2020

Unexpected Effects of Expected Sanctions

Giuseppe Dari-Mattiacci
Amsterdam Law School, gdarimat@uva.nl

Alex Raskolnikov
Columbia Law School, arasko@law.columbia.edu

Unexpected Effects of Expected Sanctions*

Giuseppe Dari-Mattiacci† and Alex Raskolnikov‡

December 19, 2020

Abstract

The economic analysis of law enforcement holds that greater expected sanctions lead to greater compliance. The literature on positive and negative incentives holds that rewards and sanctions—or carrots and sticks—have identical first-order incentive effects. We extend the basic model of law enforcement in three ways. We allow agents to opt out of the regulatory regime, we allow for enforcement errors, and we model agents who vary in at least one trait in addition to their cost of compliance. We show that following these three realistic modifications of the basic model, the two fundamental conclusions just described do not hold. Greater expected sanctions do not necessarily lead to greater compliance; carrots and sticks are not substitutes in their incentive effects. We also show that adding taxes and subsidies to the regulatory toolkit does not expand the set of achievable outcomes.

Keywords: deterrence, chilling, compliance, participation, accuracy, sanctions, rewards, carrots, sticks.
JEL codes: K10, K20, K42.

1 Introduction

Gary Becker's seminal article on the economic analysis of crime (Becker 1968) was not really about crime. Or at least not only about it. In the article's first paragraph Becker emphasized that he would consider not only murder, rape and similar felonies, but also discrimination, collusion, traffic safety, and "thousands of other activities." One page later he added tax evasion and white-collar offenses to the list. And by the time he turned to applications of his model, he quipped that "crime is apparently not so different analytically from any other activity that produces external harm[,] and when crimes are punishable by fines, the analytical differences virtually vanish" (Becker 1968: 201). Becker's model had a broad reach indeed.

The theory of optimal deterrence that grew out of Becker's foundational work embraced his ambition. That theory is also known as the theory of public enforcement of law (Polinsky and Shavell 2007)—a name that leaves no doubt that the theory's goal is to analyze *all* government

*We received valuable feedback from Scott Baker, Ryan Bubb, Gerrit De Geest, Ezra Friedman, Abe Wickelgren and workshop participants at Columbia University, New York University, University of Chicago, University of Texas, and the annual meetings of ComplianceNet and the American Law and Economics Association. We would also like to thank the editors and two anonymous referees for very constructive comments. All mistakes are ours.

†Professor of Law and Economics, University of Amsterdam Law School.
‡Wilbur H. Friedman Professor of Tax Law, Columbia Law School.

regulation. Yet, the theory's workhorse model serves as an adequate representation of only a small number of regulatory settings. The model's key limitation is that it incorporates one choice facing most regulated parties while largely ignoring another one.

The individual's decision at the center of the standard deterrence model is whether to comply with the legal rule or to violate it. Economic analysis of law enforcement has much to say on how the government can shape this decision. The most fundamental result is that higher expected sanctions reduce violations and, equivalently, improve compliance.

Yet, compliance and violation are not the only options available to individuals facing most regulatory regimes. The third option is to avoid participating in the regime altogether. The choice of whether or not to participate in a regulated activity has received much less attention from law-and-economics scholars. The analysis of how individuals optimize along the participation and compliance margins simultaneously is particularly thin.

Becker's own illustrations reveal the importance of addressing both choices. Consider his first example: discrimination, say housing discrimination to be concrete. The obvious goal of anti-discrimination laws is to induce compliance with these laws by landlords. But discrimination is difficult to define and to prove, so enforcement is never perfectly accurate. Facing a possible mistaken imposition of liability, some landlords may decide to exit rental real estate business, or may not enter it in the first place. Importantly, some of these landlords would have run a perfectly non-discriminatory, socially beneficial business. Their decision to exit or to abstain from entry is a social cost.

The same analysis applies to collusion, Becker's second example. It, too, is difficult to define and to prove. And here, too, a possible mistaken imposition of antitrust liability may induce some firms to exit concentrated industries where collusion prosecutions are more likely. Turning to traffic safety, speed limits are sometimes uncertain and are enforced with error (Brabender 2004). Fearing mistakenly imposed sanctions, some would-be drivers may switch to public transportation or just stay home. More examples are easily available, and they reveal two challenges faced by regulators in the vast majority of cases. The first challenge is to ensure compliance with legal rules by those who participate in a regulated activity. The second challenge is to induce the desired degree of participation in that activity. Ignoring the second challenge necessarily limits the usefulness of the model.[1]

The voluminous literature on the use of positive and negative incentives suffers from a similar failure to incorporate important features of real-world incentive schemes. The economic analysis of carrots and sticks has established long ago that the two instruments produce identical incentive effects (Ben Shahar and Bradford 2012). For example, both a subsidy to homeowners using energy conservation measures and a tax on homeowners who fail to use such measures induce energy conservation by home owners. In fact, carrots and sticks are so similar that some scholars argue that the two are actually one and the same (Gordon 1992).[2] This does not mean, of course, that real-world positive and negative incentives are fully interchangeable. Carrots are expensive; sticks

[1]We recognize that in some settings participation is not optional. As long as one belongs to the society, one is required to comply with its criminal laws. Even here, however, non-participation is possible. For instance, many white-collar offenses and many common crimes and misdemeanors do involve both the participation and the compliance decisions (Posner 1985, Raskolnikov 2020).

[2]Continuing with the energy conservation example, Congress can enact a tax on all homes and then add an offsetting energy conservation subsidy. Or Congress can enact no uniform home tax but, instead, adopt a tax only on homes that fail to use energy conservation measures. The first scheme features an energy conservation carrot; the second one features an energy conservation stick. But either way, owners of energy conserving homes end up neither better nor worse off, and owners of energy wasting homes end up paying more.

less so. Carrots induce participation; sticks discourage it (Wittman 1984). Carrots are given to all eligible claimants; sticks may never apply because individuals may act to avoid them (Dari-Mattiacci and De Geest 2010). Carrots work better to incentivize complex behaviors; sticks work well in simple settings (De Geest and Dari-Mattiacci 2013). These are just some of the known differences. But importantly, all these differences are predicated on the basic, fundamental insight that in terms of their incentive effects, carrots and sticks are identical. This insight, to the best of our knowledge, has never been questioned in a realistic setting where individuals have an option of staying (or opting) out of the regulatory regime altogether.

In this article, we show that the two fundamental results just described do not hold as a general matter. Higher expected sanctions do not always improve compliance (and, conversely, higher expected rewards do not always reduce violations), and carrots and sticks do not always have identical incentive effects. To be clear, the unexpected effect of sticks arise irrespective of the presence of carrots and the unexpected effects of carrots arise irrespective of the presence of sticks. Agent heterogeneity and enforcement errors are key to our analysis.[3]

1.1 An Example and Summary of the Results

To see why higher expected sanctions do not always improve compliance, consider a simple example. A principal enforces a regulation by using a sanction (or stick) equal to $s = 1$ for violators. The principal's enforcement is certain (every agent is inspected), but not perfectly accurate; she correctly assesses violations and compliance with probability $q = \frac{3}{4}$. Thus, there is a twenty five percent chance (probability $1 - q = \frac{1}{4}$) that the principal mistakenly concludes that a complying agent violated the rule or that a violating agent complied. Compliance requires agents to exert effort at a cost, e, that varies across agents.[4] Assume that e is uniformly distributed between 0 and 1.

If the agent complies, he bears the cost of compliance plus the expected cost of mistaken penalties, or $e + (1 - q)s = e + \frac{1}{4}$. If the agent violates, he bears the expected cost of correctly imposed penalties, or $qs = \frac{3}{4}$. The agent complies if the cost of compliance is less than the cost of violation, that is, if $e < \frac{1}{2}$. Given our assumption about the uniform distribution of e, half of all agents comply.

What if the principal increases the penalty from $s = 1$ to $s = 2$? The costs of compliance and violation become $e + \frac{1}{2}$ and $\frac{3}{2}$, respectively. With a higher expected penalty, an agent complies if $e < 1$, meaning that now all agents comply.[5] That compliance goes up with the magnitude of the (expected) sanction is the most fundamental tenet of the law enforcement theory. However, this fundamental conclusion turns out to hold only if the sole choice faced by the agent is whether to comply or to violate. If the agent can also choose whether to participate in the regime, the monotonic relationship between penalties and compliance breaks down.

To see why, let us now consider two subgroups in the population of agents: high-benefit agents, who derive a benefit $b_H = \frac{3}{2}$ from participation, and low-benefit agents who derive a benefit $b_L = \frac{1}{2}$ from participation. (In both subgroups, e is uniformly distributed between 0 and 1 as before.) If the agent does not participate in the regulated activity, his payoff is equal to 0. By subtracting the

[3]To be clear, The example in the next section demonstrates both points.

[4]The effort e can be interpreted as a foregone benefit of violation. In the literature on law enforcement, the benefit of violating the rule is usually denoted by g and varies among individuals, as e varies in our framework. The two narratives are interchangeable.

[5]That is, compliance yields lower costs than violation if $e + \frac{1}{2} < \frac{3}{2}$.

cost of compliance or violation from the benefit of participation for each subgroup of agents we obtain the payoffs reported in Table 1.[6]

	Compliance	Violation	No participation
High-benefit agents	$\frac{5}{4} - e$	$\frac{3}{4}$	0
Low-benefit agents	$\frac{1}{4} - e$	$-\frac{1}{4}$	0

Table 1: Payoffs with $s = 1$.

Among the three available options, each agent chooses the one associated with the highest payoff. All high-benefit agents participate because the payoff of both compliers and violators is greater than 0. Thus high-benefit agents choose—as before—between compliance and violation, and comply if $e < \frac{1}{2}$.

In contrast, a low-benefit agent earns a negative payoff if he participates and violates the rule. Hence only those low-benefit agents who would comply if they participated may decide to participate. A low-benefit agent's payoff from compliance is positive if $e < \frac{1}{4}$. Therefore, only $\frac{1}{4}$ of the low-benefit agents participate and comply, while $\frac{3}{4}$ of them—those with higher cost of participation $e \geq \frac{1}{4}$—do not participate.

What if the principal increases the penalty from $s = 1$ to $s = 2$? The parties' payoffs will change, and the new payoffs are summarized in Table 2. As before, all high-benefit agents choose to participate, but now all of them comply—a higher penalty induces more compliance among these agents. The standard deterrence result still holds. The response of low-benefit agents is very different, however. Facing a higher penalty, all low-benefit agents choose not to participate. This change reduces the number of complying agents compared to the low-penalty setting.

	Compliance	Violation	No participation
High-benefit agents	$1 - e$	0	0
Low-benefit agents	$-e$	-1	0

Table 2: Payoffs with $s = 2$.

The reason for these diverging responses is straightforward. An increase in the penalty reduces the payoff from violation more than it reduces the payoff from compliance (compare Table 1 with Table 2). This change induces more high-benefit agents to comply, as the deterrence theory predicts. Low-benefit agents, however, do not compare compliance with violation—they compare compliance with non-participation. The payoff from non-participation is fixed at 0 no matter what the penalty is, but enforcement errors reduce the payoff from compliance when the penalty increases. Given this lower payoff, fewer agents choose to participate and comply.

More generally, an increase in the penalty affects the compliance and the participation decisions differently. Depending on the relative proportions of high- and low-benefit agents in the population, increasing the penalty may lead to an overall reduction, rather than an increase, in the number of complying agents. In the example, if more than $\frac{2}{3}$ of all agents are low-benefit types, the decline in the number of compliers in the (relatively large) low-benefit group more than offsets

[6]For instance, the payoff of a complying high-benefit agent is equal to $b^H - [e + (1-q)s] = \frac{3}{2} - \left(e + \frac{1}{4}\right) = \frac{5}{4} - e$. The other payoffs are calculated in an analogous way.

the increase in the number of compliers in the (relatively small) high-benefit group. Of course, a higher penalty also decreases the number of violators in the high-benefit group. But the reduction in compliers may exceed the reduction in violators. If that happens, a familiar policy of increasing penalties would result in an unfamiliar outcome of reduced compliance.[7] Incorporating imperfect detection into the model does not change these conclusions.

Our second result modifies the long-held view that carrots and sticks have identical incentive effects. That view has persisted because the rewards-versus-sanctions literature rarely considers the possibility of avoiding both the sticks and the carrots by staying out of the incentive scheme altogether. In fact, we believe that scholars have not considered the combined effects of these complications at all. In reality, however, participation in schemes based on positive incentives is often optional and imperfect enforcement is widespread.

For example, the energy conservation subsidy for home owners (known as the non-business energy property tax credit, Internal Revenue Code, sec. 25C) is among many energy-related carrots in the Internal Revenue Code (Joint Committee on Taxation 2009). That subsidy is imperfectly enforced both because some energy saving measures are difficult to observe (for instance, the type of insulation installed inside house walls) and because the legal definition of qualifying conservation measures is vague (insulation materials and systems qualify if they are "specifically and primarily designed to reduce the heat loss or gain of a dwelling unit," Internal Revenue Code, sec. 25C(c)(3)). Improper claiming of credits may be sanctioned (Internal Revenue Code, sec. 6662). And obviously, home ownership is not mandatory.

We show that when individuals have a choice of avoiding the carrots-and-sticks scheme completely, changes in the magnitude of carrots and sticks have ambiguous effects on the number of compliers and violators. The standard result in the literature is that higher carrots increase participation in the activity (Wittman 1984). In our model, this result holds. But we show that if the carrot is large enough, some of the additional participants violate the rule rather than comply with it, hoping to capture the carrot if the regulator makes a mistake. In fact, it is possible that higher participation in response to larger carrots comes mostly from violators opting in.

