Dynamic monitoring of adaptive fraud

Alae Baha [*]

August 15, 2022

Click here for the latest version

Abstract

I study the problem of monitoring in a dynamic setting, in which the monitor's ability to detect misbehavior is endogenous: In addition to choosing the amount of fraud, a fraudster can privately develop a hiding technology that makes misbehavior undetectable, and the inspector can invest in R&D to recover her detection ability. In equilibrium, the inspector invests whenever she is sufficiently confident to be lagging technologically. However, too much deterrence of detectable fraud (e.g., high fines or more monitoring) induces the fraudster to invest in hiding technologies, which triggers an arms race and can increase the average quantity of misbehavior. The optimal policy trades off less misbehavior, when detectable, with shorter technological cycles (and higher spending in R&D). The model has applications to digital security, drug smuggling, money laundering, doping, and tax evasion.

1. Introduction

Is there a limit to feasible deterrence? More monitoring and higher fines are often seen as a solution to reducing the amount of misbehavior in society. The success of such policies requires monitors to have the ability to detect and punish misbehavior. However, fraud and crime are often characterized

[*] Toulouse School of Economics (email: baha.alae@gmail.com)

by their adaptive nature and investments in novel hiding technologies can make misbehavior undetectable. For instance, cybercriminals can develop new steganographic techniques that hide data stealing. Drug smugglers can changes their smuggling routes or use more sophisticated transportation methods (submarines, mules, etc.). Money launderers can relocate their capital to tax havens. In these examples, adopting hiding technologies allows criminals to act outside of the scope of enforcement until the monitor develops adapted detection technologies.

Designing monitoring policies in this context can be challenging as a policy that aims at more deterrence of detectable misbehavior increases incentives to develop hiding technologies. For instance, more monitoring makes detectable fraud less rewarding and can reduce misbehavior by fraudsters who do not have access to hiding technologies. However, this “deterrence effect” comes at the expense of making investments in hiding technologies more appealing, reducing the scope of enforcement.[1] As a reaction to these higher incentives, monitors are required to acquire new detection technologies more frequently in order to keep pace with the evolution of hiding technologies which leads to a technological arms race between the two parties. Designing and evaluating policies in these environments requires understanding not only their short-term effects on the deterrence of detectable fraud but also their long-run effect on the development of both fraud and detection technologies.

This paper studies the technological arms race between fraudsters and monitors in a dynamic monitoring setting. Studying the monitoring problem and the technology problem jointly contributes to the literature in three ways: First, it sheds light on the impact of standard monitoring policy tools, e.g monitoring rates and penalties, on the developments of new hiding and detection processes. Second, it allows comparing seemingly diverse policies using only their effects on short term payoffs and flow of information. Finally, the paper contributes to the reputation literature à la Board and Meyer-ter Vehn (2013) by studying an environment in which the technology state can be manipulated by both players due to the arms race.

[1] Riley (2005) reports evidence of such effects in border control where higher monitoring intensities displaced drug smuggling to unguarded portions of the border and made smugglers adopt transportation technologies that are harder to detect, such as submarines, lightplanes, mules, etc.

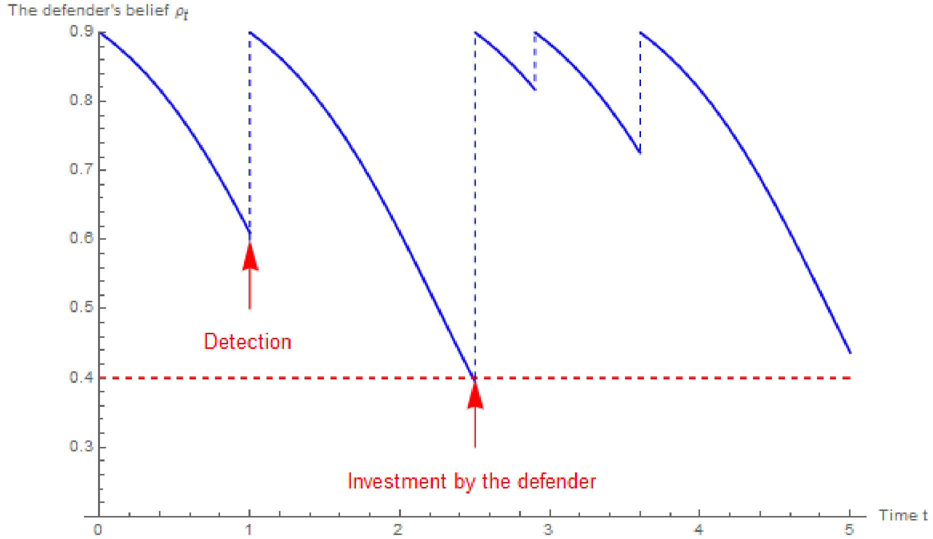
The setting is a discrete-time model where one monitor, which I interpret as a cyber-defender (henceforth the defender), seeks at reducing the harm she incurs from a cyber-attacker (henceforth the attacker). The attacker undertakes two actions: A short-term action which I interpret as the attack intensity, affects flow payoffs for both players, and a long-term action of investing in R&D, which affects the detectability of the attacks. Similarly, the defender undertakes R&D investments in each period, and her ability to detect attacks (henceforth monitoring ability) takes the form of a persistent technology state: attacks are detectable only if she invested last.

Players have asymmetric information about the monitoring ability, known only for the attacker, and the defender learns about this ability from past detections. As detection (or its absence) is informative about whether a hiding technology has been adopted, the defender can use information gathered from the past to update her beliefs about the monitoring ability and make investment decisions optimally. I study equilibria of the game that depend only on the history since the last public signal of either an investment by the defender or an attack detection.

