Moreover, socially optimal levels of use can never be achieved in the sense that there exist some agents who rationally choose to violate the regulation if the regulator sets the restricted activity level at the socially optimal level. We also discuss optimal enforcement.

Key Words: congestion, emissions cap, regulation standard, tragedy of the commons

JEL Classification Code: K42

*Corresponding author: This research was begun when the first author was visiting ISER, Osaka University in the spring of 2008. Mailing address: Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaemunku, Seoul 130-701, Korea, Tel: +822-961-0986, Fax: +822-966-7426, Email: jyookim@khu.ac.kr

[†]Nathan Berg: Associate Professor of Economics, Department of Economics, University of Otago, P.O. Box 56, Dunedin, New Zealand, Email: nathan.berg@otago.ac.nz

1 Introduction

Some activities generate such severe negative externalities that they are prohibited in any positive quantity. Other activities, however, such as using one's car, are normally regarded as socially acceptable even though they generate negative externalities that motivate governments to attempt various forms of regulation. In such cases, the entity responsible for designing and enforcing regulation sometimes chooses to enact rules that restrict quantities of the externality-generating activity.

This paper focuses on quantity regulation in the context of congestion problems that are, in essence, a tragedy of the commons resulting from congestion or other negative over-use externalities. Real-world examples of regulations aimed at mitigating over-use of a commons include restrictions on automobile use in Bogota, Sao Paulo, Mexico City, Athens, Seoul, Beijing, and Tianjin; legislated limits on carbon emissions in countries such as Australia and New Zealand; catch limits in the seafood industry; and private owners of computer bandwidth that restrict quantities of information flow (e.g., proposed restrictions on stock trades and/or network bandwidth by high-frequency traders).¹

Congestion pricing can, in theory, be implemented by using Pigouvian taxes as an alternative to quantity restrictions.² There are substantial challenges, however, facing regulatory entities seeking to implement Pigouvian taxes in real-world regulatory environments, which could explain why both public and private regulators frequently opt for quantity restrictions instead. Congestion pricing schemes have the effect of increasing costs on all users, which may limit political support for their implementation. Another challenge for regulators at-

¹Although over-use of a commons is not usually the motivation for restricting illegal drugs, laws on marijuana possession in some places in the U.S. that allow possession up to a positive quantity threshold provide yet another example of a relatively permissive quantity regulation that the model in this paper shows may counter-intuitively function well at moderating aggregate levels of use. Other relevant examples might include the permissiveness of bank regulators' approach to enforcing asset requirements (i.e., lenient accounting rules that allow for marking illiquid bonds at par) in the U.S. and Europe (under Basel III), as well as the Securities and Exchange Commission's permissiveness regarding investment firms' use of high-frequency trading algorithms that impose externalities on other users of bandwidth in electronic trading networks.

²In the dilemma analyzed by Weitzman (1974) between price regulation (e.g., Pigouvian taxes) versus quantity limits dictated by the regulator, the slopes of the marginal benefit and marginal cost curves are identified as the main determinants that decide which of these two regulatory approaches maximize expected social welfare—under uncertainty and information asymmetries that block the feasibility of first-best regulation (and, consequently, equivalence of price and quantity regulation) in a perfect-information world. Weitzman's work gave rise to a vast literature dealing with how best to regulate externalities in the presence of uncertainty (Roberts and Spence, 1976; Yohe, 1978; and many others).

tempting to implement Pigouvian taxes is the information required regarding social marginal benefit and cost curves. Uncertainty about the position and shapes of these curves leads to uncertainty about the likely responses to different pricing schemes by agents who generate the externality.³

In contrast, quantity regulation—which directly restricts activity levels—requires far less information. All that is required is the regulator’s desired level of activity (i.e., without requiring knowledge about the shapes of the benefit and cost curves that are required for congestion pricing or other approaches based on Pigouvian taxes). In cases where policy makers want to avoid exceeding a catastrophic threshold level of over-use of a commons, the attractiveness of outright quantity restrictions over the uncertain responses to taxation may be especially compelling. Regardless of the validity of policy makers’ views of the political and informational challenges to implementing Pigouvian tax schemes, the prevalence of real-world regulations that restrict quantities serves as the primary motivation for studying them.⁴

This paper analyzes the optimal quantity regulation under the assumption of a benevolent regulator who attempts to maximize aggregate payoffs associated with use of a commons while accounting for the negative externalities it generates. Over-use and congestion in the examples above are instances of the strategic interaction referred to by Hardin (1968) as the “Tragedy of the Commons.” The mechanism generating negative externalities in Hardin’s classic work serves as a frequently cited justification for regulators attempting to reduce aggregate quantities of use of the commons. The effectiveness of regulation that takes the form of an announced threshold that restricts the allowable range of an aggregate quantity

³*The Economist* (Feb. 24, 2011) observes that quantity regulation (i.e., command-and-control regulation) is preferred over Pigouvian taxes or tradable permits (e.g., cap-and-trade regulation), because: “people generally prefer rules telling them something is not allowed (quantity regulation), rather than charges making them pay for it, even if the latter are clearly more efficient at maximizing social value.” That article suggests that aversion to taxes is among the primary reasons for the preference.

⁴Quantity regulation schemes in the real world include both cap-and-trade (CAT) and command-and-control (CAC) approaches that legislate maximum or minimum quantities of particular activities by private agents. It is known that CAT provides economic incentives for private actors to engage in mitigation and, with some auxiliary conditions, can, in theory, efficiently achieve emissions targets. Insofar as there are a relatively small number of polluting firms, the information requirements for CAT schemes to achieve efficiency appears more plausible to some observers. The feasibility of such information-intensive price regulation would seem, however, relatively more limited in the case of the traffic congestion. There is a large number of potential automobile users, who typically know little about each other’s valuations and do not make frequent decisions about investing in congestion-reducing technology, although some authors such as Verhoef et al. (1997) and Kornhauser and Fehlig (2002) have made interesting proposals about how price regulation could be implemented in this setting.

of economic activity is not easy to anticipate, however, because enforcement is typically imperfect (i.e., the chance that violators are caught is strictly less than 1).

The premise for this paper's model of quantity regulation is the real-world observation that enforcement of rules restricting quantities of economic activity tends to be imperfect: a proportion less than 100 percent of those who violate the regulation is caught and required to face penalties. For any fixed expenditure on enforcement by a regulator, the probability that violators face penalties is, in many cases, a decreasing function of the number, or proportion, of violators in the population, referred to in this literature as the *crowding effect*. Intuitively, when everyone violates the standard, the enforcement expenditure is stretched thinly and the chance of getting caught decreases. Therefore, our model describes equilibrium properties of quantity restrictions in an environment where the probability that violators get caught and actually pay the fine depends on the regulator's enforcement expenditures and the proportion of other agents who violate the regulatory standard.⁵ Counter-intuitively, the model shows that more permissive quantity restrictions generate smaller violation rates and, consequently, can reduce costly over-use or congestion.

A similar result was obtained by Viscusi and Zeckhauser (1979) who studied a quantity restriction on pollution assuming incomplete enforcement. Similar to our paper, they showed that tightening of the regulatory standard may lower overall performance, but the mechanism driving their result is different. In our model, a more stringent quantity standard leads to more violators, and this lowers the probability of detection because of the crowding effect. This reduction in the probability of detection, in turn, incentivizes some agents, who optimally choose not to use the commons under the looser standard, to use the commons. In contrast, the reason why there are some firms in Viscusi and Zeckhauser's (1979) model that would have chosen to invest to meet a looser standard but instead stop investing in pollution abatement under the tighter standard is because the investment becomes more costly under the tighter standard. The mechanism favoring more permissive quantity regulation in their model is independent of the detection probability. Costs of compliance and non-compliance represent the distinct mechanisms in their model and ours. In their model, compliance costs are higher when the quantity standard is strict, whereas in our model the *non-compliance* cost is smaller when the quantity standard is strict (without any change in the cost of compliance).

