Supplemental Material (not copyedited or formatted) for: Giuseppe Dari-Mattiacci, Alex Raskolnikov. 2021. "The Unexpected Effects of Expected Sanctions." The Journal of Legal Studies 50(1). DOI: https://doi.org/10.1086/713149.

Online Appendix

Unexpected Effects of Expected Sanctions

Giuseppe Dari-Mattiacci and Alex Raskolnikov

Supplemental Material (not copyedited or formatted) for: Giuseppe Dari-Mattiacci, Alex Raskolnikov. 2021. "The Unexpected Effects of Expected Sanctions." The Journal of Legal Studies 50(1). DOI: https://doi.org/10.1086/713149.

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. A potential complier will participate

Supplemental Material (not copyedited or formatted) for: Giuseppe Dari-Mattiacci, Alex Raskolnikov. 2021. "The Unexpected Effects of Expected Sanctions." The Journal of Legal Studies 50(1). DOI: https://doi.org/10.1086/713149.

only if $\Pi_k(b,e) > 0$ at his level of e. More precisely, the agent will participate and 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 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 in which 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.

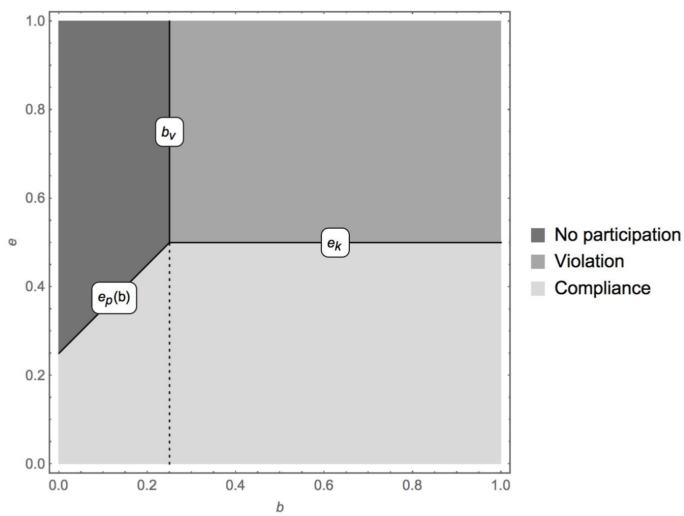


Figure OA1. Agent's behavior (simulation parameters: $q_k = q_v = \frac{3}{4}$, $c = s = \frac{1}{2}$).

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.

Supplemental Material (not copyedited or formatted) for: Giuseppe Dari-Mattiacci, Alex Raskolnikov. 2021. "The Unexpected Effects of Expected Sanctions." The Journal of Legal Studies 50(1). DOI: https://doi.org/10.1086/713149.

Proposition OA**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 OA2a). 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

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

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 (see Figure OA2b), 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 due to the fact that 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

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

Therefore, we have:

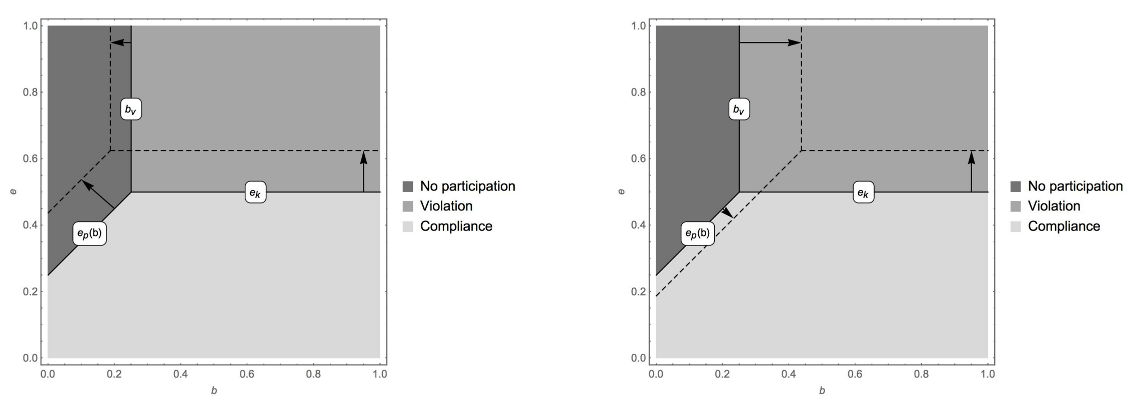
$$\begin{aligned}\frac{\partial K}{\partial s} &= \frac{\partial b_v}{\partial s}\int_0^{e_p(b_v)} f(b_v,e)\,de \\ &\quad + \int_0^{b_v}\frac{\partial e_p(b)}{\partial s} f\left(b,e_p(b)\right)db \\ &\quad - \frac{\partial b_v}{\partial s}\int_0^{e_k} f(b_v,e)\,de \\ &\quad + \frac{\partial e_k}{\partial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

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 the 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 $1 - \phi_c - \phi_s \geq 0$.[1]

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.[2]

Action	Payoff
The agent participates and complies	$\Pi_k(b,e) = b + pq_k c - p(1-q_k)s - e + (1-p)(\phi_c c - \phi_s s)$
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-

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

[2] 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.

Supplemental Material (not copyedited or formatted) for: Giuseppe Dari-Mattiacci, Alex Raskolnikov. 2021. "The Unexpected Effects of Expected Sanctions."
The Journal of Legal Studies 50(1). DOI: https://doi.org/10.1086/713149.

fect monitoring:

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

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 $(1-p)\left(\phi_c c-\phi_s s\right)$ 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 OA**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	$e_p^i\left(b\right)$
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:

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

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 OA**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

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

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, the 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 o f p articipation $b = 0$. 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 be 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.

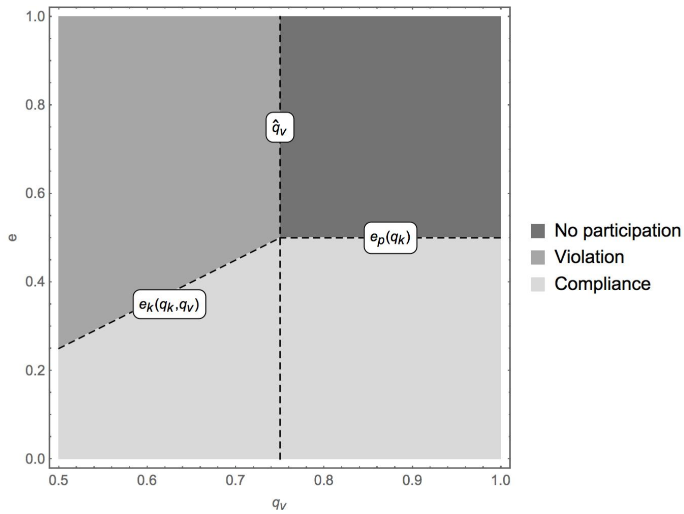


Figure OA3. Agent's behavior (simulation parameters: $q_k = \frac{3}{4}$, $c = \frac{3}{4}$, $s = \frac{1}{4}$).

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.