A first step to studying this arms race is the analysis of the attackers incentives to invest. Section 3.1 studies classes of equilibria that can emerge as a function of the gains from becoming undetectable relative to the cost of investing in hiding technologies. I show that any Markov perfect equilibrium of the game belongs to three classes: When the gains are low, the equilibrium is an "entente equilibrium" in which no player invests in R&D, and attacks are always detectable. When these gains are high, two types of equilibria can be sustained: "arms race equilibria" and "Complete hiding equilibria". In arms race equilibria, the two players engage in a perpetual arms race to determine which one has a technological advantage, and the defender is always uncertain about her ability to detect attacks. Finally, complete hiding equilibria are equilibria in which attacks are never detectable, and the attacker always has a technological advantage. When both types of equilibria can be sustained, arms race equilibria are the ones preferred by the defender whereas the attacker prefers complete hiding equilibria.

A policy implication of this result is that policy interventions such as raising fines or increasing the monitoring rates can lead players to engage in a technological arms race. This effect makes evaluating policies challenging

Figure 1: Dynamics of the defender's beliefs in equilibrium



as the global effect on deterrence depends on the cost of the arms race and its effect on the dynamics of attacks. Empirical evidence of such effects were reported in a tax context by Bustos et al. (2022) and in custom control by Yang (2008). In these two contexts, an increase in government monitoring rates made it appealing to attackers to develop technologies that are not detectable by the defender.

Section 3.2 studies the effect of monitoring policies on the intensive margin of investments in arms race equilibria. These equilibria can be described by cycles (see graph above) that start (and ends) after a public signal of a detection (time $t = 1$, $t = 3$ and $t = 3.8$) or an investment by the defender (time $t = 2.5$). Along the cycle, as the defender fails to detect attacks, she becomes more pessimistic about her monitoring ability and, after failing for a given amount of time, she invests in a novel detection technology. The attacker reacts to these investments by investing in a hiding technology with a strictly positive probability in the continuation game which implies that the defender is always uncertain about her ability.

Developing a new fraud technology allows the attacker to remain undetectable until the defender invest in a new detection technology. This implies that he is indifferent and invests with an interior probability only if the defender invests frequently enough. As an implication, policy interventions that increase gains from becoming undetectable also lead to shorter technology cycles. Intuitively, these interventions increase the attacker's gains from investments and, in order to make up for these higher incentives and restore indifference at the beginning of the cycle, the defender's investments have to be such that the technological advantage of the attacker lasts for a shorter amount of time.

Section 4 is dedicated to studying applications and the effect of policy intervention. I show that policies such as higher penalties for detected attacks lead to less intense detectable attacks at the expense of more frequent investments by both players, which creates a trade-off for policy designers. I show that this type of policy interventions can backfire: reducing the defender's payoffs when he is detectable below a cutoff (or equivalently increasing the penalties above a cutoff) leads to more intense attacks on average. As a result, this type of policies can be Pareto-dominated, and there exists an upper bound to feasible deterrence. Due to the R&D effect, I show that policies that induce an arms race can be dominated by entente policies: This is the case if the defender's investment cost in detection technologies is high or the attacker's investment cost is low.

In addition to the attacker's incentives, the intensity of the arms race also depends on the defender's speed of learning. Policies that increase the arrival rate of detections while keeping short-term payoffs unchanged, for instance, through an increase in monitoring rates and a reduction in penalties, make monitoring more informative. I show that in this case, the defender invests more aggressively; that is, for each attacker strategy, she invests more frequently in R&D. As an outcome, these policies lead to lower probabilities of investments by the attacker and a softer arms race. This policy intervention leads to a Pareto improvement as the defender strictly benefits from fewer investments by the defender, whereas the latter is indifferent between the "old" and the "new" policy. As an implication, monitoring and punishment are not perfect substitutes as a policy with higher levels of monitoring (and lower level of punishment) leads to fewer investments in R&D by fraudsters and, therefore, less fraud on average. This result contrasts with models 'a la Becker in which this policy intervention would not affect misbehavior.

1.1. Literature:

This paper relates to the literature on the optimal design of monitoring policies following the seminal work by Becker (1968) (see Polinsky and Shavell (2000) for a survey), Lazear (2006), Eeckhout et al. (2005), Gibson (2019), Blundell et al. (2020) and Telle (2013) for more recent works) that studies the effect of monitoring and levels of punishment on fraud. I contribute to this literature by extending this approach to a dynamic setting where the monitor's ability to detect misbehavior is endogenous and depends on both players' available hiding and detection technologies.

This extension allows for taking into account the attackers' outside options: As a response to a harsher monitoring policy, higher fines, for instance, they can either reduce misbehavior or adopt novel hiding technologies. I provide conditions for this outside option to be relevant. Moreover, studying the monitoring and the arms race problems jointly allows analyzing the effect of monitoring policies on the intensity of the arms race, that is, the frequency of investments by the two players. In contrast to this literature, I show that this implies that harsher monitoring policies can lead to higher levels of misbehavior under certain conditions.

Moreover, I show that higher levels of punishment are no longer a substitute for monitoring intensity (proposition 4). Monitoring has an informational benefit to the monitor as it helps to assert whether attacks are detectable. In this case, penalties are strategic complements for the attacker's investments in hiding technologies, whereas monitoring intensity can either be complements or substitutes to these investments.

More closely related to this paper, following the seminal work of Board and Meyer-ter Vehn (2013), an emerging literature studies learning in environments in which a myopic player's actions depend on his beliefs about a state that is partially controlled by a long term player's investment (See also Board and Meyer-ter Vehn (2022) and Dilmé (2019) for related works). Halac and Prat (2016), Dilmé and Garrett (2019), and Varas et al. (2020) extend this approach to environments with monitoring: A monitor's ability to detect fraud depends on the history of her investments only (Dilmé and Garrett (2019)) or this history and exogenous shocks (Halac and Prat (2016) and Varas et al. (2020)). As I focus on the effect of monitoring policies on the arms race between the attacker and the defender, I depart from these papers by developing a model in which (i) both players are forward-looking and (ii) the monitor's ability depends on both players' investments. This allows studying strategic complementaries between the short-term monitoring problem and the long-term R&D race between the two players.

On an independent work, Marinovic and Szydlowski (2022) study a setting where two forward-looking players (one principal and one agent) face uncertainty about the state and where the arrival rate of detection depends on this state and both players' actions. The authors show that in this setting, the agent has incentives to backload fraud as the principal becomes more pessimistic. I study an environment in which both players can invest in order to change the monitoring ability, which allows studying the arms race, whereas Marinovic et al. focus on the experimentation problem of an agent who wants to commit fraud and learn about this ability.