Harford (1978), Malik (1990a) and Montero (2002) also studied the effect of incomplete

⁵Bar-Gill and Harel (2001) provide detailed justifications and implications of this assumption (i.e., the probability that a violator gets caught depends on both enforcement expenditures *and* the proportion that chooses to violate the standard).

enforcement on quantity regulation. Harford showed that the direction of change in waste production with respect to a change in the allowable range of waste levels is ambiguous, depending on the curvature of the expected penalty schedule. Because he considers regulation of a monopolist, the crowding effect does not come into play. Montero (2002) compared the performance of price regulation and quantity regulation with incomplete enforcement, showing that they perform equally if benefits and costs are known, but that quality regulation performs better when benefits and costs are uncertain. Copeland and Scott (2009) analyzed a dynamic model of *renewable* resource management with incomplete or imperfect enforcement. And Villegas-Palacio and Coria (2010) examined the choice between price- versus quantity-based regulation of emissions and the effects of this choice on compliance incentives (i.e., decisions about whether to invest in abatement technology in the presence of incomplete enforcement).

As a heuristic intended to apply more generally to the regulation of many problems of over-use of a commons, we consider a continuum of agents making binary decisions about whether or not to use their cars. The greater the mass of agents who decide to use their cars, the greater is the external congestion cost facing each agent. In a decentralized environment, we first show that the unregulated Nash equilibrium divides potential drivers into two groups: those who use a car and those who do not. The subpopulation of agents who choose to use their cars consists of those whose private valuation from driving exceeds the personal costs of traffic congestion; and the non-driving subpopulation of potential drivers consists of those whose value of using their cars is lower than the cost of traffic congestion that they face.

Once the quantity regulation is imposed, two cases are considered. In the first case, the probability that a violator is caught and pays a fine depends only on enforcement expenditures, as is commonly assumed (e.g., Polinsky and Shavell, 1979, 1984, 1991; Malik, 1990b; and many others). This case serves as a benchmark against which to compare the second case. In the second case, the enforcement probability is assumed to depend on enforcement expenditures and the violation rate.

The model shows that, when the enforcement probability depends on both regulatory expenditures and the violation rate, there is a discontinuity in the population's aggregate best-response quantity, or intensity of use of the commons (as a function of the regulated quantity). In the heuristic interpretation of agents as automobile owners and the over-use externality as traffic congestion, the population's aggregate quantity of binary best-response decisions is interpreted as the volume of traffic or the congestion level. The abrupt reduction in traffic volume just to the right of the discontinuity in the mapping from regulated quantities into best-response volumes of traffic is shown to determine the near-optimal quantity

regulation.

As long as the regulator's primary objective is to maximize aggregate net payoffs by mitigating over-use (of freeways or public roads), tightening the quantity regulation is ineffective over a large and dense subset of the model's parameter space: traffic congestion counter-intuitively increases in response to the stricter quantity standard, and the population is, in the aggregate, worse off. Regardless of whether the enforcement probability depends on enforcement expenditures alone, or both enforcement expenditures and the number of violators, tightening the allowable quantity winds up incentivizing some non-violators to become violators and therefore fails to decrease aggregate levels of use of the commons. When the enforcement probability also depends on the violation rate, this increase in over-use of the commons (as a response to stricter regulation) becomes more severe, rising discontinuously in the neighborhood of the near-optimal quantity restriction, and rising continuously thereafter as the quantity standard is tightened further.

The idea that enforcement probabilities depend on violation rates as well as enforcement expenditures is not new. The dependence of the enforcement probability on the violation rate (or number of violators) is referred to as the crowding effect, and its implication of the possibility of multiple equilibria appears in previous studies such as Ehrlich (1973), Lui (1986), Schrag & Scotchmer (1997) and Bar-Gill and Harel (2001), Ferrer (2010) and Kim (2013), but none of these papers endogenized the regulator's decision of what is legal versus illegal.

Ferrer's (2010) model bears special relevance to our model. She considers two externalities: a crowding effect (i.e., enforcement probabilities decline as a result of greater numbers of violators) and a neighborhood involvement effect (i.e., higher crime rates impose external costs that decrease activity levels identified by Ferrer as lower levels of neighborhood involvement). Given these two features, Ferrer shows that multiple equilibria naturally arise in a model with an exogenous legal standard. Ferrer applies risk dominance to select among these multiple equilibria and then computes an optimal enforcement policy. In most of her analysis, however, she maintains the assumption that all individuals obtain the same benefit from a violation and use pure strategies, which lead to the crime rate equal to either 0 or 1.

Kim (2013) similarly investigates optimal enforcement policy in the presence of a crowding effect, but considers both enforcement expenditures and the fine (that violators who are caught must pay) as jointly chosen policy variables. In that setup, Kim finds that maximal fines are in general sub-optimal. Bar-Gill and Harel (2001) discuss the implication of the crowding effect in a more general setting. They also consider the problem of the optimal fine, but rather concentrate on the case in which sanctioning is costly and pay little attention

to the case where higher fines lower the crime rate.⁶ They therefore failed to recognize the general feature that the optimal fine is not maximal if there are multiple equilibria.

This paper can be distinguished from those just mentioned in two respects. First, quantity regulation is introduced as the primary policy variable of interest. And second, the main outcome that measures the effectiveness of different quantity restrictions is the aggregate quantity of the externality-generating activity rather than the violation rate *per se*. Although the question is analyzed extensively in the public economics literature as to whether over-use of a commons should be addressed by decentralized privatization or centralized regulation, the enforcement of regulation is frequently taken for granted in many such models, as if the regulator possessed a zero-cost enforcement technology.⁷ A detailed specification of imperfect enforcement technology and its effects on externality-generating behavior has yet to be incorporated into models of regulating over-use of a commons as this paper attempts to do.

It is perhaps worth emphasizing that the main issue this paper addresses is the enforcement of regulation rather than law enforcement of completely prohibited actions. The law enforcement literature typically takes as exogenously given the acts that are regarded as illegal. In contrast, this paper models the regulatory question of determining which acts should be regarded as illegal. Its main contribution is to endogenize socially acceptable acts in the context of a tragedy of the commons problem.

The paper is organized as follows. Section 2 describes the model setup and computes the unregulated equilibrium outcome and the social optimum which serve as two important benchmarks. Section 3 introduces the policy of interest, which takes the form of a regulation that limits quantities of an economic activity, motivated by the goal of maximizing social welfare (i.e., maximizing aggregate payoffs with a Benthamite social welfare function interpretation). The ineffectiveness of stricter quantity regulation is demonstrated. Section 4 addresses the issue of optimal enforcement. Section 5 presents a concluding discussion.

2 Basic Model

The model assumes there is a continuum of agents. Each agent makes a binary decision of whether to use his or her car. The more agents choose to use their cars, the more severe the

⁶See Proposition 2 of Bar-Gill and Harel (2001).

⁷Ostrom (1990) proposed a different approach attempting to resolve the common pool problem by designing durable cooperative institutions organized and governed by the resource users themselves.

congestion costs are, which should be thought of as an analogy that applies more generally to over-use of a commons.

The following definitions introduce the notation used to describe the model:

- v = private value obtained by an agent who uses a car (i.e., gross payoff without netting out congestion costs, fees and fines);
- x = proportion or volume of agents who choose to use their cars (i.e., traffic congestion or aggregate use of the commons $x \in [0, 1]$);⁸
- $c(x)$ = congestion cost function that maps delay, stress and inconvenience into payoffs forgone ($c' > 0$, $c'' > 0$ and $c \geq 0$ for all $x \in [0, 1]$);
- $g(v)$ = probability density function for v ($g > 0$ for all $v \geq 0$);
- $G(v)$ = cumulative distribution function corresponding to $g(v)$ ($G(v) \in [0, 1]$ for all $v \geq 0$).

Here, v is the value and $c(x)$ is the cost of using a car, where other costs such as fuel, insurance and capital depreciation are normalized to zero. This society consists of a continuum of agents (i.e., all potential drivers) indexed by their valuations of having access to the commons, v . We make use of the continuum version of Nash equilibrium: each agent is assumed to maximize his or her net benefit from using a car, given the mass of other agents who choose to use their cars.