It is easy to see that if we replace a stick with a carrot of the same magnitude in the example above (that is, change $s = 1$ to $c = 1$), the behavior of high-benefit agents does not change. All these agents still participate and half of them comply.[8] In contrast, while no low-benefit agent finds it advantageous to participate and violate the rule if $s = 1$,[9] half of the low-benefit agents would choose to participate and violate if $c = 1$.[10] The expected consequence of replacing sticks with carrots (here, going from $s = 1$ to $c = 1$) is increased participation. The unexpected result is

[7] Denote by h the fraction of high-benefit agents. With the original penalty of $s = 1$, total number of compliers is $\frac{1}{2}h + \frac{1}{4}(1-h) = \frac{1+h}{4}$. With the increased penalty of $s = 2$, the total number of compliers is equal to h, as only the high-benefit agents comply and all of them do so. Therefore, increasing the penalty from 1 to 2 reduces the number of compliers if $h < \frac{1+h}{4}$ that is, if $h < \frac{1}{3}$ and, hence, less than $\frac{1}{3}$ of the agents are high-benefit types while more than $\frac{2}{3}$ are low-benefit types.

[8] If $c = 1$, the payoff of a complying high-benefit agent is equal to $b_H + qc - e = \frac{3}{2} + \frac{3}{4} - e = \frac{9}{4} - e$, while the payoff of a violating high-benefit agent is $b_H + (1-q)c = \frac{3}{2} + \frac{1}{4} = \frac{7}{4}$. Comparing the two payoffs, we see that a high-benefit agent complies if $e < \frac{1}{2}$ and violates if $e > \frac{1}{2}$. This outcome is identical to the high-benefit agents' decisions with $s = 1$.

[9] Recall that with $s = 1$, $\frac{1}{4}$ of the low-benefit agents choose to participate and comply, while $\frac{3}{4}$ of them—those with higher cost of participation $e \geq \frac{1}{4}$—do not participate at all.

[10] The payoff for a complying low-benefit agent is equal to $b_L + qc - e = \frac{1}{2} + \frac{3}{4} - e = \frac{5}{4} - e$, while the payoff for a violating low-benefit agent is $b_L + (1-q)c = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$. Comparing these two payoffs we see that a low-benefit agent participates and complies if $e < \frac{1}{2}$ and participates and violates if $e > \frac{1}{2}$.

that most of the additional participants claim carrots improperly.[11]

Once it is understood that larger sticks may result in fewer compliers and larger carrots may result in more violators, it becomes clear that the two instruments are not necessarily substitutes, contrary to what is typically assumed. We demonstrate that carrots and sticks are indeed substitutes in the agents' choice between compliance and violation. But the two are complements in the agents' participation decision. While it is known that sticks deter participation and carrots encourage it,[12] no attention has been paid to the fact that carrots and sticks also select the type of agents who choose to participate: whether they are compliers or violators. This selection has an effect on the overall numbers of compliers and violators and on the compliance rate.

We show that the overall effect of carrots and sticks on the number of compliers and violators depends on the relative proportions of high-benefit agents—who make the compliance choice—and low-benefit agents—who make the participation choice—in the population. As is true for our first unexpected result, admitting a probability of detecting violations below one does not change this conclusion.

Finally, we show that the effects of carrots and sticks that we identify cannot be changed by introducing taxes and subsidies. A tax amounts to simultaneously increasing the stick and reducing the carrot by a fixed amount (and vice-versa for a subsidy) without expanding the set of outcomes that the regulator may achieve. Adding more instruments to the policy mix does not make a difference in terms of achievable results. To control two margins—the participation and compliance decisions—it does not help to use more than two instruments—a carrot and a stick.

1.2 Related Literature

Our results relate to a large literature on law enforcement, but we build on a very small number of key contributions that considered deterrence in the presence of errors and optional participation. Enforcement errors play an important role in our model. Without errors, compliers would be indifferent to changes in the stick and violators would be indifferent to changes in the carrot. Png (1986) was the first to consider the effect of errors in a law enforcement model where agents made both a compliance and a participation decision. He identified a negative effect of errors on participation and argued that carrots can be used to complement sticks to counter that effect. [13]

Kaplow and Shavell (1994) were the first to formally study the interaction between accuracy and sanctions in the legal system. Kaplow (2011) introduced chilling in the model of deterrence with imperfectly accurate law enforcement. Chilling is closely related to the decision of would-be compliers to abstain from participation in our model. Kaplow (2011) studied how the possibility of a mistaken imposition of sanctions affects the optimal burden of proof. Friedman and Wickelgren (2008) showed that the tradeoff between deterrence and chilling may make settlement of legal disputes welfare-reducing. While these contributions provide important foundations for our inquiry,

[11] Specifically, a change from $s = 1$ to $c = 1$ increases the number of participating low-benefit compliers from $\frac{1}{4}$ to $\frac{1}{2}$, while increasing the number of participating low-benefit violators from zero to $\frac{1}{2}$. Recall that the behavior of high-benefit agents does not change.

[12] This point originates with Wittman (1984). He did not consider, however, the interplay between compliance, violations, and participation, which is the focus of our analysis.

[13] Levmore (1986) also considered the joint use of carrots and sticks. He observed that a regulator trying to incentivize behavior with carrots may need to introduce sticks to prevent false claiming of carrots. That is precisely the problem we study here. Levmore's (1986) observation was brief, however. He did not offer a model, and did not analyze any heterogeneity among the agents as we do.

none of them considers heterogeneity in the agents' participation decisions, which is crucial to our analysis.

More generally, the literature on the economics of law enforcement mostly addresses the question of deterrence—that is, the choice between compliance and violation—conditional on participation. The literature has considered the participation decision in order to determine whether individuals have incentives to engage in the optimal level of activity in threshold-based (also known as negligence or fault-based) and strict liability regimes, conditional on individuals complying with the optimal rule (Polinsky and Shavell 2007: 425). In contrast, we investigate how individuals choose along both the participation and the compliance margin at the same time.[14]

A large literature on carrots and sticks mostly views them as interchangeable in their incentive effects (Gordon 1992, Ben-Shahar and Bradford 2012). Only few contributions stress the differences between the two, and even these contributions mostly retain the assumption that policymakers use either carrots or sticks (De Geest and Dari-Mattiacci 2013, Polinsky 1979, Wittman 1984). Authors considering the joint use of carrots and sticks study the effect of the resulting incentives on a single margin (Andreoni 2003; Ben Shahar and Bradford 2012; Gilpatric 2009; Moldovanu 2012). In contrast, we study the effect of combining rewards and penalties to affect two separate decisions: whether or not to participate and, conditional on participation, whether or not to comply. Dari-Mattiacci (2009) investigates how agents make decisions along both of these margins but he studies the use of a single incentive. Baker and Malani (2019) also consider both margins and show that type-one and type-two errors have different effects on welfare of third parties relying on the regulator's certification of compliance. The regulator in their model can choose the error rate, and market prices replace sanctions as negative consequences. Kleven and Kopczuk (2011) analyze government benefits (carrots) jointly with variations in complexity, which they model as costly increases in enforcement accuracy. They do not include sticks, however, and the cost of greater accuracy born by agents plays a key role in their model while it is absent from ours.

The plan for the remainder of this article is as follows. In Section 2 we introduce our general model and demonstrate our main results in a simplified version of that model while leaving general proofs for the Online Appendix. In Section 3 we show that adding taxes and subsidies adds nothing to our results. In Section 4 we discuss our assumptions of optional participation, available carrots, and heterogeneous agents. We also highlight the main analytical and policy payoffs from our results. Section 5 concludes.

2 The model

2.1 Model Setup

Consider a principal regulating an activity in which any agent can elect to participate. Each agent chooses independently whether to participate in the activity and, if so, whether to comply with the regulation. (Those who do not participate are not subject to the regulation.)[15] Agents derive a fixed

[14] The main concern of the economic analysis of accident law is that it is impossible to incentivize optimal care (that is, compliance) and optimal activity level (that is, participation) by two parties—the tortfeasor and the victim—through ordinary liability rules (Shavell 1980). Here we are concerned with the behavior of a single agent in a population of heterogeneous agents and the main problem is that the policy is not tailored to the agent-specific characteristics. The literature on torts has paid limited attention to this problem, mainly with respect to heterogeneity in the cost of care (Ganuza and Gomez 2006).

[15] In Section 3, we show that our results also apply when the principal directly regulates participation through taxes and subsidies, which are not conditional on agents' compliance.

benefit from participation $b > 0$ and face a fixed cost of compliance $e > 0$.[16] Costs and benefits vary across agents, so that the population of agents can be fully described by the probability distribution $F(b,e)$ of the agents' two-dimensional types (b,e). Agents are risk neutral and maximize their expected payoffs.

To induce agents to participate and comply, the principal rewards compliance with a carrot $c > 0$ while punishing violations with a stick $s > 0$.[17] However, the principal cannot perfectly distinguish compliers from violators. The principal correctly identifies compliers with probability $\frac{1}{2} < q_k < 1$ and, with the complementary probability $1 - q_k$, the principal erroneously classifies compliers as violators (a false positive, akin to convicting the innocent). Likewise, the principal correctly identifies violators with probability $\frac{1}{2} < q_v < 1$ and, with probability $1 - q_v$, the principal erroneously treats violators as compliers (a false negative, akin to acquitting the guilty).[18] The principal monitors participating agents with probability $0 < p \leq 1$.

2.2 Simplifying Assumptions

To make our points in the starkest possible way, we make a number of simplifying assumptions. First, we set $p = 1$, that is, we assume that monitoring takes place with certainty. Second, we assume that for all agents the cost of compliance e is uniformly distributed between 0 and 1.[19] Third, we assume that the principal is equally accurate in detecting compliance and violations, that is

$$q_k = q_v = q$$

and therefore $\frac{1}{2} < q < 1$. Finally, we assume that the population of agents is divided into two groups. A portion h of the agents are high-benefit types and they derive a benefit b_H from participation; the remaining portion $1 - h$ of the agents are low-benefit types and they derive a benefit b_L from participation. For these two groups, we assume that in the baseline case

$$b_H > p\left(qs - (1-q)\,c\right) \geq b_L \tag{1}$$

This assumption, as will be clarified below, guarantees that in the baseline case all high-benefit agents choose to participate in the regulated activity while low-benefit agents participate only if their cost of effort is low enough.[20] In the Online Appendix, we relax all these simplifying

[16] Non-participants earn no benefit and incur no cost. Some activities are costly to the agent; that is, we could allow b to be negative. Setting a zero lower bound for b, however, is without loss of generality. Allowing for negative benefits for participants is equivalent to setting a positive benefit for non-participants, which would not affect our results.

[17] We define carrots and sticks in an unambiguous way: a carrot is a monetary payment from the principal to the agent conditional on (detected) compliance, while a stick is a monetary payment from the agent to the principal conditional on (detected) violation. In the Online Appendix we generalize our setup (and hence this definition) to include cases in which the probability of monitoring, p, is less than 1 and non-monitored agents may receive a carrot or pay a stick so that only part of the carrot or the stick is conditional on detection.

[18] Positing that both probabilities are greater than $\frac{1}{2}$ simply restricts attention to the plausible scenarios where the principal's enforcement technology fares better than a coin toss.

[19] Note that, as a result of this assumption, we impose independence between the distributions of b and e. In the Online Appendix we will not assume independence between b and e.

[20] Note that this setup describes a realistic—arguably the only realistic—setting. If the first inequality does not hold, the benefit of high-benefit agents is so low that they earn a negative payoff from participating and violating the rule. In that case, all participating agents comply—even in the face of imperfect enforcement. If the second inequality does not hold, the benefit of low-benefit agents is so high that even non-complying agents participate. In this case, every single agent (both high- and low-benefit ones) participates—even though participation is not mandatory. Thus, only if both parts of Equation (1) hold does the model reflect a realistic scenario where some agents participate and comply, some agents participate and violate, and some agents stay out of the optional, imperfectly enforced regime.

assumptions, show that the results remain unchanged, and demonstrate some additional results. (In particular, relaxing the assumption that $p = 1$, shows that the model is robust to all possible treatments of non-monitored agents, including settings in which the carrot is paid to all agents — and disgorged if agents are found violating — and settings in which the carrot is paid only to monitored agents upon verification of compliance.)

2.3 Analysis

Given the general setup in Section 2.1 and the simplifying assumptions in Section 2.2, an agent of type (b, e) anticipates the payoffs reported in Table 3.

Action	Payoff
The agent participates and complies	$\Pi_k (b, e) = b + qc - (1 - q) s - e$
The agent participates but violates	$\Pi_v (b) = b + (1 - q) c - qs$
The agent does not participate	0

Table 3: Agent's payoffs

Starting with the bottom row, if the agent does not participate, his (normalized) payoff is zero. If he participates but violates the principal's rule of conduct, the agent earns the benefit b, faces a stick s with probability q (that is, if the principal correctly qualifies the agent's conduct as a violation of the rule), and faces a carrot c with probability $1 - q$ (that is, if the principal erroneously qualifies the agent's conduct as compliance). The agent incurs no cost of effort. We denote the payoff of an agent who participates but violates the rule as $\Pi_v (b)$. Finally, if the agent participates and complies, he earns the benefit b, faces a carrot with probability q (that is, if the principal correctly qualifies his conduct as compliance), faces a stick with probability $1 - q$ (that is, if the principal erroneously qualifies his conduct as a violation), and exerts costly effort. We denote the payoff of an agent who participates and complies as $\Pi_k (b, e)$.

We can now characterize the agents' behavior, starting from high-benefit agents. Assuming that an agent participates, he complies if the payoff from compliance is greater than the payoff from violation, that is, if $\Pi_k (b_H, e) > \Pi_v (b_H)$. This is the case if the agent's cost of effort is lower than a threshold value that we call the *compliance threshold*

$$e_k \equiv (c + s) (2q - 1) \tag{2}$$

Note that, given the agent's choice between compliance and violation, the payoff of complying agents must be at least as high as the payoff of violating agents. Therefore, the lowest payoff that a participating agent may earn is $\Pi_v (b_H)$. Given the assumption in Expression (1), the payoff of high-benefit agents who violate is positive, $\Pi_v (b_H) > 0$, and hence all high-benefit agents participate in the activity irrespective of whether they comply or violate. Summing up, a portion e_k of the high-benefit agents participates and complies, while the remaining portion $1 - e_k$ participates and violates.[21]

Turning to low-benefit agents, note that by the assumption in Expression (1) we have $\Pi_v (b_L) \leq 0$, hence not all low-benefit agents participate in the activity. None of the would-be violators and

[21] This is due to the assumption that e is uniformly distributed between 0 and 1, so that the share of agents with $e < e_k$ is exactly e_k and, conversely, the share of agents with $e \geq e_k$ is $1 - e_k$.

some of the would-be compliers stay out. Agents who participate must therefore be compliers and must earn a positive payoff $\Pi_k(b_L, e) > 0$. This is the case if the agent's cost of effort is lower than a threshold value that we call the *participation threshold*

$$e_p \equiv b_L + qc - (1-q)s \tag{3}$$

To sum up, a portion e_p of the low-benefit agents participates and complies, while the remaining portion $1 - e_p$ refrains from participation. Figure 1 below offers a visualization of the agents' behavior in the baseline case when the assumption in Expression (1) holds.