This paper also relates to the literature about crime displacement that studies how monitoring crime in one location/technology displaces crime to other locations/technologies (see Johnson et al. (2014) for a survey of the criminology literature). See also Yang (2008) for an application to tariffs avoidance, Ladegaard (2019) for the digital drug market, and Gonzalez-Navarro (2013) for the location of auto theft. Finally, this paper relates to the extensive literature that studies models with learning through exponential bandits initiated by Keller et al. (2005) (see Bergemann and Valimaki (2006) and Hörner and Skrzypacz (2017) for surveys). Under an arms race policy, the defender's problem in our model has the same structure as the one studied in this literature; however, payoffs from detection and investments are endogenous as they depend on the attacker's investment strategy in equilibrium.

2. The setting:

2.1. The model:

Consider a game where one defender (player D) and one attacker (player A) repeatedly interact at a fixed time interval Δ. Time $t \in \{0, \Delta, 2\Delta ...\}$ is discrete and the horizon infinite and players discount the future at the same rate e^{-r}.

Actions, policies and states: For each $t \geq 0$, players play a stage game where the defender's ability to detect ongoing attacks depends on a persistent technology state $\theta_t \in \{0, 1\}$ to which I refer as the monitoring ability: Attacks are detectable only if $\theta_t = 1$. Without loss of generality, set $\theta_0 = 1$; That is, attacks are detectable at the beginning of the game.

At the beginning of time t, the state $\theta_{t-\Delta}$ is inherited from the past. Then, players make simultaneous investment decisions $\alpha_t \in \{0, 1\}$ for the attacker and $\delta_t \in \{0, 1\}$ for the defender. Investment $\alpha_t = 1$ allows the development of a new hiding technology that makes the attacks undetectable and shift the monitoring ability to $\theta_t = 0$ at a cost F^A. [2] Similarly, $\delta_t = 1$ is an investment in a detection technology that costs F^D and allows the defender to "regain" her ability to detect attacks by shifting her monitoring ability to $\theta_t = 1$. Not investing $\alpha_t = 0$ and $\delta_t = 0$ is costless.

Following this investment stage, the attacker chooses the intensity of his attack $a_t \in [0, \bar{a}]$ which generates flow payoffs that depend on the monitoring policy. Under a policy π, an attack intensity a generates an expected flow utility $u^A_\pi(a, \theta)\Delta$ for the attacker, an expected flow utility $u^D_\pi(a, \theta)\Delta$ for the defender and leads to detection at a rate $\theta\lambda_\pi(a)\Delta$. We assume that: $u^A_\pi(a, 0)$ and $u^A_\pi(a, 1)$ are single peaked.

The timing of the game at time t is the following:

- Stage 0: The state $\theta_{t-\Delta}$ is inherited from the past, and the defender updates her beliefs ρ_t about it,
- Stage 1: Both players simultaneously make investment decisions in hiding and detection technologies,
- Stage 2: The state θ_t is determined and observed by the attacker,
- Stage 3: The attacker chooses an intensity of attack a_t,
- Stage 4: The outcome of detection is publicly observed, and stage payoffs are realized.

[2] We refer the interested reader to Cabaj et al. (2018) for a review of the techniques that can be used in the Cybersecurity context, for instance, and to Riley (2005) for an application to border control

Law of motion of the defender's monitoring ability: The monitoring ability θ_t is persistent and is determined by the last player who invested in R&D (see figure 1 below): The defender can only detect attacks if she invested last. More formally, denote by $t^D = max\{\tau \leq t : \delta_\tau = 1\}$ the period of last investment by the defender and by $t^A = max\{\tau \leq t : \alpha_\tau = 1\}$ the period of last investment by the attacker. We have:

$$\theta_t = \begin{cases} 1 \text{ If } t^D > t^A \\ 0 \text{ Otherwise} \end{cases}$$

.

Here, we set as a tie-breaking rule that if both players invest in the same period, attacks are not detectable and $\theta_t = 0$.[3]

Information structure and strategies: Set $\omega_t \in \{0, 1\}$ a variable that takes a value of $\omega_t = 1$ if the attack is detected at time t and $\omega_t = 0$ otherwise. A public history at time t in this game consists of detections $\{\omega_0, \omega_\Delta, ..., \omega_{t-\Delta}\}$ and the defender's investments $\{\alpha_0, \alpha_1, ...\alpha_{t-\Delta}\}$.

Let h^{it} be player i's private history at the beginning of time t. The private history for the defender h^{Dt} consists of the public history up to (but not including) time t. A pure strategy for her is a choice of an investment decision δ_t for each t and history h^{Dt}. A (pure) Markov strategy for the defender consists on investment decisions δ_t as a function of her belief ρ_t about θ_t. More formally, a pure Markov strategy for the defender σ^D is:

$$\begin{aligned} \sigma^D :& [0, 1] \to \{0, 1\} \\ & \rho \to \delta \end{aligned}$$

The attacker's private history at the beginning of time t, h^{At}, consists of the public history, investments α, the intensity of the attack a and the state θ up to (but not including) time t. A strategy for the attacker consists of a choice of the investment decision α_t and the intensity of attack for each t and

[3] This tie-breaking rule has no qualitative impact on the equilibrium in the discrete-time version of the game. However, it ensures that the limit case of the equilibrium as Δ goes to zero is equivalent to the equilibrium of the continuous-time version of the game.