First, consider the private optimum achieved in a decentralized environment with no regulation. Let x^N be the (Nash) equilibrium proportion of agents who choose to use their cars as a best response to the traffic volume in the commons represented by x^N . In the absence of fines or fees required of agents who use their cars, an agent chooses to drive iff $v \geq c(x^N)$. If an equilibrium exists, then it can be characterized as the threshold valuation v^N such that all agents with $v \geq v^N$ choose to drive, where:

$$v^N = c(x^N) \text{ and } x^N = 1 - G(v^N). \quad (1)$$

⁸The capacity of the commons is normalized to one.

Now we consider the properties of a social optimum, denoted x^S . The first efficiency requirement is that, given any mass of traffic x^S , the agents who use their cars should be those with the largest valuations. Therefore, the aggregate level of use of the commons, x^S , must be generated by an action profile in which this mass of agents is comprised of agents who have the highest valuations for driving. Given that the infimum valuation among agents whose driving decisions comprise x^S is given by v^S , the social optimum, (v^S, x^S) , is determined by maximizing the sum of individual valuations net of total external costs, that is, by solving the following optimization problem:

$$\max_{\hat{v}} V = \int_{\hat{v}}^{\infty} (v - c(x))g(v)dv, \text{ such that } x = 1 - G(\hat{v}). \quad (2)$$

The objective function is the total net surplus from driving.

Substituting the constraint $x = 1 - G(\hat{v})$ into the objective function, the regulator's choice variable \hat{v} then appears twice. Taking the derivative with respect to \hat{v} , the first order condition can be written as:

$$\frac{\partial V}{\partial \hat{v}} = \left[c(x^S) + c'(x^S) \int_{\hat{v}}^{\infty} g(v)dv - v^S \right] g(v^S) = 0.$$

Because the term in brackets must be zero at an interior optimum, the solution to the benevolent regulator's problem is defined implicitly by the shapes of $c(x)$ and $g(v)$ in the following relation:

$$v^S = c(x^S) + c'(x^S) \int_{v^S}^{\infty} g(v)dv. \quad (3)$$

Equation (3) has the usual interpretation that the marginal social benefit from an increase in traffic equals the marginal social cost due to congestion. The second term on the right-hand side is the aggregate cost of negative externalities from legal driving. Thus, $c^S(x) \equiv c(x) + c'(x) \int_{v^S}^{\infty} g(v)dv$ is the social cost associated with traffic generated by agents choosing to use their cars.

Comparing the social optimum x^S with the Nash equilibrium x^N , it is easy to see that $x^N > x^S$. This follows from equations (1) and (3). Because $c'(x) > 0$ for all x , equation (3) implies that the minimum value of driving among agents who choose to drive in the social optimum (v^S) is strictly greater than the personal cost ($c(x^S)$). This inequality, $v^S > c(x^S)$, in turn, implies that $x^N > x^S$, because $v^N = c(x^N)$ from equation (1) and $c'(x) > 0$. Therefore, there is too much driving (i.e., congestion or over-use of the commons) in the Nash equilibrium relative to the social optimum, as illustrated in Figure 1.

Figure 1 plots the convex cost function $c(x)$ and the valuation function $v(x) = G^{-1}(1-x)$, based on the definition of the distribution function and the equilibrium relationship given in

equation (1) (i.e., $x = 1 - G(v)$). The unregulated Nash equilibrium in which agents do not internalize the external costs of their driving decisions is given by the intersection of the two curves at $x = x^N$. The social optimum depicted in Figure 1 x^S , which reflects the inequality $v^S > v^N$, corresponds to a strictly smaller volume of traffic $x^S < x^N$. This result is simply a continuum version of the tragedy of the commons.⁹

3 Regulation

We consider a regulator who wants to mitigate over-use of the commons (which shows up in the context of driving as traffic congestion) by directly controlling the aggregate quantity of agents who use the commons. The problem with quantity regulation is that it is difficult to tell who should be penalized when traffic exceeds the regulated quantity. A theoretical solution is to give permits only to individuals with sufficiently high valuations. But agents' individual valuations may not be easy to observe *ex ante*. This would, however, not cause any serious problem so long as individual valuations are *ex post* observable. Thus, for our analysis of quantity regulation, we assume that the regulatory authority knows the valuation of each heterogeneous agent.¹⁰

The following notation is introduced to analyze the regulation problem in more detail:

- z = the regulator's policy variable that sets the allowed quantity (or proportion) of agents permitted to drive;

- p = the probability that a violator is detected and caught;

⁹Strategic interactions involving costly contributions of information that facilitate group payoffs (shared by all players) are interpreted as an extended version of the tragedy of the commons problems in the social psychology and biology literatures. See Kameda, Tsukasaki, Hastie and Berg (2011) for further examples of interpretations of tragedy of the commons in strategic problems in social groups relevant to the model presented in this paper.

¹⁰Consider a situation in which the regulator randomly inspects drivers and can determine based on their valuations, which are revealed *ex post*, whether or not they violate the regulation (i.e., whether their valuations exceed the regulatory requirement). For example, on College Entrance Examination day in Korea, to reduce congestion, the government prohibits everyone from operating automobiles from 8 a.m. to 9 a.m. except for those dropping off students to their exams. Whether or not a particular vehicle carries a student is not observable *ex ante*; it is, however, observable *ex post*, which makes the regulation enforceable.

- r = the quantity of violators or population-wide rate of violation (which measures the mass of agents who violate the regulation); ¹¹
- e = enforcement expenditures exogenously allocated to the regulator for enforcement of the quantity regulation;
- f = the fine paid by violators who are caught;
- $w(x) \equiv c(x) + pf$, the expected total cost of driving for an agent who chooses to drive illegally;
- $x^*(z)$ = equilibrium traffic volume when the regulation level is z .

The benevolent regulator chooses the allowable quantity of drivers by selecting the infimum valuation that solves the following optimization problem:

$$\max_z V = \int_{v(x^*(z))}^{\infty} (v - c(x^*(z)))g(v)dv. \quad (4)$$

Under a regulatory regime that imposes a fine of f and detects violators with probability p , agents who are prohibited from driving face the decision of whether to violate the regulation. Prohibited agents compare their personal valuation v with the expected cost $w(x)$. By definition of the regulation z , quantities of traffic that exceed it ($x > z$) are not allowed, while quantities of traffic below the regulatory limit ($x \leq z$) are allowed.

The quantity of violators r is defined by $r = x - z$ if $x - z \geq 0$ and zero otherwise: $r = \max\{x - z, 0\}$. Although it is usual to define the violation rate by the number of violators divided by the number of potential violators, with a continuum of agents on the

¹¹Because the continuum of agents who are potential drivers (i.e., the population) is normalized to a mass of measure 1, the quantity of agents who drive (denoted x) and the quantity of agents who drive in violation of the regulation (denoted r) can be described as quantities or rates. These variables measure quantities as the sizes of masses of agents. They simultaneously measure population-wide rates. The population-wide violation rate (averaging the mass of violators over the mass of both drivers and non-drivers, which is unity) is not the same as the violation rate among those who drive, which would be given by the formula r/x . The term “number of agents” is not technically correct, since the model features infinite numbers of agents in each subpopulation category with strictly positive mass.

unit interval (i.e., with mass of 1), the quantity of violators and violation rate are one in the same. This distinction does not matter in the law enforcement literature because the entire population is considered to be potential criminals. In the context of regulation, however, the number of potential violators (i.e., the mass of agents in the denominator) varies with the standard. Since it is the difference $x - z$ that directly affects the detection probability, we adopt this definition of the measure of violation and refer to it interchangeably as the number, volume or rate of violation.