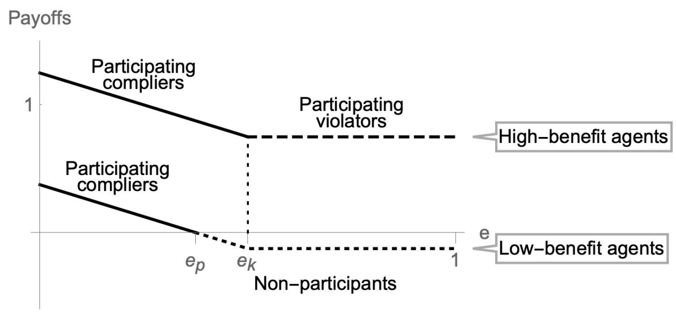


Figure 1: The baseline case (Simulation values: $c = s = \frac{1}{2}$, $q = \frac{3}{4}$, $b_H = 1$, $b_L = \frac{1}{8}$)

2.4 Results

Having specified the behavior of high- and low-benefit agents, we begin by considering whether under any conditions our model yields the well-known results from the theory of deterrence and the economic analysis of carrots and sticks. Unsurprisingly, it does. The basic model of deterrence rests on the implicit assumption that all agents (must) participate in the regulated activity. In our simplified model, this approach is equivalent to focusing on the behavior of high-benefit agents whose choices are controlled by the compliance threshold e_k defined in Expression (2). The only choice that these agents have is between compliance and violation. A change in any enforcement parameter that reduces the number of violators necessarily increases the number of compliers. An increase in the stick (sanction) is one such change. Moreover, the compliance threshold e_k depends on the sum of the carrot c and the stick s. So the two are interchangeable in how they affect the decisions of high-benefit agents. Thus, if we focus only on these agents, expected sanctions do indeed have expected effects:

1. Compliance and violations are complementary outcomes: an increase in compliance is mirrored by a decrease in violations and vice versa.
2. Higher expected sanctions always increase compliance: They increase the number of compliers, reduce the number of violators, and increase the compliance rate—that is, the number of complying agents over the total number of participants.
3. Carrots and sticks are substitutes: as long as their sum stays constant, any combination of c and s results in the same level of e_k and hence in the same level of compliance and violations.

Considering the behavior of low-benefit agents—in addition to the behavior of high-benefit agents—reveals the unexpected effects of expected sanctions.

2.4.1 Compliance and Violations are not Complementary Outcomes

Consider our baseline case when the assumption in Expression 1 holds, as illustrated in Figure 1. Focusing on the behavior of low-benefit agents, note that it is controlled by the participation threshold e_p, not the compliance threshold e_k. At the margin, low-benefit agents do not choose between compliance and violation; they choose between compliance and non-participation. So for these agents compliance and violation are clearly not complementary outcomes. As clearly, any policy evaluation of a setting similar to our baseline must consider all three possible outcomes—compliance, violation, and participation.

This discussion highlights a critical ambiguity of the term "compliance." One possible meaning of this term is the number of complying agents. Another meaning, however, is the share of complying agents among all participating agents—compliance as compliance rate. In any setting where changes in enforcement parameters may change the number of participants, these two meanings are by no means equivalent. An increase in the number of compliers may coincide with an increase or a decrease in the rate of compliance, depending on the change in the number of violators. This complication does not exist in the traditional law enforcement setting where the number of participants is fixed. In contrast, in the more general model that we study here, a change in the number of compliers may lead to a change in both the number of violators and the number of participants. Compliance and violations are no longer the only two possible, complementary outcomes.

2.4.2 Increasing the Stick May Reduce Compliance

An increase in the stick induces some high-benefit violators to start complying. But a larger stick also induces some complying low-benefit agents to opt out of the regulated activity. If there are enough low-benefit agents in the population, the decrease in participation of complying low-benefit agents will dominate the increase in compliance among high-benefit agents. The ambiguous effect of sticks on compliance comes from the fact that an increase in sticks affects the compliance threshold e_k and the participation threshold e_p in opposite ways.

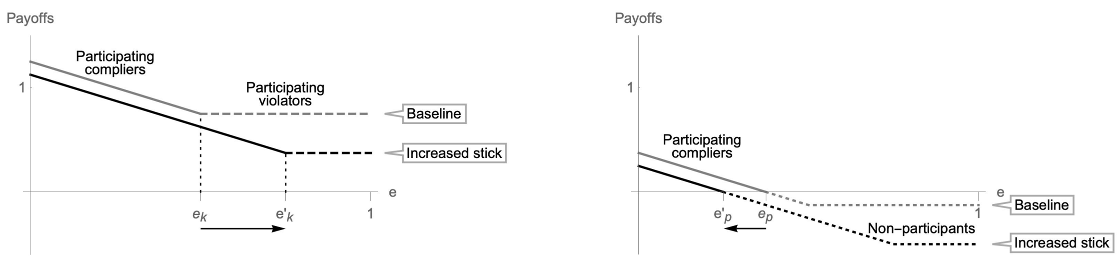
Formally, if the stick is increased from s to $s' > s$ and the assumption in (1) remains satisfied—that is, if the changes are marginal[22]—compliance among high-benefit agents increases from $e_k = (c+s)(2q-1)$ to $e_k' = (c+s')(2q-1)$. The increase in the number of compliers in the high-benefit group is equal to $e_k' - e_k = (s'-s)(2q-1)$. In contrast, the number of compliers in the low-benefit group declines from $e_p = b_L + qc - (1-q)s$ to $e_p' = b_L + qc - (1-q)s'$, with a reduction equal to $e_p - e_p' = (1-q)(s'-s)$. Overall, the number of compliers in the population decreases if the latter effect dominates the former, that is if $h(e_k' - e_k) < (1-h)\left(e_p - e_p'\right)$, which

[22]By marginal changes we mean changes in sticks for which the assumption that $b_H > p(qs' - (1-q)c) \geq b_L$ in Expression (1) continues to hold. This guarantees that the high-benefit agents still fully participate (their payoff line in Figure 1 is completely above 0) and the low-benefit agents participate only if they also comply (their payoff line in Figure 1 is below 0 for high levels of e). We focus on marginal effects because they generate the interesting ambiguity in the effect of sticks: they may increase or reduce the number of compliers. This would not be the case if we considered "inter-marginal" effects. The inter-marginal effect of sticks reveals itself in the possibility that high-benefit agents may decide not to participate—that is, an increase in s may imply $p(qs' - (1-q)c) \geq b_H > b_L$—which in turn unambiguously reduces the number of compliers. Effectively, an increase in the stick that is large enough to violate the assumption in (1) in the direction just described turns high-benefit agents into low-benefit ones. In the general model in the Online Appendix, both the marginal effects (on the compliance and participation thresholds) and the inter-marginal effects (on which threshold controls behavior) are continuous because we consider a continuum of type. In the simplified model presented in the text above, the inter-marginal effects are discrete because we only have two types.

is the case if there are enough low-benefit agents in the population, that is if[23]

$$h < \frac{1-q}{q}$$

In sum, larger sticks (higher sanctions) clearly reduce the number of violators. However, larger sticks may also reduce the number of complying agents if the share of low-benefit agents is large enough. So the first unexpected effect of expected sanctions is that higher sanctions may reduce—rather than increase—the number of compliers. We show below that greater sanctions may lead to a lower compliance rate as well. Figure 2 illustrates the effect of doubling the stick compared to the baseline of Figure 1. This change results in a decrease in the total number of complying agents if $h < \frac{1}{3}$, that is if low-benefits agents make up for more than the two thirds of the population.



(a) High-benefit agents

(b) Low-benefit agents

Figure 2: Increased stick (from $s = \frac{1}{2}$ to $s' = 1$, with $c = \frac{1}{2}$, $q = \frac{3}{4}$, $b_H = 1$, $b_L = \frac{1}{8}$).

2.4.3 Increasing the Carrot May Increase Violations

An increase in carrots induces high-benefit agents to reduce violations, just as an increase in sticks does. Larger sticks reduce the expected payoff from violations; larger carrots increase the expected payoff from compliance.[24] An increase in carrots, however, may also induce broader participation by low-benefit agents.

In particular, a larger carrot may induce all[25] non-participating low-benefits agents to participate, effectively turning them into high-benefit agents, an effect that we call "inter-marginal".[26] Some of these former low-benefit agents will participate and comply; others will participate and violate. If there are enough low-benefit agents in the population, and if enough of them violate once they choose to participate, the resulting increase in violators would dominate the reduction in violators in the high-benefit group.[27]

[23]The inequality in the text can be rewritten as $h < \frac{(1-q)(s'-s)}{(2q-1)(s'-s)+(1-q)(s'-s)} = \frac{1-q}{q}$.

[24]These results follow directly from Expression 2 for the compliance threshold.

[25]In the Online Appendix, where we allow for a continuous distribution of the benefit b in the population, only some of the low benefit agents will be induced to fully participate.

[26]The change is inter-marginal because low-benefit agents go from optimizing along the extensive margin (participation) to optimizing along the intensive margin (compliance).

[27]Given our simplifying assumption of only two agent types, such inter-marginal change would lead to an implausible

Formally, if the increase in the carrot is large enough to violate the assumption in (1),[28]—that is, if the increase leads to an inter-marginal change in behavior—all low-benefit agents will participate, becoming new high-benefit agents.[29] Some of these new participants will violate the rule. For that group, the number of violators will increase from zero to $1 - e'_k = 1 - (c' + s)(2q - 1)$. At the same time, a marginal effect of a larger carrot is to decrease violations among the original high-benefit agents. If the carrot is increased from c to $c' > c$, violations in this group decrease from $1 - e_k = 1 - (c + s)(2q - 1)$ to $1 - e'_k = 1 - (c' + s)(2q - 1)$. Consequently, the reduction in violations in the original high-benefit group is equal to $e'_k - e_k = (c' - c)(2q - 1)$. Overall, violations in the population increase if the former effect dominates the latter, that is if $h\left(e'_k - e_k\right) < (1 - h)\left(1 - e'_k\right)$, which is the case if there are enough low-benefit agents in the population, that is if[30]

$$h < \frac{1-(c'+s)(2q-1)}{1-(c+s)(2q-1)}$$

Figure 3 illustrates the effect of doubling the carrot as compared to the baseline case depicted in Figure 1. In this example, the policy results in a increase in violations if $h < \frac{1}{2}$, that is if low-benefits agents make up for more than the half of the population.[31]

2.4.4 Increasing the Carrot or the Stick May Reduce the Compliance Rate

We have shown that increasing the stick may result in fewer compliers and increasing the carrot may result in more violators. Yet, these effects do not necessarily lead to a lower compliance rate. Sticks reduce violations, and hence the compliance rate falls only if the reduction in the number of compliers outweighs the decline in violations. Similarly, carrots increase the number of compliers and hence the compliance rate falls only if the increase in violations outweighs the increase in the number of compliers. Therefore, the unexpected effect of sticks or carrots must be large enough for the compliance rate to fall following an increase in the stick or the carrot. We show here that this is indeed a possibility.

---

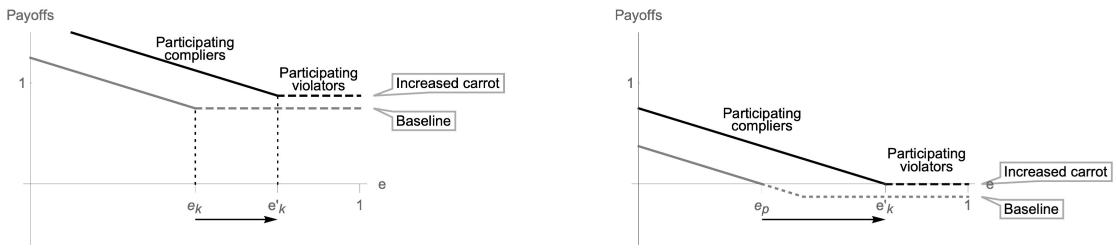
outcome when all agents participate, even though participation is not mandatory and erroneous sanctions are possible. In the general model presented in the Online Appendix we consider a continuous distribution of agents' benefits. In that realistic case, an inter-marginal change turns some low-benefit agents (those with relatively high benefits) into high-benefit ones, while other low-benefit agents remain in the low-benefit group.

[28] That is if c' if such that $b_L > qs - (1 - q)c'$.

[29] An increase in the carrot may have both marginal and inter-marginal effects, just as an increase in the stick may, as we pointed out in footnote 22. With sticks, it is the marginal effect that leads to interesting results. With carrots it is the inter-marginal effect. The marginal effect of higher carrots is the change in behavior within groups: high-benefit agents react to a larger carrot by complying more and violating less (due to the increase in e_k), while low-benefit agents react to a larger carrot by participating and complying more (due to the increase in e_p). These effects are straightforward and do not yield surprising results. The overall outcome is more compliers and fewer violators. The inter-marginal effect of a larger carrot is that low-benefit agents may be induced to behave like high-benefit agents—that is, to fully participate. When this happens, the controlling threshold for low-benefit agents changes from e_p—which controls behavior if agents who do not fully participate—to e_k—which controls the behavior of agents who fully participate.

[30] The inequality in the text can be rewritten as $h < \frac{1-\left(c'+s\right)(2q-1)}{(c'-c)(2q-1)+1-(c'+s)(2q-1)} = \frac{1-\left(c'+s\right)(2q-1)}{1-(c+s)(2q-1)}$.