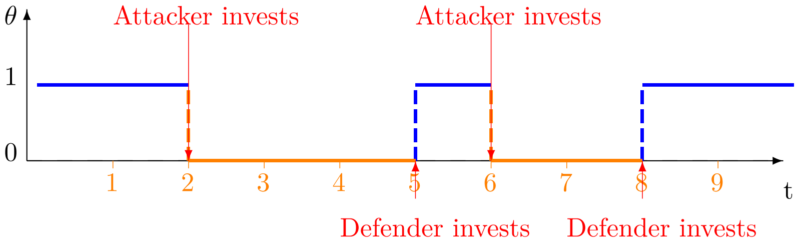


Figure 1: The monitoring ability as a function of time and investments

private history h^{At}. A Markov strategy for the attacker consists of investment decisions and intensity of the attack as a function of (ρ_t, θ_t). Formally, a pure Markov strategy for the attacker is a function:

$$\begin{aligned} \sigma^A :& [0,1] \times \{0,1\} \to \{0,1\} \times [0,\bar{a}] \\ & \rho \times \theta \to \alpha \times a \end{aligned}$$

The payoffs: Players are forward looking and discount future at the same rate $e^{-r\Delta}$ where $r > 0$ is a discount factor. The defender's expected payoffs at time $t = 0$ are:

$$U_t^D = E_{a,\theta,\delta}\left[\sum_{\tau=0}^{\infty} e^{-r(\tau-t)\Delta}\Big(u_\pi^D(a_\tau, \theta_\tau)\Delta - \delta_\tau F^D\Big)\right] \tag{1}$$

The defender's expected instantaneous payoffs at time t in equation (1) can be decomposed into the flow utility given the attacker's action and the state $u_\pi^D(a_\tau, \theta_\tau)\Delta$ and the cost of investing in detection technologies $\delta_t F^M$. Similarly, denote by U_t^A the attacker's value function. We have:

$$U_t^A = E_{\alpha,\delta}\left[\sum_{\tau=t}^{\infty} e^{-r(\tau-t)}(u_\pi^A(a_\tau, \theta)\Delta - \alpha_\tau F^A\right] \tag{2}$$

Where $u_\pi^A(a_\tau, \theta)\Delta$ is the expected benefit from the attack at time t and $\alpha_\tau F^A$ is the cost of investment in hiding technologies.

2.2. Illustrations and policy interventions:

In the model, we allowed for a general definition of utility functions and arrival rate of detection. This allows for flexibility both in terms of policies and economic environments that could be studied. As stated above, a policy is a set of functions $(u_\pi^A(a,0), u_\pi^A(a,1), u_\pi^D(a,0), u_\pi^D(a,1), \lambda(a))$ which determine the arrival rate of detection and players payoffs as a function of the attack intensity a and the state θ. In general, policy interventions can affect one or many of these functions. As this paper aims to disentangle their effects on investment and deterrence, I find it convenient to introduce the two following examples as illustrations for the setting and the effect of policy interventions on these functions.

Application to cybersecurity: The first application of interest is cybersecurity. In this context, a service provider (the defender) seeks to reduce the amount of data stolen by a cyberattacker. In each period, the defender decides whether to invest in improving the ability of her system to detect new types of attacks, whereas the attacker decides the amount of data to steal and the whether to develop a new attack hiding technology. It is of interest s to understand how changes in monitoring policies such as increasing the punishment for detected attacks or increasing the monitoring capacity, or instance by increasing the number of inspectors (cybersecurity officers), for implementing a more stringent red-flags system affect the arms race between these two players.

A first important characteristic of this environment is that it is hard to punish cyberattackers legally as most of them use algorithms to hide their identity to avoid punishment even in case of detection. Moreover, some countries can be more lenient in terms of punishment, and the lack of international cooperation in this context makes judiciary punishment very rare. For this reason, I consider punishment which is independent of the intensity of the attack. [4] Formally consider the following starting policy:

$$\lambda_\pi(a) = am\Delta$$

[4] In section 4, I show the equivalent between a choice of security level and this type of punishment

Where m is the monitoring intensity. Players get flow payoffs:

$$u_\pi^A(.,\theta) = \Big[2\sqrt{a} - \theta m a P\Big]\Delta \mathbf{1}_{a>0}$$
$$u_\pi^D(.,1) = -hE_\theta[a(\rho,\theta)]\Delta$$

1- The effect of punishment: A first policy intervention of interest is a rise in the cost of being detected for the attacker. This policy intervention impacts only the utility function of the attacker when attacks are detectable. The objective is to study how higher levels of punishment increase her incentives to invest for any given strategy by the defender and its long-run impact on the arms race between the two players.

2- The effect of monitoring: Similarly, consider a change of policy that increases the arrival rate of detection. From the attacker's perspective, monitoring and punishment are perfect substitutes: his payoffs depend only on mP. As an implication, a new policy that reduces the punishment and increases monitoring can lead to the same payoffs for both players. In practice, increases in monitoring can be achieved by hiring new inspectors (or cybersecurity officers), improving some aspects of the software, a more stringent red-flags system, etc. [5]

In a one shot-game, this policy intervention has no impact on payoffs. However, from a dynamic perspective, as detections occur more frequently when m is high, this policy intervention makes the defender become pessimist faster (or equivalently, learn faster) under the new policy. It is interesting to study the effect of introducing a more informative policy on the arms race. [6]

3- The effect of the cost of cyberattacks: Finally, as stated above, the cost of being detected often takes the form of a cost of intruding again in the system. This aspect will be studied, and the extent to which punishment and security are equivalent will be developed further.

[5] Note that as opposed to investments which are a qualitative increase in the ability to detect fraud, monitoring is a quantitative shifter where, only when attacks are detectable, higher monitoring leads to more detection

[6] Due to the impact on learning, and as opposed to the literature to the best of my knowledge, this effect makes the choice of monitoring a qualitatively different decision compared to the choice of punishment.

Example 2: Border control:

The second application which is of interest is the problem of border control. Consider a border control agency (the defender) that seeks to detect the smuggling of illicit products to the country (drugs, weapons, etc.). In each period, the defender decides whether to invest in acquiring novel detection technologies, whereas the attacker makes an attack and investment decisions. Investment by the border control agency can be interpreted as acquiring new detection tools such as radars, satellites, patrol vehicles, etc., or acquiring knowledge about more recent smuggling techniques. On the other hand, drug smugglers can acquire vehicles such as submarines or light planes that are hard to detect or change the route that they use.