Regarding the probability that violators are detected and forced to pay the fine, we consider two cases. In case 1, the detection rate depends only on enforcement expenditures: $p = p(e)$. In case 2, the detection rate depends on both expenditures and the mass of agents who violate the standard: $p = p(e, r)$. In both cases, we assume that greater enforcement expenditures, all else equal, lead to higher rates of detection: $p_1 = \partial p / \partial e > 0$. And in case 2, we assume that $p_2 = \partial p / \partial r < 0$. The assumption that $\partial p / \partial r < 0$ is referred to in the literature mentioned earlier as the crowding effect. This captures the idea that the detection rate declines, the more violators there are.

Case 1 (in which the probability of detection depends only on enforcement expenditures) serves as a benchmark against which to compare case 2 (where the probability of detection depends on both enforcement expenditures and the violation rate). Case 2 is the main focus of this paper. Additionally, we assume that $p = p(e, r)$ satisfies the Inada conditions: $p_{22}(e, r) > 0$, $\lim_{r \rightarrow 0} p_2(e, r) = -\infty$ and $\lim_{r \rightarrow 1} p_2(e, r) = 0$ for any given $e > 0$.

We assume that e and f are fixed quantities that all agents regard as exogenously given. This enables us to focus on the regulated quantity z as the primary policy variable under consideration, which our model endogenizes, and provides a context for analyzing population-level reactions to changes in the regulated quantity based on individual-level best-response behavior. If e and f were simultaneously chosen in the regulator's optimization problem, then we would already know there exists a unique value of $p(e)f$ that implements the social optimum x^S . By the well-known argument of Becker (1968), the optimal fine f should be set at the maximum level possible. In the real-world policy contexts addressed by our model that were discussed in the introduction, however, the allocation of resources for enforcement is determined by someone other than the entity responsible for enforcement.

3.1 Detection probability that depends only on enforcement expenditures

To focus on the role of the quantity regulation (z) as a tool for limiting over-use of the commons, we assume that e and f are sufficiently small to guarantee that $v(x^S) > w(x^S)$. This inequality implies that there exists a mass of violators whenever the regulator attempts to set z at the socially optimal level. We refer to it (the inequality, $v(x^S) > w(x^S)$) as the assumption of *weak enforcement*, drawing motivation from the many real-world settings in which the regulator's problem is challenging in the sense that his or her first-best solution cannot be implemented as a straightforward legislative command. The weak enforcement inequality also implies that $p(e)f < v(x^S) - c(x^S)$.¹² If law enforcement is strong in the sense that law enforcement policies e and f can be so chosen to satisfy $p(e)f = v(x^S) - c(x^S)$, then the social optimum is trivially achieved.

Figure 2 illustrates that, if agents face expected fines (pf) in addition to congestion costs, then quantity regulation succeeds at reducing aggregate use of the commons below the level that prevails in the un-regulated Nash equilibrium, but generally fails to reduce congestion to the socially optimal quantity. There is a threshold value of the regulation, denoted $z = x_0$, that is the strictest regulation (i.e., smallest value of z) that can be implemented with a zero rate of violation. Consider the quantity restriction set at this value, $z = x_0$, which lies somewhere between the larger unregulated Nash level and the smaller socially optimal level, where $x^S < x_0 < x^N$. If this regulation achieves a zero violation rate, then $x^* = z$, which results from agents with valuation v choosing to drive if and only if $v \geq v_0$, where $x_0 = 1 - G(v_0)$. This can be seen in Figure 2 by virtue of the position of the $w(x)$ curve. For agents with $v < v_0$, the value of using a car is less than the expected cost of driving illegally, because $v < v_0 = w(x_0)$. In the regulatory equilibrium depicted in Figure 2, no one violates the regulation.¹³

¹²We consider the case of weak enforcement, because it conforms with observed real-world problems of coordinating policy-making and enforcement, although it turns out not to be essential for the model to work. Budgets set by legislatures that meet infrequently, when combined with political pressure and other factors that slow decision making by those responsible for setting enforcement budgets, would imply that e is difficult and/or slow to adjust.

¹³The decision of the agent whose valuation is precisely on the threshold value v_0 is not straightforward. This agent is not indifferent between driving and not driving, even if $v_0 = w(x_0)$. In the absence of expected regulatory fines, this agent would strictly prefer to drive, because $v_0 = w(x_0) > c(x_0)$. Under the quantity regulation's threat of being fined, however, the agent with $v < v_0$ strictly prefers not driving (i.e., not violating) to driving (i.e., violating), because $v < v_0 = w(x_0)$. There is a discontinuity in the cost of driving at v_0 , because this is where the regulation begins to bind and expected fines jump from zero (for those with

We now consider a stricter regulatory standard, $z = x_1$, where $x^S < x_1 < x_0$, as depicted in Figure 2. It is straightforward to see that tightening the quantity regulation z without changing fines or enforcement expenditures leaves $w(x)$ in exactly the same position and, consequently, equilibrium traffic remains unchanged at x_0 . The policy tightens from $z = x_0$ to $z = x_1 < x_0$, but the equilibrium traffic or congestion level does not respond. In the range $z < x_0$, changes in quantity regulation have no effect on behavior. The result is that the mass of agents with measure $x_0 - x_1$ chooses to violate the regulation, which can be explained as follows. Let v_1 satisfy $x_1 = 1 - G(v_1)$, and take any agent with $v \in (v_0, v_1]$. When this agent (with valuation $v \in (v_0, v_1]$) considers driving in violation of the regulation, the agent sees that her valuation exceeds cost, because $v > v_0 = w(x_0)$. Therefore, this agent chooses to violate the regulation. Because any regulation standard z stricter than x_0 (i.e., $z < x_0$) has the same effect of increasing congestion, we refer to v_0 and x_0 as the *marginal individual* and *marginal quantity*, respectively, and refer to the regulation $z = x_0$ as *marginal regulation*.

Figure 3 shows the equilibrium traffic volume as a function of the regulator's choice of z . The graph has a flat region representing the unresponsiveness of traffic volume to regulation when the regulated quantity is set strictly below the marginal regulation x_0 and a positively sloped region with zero violation when $z > x_0$. Figure 3 demonstrates that the social optimum x^S cannot be implemented for any quantity regulation z , so long as enforcement is weak in the sense that $p(e)f < v(x^S) - c(x^S)$.

This result suggests that regulating traffic more strictly than the marginal regulation is ineffective, because more stringent restrictions on quantities do not reduce over-use of the commons. Therefore, the marginal regulation $z = x_0$ is the best regulation the regulator can choose (assuming, as always, that the regulator regards regulation-violating behavior as socially undesirable and wants to reduce it).

3.2 Detection probability that depends on the violation rate and enforcement expenditures

We now assume that the detection probability depends on the violation rate, implying that the detection probability also depends on the regulation z . In this case, the expected cost of violation ($w(x)$) also depends on z .

valuations nearby, but driving legally) to strictly positive expected fines (for those with arbitrarily nearby valuations, but driving illegally because their valuations happen to lie on the wrong side of the epsilon ball about the regulatory threshold). Such discontinuities are typical in economic models of legal standards where an exogenously given costly sanction kicks in abruptly for agents who cross a critical threshold.

This crucial feature whereby p and $w(x)$ both depend on z is what distinguishes this version of the model from those analyzed previously and leads to this paper's main results. When p and $w(x)$ depend on z , the regulator can no longer choose z simply by consulting $w(x)$, because $w(x)$ represents the expectation of agents who first condition their expected cost (of illegally using the commons) on the regulator's choice of z . The regulator must therefore try to anticipate how the aggregate level of use of the commons will respond to the announcement of z . Thus, we introduce the notation $w(x | z)$, which is defined as an un-permitted agent's expected cost of driving, conditional on the regulation policy z , which is announced by the regulator before agents make decisions about whether or not to drive.

The shape of $w(x | z)$ provides the mechanism that drives the counter-intuitive result that follows. Because of the Inada conditions given earlier, we know that as x approaches the regulated quantity z from the right, the slope of $p(e, x - z)$ with respect to x goes to $-\infty$ and so, too, does the slope of $w(x | z)$, since $w(x | z) = c(x) + pf$. As x increases, the slope of $p(e, x - z)$ with respect to x eventually converges to zero. Therefore, the slope of $w(x | z)$ converges to the slope of $c(x)$. Since $c'(x) > 0$, the curve of $w(x | z)$ must eventually rise, implying that $w(x | z)$ is U -shaped.