[31] With the parameter values used in the example in Figure 3 the assumption in (1) is satisfied as an equality, therefore the example should be interpreted as showing the effect of increasing the carrot from $\frac{1}{2}$ to just above 1. Summing up, the unexpected effect of sticks is due to the marginal effect of sticks on the participation threshold e_p, while the unexpected effect of carrots arises because some agents who previously reacted to e_p start reacting to e_k. In the general model discussed in the Online Appendix, we dispose of our simplifying assumptions and allow marginal and inter-marginal effects to play out simultaneously. Our results remain unchanged and the logic governing the general results is the same as explained here.



(a) High-benefit agents

(b) Low-benefit agents

Figure 3: Increased carrot (from $c = \frac{1}{2}$ to $c' = 1$, with $s = \frac{1}{2}$, $q = \frac{3}{4}$, $b_H = 1$, $b_L = \frac{1}{8}$).

In the baseline case, the compliance rate is equal to $\frac{he_k+(1-h)e_p}{h+(1-h)e_p}$, where the numerator is the total number of compliers in each group—that is, e_k for the high-benefit agents and e_p for the low-benefit agents—multiplied by the size of the relevant group—that is, h and $1-h$, respectively. The denominator is similarly the total number of participants in each group—that is, 1 for the high-benefit agents and e_p for the low-benefit agents—multiplied by the size of the relevant group—that is, again, h and $1-h$, respectively. We need to compare this baseline compliance rate with the compliance rate that obtains after increasing the stick or the carrot.

We start with carrots. Increasing the carrot to $c' > c$ results in a compliance rate equal to $e'_k > e_k$ because, as we have observed above, now all agents fully participate and choose to comply if $e < e'_k$. Therefore, the compliance rate is the number of compliers, e'_k, over the total size of the population, which is 1. The compliance rate falls as a result of the increase in the carrot if $e'_k < \frac{he_k+(1-h)e_p}{h+(1-h)e_p}$, which can be rewritten as

$$h < \frac{e_p\left(1-e'_k\right)}{e_p\left(1-e'_k\right)+e'_k-e_k} \tag{4}$$

In the example underlying Figures 1 and 3, we have $e_k = \frac{1}{2}$, $e_p = \frac{3}{8}$, and $e'_k = \frac{3}{4}$. Plugging these values in (4), we have the following condition: $h < \frac{3}{11}$. Therefore, in the example, if $h < \frac{3}{11}$ increasing the carrot results in an increase in violations and a decrease in the compliance rate. Note that if $\frac{3}{11} < h < \frac{1}{2}$, increasing the carrot results in an increase in violations and an increase (rather than a decrease) in the compliance rate. This is because, the increase in violations in the low-benefit group is more than compensated by an increase in the number of compliers in both groups in this case. Finally, if $h > \frac{1}{2}$, violations decrease and the compliance rate increases with larger carrots.

We can now turn to sticks. The logic is similar, while the analysis is slightly more involved. As we have shown, increasing the stick from s to s' results in an increase in the compliance threshold from e_k to e'_k and a decrease in the participation threshold from e_p to e'_p. The compliance rate increases with sticks if $\frac{he'_k+(1-h)e'_p}{h+(1-h)e'_p} < \frac{he_k+(1-h)e_p}{h+(1-h)e_p}$, that is, if

$$h < \frac{e_p\left(1-e'_k\right)-e'_p\left(1-e_k\right)}{e_p\left(1-e'_k\right)-e'_p\left(1-e_k\right)+e'_k-e_k} \tag{5}$$

Note that the latter condition is within the feasible range of h (which has to be positive) only if $e_p\left(1-e'_k\right) > e'_p\left(1-e_k\right)$, which is not the case in our running example.[32] This shows that larger sticks may result in fewer compliers and yet not in a decrease in the compliance rate. Changing the example slightly reveals the case where an increase in the stick leads to both fewer compliers and a lower compliance rate. In this case, higher sanctions harm compliance in both senses of that term.[33]

2.4.5 Carrots and Sticks are not (Perfect) Substitutes

The analysis in the previous two sections makes it clear that carrots and sticks are not perfect substitutes in our setting. The two instruments are indeed substitutes for high-benefit agents making a marginal choice between complying and violating, as is clear from Expression 2 for the compliance threshold. Both larger carrots and larger sticks reduce violations in that group. In contrast, carrots and sticks are complements for low-benefit agents making a marginal choice between complying and abstaining from participation, as is clear from Expression 3 for the participation threshold. Larger carrots induce participation while large sticks depress it for that group.

Moreover, inter-marginal changes are also possible. With respect to those, carrots and sticks are complements. A sufficiently large increase in the stick or a sufficiently large decline in the carrot turns high-benefit agents into low-benefit ones, causing former high-benefit violators to opt out of the regime. Vice versa, a sufficiently large decline in the stick or a sufficiently large increase in the carrot turns low-benefit agents into high-benefit ones, causing former low-benefit non-participants to participate and violate the rule.

Overall, an increase in carrots and an increase in sticks will typically results in different outcomes. More starkly, they may result in the opposite outcomes. While increasing the carrot unambiguously leads to more compliers, increasing the stick may result in fewer of them. Similarly, while an increase in the stick reduces the number of violators, an increase in the carrot may increase that number. So the second unexpected effect of expected sanctions (or sticks) is that they are not substitutes for expected rewards (or carrots) in their incentive effects, as is commonly assumed.

3 Taxes and Subsidies

Let us now examine whether adding a tax to the model expands the set of outcomes that the principal can reach. Assume that that the principal perfectly observes participation and non-participation. Assume further that the principal taxes non-participation in the activity by a fixed amount τ. Obviously, and in contrast with the carrot and the stick in our model, this tax does not depend on the compliance decisions of participating agents. Alternatively, the principal subsidizes the regulated activity (again, regardless of the principal's findings of compliance or violation) by

[32] This is because with the parameter values used in the example, we have $\frac{3}{8}\left(1-\frac{3}{4}\right) < \frac{1}{4}\left(1-\frac{1}{2}\right)$.

[33] Start with a different carrot, $c=\frac{1}{4}$, and increase the stick from $s=\frac{1}{2}$ to $s'=1$ as we did before (and retain $q=\frac{3}{4}$). This new parameter value gives the following outcomes: $e_k=\frac{3}{8}$, $e_p=\frac{3}{16}$, $e'_k=\frac{5}{8}$, and $e'_p=\frac{1}{16}$. Plugging these values in (5), we have that if $h<\frac{1}{9}$, both the number of compliers and the compliance rate fall as a result of increasing the stick, yielding the result described in the text. If $\frac{1}{9}<h<\frac{1}{3}$ the number of compliers falls but the compliance rate increases if the stick is increased, owing to the lower number of violators. Finally, if $h>\frac{1}{3}$ increasing the stick has the expected effect of improving both compliance and the compliance rate.

the same amount τ. (If τ is negative we have a tax for participants or, equivalently, a subsidy for non-participants.) The new payoff matrix is as follows:

Action	Payoff
The agent participates and complies	$\Pi_k(b,e) = b + qc - (1-q)s - e + \tau$
The agent participates but violates	$\Pi_v(b) = b + (1-q)c - qs + \tau$
The agent does not participate	0

Table 4: Agent's payoffs

We have now the following thresholds:

$$\begin{aligned} e_k^\tau &= (c+s)(2q-1) \\ e_p^\tau &= b_L + qc - (1-q)s + \tau \end{aligned}$$

Clearly the compliance threshold is unaffected by τ, because every participant receives τ whether he complies or not. In contrast, participation is incentivized by τ, both for compliers and for violators, given that e_p^τ increases in τ.

The addition of τ, however, does not expand the set of outcomes that the principal can reach. To see why, consider that the principal could replicate the same compliance and participation thresholds as above by simply adjusting the carrot and the stick instead of implementing a tax or a subsidy. Specifically, the principal could increase the carrot by an amount equal to τ and reduce the stick by the same amount; that is, the principal could set $c' = c + \tau$ and $s' = s - \tau$ to mimic the thresholds reported above. The compliance threshold is clearly met, because $c' + s' = c + s$. The participation threshold is also met because we have

$$\begin{aligned} e_p &= b_L + qc' - (1-q)s' \\ &= b_L + q(c+\tau) - (1-q)(s-\tau) \\ &= b_L + qc - (1-q)s + \tau \qquad = \quad e_p^\tau \end{aligned}$$

In essence, the principal can replicate the effects of taxes and subsidies by keeping the sum of the carrot and the stick constant (so as to keep the compliance threshold constant) while altering their ratio $\frac{c}{s}$. That ratio should increase if the activity is subsidized (that is, if τ is positive) and decrease if the activity is taxed (that is, if τ is negative).[34]

The conclusion that adding taxes or subsidies offers no benefit to the regulator seemingly contradicts the classic result in the economic analysis of tort law: Because no liability rule can assure both the optimal level of care (compliance) and activity (participation), taxing or subsidizing the activity improves the regulator's ability to control the agent's behavior along both margins at the same time (Png 1986). The reason for this improvement is that the regulator in the tort model has only one instrument—damages—to control both the care and the activity levels. A tax or a subsidy amounts to a second instrument at the regulator's disposal. In contrast, our model features two instruments—the carrot and the stick. These two instruments are all the regulator needs in order to control two behavioral margins. Adding more instruments does not improve the results.[35]

[34] This simple transformation rule works in the simplified model and is a special case of a more general transformation rule presented in the Online Appendix.

[35] Carrots and sticks have opposite effects on participation. Since the compliance margin depends on their sum, they can be adjusted to vary participation as desired without affecting compliance.

4 Discussion

Our results modify some long-held views about the basic workings of deterrence mechanisms and the interplay between positive and negative incentives. Yet our analysis hews closely to Becker's (1968) canonical model. We modify that model in only three ways. First, we add the participation margin to the analysis of compliance. Second, we add carrots to the law enforcement setting.[36] And third, we introduce agents who differ in at least one dimension in addition to their varying costs of complying with the law. In this section, we explain that our modifications are neither minor nor artificial. Rather, they make the law enforcement model more realistic and more useful to real-world regulators. We also alert regulators to the conditions that may lead to unintended effects of familiar regulatory incentives. Finally, we reflect on the normative implications of the ambiguities that we have identified in this article.

4.1 Optional Participation

Consider some of the canonical examples of socially costly acts in the law and economics literature: pollution (Coase 1960, Calabresi and Melamed 1972), collusion and monopolies (Becker 1968, Stigler 1970, Tullock 1967), speeding (Becker 1968, Stigler 1970), theft (Becker 1968, Calabresi and Melamed 1972, Tullock 1967), and tax evasion (Allingham and Sandmo 1972). All these acts (and many others) raise the question of deterring socially harmful conduct—the question which gave rise to the deterrence theory itself.

But all of the same examples also raise the question of participation in the named activity. A firm potentially liable for pollution may exit the "dirty" industry; a firm potentially liable for collusion may exit the concentrated sector of the economy; a driver potentially liable for speeding may stay home or get on a bicycle; a diner afraid of being charged with theft after mistakenly taking someone else's umbrella on the way out may stay clear of restaurants on rainy days, a taxpayer concerned with sanctions for non-compliance may avoid transactions, structures, and financial instruments with uncertain tax consequences. It is not an overstatement to say that regulatory settings where opting out of the regime is impossible are the exception, not the rule.

Moreover, legal uncertainty and enforcement errors are hardly uncommon in all of the areas just mentioned (and in many others). Environmental laws are often vague (*Sackett v. EPA*, 132 S.Ct. 1367, 1375 (Alito, J., concurring)), determinants of collusion are "difficult to articulate" (Kaplow 2012b: 479), liability for theft and many other theft-like activities turns on the defendant's (unobservable) state of mind (Raskolnikov 2014), and one hardly needs to elaborate on the complexity and uncertainty of the tax law. Mistaken imposition of liability is far from a remote possibility in these and many other legal regimes. Given this concern, opting out is a real choice even for agents who would have complied with the law if they participated. A model that ignores this choice necessarily has only a limited explanatory power when applied to the real world.

In fact, the limitation of such a model is even more significant than the examples offered thus far suggest. All these examples involve potentially harmful acts that yield private benefits in excess of private costs. There is, however, another category of acts—those producing private benefits that are too small (compared to private costs) to induce individuals to act. If the government wants these acts to occur—perhaps because they produce positive externalities—the government needs to offer positive incentives to induce individuals to participate in these socially beneficial activities.

[36] As we stressed in note 3 and accompanying text, the co-existence of carrots and sticks in only crucial to our comparative results but not to the results on the effects of carrots and sticks taken in isolation.

When regulating these acts, inducing participation is the institutional designer's main objective, not an afterthought. Overall then, fully incorporating the participation decision into the canonical model of deterrence is essential to improving the model's realism and usefulness.

4.2 Carrots, not Just Sticks

The traditional analysis of law enforcement instructs the government to deter privately beneficial but socially costly acts by imposing sanctions equal to the acts' expected external harms. Thus the key analytical variables are the private benefit, the external harm, and the expected sanction. Positive incentives appear to have no role in this setup. Png (1986) suggested that carrots should be used to counter the erroneous imposition of sticks on complying agents. But Png's (1986) brief analysis considered a highly simplified setting, his application was unrealistic,[37] and his insight has remained undeveloped.

A voluminous literature on carrots and sticks certainly does investigate the effects of rewards in optional incentive schemes. But that literature generally does not address the problem of improper (illegal) claiming of carrots. Taken together, the deterrence scholarship focuses on sticks and enforcement errors while mostly ignoring carrots, and the literature on carrots and sticks mostly ignores imperfect enforcement. Yet a moment's reflection leads to two conclusions. First, optional incentive-based regimes do raise noncompliance concerns. Second, traditional enforcement settings studied in the deterrence literature do feature carrots, though usually of an unintended, quasi-carrot variety.