In addition to the effect of monitoring and punishment discussed above, policy design affects other aspects of the fraud environment: First, policies shape the defender's gains from detection by designing rewards for detected smuggling. Moreover, as often in organized crime, the attacker's payoffs depend on the size of their market/territory. This size can be reduced, for instance, by increasing monitoring in cities or the final consumers of illicit goods. This type of policy measure is complementary to border control as it reduces the demand for criminal activities. However, this reduction is independent of the smuggling technology used, which implies different effects on the arms race. To study the effect of these policy measures, consider a starting policy π such that detection arrives at a rate:

$$\lambda(a) = m\mathbf{1}_{a>0}$$

Here, for simplicity we are assuming the arrival rate of detection $m >$ to be independent from a whenever $a > 0$. Flow payoffs under policy π are:

$$u^A_\pi(a, \theta) = \Big[2\sqrt{\alpha a} - \theta maP\Big]\Delta\mathbf{1}_{a>0}$$

Where $\alpha > 0$ captures the size of the demand (for instance, the territory controlled by the cartel). Note that as opposed to the previous example, punishment P depends on the intensity of attacks. This captures the fact that in practice, the punishment depends on the quantity of seized drugs or weapons. On the other hand, the defender is interested in detecting attacks, in which case she receives a lump sum reward R. Her flow utility is:

$$u_\pi^A(a,\theta) = \theta m R \Delta$$

It is of interest to study the following changes of policies:

1- Downstream policies: Now consider a policy intervention which consists of decreasing the downstream demand for illicit goods by decreasing α.[7] This policy is a demand shifter that affects the attacker's payoffs in both technology states; therefore, it deters attacks with a limited impact on the attacker's investment incentives.

2- Providing incentives to monitors: Finally, consider an increase in the reward for detected smuggling R, and for the sake of exposition, we will interpret investments in detection technologies as the defender's private effort to acquire knowledge about the latest smuggling techniques or newer routes. In this case, the policy intervention affects the defender's payoffs only when attacks are detectable, leading to a change in her incentives to invest for any given strategy by the attacker.

3. Preliminary analysis

Studying equilibria of this game requires (i) understanding when investments can emerge as an equilibrium outcome as a function of the policy choice and investment costs and, (ii) in equilibria with investments, the patterns of these investments and the evolution of the defender's beliefs on the equilibrium path. The objective of this section is to determine these patterns and the determinants of attack intensities in equilibrium.

As investments can be wasteful, the first type of equilibria of interest is such that the attacker has no incentives to invest and, therefore, no player undertakes R&D investments. I refer to these equilibria as “entente equilibria". Formally:

Definition 1. ***(Entente equilibria)*** *An equilibrium is an entente equilibrium if for any time t, and any histories (h_t^A, h_t^D) reached with a strictly*

[7] Examples of such policies can be a large scale intervention to reduce the cartel's area of influence, an increase in monitoring in cities, awareness-raising, partial legalization, etc.

positive probability; we have $\alpha_t = \delta_t = 0$

The opposite of an entente equilibrium is equilibria, in which both players engage in a perpetual arms race where investments never stop. That is, for each point in time, both players invest in R&D in some future period for all possible histories. This can be, for instance, the case when investment costs are low for both players or their incentives, given the monitoring policy, are high. I refer to these equilibria as arms race equilibria. More formally:

Definition 2. ***(Arms race Equilibria)*** *An equilibrium is an arms race if for any private history* h^i_t *reached with a strictly positive probability and for each player* $i \in \{A, D\}$*, there exists a continuation history, reached with a strictly positive probability, such that player* i *invests. That is:* $\forall t : \exists \tau_1, \tau_2 > t : E_{h^{\tau_1}}\Big[\alpha_{\tau_1}\Big], E_{h^{\tau_2}}\Big[\delta_{\tau_2}\Big] > 0$

Finally, for some policies or strategies by the attacker, the defender might not have any incentives to engage in R&D, leading the attacker to always have a technological advantage. These equilibria are referred to as "complete hiding equilibria". Formally:

Definition 3. ***(Complete hiding equilibria)*** *An equilibrium is a complete hiding equilibrium if for any public history* (h^t) *reached with a strictly positive probability attacks are not detectable* $(\theta_t | h^t = 1\ )$*, the defender does not invest* $(\delta t | h^t = 1)$*, and the attacker only invests at the beginning of the game:* $\alpha_0 = 1$*.*

Denote by $a^*(\theta) = argmax_a u^A_\pi(a, \theta)$ the myopic attack intensity as a function of the defender's monitoring ability. We have:

Proposition 1. ***(Types of equilibria)*** *For any policy* π*, an equilibrium exists, moreover, the equilibrium is:*

- ***An entente equilibrium*** *if:*

$$u^A_\pi(a^*(0), 0) - u^A_\pi(a^*(1), 1) < (1 - e^{-r\Delta})F^A$$

- ***An arms race policy or a complete hiding policy*** *if:*

$$u^A_\pi(a^*(0), 0) - u^A_\pi(a^*(1), 1) > (1 - e^{-r\Delta})F^A$$

Proposition 1 describes investments in equilibria from an extensive margin perspective. When the cost of investing in hiding technologies is sufficiently high, the attacker has no incentives to invest in hiding technologies and no player invests in R&D in equilibrium. When this cost is low relative to gains from investments, the attacker engages in R&D, and two types of equilibria can emerge: Arms race equilibria which are preferred by the defender, and a complete hiding equilibrium which is the equilibrium preferred by the attacker. While the complete hiding equilibrium has trivial dynamics, investments and intensity of attacks under an arms race equilibrium depend on the defender's belief and will be analyzed intensively in the rest of the paper.