Figure 4 illustrates the consequences of the U -shaped expected cost functions $w(x | z)$. The three bolded plots of $w(x | z)$ in the figure represent conditional expected cost functions corresponding to three different choices by the regulator of z (which, as always, represents the maximum allowable aggregate quantity of use of the commons). We consider three increasingly tighter (i.e., stricter) policies: $z = z^A$ corresponding to equilibrium point A in Figure 4; $z = z^B$ corresponding to equilibrium point B; and $z = z^C = x^S$ corresponding to equilibrium point C. Figure 4 shows that the policy choices and equilibrium traffic volumes move in opposite directions. The figure demonstrates the counter-intuitive result that equilibrium congestion increases ($x^A < x^B < x^C$) in response to stricter regulatory policies ($z^A > z^B > z^C$).

The bottom-most bolded curve in Figure 4, $w(x | z = x^S)$, is generated by a regulator who announces the policy $z^C = x^S$, which is violated severely so that equilibrium traffic (the x-component of the point labeled C) is much greater than the maximum allowable quantity chosen by the regulator. The policy $z = x^S$ therefore represents a policy maker who announces that the socially optimal level of traffic will be permitted under the naive expectation that all unpermitted agents will follow the regulation. If $z = z^C = x^S$, then equilibrium traffic volume (as determined by the intersection of $v(x)$ and $w(x | x^S)$) is denoted x^C , which lies far to the right of x^S (and just to the left of x^N) in Figure 4. The distance by which actual traffic x^C exceeds permitted traffic x^S measures the corresponding

violation rate generated by the policy: $x^C - x^S$.

The middle bolded curve in Figure 4 labeled $w(x | z^B)$ (just above the lowest bolded curve) corresponds to the more permissive policy $z^B > z^C = x^S$. If the regulator decides to increase the quantity of permitted traffic from z^C to z^B , then equilibrium traffic x^B is determined by the intersection of $v(x)$ and $w(x | z^B)$; the violation rate decreases from $x^C - z^C$ to $x^B - z^B$; and the equilibrium traffic volume decreases from the horizontal component of point C (x^C) to that of point B (x^B). Note that x' , which lies on the curve $w(x|z^B)$, cannot be an equilibrium traffic volume by the following proposition.

Proposition 1 *Given z , the equilibrium traffic volume satisfies (i) $v(x) = w(x | z)$ and (ii) $v'(x) < w'(x | z)$.*

If $v'(x) > w'(x | z)$, as is the case at x' , then the marginal driver prefers driving because the marginal benefit is greater than marginal cost.

Figure 4 shows that the less stringent regulation $z^B > z^C$ winds up reducing aggregate use of the commons (i.e., leads to a smaller equilibrium traffic volume and smaller volume of violators) than under the strictest policy (z^C) shown in Figure 4. In this instance, congestion is mitigated by more permissive quantity regulation. Decreases in traffic volume as desired by the regulator are achieved when the regulator chooses more permissive regulated quantities, up to the traffic level at which the two curves, $v = 1 - G(x)$ and $w(x) = c(x) + p(e, r)f$, are tangent to each other. The tangent point is illustrated as A in Figure 4. Let the regulation quantity inducing the tangent point be denoted z^* . Then, in this figure, $z^* = z^A$. We observe that the violation rate is continuously decreasing along the policy path of successively more permissive policies that approach z^A from the left. Just to the right of the policy that induces equilibrium A , however, there is an interesting discontinuity in traffic volume (not shown in Figure 4). If the regulator chooses a policy slightly more permissive than z^A , say $z = z^A + \epsilon$, $\epsilon > 0$, then traffic volume falls discontinuously, and thereafter increases smoothly as an increasing function of z for $z > z^A$ as shown in Figure 5.

The reasoning just described leads to the conclusion that a benevolent regulator should choose $z = z^* + \epsilon$, with ϵ chosen as small as possible, where z^* is the regulation quantity that achieves the tangency condition. We will refer to this policy as “near-optimal” regulation, because there exists no maximum aggregate payoff. The discontinuity at z^* forces the regulator to choose ϵ as small as possible to achieve an aggregate payoff that is arbitrarily close to the least upper bound on the aggregate payoff function.

Beyond this small technicality, the key point is that the optimal (i.e., near-optimal) quantity regulation cannot be x^S , and in fact should be much more permissive, so long as

$$w(x^S | x^S) < v(x^S).$$

Proposition 2 *If enforcement is weak (i.e., $p(e)f < v(x^S) - c(x^S)$), the socially optimal (i.e., aggregate-payoff-maximizing) regulation standard cannot be implemented using any quantity regulation regardless of whether the probability of detection depends on enforcement expenditures alone or on both expenditures and the number of violators.*

The intuition for this proposition is that any agent whose valuation is very close but just below the threshold v^S , which is the minimum valuation among all agents permitted to drive under the quantity regulation, would rather violate the regulation because of weak enforcement technology. Agents' best response function conditional on stricter quantity regulation results in traffic levels that, on the range of z to the left of the discontinuity point z^* , are either non-responsive (when the probability of detection depends only on enforcement expenditures) or, counter-intuitively, increasing in strictness (when the probability of detection depends on both expenditures and the violation rate).

Proposition 3 *If enforcement is weak, equilibrium traffic volume $x^*(z)$ is nonmonotonic with respect to the quantity regulation z . Beginning from the strictest policy $z = 0$, equilibrium traffic volume is a decreasing function of permissiveness for $0 \leq z < z^*$; it then drops discontinuously at $z = z^*$; and to the right of z^* , traffic volume increases in z for $z > z^*$.*

Proof. From the discussion above, it suffices to show that, for all $z > z^*$, equilibrium traffic volume is $x^*(z) = z$. It is clear that all agents with $v \geq v(z)$ choose to drive cars. For any agent with $v < v(z)$, his or her expected cost as a violator exceeds the benefit of driving, since $w(x | z) > v(x)$ for all $x > z$. This implies that agents who are not permitted to drive have no incentive to deviate when z is strictly greater than z^* . Therefore, equilibrium traffic level is $x^*(z) = z$.

Figure 5 illustrates the equilibrium traffic level over the full range of quantity regulations (z) for the case of detection technology p that depends on both expenditures and the mass of violators. The figure shows that more stringent limits on quantities of use (i.e., lower values of z) result in nonmonotonic responses in traffic levels as described in the proposition.

Proposition 4 *If enforcement is weak, the near-optimal regulation minimizing traffic volume is $z = z^* + \epsilon$, where z^* is the regulation inducing the tangent point defined by $v'(x^*) = w'(x^* | z^*)$ denoted as point A and ϵ is a positive number chosen as small as possible.*

Proof. It directly follows from Proposition 3.

This proposition holds even when there are multiple equilibrium traffic volumes. Figure 6 illustrates the possibility of multiple equilibria and the resulting optimal policy,¹⁴ and Figure 7 shows the equilibrium correspondence as the policy changes.

Because traffic volume varies non-monotonically as a function of the regulation standard z , identifying the optimal standard is not as straightforward as in models where maximum harshness or stringency is found to be optimal. Nevertheless, the incentive structure that determines this near-optimal regulation is quite clear. If the regulated quantity is sufficiently large (i.e., a relatively permissive allowable aggregate traffic volume z), then the detection probability remains relatively large because there are few violators. To maintain this sufficiently large detection probability and its consequence of a sufficiently large expected fine that would-be violators face, the near-optimal regulation requires the regulator to select the harshest quantity standard that induces zero violation.

It is worthwhile stressing that the assumption of weak enforcement maintained in the hypotheses of all propositions stated above (in combination with other assumptions in each proposition) is a sufficient condition for the results. The next section suggests that this assumption is not a necessary condition, however.