Starting with the optional, incentive-based regime, note that the US government spends multiple billions each year on cash grants (Goodwin and Smith 2018; Theodos, Stacy and Ho 2017), loan assistance (Goodwin and Smith 2018), price and revenue protection programs (Glauber 2018; Sumner 2018), insurance subsidies (Glauber 2018; Sumner 2018), and other carrots designed to induce participation in a wide variety of government schemes aimed at anything from peanut farming (Goodwin and Smith 2018) to community development (Theodos, Stacy and Ho 2017) to workplace diversity (Sturm 2006). Small businesses alone benefit from fifty-two different federal assistance programs administered by the Departments of Agriculture, Commerce, Housing and Urban Development, and the Small Business Administration (US Government Accountability Office 2012). US states spend billions more on similar inducements (Patrick 2014).

Some of the carrot-granting programs challenge the recipients (and the regulators) with "a complicated labyrinth of criteria," making enforcement errors all but certain (Howells 1970: 393-94, 412). Some carrot-claimers skirt the rules, failing to comply either with their letter or their spirit (Goodwin and Smith 2018; Howells 1997; Theodos, Stacy and Ho 2017). In response, regulators use penalties to deter improper carrot-claiming (Howells 1997). Thus simultaneous presence of carrots, sticks, and enforcement errors is a typical feature of many US incentive programs, not an aberration.

The US government also spends billions more in positive incentives outside of the federal budget. The so-called tax expenditures—special tax breaks in the Internal Revenue Code designed to induce participation in a particular activity—offer potential participants a carrot just like outright cash grants do (US Government Accountability Office 2013). [38]In fact, three of the largest corpo-

[37] Png suggested a carrot in the form of payments to all drivers involved in accidents, whether they exercise due care or not, Png (1986: 103).

[38] Note that for the largest eight hundred or so companies under a continuous IRS audit (Elliott 2012), the only uncertainty surrounding the carrot arises from inaccurate enforcement by the government auditors that is key to our model. In other

rate tax expenditures of the recent past—the deduction for accelerated depreciation, the domestic production activities deduction (repealed at the end of 2017), and the research and development (R&D) tax credit—illustrate every feature of our model (US Government Accountability Office 2013). The R&D credit, to take one example, is a carrot. Its false claiming is punished with a tax-penalty stick. The rules determining a taxpayer's eligibility for the tax credit are complex and uncertain, so enforcement errors are inevitable. Thus both the participation and the compliance decisions are in play.

Two further insights highlight the wide applicability of our results. First, the carrot need not take the form of a payment or a reduction in a price or a tax burden. Any government-created benefit is a carrot in terms of our model. Second, the carrot need not be intended as such by the regulator.

Consider the Federal Reserve's so-called discount window that offers eligible banks access to short-term liquidity under certain conditions. The discount window was put in place in 1913 to establish the Federal Reserve as the lender of last resort, and possibly to incentivize small-business lending by local banks (Hunt 2014). It was certainly not created to incentivize trading firms to become banks. Yet, this was exactly how it worked when Morgan Stanley and Goldman Sachs chose to become banks in the midst of the financial crisis (Sorkin and Bajaj 2008). The discount window amounted to an unintended quasi-carrot.

The Volcker Rule enacted in the aftermath of the crisis in order to stop proprietary trading by deposit-taking institutions completed the picture. The discount window is a quasi-carrot. The broad enforcement powers of the Fed are a stick (Hamilton 2017). And the purpose-based definition of proprietary trading, as well as the hedging and market-making exceptions, in the Volcker Rule inevitably give rise to false positives and false negatives (Bubb and Kahan 2018). The overall incentive structure is exactly the one we model.

Tax quasi-carrots are the focus of much day-to-day corporate tax planning. The tax-free reorganization provisions of the Internal Revenue Code, to take one example, were not enacted to induce tax-free reorganizations. Rather, Congress decided that if a corporate enterprise continues in a different form but with little change, there is no reason to impose the tax.[39] Yet there is no doubt that massive tax savings from securing a tax-free treatment for corporate acquisitions and separations are of first-order significance in momentous business decisions that shape financial markets and the US economy (Elliott 2015). And conditioning the tax-free treatment on whether transactions take place pursuant to a "plan," whether "substantially all" of the assets are transferred, and whether the entire reorganization has "economic substance," along with other vague tests, makes it quite clear that the application of tax-free provisions involves plenty of uncertainty.

Likewise, Congress enacted international tax rules—including the foreign tax credit provisions—to enable source-based taxation of income earned by US companies abroad without double-taxing that income (Graetz and O'Hear 1997). The goal was not to induce US companies to move their operations or corporate domicile offshore. In fact, the major reform of the US international tax regime in 2017 was motivated by the desire to prevent these very shifts, including the so-called corporate inversions. Yet, it is beyond doubt that foreign tax credits do serve as a quasi-carrot—an unintended inducement to move operations, headquarters, and income offshore. At the same

words, $p = 1$ in this case as in the simplified model presented in the text.

[39]In the words of the leading treatise, these provisions were put in place under "the assumption [...] that the new enterprise [...] that may hold the corporate assets, and the new stock or securities received in exchange for old stock or securities, are substantially continuations of, and interests in, the old corporations, still alive but in different form" (Bittker, Eustice, and Goldstein 2015, Par.12.00[1]).

time, the complexity of US international tax rules makes it impossible to enforce them error-free.

Overall, carrots are pervasive. Numerous regulatory regimes are designed around positive incentives. Many more feature quasi-carrots that were not put in place to incentivize a particular activity but that do so just the same. An economic analysis of regulation that ignores carrots is of limited practical use indeed.

4.3 Heterogeneous Agents

Our last modification of the standard deterrence model is the introduction of agents who vary along a dimension other than their cost of compliance. Thus far, that dimension was the agents' private benefit from participation. One hardly needs to spend much time to defend this assumption. In fact, the opposite assumption would be wildly unrealistic. Different polluting plants do not derive the same benefits from running their operations or from installing identical pollution control devices. Drivers do not benefit equally from getting to their destinations. Reorganizing firms—even while undertaking the same type of a tax-free reorganization—do not capture identical gains from their business combinations. Private benefits are, indeed, heterogeneous.

But heterogeneity of benefits is just one possible dimension of additional heterogeneity. As we show in the Online Appendix, our results hold in a setting in which all agents derive the same benefit from participation but hold different beliefs about the enforcement accuracy q_k and q_v. Agents' beliefs may vary for at least two reasons. First, the principal may not be equally accurate in evaluating compliance of different agents. Second, whatever the principal believes about her own accuracy, it is difficult for the principal to convey this belief to the agents, and all but impossible to do so credibly. We do not assert that agents have no information about the matter. Litigation, published official interpretations, and informal guidance all provide agents with valuable information. But it is implausible to assume that this information leads all agents to arrive at the same conclusion regarding the government's enforcement error rates.

The two dimensions of additional heterogeneity that we model (additional, that is, to the variation in the compliance costs) are just the tip of the iceberg. Any idiosyncratic characteristic that affects an agent's participation decision will have similar effects to those revealed in our model. One important source of heterogeneity could be variation in the agents' private experiences with the nominally uniform policy variables. Punishments, to take one example, are often non-monetary, and the same punishment may be felt differently by different agents (Galle and Mungan 2019). Suspension of a driving license, for instance, may impact rich and poor agents differently. Even monetary fines may have harsher effects on liquidity-strapped agents. If so, a portion of the population may effectively face high sticks, while the remaining fraction faces low sticks, even though the nominal sanction is the same for everyone. Similar results obtain when agents experience the carrot differently. All such differing experiences would generate results analogous to ours.[40]

In addition to having different private realizations of uniform carrots and sticks, agents may differ in their perceptions of the variables we consider. Identical carrots and sticks may have different salience for different agents. Different agents may fall victims to well-known probability misperceptions to a different extent (Tversky and Kahneman 1974, Williams 2014). Any sanction or reward that is not immediate would have a different present value for agents with varying private discount rates (Frederick, Loewenstein, and O'Donoghue 2004). In fact, a variation related to

[40]Each different source of heterogeneity may in fact generate different specific results, but our central contention—that the participation decision is affected by changes in policy variables in a way that interacts with the decision to comply or violate—would remain valid.

almost any bias identified in the behavioral economics literature is likely to produce heterogeneity that would give rise to our results.

In reality, it is almost certainly true that people and firms subject to the kinds of regulation that we model here vary along many different dimensions at the same time. We do not model such multi-dimensional heterogeneity. But our results make it clear that the canonical conclusions that we question here are even less likely to be true in a complex regulatory environment with agents varying along multiple dimensions in different ways.

4.4 Implications

Our analysis does not produce simple and easily testable insights; it undermines the existing ones. We show that a simple compliance-enhancing effect of expected sanctions does not generally hold, and neither does a simple relationship between the incentives produced by carrots and sticks. Negative results are informative, but our inquiry has positive implications as well—both practical and conceptual ones.

On the practical side, the crucial payoff of our analysis is to alert real-world regulators about settings where they ought to be especially concerned about the unexpected effects of familiar enforcement instruments. When regulators focus on enforcement—of competition law, environmental law, discrimination law, tax law, and many other regulatory settings—they should ask themselves whether regulated parties are likely to consider exiting (or not entering) the regulatory regime. Is it likely that imperfectly administered sanctions for housing discrimination would lead potential compliant landlords to invest in other assets? Is it likely that sanctions for tax noncompliance would drive taxpayers away from particular business transactions or securities? Of course, regulators will never know answers to these questions perfectly. But just asking the question would be a step forward compared to the sole focus on the magnitude of statutory penalties and the probability of detection.

Regulators have long-known that increasing expected sanctions may reduce participation in an optional regime. We show that the same increase may lead to an outcome that, we believe, most regulators would view as much more problematic than lower participation: it may reduce compliance. This in turn calls for monitoring the consequences of an increase in expected sanctions with greater vigilance.

Turning to regimes based on positive incentives or rewards, regulators need to pay close attention both to the possible noncompliance and to the possible exit. If a regulator interested in energy conservation by home owners can perfectly observe whether an owner claiming to engage in such conservation actually does so—and if everyone owns a home—the regulator can safely assume that a carrot (such as the tax credit) for energy-saving homes and a stick (such as a tax) on energy-wasting homes would produce identical incentives to convert energy-wasting homes into energy-saving ones. The regulator may increase the carrot, or the stick, or both; only the sum of the two instruments would matter in terms of the magnitude of the incentive.

But if the regulator cannot perfectly observe the appropriateness (legality) of claiming the carrot, if the regulator uses the stick to punish those who claim the carrot inappropriately, and if home owners may exit the entire energy saving regime by switching to rentals, the regulator better be careful. Relying on the tried-and-true assumption that carrots and sticks are interchangeable may lead to unexpected and highly objectionable outcomes. A larger carrot may lead to many individuals buying homes and claiming the generous carrot illegally—an undesirable result that a larger stick would not produce.

Recognizing this problematic effect of carrots is particularly important because the pairing of carrots with sticks is not always deliberate or easily seen by the regulator. Returning to the housing discrimination example, US law has long-offered tax incentives to induce landlords to invest in low-income housing (Congressional Research Service 2019). Congress recently enacted another tax incentive program—the Opportunity Zone tax credit—that may also subsidize construction of rental housing in economically disadvantaged areas (Tax Policy Center 2018). Needless to say, Congress enacted these incentive programs not to subsidize discriminatory low-income housing. But our findings show that regulators must be vigilant to make sure that greater housing incentives do not lead to this very result.

Another and particularly important example of the relevance of our findings comes from the largest US means-tested transfer program for low-income workers. This program—the earned income tax credit (EITC)—offers refundable credits to workers meeting the program's eligibility requirements. In 2015, the EITC disbursed over sixty billion dollars to more than twenty five million workers (Crandall-Hollick 2018: 6). Unfortunately, some of the eligibility requirements are complex, making compliance costly and enforcement errors likely.[41] As a result, about a fifth of eligible workers are estimated to fail to claim EITC (Goldin 2018). Among those who do claim, noncompliance is a significant problem.[42] Thus, the participation and compliance challenges are readily apparent.

Congress and the IRS have long recognized these dual challenges, so it is useful to consider regulatory responses. Dissatisfied with significant noncompliance, Congress followed the standard deterrence prescription and raised expected penalties for inappropriate claiming of the EITC in 1997, 1998, 2002 (Brown 2007: 806-07), and 2015 (Chen and Nellen 2017). But at the same time—and despite its concerns about significant noncompliance—Congress made EITC more generous in 1990, 1993, 2001, 2009, and 2015 (Crandall-Hollick 2018). Thus, Congress increased both carrots and sticks over time while facing both the compliance and the participation challenges.

We do not suggest that over the past several decades, multiple sessions of Congress and IRS administrations changed EITC sticks and carrots following the logic of our findings. This would be naïve as a political economy matter, and as an economic hypothesis as well. We predict that raising sanctions may reduce the number of compliers and the compliance rate in an optional participation regime like the EITC. But the number of compliers and the compliance rate are impossible to measure precisely, and the total number of eligible participants is at best an educated guess (Goldin 2018). Measuring *changes* in these outcomes in response to changes in sanctions and rewards is even more difficult. It is worth remembering that even the most basic predictions of the deterrence model—that higher sanctions and detection probabilities increase deterrence in mandatory regimes—are by no means undisputed as an empirical matter or as a policy prescription (Raskolnikov 2020, Tonry 2008). Our results are more subtle and more difficult to prove empirically than these standard prescriptions are. These results are also new to the literature. So we do not want to overstate their connection to prior policies. We do believe, however, that the history of EITC reforms shows clearly that policymakers recognize that focusing on deterrence and

[41] The costs arise because in order to comply with relevant requirements a taxpayer needs to exert effort to understand the rules and to keep records throughout the year (Goldin 2018:77-78). In contrast, a violator may essentially assert that he or she meets the requirements without incurring these costs. The idea that tax underpayment may result from taxpayers' failure to exert effort in order to understand the rules is not new to the tax compliance literature (DeBacker et al. 2018). As for enforcement errors, IRS data show that in 2011 the tax agency improperly penalized EITC recipients 40% of the time (National Taxpayer Advocate 2013).

[42] The EITC rate of improper payments ranged between 20% and 30% since 2003—significantly greater than the average overall noncompliance rate of about 17% (Drumbl 2016).

expected sanctions while ignoring the participation margin is a poor way of regulating optional participation regimes. Our findings not only support this intuition, but inform it as well.