The defender's beliefs: First, note that detection at time t is fully informative about the state being $\theta_t = 1$. Therefore, in any MPE, beliefs depend only on history since the last detection. I abuse notation and denote by time $t = 0$ the first period after a detection. Finally, I anticipate that, in equilibrium, the attacker invests in a hiding technology only at this period ($t = 0$) and describe only the relevant law of motion of beliefs. Denote by ρ_t the probability that attacks are detectable at the beginning of period t. We have:

$$\rho_t = \begin{cases} 1 & \text{If } t = 0 \\ \rho_{t-\Delta} \frac{1-\lambda_\pi(\hat{a}_{t-\Delta}(1))\Delta}{1-\rho_{t-\Delta}\lambda_\pi(\hat{a}_{t-\Delta}(1))\Delta} & \text{Otherwise} \end{cases}$$

Where $\hat{a}_t(1)$ is the defender's belief about the intensity of attack at time t if attacks are detectable. When Δ goes to zero, beliefs at time t evolve according to:

$$\dot{\rho_\tau} = -\rho_\tau(1-\rho_\tau)\lambda_\pi(\hat{a}_\tau(1)) \tag{3}$$

This equation captures the fact that as the defender fails to detect attacks, her belief about her monitoring ability θ_t decreases. As detection is more likely when the arrival rate of detection is high, beliefs decrease faster for high values of $\lambda_\pi(\hat{a}_\tau(1))$. Note that this law of motion depends on the intensity of the attack at each time τ. Therefore, these intensities have to be determined in the equilibrium path.

Proposition 2. ***(The intensity of attacks)*** *In any Markov Perfect Equi-*

librium, the intensity of the attack is chosen myopically

Proposition 2 means that when attacks are detectable, the attacker undertakes attack decisions without incorporating their effect on the continuation game. As $u_\pi^A(a, \theta)$ is single-peaked, this implies that this intensity depends only on the state in equilibrium. First, note that this result is trivial as long as we consider entente or complete hiding equilibria: In this equilibria, the defender never invests, and the attacker faces a stationary problem whose solution is the same as a one shot game.

Sketch of the proof: Consider any putative arms race equilibrium and assume that in this equilibrium, the attacker invests with a probability 1 for some history reached with a strictly positive probability. This implies that the defender's continuation belief is 0 and gains from investing are the highest. Therefore, either she invests with probability 1, in which case the attacker could benefit from postponing his investments, or the defender invests with probability 0 for all other beliefs, in which case one can construct a deviation where the attacker invests earlier (see appendix). As a result, in any arms race equilibrium, it has to be that $\alpha_\tau < 1$ for all τ and associated histories $h^{A\tau}$.

Intuitively, this implies that never investing is also the best response for the attacker in any arms race equilibrium, which implies that attacks are detectable. As a result, his payoffs are the same as in a situation in which the cost of investing is infinite, in which case, the unique best response is to play the myopic action and never invest. Equivalently, this implies that when detectable, the attacker can get no more than his payoffs from playing his short time attack intensity forever.

In addition to simplifying the dynamics of beliefs in arms race equilibria, proposition 2 implies that any two policies leading to the same short-term payoffs and arrival rates of detection lead to the same investment profiles in equilibrium.

4. Results:

An arms race equilibrium can arise when both players have incentives to invest in R&D in equilibrium. In this section, I study the determinants of investments in these equilibria from an intensive margin and comparative statics and the effect of policies changes on investments and payoffs.

4.1. The arms race equilibrium:

Proposition 3. ***(Arms race equilibria)*** *If $u_\pi^A(a^*(0),0) - u_\pi^A(a^*(1),1) > (1-e^{-r\Delta})F^A$ and $F^D < F^{D^*}$, an arms race equilibrium exists.*
Any such an equilibrium is characterized by an initial belief $\rho_0 \in (0,1)$ and a stopping belief ρ^* such that:

(i) The investment by the attacker $\alpha_0 \in (0,1)$ is :

- $\alpha(\rho) = \begin{cases} 0 \ \forall \rho \in (\rho^*, \rho_0) \\ 1-\rho_0 \ \textit{if } \rho \leq \rho^* \\ 1-\frac{\rho_0}{\rho} \ \textit{if } \rho \geq \rho_0 \end{cases}$

(ii) The investment strategy by the defender:

- $\delta(\rho) = \begin{cases} 1 \ \textit{if } \rho \leq \rho^* \\ 0 \ \textit{otherwise} \end{cases}$

(iii) An equilibrium length of the cycle: $t^A = \frac{1}{r} ln(1 + \frac{rF^A}{u_\pi^A(a^*(0),0) - u_\pi^A(a^*(1),1) - rF^A})$

(iv) The stopping belief $\rho^*(\rho_0)$ is reached at time t^D such that:

$$r + \frac{X}{\rho_0} = \frac{\lambda_\pi(a^*(1))(1-e^{-rt^D}) + \frac{X}{\rho_0(1-\rho_0)}}{e^{\lambda_\pi(a^*(1))t^D} - 1}$$

(v) The initial belief ρ_0 is such that $t^* = t^A = t^D$

Equilibrium cycles: Proposition 2 allows us to describe the beliefs and technology cycles of this game (see figure 3). A cycle starts when both players receive an informative signal about the state of the monitoring ability due to detection of an attack (time t=1, t=2.9 and t=3.6 in figure 3) or to an investment in detection technologies (time t=2.5). The attacker invests with a strictly positive probability $\alpha(\rho_0)$ whenever a new cycle starts. This later probability determines the defender's initial belief $\rho_0 = 1 - \alpha(\rho_0)$.

In the continuation game, the defender learns about her monitoring ability through detection and its absence: As attacks can only be detected if the attacker did not invest, failure to detect attacks makes the defender increasingly pessimistic, and her belief decreases until it reaches a threshold ρ^* in which case she invests with probability $\delta(\rho^*) = 1$ and the cycle ends.