4 Discussion on Optimal Enforcement

So far, we assumed that the expected penalty for being caught while using the commons illegally is determined by exogenously given levels of expenditures on enforcement and the fine. In addition, we assumed weak enforcement (i.e., $pf < v(x^S) - c(x^S)$) to focus on the case of imperfect law enforcement.

We now consider the question of what happens in our model if the government can choose e and f as policy variables jointly when choosing z in the case where the detection probability depends on the violation rate as well. One might conjecture that the social optimum x^S could be achieved by choosing the regulation $z = x^S$ and large enough values of e and f to support the social optimum with strong enforcement satisfying $p(e, 0)f = v(x^S) - c(x^S)$, as illustrated in Figure 8. This policy does not induce an equilibrium, however.

Under the regulation $z = x^S$, the traffic volume $x = x^S$ cannot be an equilibrium. To see this, consider the driving decision of an individual with $v = v^S - \epsilon$ for some small $\epsilon > 0$. If

¹⁴By virtue of Proposition 1, we know that there are two equilibria. The slope condition excludes the other two intersection points as possible equilibria. An equilibrium in this model requires this extra condition, unlike previous models in this literature (e.g., Ferrer [2010], Kim [2013]), because one of the axes in Figure 6 measures the valuation of an individual.

she drives (illegally), the small increase in x when she is added to the mass of legal drivers results in a disproportionately large drop in the detection probability because of the Inada Condition. The result is that $v > w(x)$: therefore, this agent finds it rational to drive illegally, and the conjectured equilibrium unravels (i.e., is not an equilibrium). As long as the slope of $w(x)$ is smaller than $v(x)$ at x^S , this is true until the traffic volume reaches x^* . Therefore, the policy pair (e, f) satisfying $p(e, 0)f = v(x^S) - c(x^S)$ cannot implement the desired no-violation equilibrium. This problem would not occur if the enforcement probability were independent of the violation rate.

If the regulator chooses a stricter regulation $z < x^S$ together with an even stronger sanction such that $p(e, 0)f > v(x^S) - c(x^S)$, then the social optimum could be implemented. To see this, we start from a very strict regulation $z_0 < x^S$ and a very low enforcement e_0 expenditure at which there is severe congestion. Then, either a less stringent regulation $z' > z_0$ or a more intensive enforcement activity $e' > e_0$ will reduce congestion. It will be more efficient, however, for the regulator to use the option of choosing a more permissive (i.e., larger value of) z , because z and e are substitutes, and z is less costly than e .

This insight enables us to conjecture that the regulation could implement an equilibrium very near the social optimum even with minimal enforcement activity. The possibility of multiple equilibria given z might complicate the problem, but we believe that the insight will remain qualitatively unaffected as long as there is a unique optimal value of z . This suggests that the effect of quantity regulation crucially depends on whether regulation enforcement is determined exogenously or endogenously.

5 Conclusion and Caveats

Traffic and network congestion are serious real-world concerns in many sectors of the economy. The intuition behind the model presented in this paper is that whenever the probability of catching violators of a quantity regulation moves inversely with the number of violators, then tighter restrictions on access to congestion-prone commons may counterintuitively lead to increased congestion. If tighter regulation implies larger numbers of violators and therefore smaller chances that they will face sanctions, then permissive regulation can, in the context of our model, reduce congestion more effectively than strict limits on levels of economic activity. This intuition potentially applies to resources used in common, such as public roads, fisheries, computer bandwidth and electronic price quotation systems in financial markets. The model in this paper suggests that quantity restrictions in the contexts just mentioned may counter-intuitively aggravate congestion problems due to the crowding effect.

On the other hand, quantity regulation has potential as a cost-effective tool for regulating externality-generating activity, because it is relatively simple (compared to price regulation schemes such as Pigouvian taxes) and easily understood by those affected.

The potentially unstable effectiveness of enforcement, however, raises the possibility investigated in this paper that stricter regulation may lead to greater-than-anticipated use of the commons, thereby imposing unnecessarily high external costs on other users of the commons. It will therefore be worthwhile to consider the characterization our model provides of near-optimal quantity, which requires a regulatory standard that is sufficiently permissive to elicit very few attempts at violating the regulation. As a result, the limit on activity that is sufficiently permissive, but no more so than necessary, elicits a very high rate of compliance and, consequently, high rates of detection of those violating the regulatory standard.

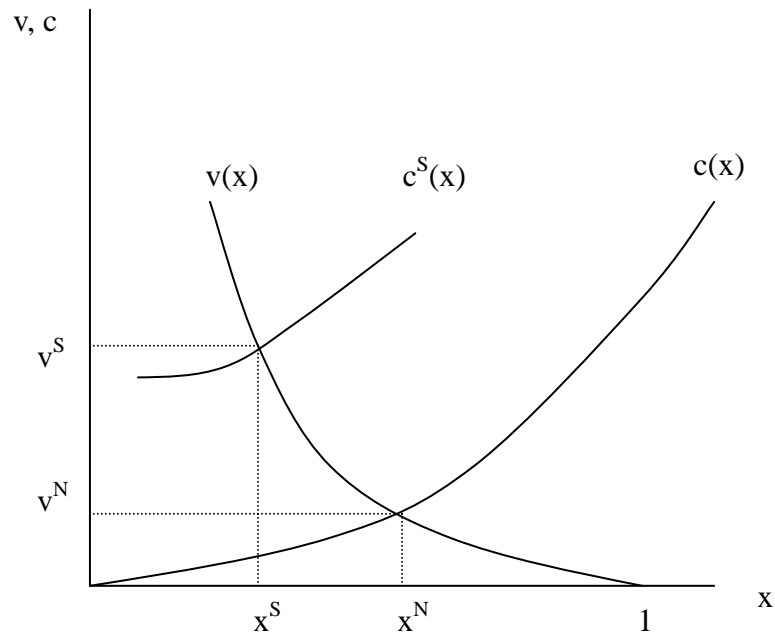
References

- [1] Bar-Gill, O. and A. Harel, 2001, Crime Rates and Expected Sanctions: The Economics of Deterrence Revisited, *Journal of Legal Studies* 30, 485-501
- [2] Becker, G., 1968, Crime and Punishment: An Economic Approach, *Journal of Political Economy* 76, 169-217
- [3] Copeland, B.R. and Scott M.S., 2009, Trade, Tragedy, and the Commons, *American Economic Review* 99, 725-749
- [4] Ehrlich, I., 1973, Participation in Illegitimate Activities: A Theoretical and Empirical Investigation, *Journal of Political Economy* 81, 521-565
- [5] Ferrer, R., 2010, Breaking the Law when Others Do: A Model of Law Enforcement with Neighborhood Externalities, *European Economic Review* 54, 163-180
- [6] Hardin, G., 1968, Tragedy of the Commons, *Science* 162, 1243-1248
- [7] Harford, J., 1978, Firm Behavior Under Imperfectly Enforceable Pollution Standards and Taxes, *Journal of Environmental Economics and Management* 5, 26-43
- [8] Kameda, T., Tsukasaki, T., Hastie, R., and Berg, N., 2011, Democracy Under Uncertainty: The Wisdom Of Crowds And The Free-Rider Problem In Group Decision Making, *Psychological Review* 118, 76-96