The final implication of our analysis is to highlight the fundamental ambiguity of the common regulatory objective of increasing "compliance." In a typically studied enforcement setting where participation is mandatory or otherwise universal, reducing the number of violators necessarily means increasing the number of compliers. Moreover, because the number of participants is fixed (everyone participates), increasing the number of compliers is equivalent to increasing the compliance rate. Given these relationships, it is no surprise that the deterrence literature treats reducing violations as unquestionably desirable. The only question is how best to achieve this reduction at the least social cost.

But when participation is neither mandatory nor universal—that is, in the vast majority of real-world regulatory schemes—a reduction in the number of violators is not longer unambiguously desirable. Once this is understood, new normative questions immediately arise.

If a regulatory intervention reduces the number of violators while also reducing the number of compliers even more, is the intervention worthwhile? Even in purely descriptive terms, if the number of violators declines but the compliance rate declines as well, has compliance improved or deteriorated? Similarly, if the number of compliers goes up, but the number of violators increases even more so that the compliance rate declines, should we think of the result as greater compliance or greater noncompliance? And in any case, is the result socially desirable?

These questions are beyond this article's scope. In fact, they open a new research agenda in the study of optimal deterrence and regulatory compliance. This article's contribution is to dislodge the accepted views about the relationships between compliance and violations as well as between carrots and sticks. Once these views are shown to be incorrect in many real-world legal regimes, scholars may begin the search for new answers to questions about the meaning of compliance—questions that until now seemed long-settled.

5 Conclusion

In this article, we identify some unexpected effects of familiar regulatory instruments. The deterrence theory holds that higher expected sanctions lead to fewer violations and greater compliance. We agree with the first part of this statement but not the with the second one. When agents choose both whether or not to comply and whether or not to participate in an imperfectly enforced regulatory scheme, higher expected sanctions may lead to both fewer compliers and a lower compliance rate. The literature on carrots and sticks holds that the two instruments create identical incentives and vary only in their second-order effects. We show that if improper claiming of carrots is possible and if the regulator cannot deter it by sanctioning violators with perfect accuracy, the incentive effects of carrots and sticks are not the same.

Inevitably, our analysis has limitations. We discuss six of them here. The first limitation reflects a well-known and persistent challenge to the economic analysis of law. Agents in our model cannot affect the principal's accuracy by adjusting their behavior. Agents either comply or violate; they cannot barely cross the line or engage in egregious noncompliance. This assumption is more realistic in some settings than in others.

The assumption of agents' binary behavior is standard in law-and-economics. The reason for it is well-known: eliminating this assumption leads to indeterminate results (Craswell and Calfee 1986; Shavell 1987). Several recent contributions show that in certain settings this indeterminacy disappears (Baker and Raskolnikov 2017; Dari-Mattiacci 2005). Our analysis does not contribute to further resolving this long-standing challenge.

The second limitation is our principal's inability to tailor its instruments to individual agents. This inability is one of the key distinctions between a constrained model like ours and the mechanism design literature. In reality, regulators usually deploy crude, uniform instruments rather than individually-tailored incentives (Kaplow 2017: 4-5, Montero 2005, Stavins 2003). So our assumption of uniform carrots, sticks, and the likelihood of detection is quite realistic in many settings.

Third, participants in our model can choose to participate or not, but they cannot choose the extent of participation. Incorporating this variable into the model would necessarily complicate it. Our goal is to show that extending the standard model to reflect a small number of realistic features is sufficient to modify the canonical conclusions of the deterrence theory and the literature on carrots and sticks.

Fourth, we do not model costs of various regulatory tools that we consider. These costs are likely to vary. Carrots are usually more expensive for the principal than sticks are. Increasing the rate of monitoring may consume more of the principal's resources than adjusting a monetary sanction. Our contribution is to clarify the effects of changing different enforcement instruments, leaving it to the principals (and to future work) to balance these effects against related costs and benefits.

Fifth, and at the risk of stating the obvious, our model is a simplified version of reality. Not only agents' benefits and principal's accuracy vary simultaneously in real-world settings, other factors such as risk aversion, uncertainty aversion, various heuristics and biases, as well as non-monotonic responses to incentives (Gneezy and Rustichini 2004) likely come into play. Our decision not to model all these complications is deliberate. We re-examine the foundational conclusions of canonical models, and we show that these conclusions do not generally hold even without accounting for risk aversion, behavioral factors, or other deviations from rationality. Extending the model to incorporate such deviations is likely to raise further questions about the general applicability of the model's basic payoffs.

Finally, we offer a positive theory of behavior, not a normative theory of optimal regulation. We do not specify the social welfare function and do not discuss efficiency. While the choice to limit the model in this way differs from most of the deterrence scholarship, our choice is deliberate. The main message of our analysis is that the canonical findings of the deterrence theory are, in essence, misleadingly simple even on the positive side. We believe that the shortcomings of the deterrence theory on the normative side are even more substantial.

The standard normative framework of the optimal deterrence model consists of an activity that produces an external harm, and a regulator who aims to force the agent producing this harm to internalize it at the lowest cost to the regulator. The negative externality is the only cause of inefficiency and the only motivation for government action. This framework is not particularly realistic, especially in the numerous optional-participation regimes that we study. Where carrots are present by design, the government obviously aims to induce some agents to participate. There must be some social benefit of this participation or social cost of non-participation. The benefit or cost must vary among the agents (or the government would issue a mandate rather than an inducement). Moreover, as soon as carrots enter the picture, one needs to consider the source of the funds and the deadweight loss of raising them. There has been very little effort to extend the basic external harm model to account for any of these additional social costs and benefits. Making progress here would not be easy, and we leave an effort to make such progress for the future.

Another reason not to commit to any specific welfare function is that doing so would limit one of this article's main takeaways. We offer a positive theory of agents' behavior that shows that the fundamental tenets of the law enforcement scholarship and the literature on positive and negative incentives do not hold in general, and perhaps not even in most cases. These results are valid given *any* welfare function—an important point given the complexity of the welfare analysis and the controversy surrounding it.

In the end, our unexpected findings—as well as all of the limitations just discussed—sound a note of caution for real-world regulators. Rather than relying on the seemingly straightforward prescriptions of the basic deterrence model, they would do well by closely monitoring the actual responses to changes in actual expected sanctions and rewards, while standing ready to respond to their unexpected effects (Duflo 2017).

References

[1] Allingham, Michael G., and Agnar Sandmo. 1972. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics* 1:323-38.

[2] Andreoni, James, William Harbaugh, and Lise Vesterlund. 2003. The Carrot or the Stick: Rewards, Penalties, and Cooperation. 93 *American Economic Review* 93:893-902.

[3] Baker, Scott, and Anup Malani. 2019. Compromising Accuracy to Encourage Regulatory Participation. 2019. Working Paper.

[4] Baker, Scott, and Alex Raskolnikov. 2017. Harmful, Harmless, and Beneficial Uncertainty in Law. *Journal of Legal Studies* 46:281-307.

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

[6] Ben-Shahar, Omri, and Anu Bradford. 2012. Reversible Rewards. *American Law and Economics Review* 15:156-80.

[7] Bittker, Boris I., James S. Eustice, and Gersham Goldstein. 2015. Federal Income Taxation of Corporations and Shareholders. Valhala, NY: Thomson Reuters.

[8] Brabender, Allen M. 2004. The Misapplication of Minnesota's Speeding Statute and the Need to Raise the Posted Limit or Expand Use of the Dimler Amendment. *Hamline Law Review* 27:1-14.

[9] Brown, Dorothy A. 2007. Race and Class Matters in Tax Policy. *Columbia Law Review* 107:790-831.

[10] Bubb, Ryan, and Marcel Kahan. 2018. Regulating Motivation: A New Perspective on the Volker Rule. *Texas Law Review* 96:1019-52.

[11] Calabresi, Guido, and Douglas Melamed. 1972. Property Rules, Liability Rules, and Inalienability: One View of the Cathedral. *Harvard Law Review* 85:1089-1128.

[12] Chen, Caroline, and Annette Nellen. 2017. Penalties: Dealing with Recent Legislative Changes. Practical Tax Strategies March 2017:76-84.

[13] Coase, Ronald. 1960. The Problem of Social Cost. *Journal of Law and Economics* 3:1-44.

[14] Crandall-Hollick, Margot L. 2018. The Earned Income Tax Credit (EITC): A Brief Legislative History. *Congressional Research Service* R44825.

[15] Craswell, Richard, and John E. Calfee. 1986. Deterrence and Uncertain Legal Standards. *Journal of Law, Economics and Organization* 2:279-303.

[16] Dari-Mattiacci, Giuseppe. 2005. Errors and the Functioning of Tort Liability. *Supreme Court Economic Review* 13:165–87.

[17] Dari-Mattiacci, Giuseppe. 2009. Negative Liability. *Journal of Legal Studies* 38:21-60.

[18] Dari-Mattiacci, Giuseppe and Gerrit De Geest. 2010. Carrots, Sticks, and the Multiplication Effect. *Journal of Law, Economics, and Organization* 26:365-384.

[19] De Geest, Gerrit and Giuseppe Dari-Mattiacci. 2013. The Rise of Carrots and the Decline of Sticks. *University of Chicago Law Review* 80:341-93.

[20] DeBacker, Jason, Bradley T. Heim, Anh Tran, and Alexander Yuskavage. 2018. The Effects of IRS Audits on EITC Claimants. *National Tax Journal* 71:451-84.

[21] Drumbl, Michelle L. 2016. Beyond Polemics: Poverty, Taxes, and Noncompliance. *eJournal of Tax Research* 14:253-90.

[22] Duflo, Esther. 2017. The Economist as Plumber. *American Economic Review: Papers & Proceedings* 107:1-26.

[23] Elliott, Amy. 2012. Audit Proof? How Hedge Funds, PE Funds, and PTPs Escape the IRS. *Tax Notes* 136:351-355.

[24] Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2004. Time Discounting and Time Preference: A Critical Review. Pp. 162222 in *Advances in Behavioral Economics*, edited by Colin F. Camerer, George Loewenstein, and Matthew Rabin. New York NY: Russell Sage Foundation.

[25] Friedman, Ezra and Abraham Wickelgren. Chilling, Settlement, and the Accuracy of the Legal Process. *Journal of Law, Economics & Organization* 26:144-157.

[26] Galle, Brian and Murat Mungan. 2019. Optimal Enforcement with Heterogeneous Private Costs of Punishment. George Mason University Law & Economics Research Paper 19-11.

[27] Ganuza, Juan José and Fernando Gomez. 2005. Caution, Children Crossing: Heterogeneity of Victim's Cost of Care and Negligence Rule, *Review of Law and Economics* 1:365-297.

[28] Gilpatric, Scott M. 2009. Risk Taking in Contests and the Role of Carrots and Sticks, *Economic Inquiry* 47:266-77.

[29] Glauber, Joseph A. 2018. Unraveling Reforms? Cotton in the 2018 Farm Bill. *American Enterprise Institute*.

[30] Gneezy, Uri, and Aldo Rustichini. 2004. Incentives, Punishments, and Behavior. Pp. 573-89 in *Advances in Behavioral Economics*, edited by Colin F. Camerer, George Loewenstein, and Matthew Rabin. New York NY: Russell Sage Foundation.

[31] Golding, Jacob. 2018. The Benefit Complexity and Take-up: Lessons from the Earned Income Tax Credit. *Tax Law Review* 72:59-110.

[32] Goodwin, Barry K., and Vincent H. Smith 2018. Reflections on the US Peanut Program. *American Enterprise Institute*.

[33] Gordon, Wendy J. 1992. Of Harms and Benefits: Torts, Restitution, and Intellectual Property. *Journal of Legal Studies* 21:449-82.

[34] Graetz, Michael J., and Michael M. O'Hear. 1997. The "Original Intent" of U.S. International Taxation. *Duke Law Journal* 46:1021-1109.

[35] Hamilton, Jesse. 2017. Deutsche Bank Is First Bank Busted for Breaking Volcker Rule. *Bloomberg*, April 20.

[36] Howells, Louise A. 1997. Looking for The Butterfly Effect: An Analysis of Urban Economic Development Under the Community Development Block Grant Program. *Saint Louis University Public Law Review* 16:383-417.

[37] Hunt, Daniel J. 2014. Just Grin and Bear It: Why Consistent Use of Individual Bailouts Under Section 13(3) of the Federal Reserve Act Is a Necessary Evil to Combat Economic "Mass Destruction." *George Mason Journal of International Law* 6:59-105.

[38] Joint Committee on Taxation. 2009. Tax Expenditures for Energy Production and Conservation. JCX-25-09R.

[39] Kaplow, Louis. 2011. Optimal Proof Burdens, Deterrence, and the Chilling of Desirable Behavior. *American Economic Review: Papers & Proceedings* 101:277-80.

[40] ____________. 2012a. Burden of Proof. *Yale Law Journal* 121:738-859. ***

[41] ____________. 2012b. Direct Versus Communications-Based Prohibitions on Price Fixing. *Journal of Legal Analysis* 3:449-538.

[42] ____________. 2017. Optimal Regulation with Exemptions. *International Journal of Industrial Organization* 66:1-39.

[43] Kaplow, Louis, and Steven Shavell. 1994. Accuracy in the Determination of Liability. *Journal of Law and Economics* 37:1-15

[44] Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. Transfer Program Complexity and the Take-Up of Social Benefits. *American Economic Journal: Economic Policy* 3:54-90.

[45] Levmore, Saul. 1986. Waiting for Rescue, An Essay on the Evolution and Incentive Structure of the Law of Affirmative Obligations. *Virginia Law Review* 72:879-941.

[46] Moldovanu, Benny, Aner Sela, and Xianwen Shi. 2012. Carrots and Sticks: Prizes and penalties in Contests. *Economic Inquiry* 50:453-62.

[47] Montero, Juan-Pablo. 2005. Pollution Markets with Imperfectly Observed Emissions. *RAND Journal of Economics* 36: 645-660.

[48] National Taxpayer Advocate. 2013. *Annual Report to Congress*. https://taxpayeradvocate.irs.gov/2013-Annual-Report/full-2013-annual-report-to-congress.html.