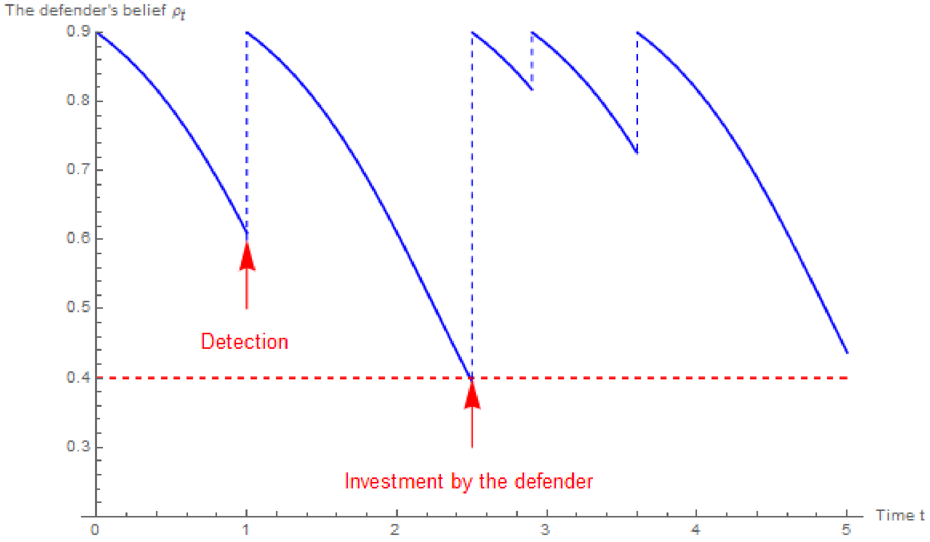
The length of the cycles: I refer to the duration of this learning phase as the length of the cycle t^*, which represents the amount of time that the defender needs to be pessimistic enough to invest. When the attacker invests, he can benefit from undetectable attacks for exactly t^* periods. The length of the cycles has to make him indifferent between his investment decisions (iii).

On the other hand, the defender's incentives to invest in detection technologies depend on her beliefs. As the continuation game after investing is independent of the past histories, her incentives are higher when she is more pessimistic. As an implication, for each initial belief ρ_0, her investment strategy is defined by a unique cutoff $\rho^*(\rho_0)$ such that she invests at the period in which the belief $\rho^*(\rho_0)$ is reached. In other words, given her initial belief, (iv) she experiments for t^D periods before investing. Finally, in order to be in equilibrium, it has to be that (v) $t^A = t^D$.

As the defender's investment strategy is a cutoff strategy, the attacker prefers investing "earlier" in the cycle in order to benefit from undetectable attacks for longer. Therefore, he instantaneously reacts to the increase in the defender's beliefs by investing in hiding technology with (strictly) positive probability.

The initial investment: In arms race equilibria, t^D determines, for each initial investment α_0, the distance between the initial belief ρ_0 and the stopping belief ρ^*. On the other hand, t^A determines the time at which the stopping belief has to be reached in equilibrium in order for the attacker to be indifferent. (vi) links, therefore distance $\rho_0 - \rho^*$ to time t^A and can therefore be interpreted as a condition about the speed of learning: Investments at the

Figure 1: Dynamics of the defender's beliefs in equilibrium



beginning of the cycle have to be such that the defender learns sufficiently fast to invest exactly at time t^A.

Note that (vi) admits at most two solutions, and the attacker is indifferent between the two equilibria. I use as an equilibrium selection that players play the equilibrium preferred by the defender, which is also the Pareto dominant one.

4.2. Comparative statics in arms race equilibria:

In an arms race equilibrium, monitoring policies determine the intensity of attacks and the frequency of investments studied in proposition 3. In order to assess the effect of policy interventions, let us assume from here onward that for all policies $\pi \in \Pi$, flow payoffs and arrival rate of detection are continuously differentiable in the intensity of the attacks and that the arrival rate of detection is non decreasing in the intensity of the attack $(\frac{d\lambda_\pi(a)}{da} \geq 0)$. Moreover, we assume that when there is no attack $(a = 0)$, players get zero payoffs and no detection can occur: $u_\pi^A(0, \theta) = u_\pi^D(0, \theta) = \lambda_\pi(0) = 0$.

A. The effect of more informative policies: The first question of interest in this type of environment is the effect of the arrival rate of detection on the arms race. This arrival rate affects the defender's investments through two effects: First, it increases the likelihood of avoiding wasteful investments by restarting cycles through detection rather than investments. The second effect of the arrival rate is that it affects the defender's speed of learning and, therefore, her investment strategy. In order to study the global effect on the equilibrium, consider two policies π and π' which lead to an arms race.

Definition 4. *Fix $a^*(1)$ and $a^{*\prime}(1)$ the equilibrium intensities of detectable attacks under policies π and π'.The policy π is more informative than the policy π' if it leads to a higher arrival rate of detection:*

$$\lambda_\pi(a^*(1)) > \lambda_{\pi'}(a^{*\prime}(1))$$

.

Now, consider two policies that lead to the same flow payoffs; however, one of them is more informative. These policies can be ranked as follows:

Proposition 4. ***(Pareto dominance of more informative policies)*** *Consider any two policies π and π' which lead to the same equilibrium flow payoffs $\forall \theta, i$: $u^i_\pi(a^*(\theta), \theta) = u^i_{\pi'}(a^{*\prime}(\theta), \theta)$ and denote by π the most informative policy. We have:*
(i) The policy π induces less investments in R&D by the attacker: $\alpha_0^\pi < \alpha_0^{\pi'}$
(ii) The policy π Pareto-dominates the policy π'

Proposition 4 allows comparing policies that are similar from a short-term perspective. Under a more informative policy, the defender learns faster about her monitoring ability. As a result, these policies induce her to invest more aggressively in R&D, that is, for each initial belief ρ_0, her stopping belief $\rho^*(\rho_0)$ is reached earlier. However, from proposition 3, we know that as the attacker's short-term incentives to invest did not change, these two policies entail the same length of the cycles. As a result, (i) the attacker has to invest with a strictly lower probability under the most informative policy. As an implication, flow payoffs given the state are similar under both policies. However, the defender strictly benefits from having fewer investments in hiding technologies in equilibrium, leading the policy π to Pareto dominate the policy π'. [8]

[8] Note that as one best response for the attacker is to never invest in any arms race equilibrium, he is indifferent between the two policies.