- [9] Kim, J.-Y., 2013, A Note on the Non-Maximality of the Optimal Fines when the Apprehension Probability Depends on the Offense Rate, *European Journal of Law and Economics* 36, 131-138
- [10] Kornhauser, A., Fehlig, M., 2002, Marketable Permits for Peak Hour Congestion in New Jersey's Route 1 Corridor, Mimeo
- [11] Lui, F., 1986, A Dynamic Model of Corruption Deterrence, *Journal of Public Economics* 32, 215-236
- [12] Malik, A., 1990a, Markets for Pollution Control when Firms are Noncompliant, *Journal of Environmental Economics and Management* 18, 97-106
- [13] Malik, A., 1990b, Avoidance, Screening and Optimum Enforcement, *Rand Journal of Economics* 21, 341-353
- [14] Montero J.-P., 2002, Prices versus Quantities with Incomplete Enforcement, *Journal of Public Economics* 85, 435-454
- [15] Ostrom, E., 1990, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press
- [16] Polinsky, M. and S. Shavell, 1979, The Optimal Tradeoff between the Probability and Magnitude of Fines, *American Economic Review* 69, 880-891
- [17] Polinsky, M. and S. Shavell, 1984, The Optimal Use of Fines and Imprisonment, *Journal of Public Economics* 24, 89-99
- [18] Polinsky, M. and S. Shavell, 1991, A Note on Optimal Fines When Wealth Varies Among Individuals, *American Economic Review* 81, 618-621
- [19] Roberts, M. and Spence, M., 1976, Effluent Charges and Licenses under Uncertainty, *Journal of Public Economics* 5, 193-208
- [20] Schrag, J. and Scotchmer, S., 1997, The Self-Reinforcing Nature of Crime, *International Review of Law and Economics* 17, 325-335
- [21] *The Economist*, 2011, The Irony of the Tragedy of the Commons, Feb 24

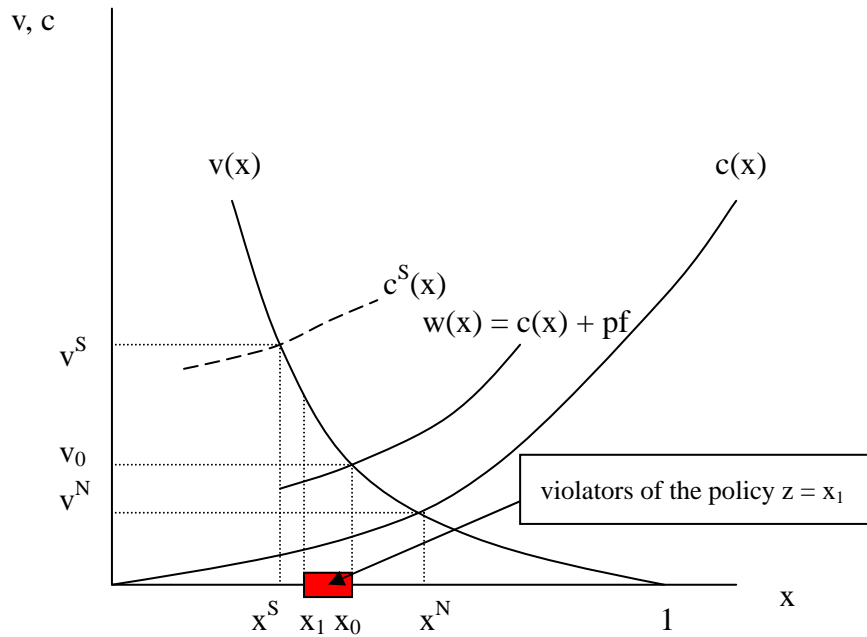
- [22] Verhoef, E.T., Nijkamp, P. and Rietveld, P., 1997, Tradeable Permits: Their Potential in the Regulation of Road Transport Externalities, *Environment and Planning B: Planning and Design* 24, 527-548
- [23] Villegas-Palacio, C. and Cirroa J., 2010, On the Interaction between Imperfect Compliance and Technology Adoption: Taxes versus Tradable Emissions Permits, *Journal of Regulatory Economics* 38, 274-291
- [24] Weitzman, M., 1974, Prices vs. Quantities, *Review of Economic Studies* 41, 477-491
- [25] Yohe, G.W., 1978, Towards a General Comparison of Price Controls and Quantity Controls under Uncertainty, *Review of Economic Studies* 45, 229-238

Figure 1: Tragedy of the commons with continuum-valued cardinality (e.g., the volume of traffic congestion)



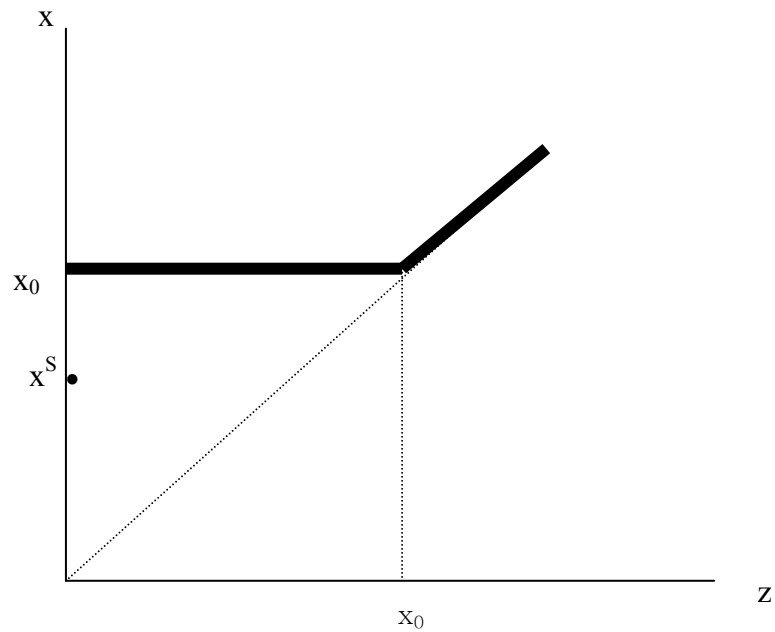
The decentralized Nash equilibrium x^N occurs at the intersection of the private value and cost functions where $v(x) = c(x)$. The social optimum x^S occurs at the intersection of the private value and social cost functions: $v(x) = c^S(x)$. Because $c^S(x) > c(x)$ at each value of x , the figure demonstrates that $x^S < x^N$: there is always over-use of the commons (i.e., x is too high) from the point of view of a regulator whose first-best outcome is x^S .

Figure 2: The effect of quantity regulation when the probability of detection depends only on enforcement expenditure: $p = p(e)$



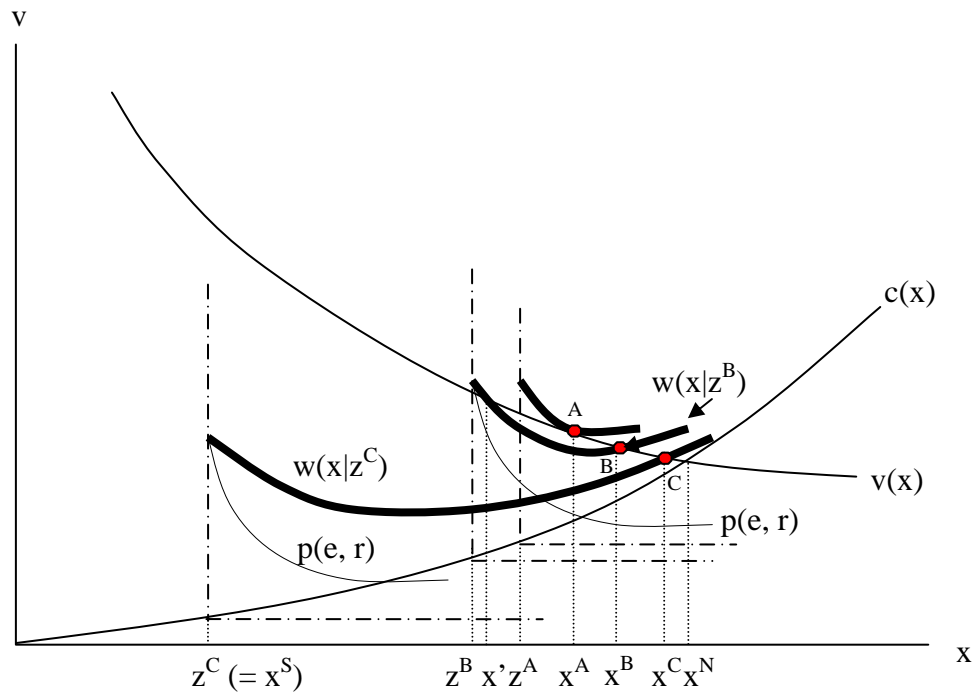
The regulator tightens the quantity regulation by shifting from x_0 leftward to x_1 . Because the probability of detection p and fine f are, by assumption, unaffected, the equilibrium traffic volume remains at x_0 . Thus, p and f , together with the shapes of $c(x)$ and $v(x)$, jointly determine the equilibrium traffic volume, under the assumption that $p = p(e)$ (i.e., p is independent of x). The assumed independence of p from x implies that tightening the quantity regulatory standard has no effect on congestion, which cannot be brought below x_0 .