[49] Patrick, Carlianne. 2014. Does Increasing Available Non-Tax Economic Development Incentives Result in More Jobs? *National Tax Journal* 67:351-86.

[50] Png, I.P.L. 1986. Optimal Subsidies and Damages in the Presence of Judicial Error. *International Review of Law and Economics* 6:101-105.

[51] Polinsky, Mitchell A. 1979. Notes on the Symmetry of Taxes and Subsidies in Pollution Control. *Canadian Journal of Economics* 12:75-83.

[52] Polinsky, Mitchell A., and Steven Shavell. 2007. The Theory of Public Enforcement of Law. Pp. 403-454 in vol. 1 of *The Handbook of Law and Economics*, edited by Mitchell Polinsky and Steven Shavell. Amsterdam: North Holland.

[53] Posner, Richard A. 1985. An Economic Theory of the Criminal Law. *Columbia Law Review* 85:1193-1231.

[54] Raskolnikov, Alex. 2014. Irredeemably Inefficient Acts: A Threat to Markets, Firms, and the Fisc. *Georgetown Law Journal* 102:1133-89.

[55] Raskolnikov, Alex. 2020. Criminal Deterrence: A Review of the Missing Literature. *Supreme Court Economic Review* 28:1-59.

[56] Shavell, Steven. 1987. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.

[57] Sorkin, Andrew Ross, and Vikas Bajaj. 2008. Shift for Goldman and Morgan Marks the End of an Era. *New York Times*, September 21.

[58] Stavins, Robert N. 2003. Experience with Market-Based Environmental Policy Instruments. Pp. 355-435 in Vol. 3 of *Handbook of Environmental Economics*, edited by Karl-Göran Mäler and Jeffrey Vincent. Amsterdam: Elsevier Science.

[59] Stigler, George J. 1970. The Optimum Enforcement of Laws. *Journal of Political Economy* 78:526-36.

[60] Sturm, Susan. The Architecture of Inclusion: Advancing Workplace Equity in Higher Education. *Harvard Journal of Law and Gender* 29:274-334.

[61] Sumner, Daniel A. 2018. Dairy Policy Progress. *American Enterprise Institute*.

[62] Theodos, Brett, Christina Plerhoples Stacy, and Helen Ho. 2017. *Taking Stock of the Community Development Block Grant*. Washington DC: Urban Institute.

[63] Tonry, Michael 2008. Learning from the Limitations of Deterrence Research. *Crime and Justice* 37:279-311.

[64] Tullock, Gordon. 1967. The Welfare Costs of Tariffs, Monopolies, and Theft. *Economic Inquiry* 5:224-32.

[65] Tversky, Amos and Daniel Khaneman. 1974. Judgment Under Uncertainty: Heuristics and Biases. *Science* 185:1124-31.

[66] US Government Accountability Office. 2012. *Entrepreneurial Assistance: Opportunities Exist to Improve Programs' Collaboration, Data-Tracking, and Performance Management* GAO-12-819. Washington, DC: US Government Accountability Office.

[67] US Government Accountability Office. 2013. *Corporate Tax Expenditures: Information on Estimated Revenue Loses and Related Federal Spending Programs* GAO-13-339. Washington, DC: US Government Accountability Office.

[68] Williams, Sean H. 2014. Probability Errors. Pp. 335-53 in *Behavioral Economics and the Law*, edited by Eyal Zamir and Doron Teichman. New York, N.Y: Oxford University Press.

[69] Wittman, Donald. 1984. Liability for Harm or Restitution for Benefit? *Journal of Legal Studies* 13:57-80.

Online Appendix

OA1 General Model

In this Online Appendix we prove that the results presented in the main text hold generally in the model setup presented in Section (2.1), that is, if we relax all of the simplifying assumptions made in Section (2.2). In addition, we generalize these results and provide additional technical insights.

OA1.1 The Model with Perfect Monitoring ($p = 1$).

We start by proving our results for the case in which the probability of monitoring is $p = 1$ and later extend the analysis to $p < 1$. Given the setup in Section (2.1), an agent of type (b, e) anticipates the payoffs reported below in Table OA1, which follows the same logic as Table 3 presented in the main text.

Action	Payoff
The agent participates and complies	$\Pi_k(b,e) = b + q_k c - (1 - q_k) s - e$
The agent participates but violates	$\Pi_v(b) = b + (1 - q_v) c - q_v s$
The agent does not participate	0

Table OA1: Agent's payoffs

It is useful to begin the analysis by defining the following threshold level for b

$$b_v \equiv q_v s - (1 - q_v) c$$

which is the level of b such that $\Pi_v(b) = 0$. Specifying this benefit threshold b_v allows us to partition the population of agents into two groups:

- High-benefit agents (with $b > b_v$) participate in the activity irrespective of whether they comply or violate. To see why, note that if $b > b_v$, then $\Pi_v(b) > 0$ and hence the payoff of violators is positive. If the agent complies, he must also earn a positive payoff because he will comply only if $\Pi_k(b,e) > \Pi_v(b)$, and we know that the latter is positive. Overall, high-benefit agents form the full participation group—they all take part in the activity. Whether the agent complies or violates depends on his cost of effort e. More precisely, the agent will comply if $e < e_k$ where the compliance threshold e_k is

$$e_k \equiv (c + s)(q_k + q_v - 1)$$

  which is the level of e such that $\Pi_k(b,e) = \Pi_v(b)$. The agent will violate if $e \geq e_k$. When the compliance threshold controls, carrots and sticks are substitutes.

- Low-benefit agents (with $b \leq b_v$) participate in the activity only if they comply. To see why, note that if $b \leq b_v$, then $\Pi_v(b) \leq 0$ and hence the payoff of violators is (weakly) negative. The payoff for compliers may be positive or negative. Overall, low-benefit agents form the partial participation group where all participants comply, all violators abstain from participation, and some would-be compliers choose not to participate as well. A potential complier will participate only if $\Pi_k(b,e) > 0$ at his level of e. More precisely, the agent will comply if $e < e_p$ where the participation threshold e_p is

$$e_p(b) \equiv b + q_k c - (1 - q_k) s$$

which is the level of e such that $\Pi_k(b,e) = 0$. The agent will participate and comply if $e < e_p(b)$ and will not participate if $e \geq e_p(b)$. Note that for low-benefit agents, $e_p(b)$ controls both compliance and participation because an agent who does not have incentives to comply would earn a negative payoff by participating and violating. When the participation threshold controls, carrots and sticks are complements.

In sum, e_k is the compliance threshold for participants, which determines the choice between complying and violating, on the assumption that the agent participates. This threshold controls the agent's compliance decision if $b > b_v$, that is, in cases where both complying and violating agents participate. In contrast, $e_p(b)$ is the participation threshold for complying agents. It controls behavior when $b \leq b_v$, that is, when violating agents do not find it advantageous to participate. In this case, the alternative to compliance is abstention rather than violation.

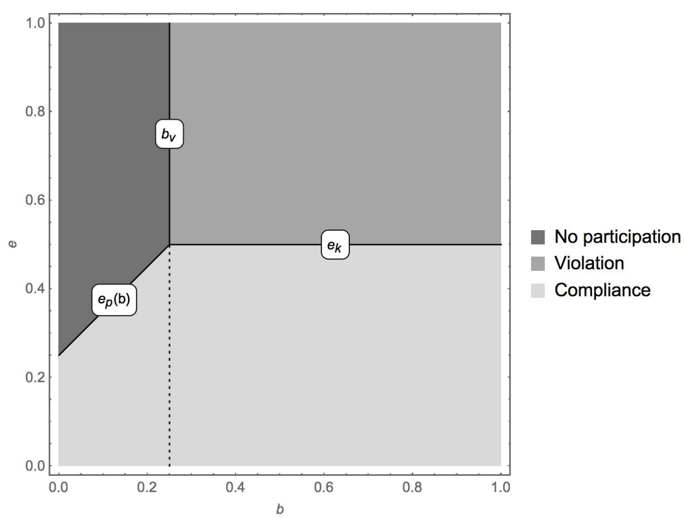


Figure OA1: Agent's behavior (simulation parameters: $q_k = q_v = \frac{3}{4}$, $c = s = \frac{1}{2}$).

We impose no constraints on the possible benefits of heterogeneous agents in our model. Therefore, some agents in our model belong to the full participation group while others belong to the partial participation group. Figure OA1 characterizes the agents' behavior. Quite intuitively, agents with low costs of effort participate and comply. Those with large benefits and high effort costs participate but violate. The remaining agents (those with low benefits and high effort costs) abstain from participation.

The next proposition shows how changes in enforcement variables affect the absolute number of complying agents ("compliance"), the absolute number of violating agents ("violations"), overall participation (compliance plus violations), and the ratio of the number of compliers over the number of participants ("compliance rate"). The results are summarized in Table OA3.

Proposition 1. *An increase in the carrot, c, results in an increase in compliance and participation, while the effect on violations and the compliance rate is ambiguous. An increase in the stick, s, results in a decrease in violations and participation, while the effect on compliance and the compliance rate is ambiguous.*

Proof. First note the comparative statics results reported in Table OA2, which is trivial to verify. The results on the effects of increased carrots on compliance and participation follow trivially from

Policy variable	b_v	e_k	$e_p(b)$
c	Decrease	Increase	Increase
s	Increase	Increase	Decrease

Table OA2: Comparative statics

the fact that e_k and $e_p(b)$ increase in c, while b_v decreases in c. The effect on violations is ambiguous because the violation region shifts over different types as c increases, which in turn is due to the fact that e_k moves up and b_v moves to the left (see Figure OA1). Hence, the result depends on the relative frequency of those types in the population of agents, that is, on $f(b,e)$. Formally, the absolute number of violating agents is given by

$$V \equiv \int_{b_v}^{\infty} \int_{e_k}^{\infty} f(b,e)\, de db$$

Therefore, we have:

$$\begin{aligned} \frac{\partial V}{\partial c} &= -\frac{\partial b_v}{\partial c} \int_{e_k}^{\infty} f(b_v,e)\, de \\ &\quad -\frac{\partial e_k}{\partial c} \int_{b_v}^{\infty} f(b,e_k)\, db \\ &= (1-q_v) \int_{e_k}^{\infty} f(b_v,e)\, de \\ &\quad -(q_k+q_v-1) \int_{b_v}^{\infty} f(b,e_k)\, db \end{aligned}$$

To show that the sign of $\frac{\partial V}{\partial c}$ is ambiguous is enough to show that there is a distribution $f(b,e)$ and parameter values that generate ambiguous results. Assume that b and e are independently and uniformly distributed on the unit interval—that is, $f(b,e)=1$—we have that $\frac{\partial V}{\partial c} > 0$ iff

$$\frac{1-e_k}{1-b_v} > \frac{q_k+q_v-1}{1-q_v}$$

which proves the result.

Moving on to sticks, the results on violations and participation follow trivially from the fact that e_k and b_v increase in v, while $e_p(b)$ decreases in s. The effect on compliance is ambiguous because the compliance region shifts over different types as s increases because e_k moves up and b_v moves to the right; hence the result depends on $f(b,e)$. More formally, the absolute number of complying agents is given by

$$K \equiv \int_0^{b_v} \int_0^{e_p(b)} f(b,e)\, de db + \int_{b_v}^{\infty} \int_0^{e_k} f(b,e)\, de db$$

Therefore, we have:

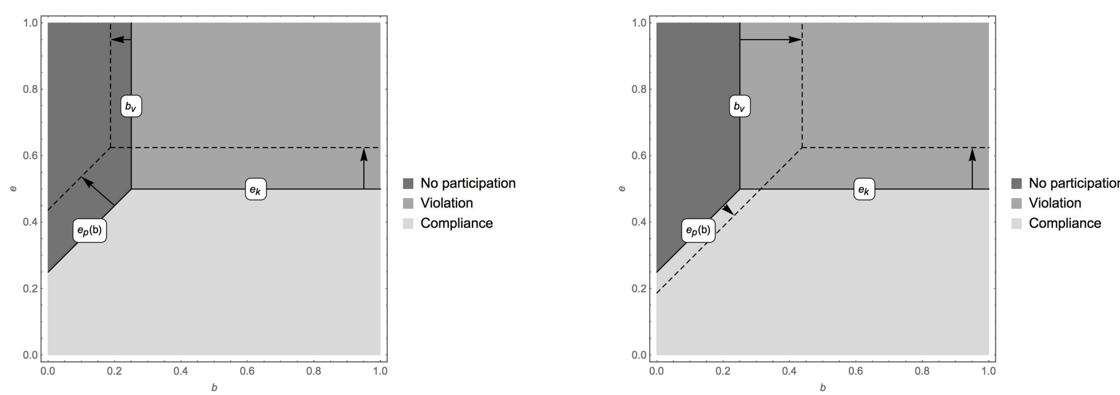
$$\begin{aligned}\frac{\partial K}{\partial s} &= \frac{\partial b_v}{\partial s}\int_0^{e_p(b_v)} f(b_v, e)\,de \\ &\quad + \int_0^{b_v} \frac{\partial e_p(b)}{\partial s} f\left(b, e_p(b)\right) db \\ &\quad - \frac{\partial b_v}{\partial s}\int_0^{e_k} f(b_v, e)\,de \\ &\quad + \frac{\partial e_k}{\partial s}\int_{b_v}^{\infty} f(b, e_k)\,db \\ &= -(1-q_k)\int_0^{b_v} f\left(b, e_p(b)\right) db \\ &\quad + (q_k + q_v - 1)\int_{b_v}^{\infty} f(b, e_k)\,db\end{aligned}$$

where note that the first and the third line cancel each other out because $e_p(b_v) = e_k$. As above, assuming that b and e are independently and uniformly distributed on the unit interval, we have that $\frac{\partial K}{\partial s} > 0$ iff

$$\frac{1-b_v}{b_v} > \frac{1-q_k}{q_k + q_v - 1}$$

which again shows that the ambiguity arises. □

Carrots and sticks have predictable effects on participation, which increases with carrots and decreases with sticks. However, while carrots unambiguously increase compliance, they may or may not reduce violations. The reason is that, by drawing more agents to the activity, carrots also increase the payoff for those who choose to participate and violate the rule. Similarly, sticks unambiguously reduce violations but may or may not increase compliance. This is because while sticks induce some violators to comply, sticks also push some compliers away from the regulated activity. The balance of these opposing effects depends on values of the various policy parameters.