Figure 3: Equilibrium traffic volume as a function of the regulator's choice of quantity regulation z , assuming the probability of detection, $p = p(e)$, depends on expenditures but not on the mass of violators



For sufficiently permissive policies ($z > x_0$), the simple enforcement technology enables the regulator to implement whichever quantity limit is desired in this range. Below the threshold x_0 , however, the regulator's tightening of the policy has no effect on the equilibrium traffic volume. In other words, changes in quantity regulation in the range ($z < x_0$) have no effect on over-use of the commons.

Figure 4: The effect of quantity regulation when the probability of detection, $p = p(e, r)$, depends on the quantity of violators r and, consequently, on the regulatory standard z



This Figure demonstrates the counter-intuitive result that holds whenever the probability of detection depends on the regulatory standard z : moving from weaker to stricter regulated quantities $z^A > z^B > z^C = x^S$ corresponds to equilibria A, B and C, which exhibit progressive worsening levels (from the regulator's point of view) of congestion $x^A < x^B < x^C$. These three points of intersection between $w(x|z)$ and $v(x)$ move in the opposite direction relative to the verticals marking the regulator's choice of the maximum allowable aggregate quantity. In other words, tightening the maximum allowable quantities results in the unintended effect of increased congestion. Equivalently, when the allowable regulated quantity becomes more permissive, then aggregate use of the commons in equilibrium decreases.

Figure 5: Discontinuity in equilibrium traffic volume as a function of the regulator's choice of the maximum allowable volume z , when the detection probability $p = p(e, r)$ depends on the violation rate r

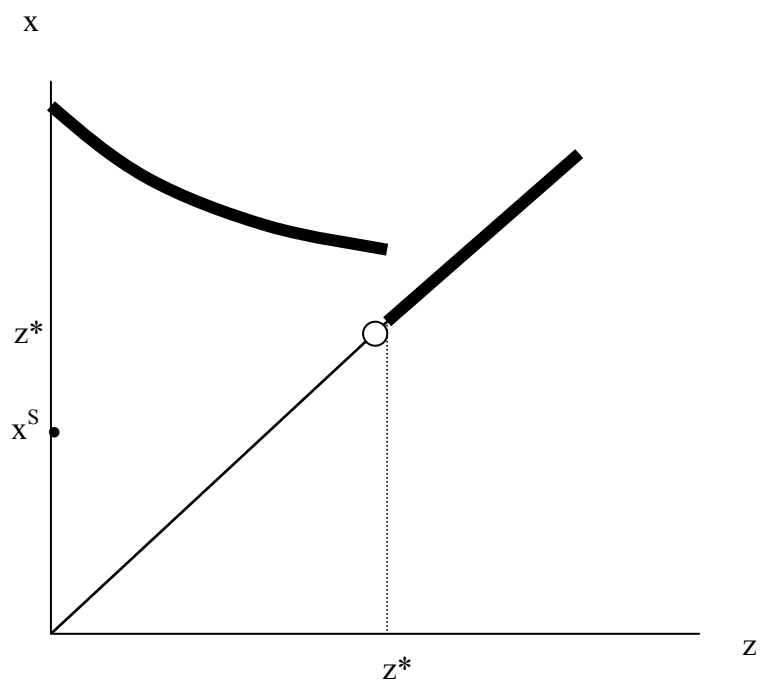


Figure 6: The possibility of multiple equilibria (x^* and x^{**} are both equilibrium traffic volumes)

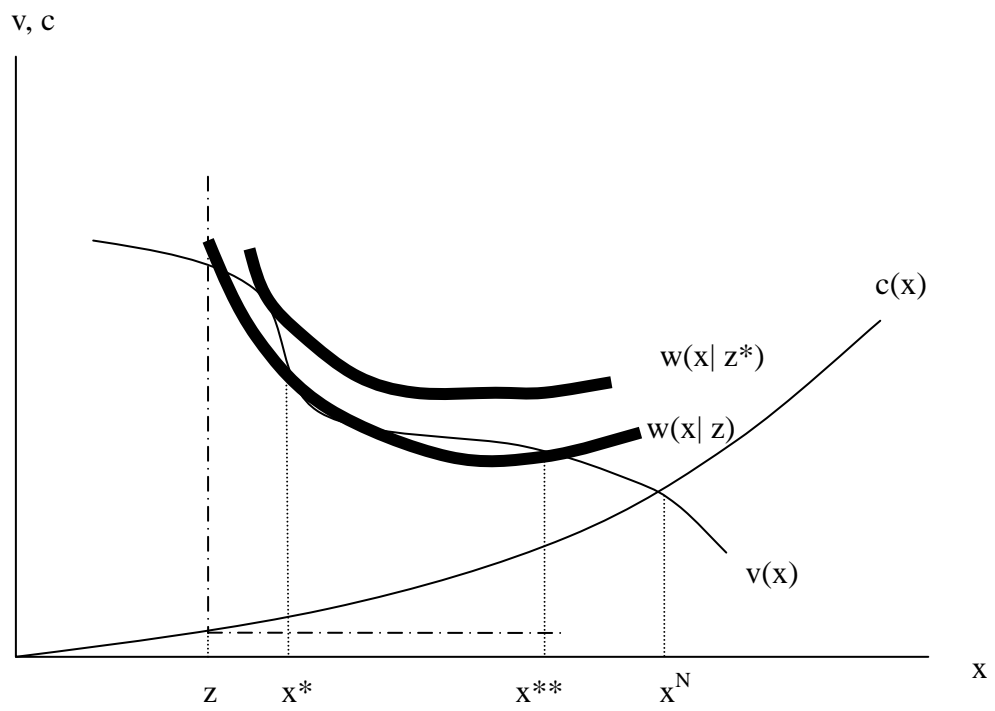


Figure 7: Equilibrium traffic volume correspondence at different values of the regulatory standard z when $p = p(e, r)$

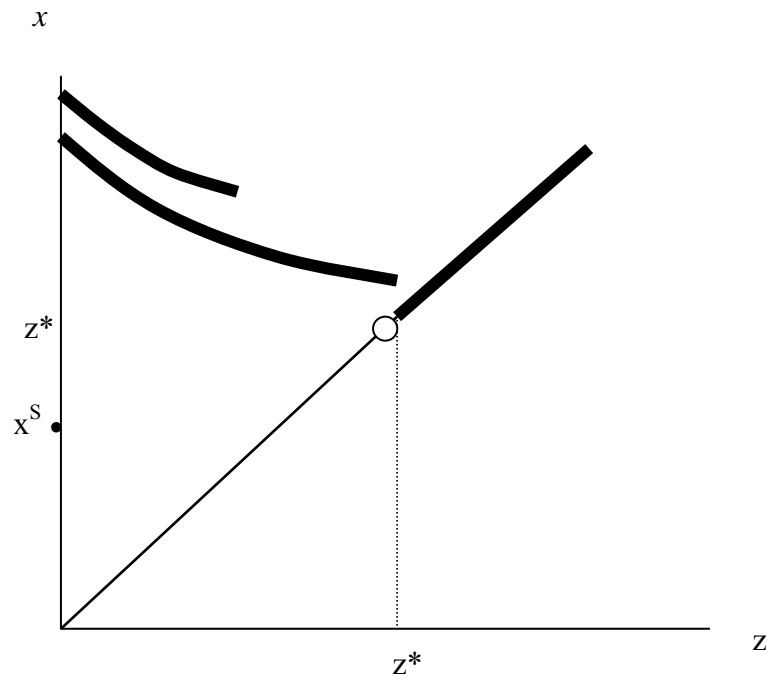


Figure 8: Strong enforcement