(a) The ambiguous effect of carrots (from $c = \frac{1}{2}$ to $c' = \frac{3}{4}$) (b) The ambiguous effect of sticks (from $s = \frac{1}{2}$ to $s' = \frac{3}{4}$)

Figure OA2: Effects of an increase in the carrot or the stick (simulation parameters: $q_k = q_v = \frac{3}{4}$, $c = s = \frac{1}{2}$).

Figure OA2 visualizes these results. An increase in the carrot pushes up both the compliance and the participation thresholds but reduces the benefit threshold b_v. As a result, the compliance area expands, the no-participation area contracts (resulting in more participation), but the effect on the violation area is ambiguous. The increase in the compliance threshold induces some violators to comply. The decrease in b_v, however, enlarges the full participation group, causing some non-participants to become violators. From a different perspective, as the threshold b_v moves towards the left, some agents that were labeled as low-benefit become high-benefit agents after the increase in the carrot. In the main text, we refer to this effect as the inter-marginal effect of carrots.

Similarly, an increase in the stick increases e_k and b_v, but reduces $e_p(b)$. As a result, the violation area contracts, the no-participation area expands (resulting in less participation), while the effect on the compliance area is ambiguous. The increase in e_k induces some violators to comply, while the reduction in $e_p(b)$ leads some compliers to abstain from participation; in the main text, we refer to this effect as the marginal effect of sticks.

Policy variable	Compliance	Violations	Participation	Compliance rate
c	Increase	Ambiguous	Increase	Ambiguous
s	Ambiguous	Decrease	Decrease	Ambiguous

Table OA3: Effects of an increase in the policy variables on compliance, violations, compliance rate, and overall participation

OA1.2 The Model with Imperfect Monitoring ($p < 1$)

In the previous section, we assumed that the principal monitors each agent with certainty. Here, we introduce imperfect monitoring (or auditing) in the model, $p \in (0,1)$, thereby completing the analysis of the general setup of Section (2.1). As a result of imperfect monitoring, agents who are not monitored may—in the general case—receive a carrot with probability $\phi_c \geq 0$, a stick with probability $\phi_s \geq 0$, or neither of the two with the residual probability $1 - \phi_c - \phi_s \geq 0$.[43]

To illustrate, if $\phi_c = 1$, then non-monitored agents receive a carrot with certainty. This case corresponds to a subsidy given to agents who engage in the activity and are not inspected and found in violation of the rule. Similarly, if $\phi_s = 1$, then non-monitored agents are subject to a stick with certainty. This amounts to taxing the activity unless the agent is inspected and found to comply. Finally, if $\phi_c = \phi_s = 0$, then carrots and sticks are applied only to monitored agents. In the general case, non-monitored agents may be taxed or subsidized with some probability.[44]

Action	Payoff
The agent participates and complies	$\Pi_k(b,e) = b + pq_kc - p(1-q_k)s - e + (1-p)(\phi_c c - \phi_s s)$
The agent participates but violates	$\Pi_v(b) = b + p(1-q_v)c - pq_v s + (1-p)(\phi_c c - \phi_s s)$
The agent does not participate	0

Table OA4: Agent's payoffs with imperfect monitoring

From the payoffs reported in Table OA4, we can calculate the relevant thresholds with imper-

[43] See Dari-Mattiacci et al. (2009) examining the incentive problems that arise from rewarding or punishing non-monitored agents.

[44] In Section OA2 we examine the classical case of taxes and subsidies that are not conditional to the agent's behavior, in addition to (partially) conditional carrots and sticks.

fect monitoring:

$$\begin{aligned} b_v^i &= p\left(q_v s-(1-q_v)\,c\right)-(1-p)\left(\phi_c c-\phi_s s\right)\\ e_k^i &= p\,(c+s)\,(q_k+q_v-1)\\ e_p^i(b) &= b+p\left(q_k c-(1-q_k)\,s\right)+(1-p)\left(\phi_c c-\phi_s s\right) \end{aligned}$$

Imperfect monitoring has two effects on the thresholds that we consider in the analysis. First, for all thresholds, imperfect monitoring dilutes the effect of carrots and sticks proportionally to the probability of monitoring p. Second, while the compliance threshold e_k^i does not depend on the treatment of non-monitored agents, the other two thresholds do. The factor $(1-p)\,(\phi_c c-\phi_s s)$ is the net expected payment that non-monitored agents receive, which is positive if $\phi_c c>\phi_s s$, that is, if non-monitored agents are rewarded in expectation, and negative otherwise.

It is easy to verify that the expanded model including imperfect monitoring generates exactly the same results as the basic model with respect to the effects considered there.

Proposition 2. *Proposition 1 holds also with imperfect monitoring, $p<1$.*

Proof. The results follow from the comparative statics reported in Table OA5, to be compared with Table OA2. Accordingly, we can easily derive Table OA6, which proves the proposition. In

Policy variable	b_v^i	e_k^i	$e_p^i(b)$
c	Decrease	Increase	Increase
s	Increase	Increase	Decrease

Table OA5: Comparative statics

particular, this is true also in the realistic sub-case in which non-monitored agents receive a carrot with certainty, that is, where $\phi_c=1$ and $\phi_s=0$. □

Policy variable	Compliance	Violations	Participation	Compliance rate
c	Increase	Ambiguous	Increase	Ambiguous
s	Ambiguous	Decrease	Decrease	Ambiguous
p	Ambiguous	Ambiguous	Ambiguous	Ambiguous

Table OA6: Effects of an increase in the policy variables on compliance, violations, compliance rate and overall participation with imperfect monitoring

OA2 Taxes and Subsidies

Let us now examine whether adding a tax to the model expands the set of outcomes that the principal can reach. This section generalizes the results illustrated in Section 3. The new payoff matrix is given in Table OA7.

We have now the following thresholds:

$$\begin{aligned} b_v^\tau &= p\left(q_v s-(1-q_v)\,c\right)-(1-p)\left(\phi_c c-\phi_s s\right)-\tau\\ e_k^\tau &= p\,(c+s)\,(q_k+q_v-1)\\ e_p^\tau(b) &= b+p\left(q_k c-(1-q_k)\,s\right)+(1-p)\left(\phi_c c-\phi_s s\right)+\tau \end{aligned}$$

Action	Payoff
The agent participates and complies	$\Pi_k(b,e) = b + pq_k c - p(1-q_k)s - e + (1-p)(\phi_c c - \phi_s s) + \tau$
The agent participates but violates	$\Pi_v(b) = b + p(1-q_v)c - pq_v s + (1-p)(\phi_c c - \phi_s s) + \tau$
The agent does not participate	0

Table OA7: Agent's payoffs with imperfect monitoring and taxes / subsidies

Clearly the compliance threshold is unaffected by τ, because every participant receives τ whether he complies or not. In contrast, participation will be incentivized by τ, both for compliers and for violators, since τ reduces b_v^τ and increases $e_p^\tau(b)$. The addition of τ, however, does not expand the set of outcomes that the principal can reach. To see why, consider that the principal could replicate the same compliance and participation thresholds as above by simply adjusting the carrot and the stick instead of implementing a tax or a subsidy. In short, the principal can replicate the effects of taxes and subsidies by keeping the sum of the carrot and the stick constant (so as to keep the compliance threshold constant) while altering their ratio $\frac{c}{s}$. That ratio should increase if the activity is subsidized (that is, if τ is positive) and decrease if the activity is taxed (that is, if τ is negative), leading to the following proposition:

Proposition 3. *Complementing carrots and sticks with a tax or a subsidy on the regulated activity does not affect any of the results; the principal can replicate the effect of any tax or subsidy, τ, by setting $c' = c + T$ and $s' = s - T$ where:*

$$T = \frac{\tau}{p + (1-p)(\phi_c + \phi_s)}$$

Proof. Let $c' = c + T$ and $s' = s - T$. It is easy to see that $c' + s' = c + s$ and hence $e_k^i = e_k^\tau$. Note also that the value of T that guarantees that $b_v^i = b_v^\tau$ also necessarily yields $e_p^i(b) = e_p^\tau(b)$. To derive the value of T note that

$$\begin{aligned}
b_v^i &= b_v^\tau \\
&\iff \\
p\left(q_v s' - (1-q_v)c'\right) - (1-p)\left(\phi_c c' - \phi_s s'\right) &= p(q_v s - (1-q_v)c) - (1-p)(\phi_c c - \phi_s s) - \tau \\
&\iff \\
p(q_v s - (1-q_v)c) - (1-p)(\phi_c c - \phi_s s) \\
-pT - (1-p)(\phi_c + \phi_s)T &= p(q_v s - (1-q_v)c) - (1-p)(\phi_c c - \phi_s s) - \tau \\
&\iff \\
T(p + (1-p)(\phi_c + \phi_s)) &= \tau
\end{aligned}$$

which yields the result. □

Note that if monitoring is perfect ($p = 1$) then $T = \tau$. Even if monitoring is imperfect ($p < 1$) we still have $T = \tau$ if $\phi_c + \phi_s = 1$, that is, if non-monitored agents are always either rewarded or punished. The common case where all unmonitored participants receive a carrot is one particular example when this condition is met. In the general case, the adjustment to the ratio of carrots and sticks needs to account for the treatment of non-monitored agents. The intuition is that, if there is a positive probability that a non-monitored agent is subject neither to a carrot nor to a stick, that agent fails to be compensated for the effect of the tax or subsidy and hence the correction on the carrot and the stick needs to take that eventuality into account.

OA3 Agents' Perceptions of Accuracy

Here we consider an alternative model where all agents have a benefit of participation $b = 0$. Agent's, however, vary in their perception of the accuracy of enforcement by the principal so that the three-dimensional agent type is (e, q_k, q_v), distributed according to $f(e, q_k, q_v)$. To elaborate: each agent has an idiosyncratic perception of the accuracy of enforcement, which is distributed around the true levels of accuracy $\left(q_k^g, q_v^g\right)$ and may or may not correlated with the agent's costs of effort e. We will show that all the results obtained in the basic setup are confirmed in this alternative setup. The agents' payoffs are reported in TableOA8

Action	Payoff
The agent participates and complies	$\Pi_k(e, q_k) = q_k c - (1 - q_k) s - e$
The agent participates but violates	$\Pi_v(q_v) = (1 - q_v) c - q_v s$
The agent does not participate	0

Table OA8: Agent's payoffs

As in the basic model, we can define three thresholds:

$$\begin{aligned} \hat{q}_v &\equiv \tfrac{c}{c+s} \\ e_k(q_k, q_v) &\equiv (c+s)(q_k + q_v - 1) \\ e_p(q_k) &\equiv q_k c - (1 - q_k) s \end{aligned}$$

The threshold $\hat{q}_v$ is the level of an agent's perception of the enforcement accuracy, q_v, that makes the agent indifferent between violation and abstention, that is, such that $\Pi_v(q_v) = 0$. This threshold is analogous to the threshold b_v in the basic model and partitions the agent's population into two groups: the full-participation group, with $q_v < \hat{q}_v$—whose compliance decision is controlled by the compliance threshold $e_k(q_k, q_v)$—and the partial-participation group, with $q_v \geq \hat{q}_v$—whose compliance decision is controlled by the participation threshold $e_p(q_k)$.

If $c > s$, we have $\frac{1}{2} < \hat{q}_v < 1$. In this case, the threshold $\hat{q}_v$ lies in the interval of admissible values for q_v. This implies that, stochastically, some agents are in the full-participation group and others are in the partial-participation group. If instead $c \leq s$, then $\hat{q}_v \leq \frac{1}{2}$ and hence all agents are in the group with $q_v > \hat{q}_v$, that is, all agents are in the partial-participation group and either participate and comply or stay away from the regulated activity. Figure OA3 illustrates how agents of different types behave when $c > s$.

We are interested in establishing the expected levels of compliance and participation in the population. We start by summarizing the comparative statics of the thresholds of interest in Table OA9.

Policy variables	$\hat{q}_v$	$e_k(q_k, q_v)$	$e_p(q_k)$
c	Increase	Increase	Increase
s	Decrease	Increase	Decrease

Table OA9: Comparative statics

Accordingly, we can derive the results in Table OA10, which are in accordance with those obtained in the basic model.

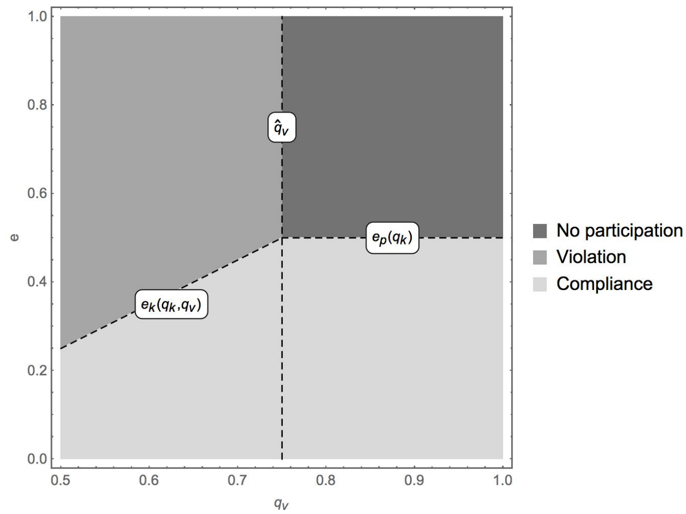


Figure OA3: Agent's behavior (simulation parameters: $q_k = \frac{3}{4}$, $c = \frac{3}{4}$, $s = \frac{1}{4}$).

Policy variable	Compliance	Violations	Participation	Compliance rate
c	Increase	Ambiguous	Increase	Ambiguous
s	Ambiguous	Decrease	Decrease	Ambiguous

Table OA10: Effects of an increase in the policy variables on compliance, violations, and overall participation with imperfect monitoring.