From a policy perspective, an example of policies that lead to the same short-term payoffs but differ in their informativeness is policies that entail the same expected punishment for misbehavior using different monitoring intensities. Consider the policies defined in example 1, that is, for $a \in [0,1]$ a policy π is characterized by:

$$\begin{aligned}\lambda_\pi(a) &= am\Delta \\ u^A_\pi(.,\theta) &= \left[2\sqrt{a} - \theta maP\right]\Delta \mathbf{1}_{a>0} \\ u^D_\pi(.,1) &= -hE_\theta[a(\rho,\theta)]\Delta\end{aligned}$$

All policies π and π' with associated monitoring rates and penalties (m,P) and (m',P') respectively such that: $mP = m'P'$ satisfy this condition. Proposition 4 imply that if $m > m'$, then π Pareto dominates π'. Another implication of this result is, as opposed to a situation without investments, the fact that detection is informative about whether the attacker has access to an undetectable attack technology, monitoring and punishment are not perfect substitutes.

B. The (non) deterrence effect of raising penalties: Raising penalties is a policy intervention that makes detection more costly for the attacker. As such, these policies make his flow payoffs lower for any given attack intensity. On the other hand, this policy intervention does not affect the defender's payoffs nor the arrival rate of detection for a given attack intensity. From an equilibrium perspective, this changes the optimal myopic attack intensity, and by extension, given the result in proposition 2, it will change the intensity of detectable attacks. More generally, define by "Purely deterrent policy intervention," any policy intervention whose unique effect is reducing gains from increasing the intensity of detectable attacks. Formally, fix π to be an initial policy and consider a policy π'.

Definition 5. *A policy intervention is purely deterrent if:*
(i) $\forall a:\ (u^A_{\pi'}(a,0), u^D_{\pi'}(a,0), u^D_{\pi'}(a,1), \lambda_{\pi'}(a)) = (u^A_\pi(a,0), u^D_\pi(a,0), u^D_\pi(a,1), \lambda_\pi(a))$ *and,*
(ii) $\forall a$*:* $u^A_{\pi'}(a,1) = u^A_\pi(a,1) + f(a)$ *with* $f(a) \geq 0$ *and strictly decreasing in a*

Here (i) means that the policy intervention does not affect the defender's payoff functions, the arrival rate of detection, and the attacker's flow payoffs when attacks are not detectable. (ii) implies that gains from increasing the intensity of the attack are strictly decreasing under the new policy. We have:

Proposition 5. ***(Effect of purely deterrent policy interventions)*** *Consider any purely deterrent policy intervention $\pi' \neq \pi$ and denote by $a^*(\theta)$ and $a^{*\prime}(\theta)$ the intensities of attacks prior and after this change, we have:*
(i) This change induces less intense detectable attacks: $a^*(1) > a^{*\prime}(1)$
(ii) Technology cycles are shorter under π'
(iii) A limit to deterrence: If $a^*(0) > a^*(1)$, then $\exists \underline{a}$ such that is $a'(1) < \underline{a}$ the policy intervention leads to higher average intensity of attacks

Purely deterrent policy interventions imply that the attacker gains less from increasing his intensity of detectable attacks. As such, these policies induce a "deterrence effect," which is captured in (i). As a counterpart to this deterrence effect, as these policies do not impact flow payoffs for undetectable attacks, they lead to higher short-term gains from investing in hiding technologies. As these gains are higher, the defender will need to invest more frequently in order to keep the attacker indifferent (from proposition 3) which leads to (ii) shorter technology cycles.

Finally, under certain conditions, shorter cycles entail an increase in investments high enough to offset any possible gains from higher levels of short-term deterrence (iii) of detectable attacks. In order to illustrate this effect, consider first policy interventions which lead to lower equilibrium flow payoffs for the defender when attacks are detectable. In this case, she invests more aggressively for any initial belief ρ_0. However, for a low enough, a more aggressive best response is not sufficient to implement short enough technology cycles. As a result, the attacker also changes his investment strategy and invests more frequently in R&D, making the attack less likely to be detectable in which case, the attacks are more intense.

The opposite case in which the defender earns higher flow payoffs when attacks are detectable is more straightforward: The defender's best response is less aggressive, therefore, the two effects always drive the attacker's investments in equilibrium to be higher.

From a policy perspective, this result implies that there is a limit to deterrence which could be achieve through these policies. In particular, raising penalties is a special case of purely deterrent policy interventions and, if too high, they can lead to an intensification of the arms race, inefficient investment and higher and more sophisticated attacks.

To illustrate these effects, consider example 1 and a change of policy to π' which consists of setting $P' = 2P$. This change of policy has: (i) A "deterrence effect" which decreases the intensity of detectable attacks from $a^*(1) = (\frac{2}{mP})^2$ to $a^{*\prime} = (\frac{1}{mP})^2$.
(ii) The second effect of this policy is that it reduces the length of the technology cycle. As $u^A_\pi(a^*(1), 1) > u^A_{\pi'}(a^{*\prime}(1), 1)$, gains from investing are strictly higher. As an implication, the equilibrium under the new policy requires the defender to invest more frequently in order to compensate for this increase in incentives.

Discussion on the equivalence between punishment and security: In a setting in which the attacker pays an intrusion cost whenever he starts a new attack, either these intrusions are not profitable in some state θ which is equivalent to saying $u^A_\pi(\theta) = 0$, or intruding is always profitable. In this situation, each time he is detected, the attacker pays a new intrusion cost to start an attack. In this case, the setting is similar to one in which the attacker pays a penalty which is independent from the intensity of the detected attack.

C. Downstream policies: In many environments, especially the ones related to organized crime, monitors can intervene in many layers of the production of the crime. For instance, consider drug smuggling: A country can monitor smuggling at the borders and at the city level. The key difference between these two modes of monitoring is that monitoring in cities does not induce a technological response in terms of hiding technologies. More generally, I refer to policy interventions that affect payoffs in both states as a downstream deterrence policy. Formally:

Definition 6. *A policy intervention π' is a "downstream deterrence policy" if $\exists f(a), g(a) \geq$ and strictly increasing such that $\forall a$:*

$$
\begin{aligned}
&u^D_{\pi'}(a,0) = u^D_\pi(a,0) - f(a)\\
&u^D_{\pi'}(a,0) = u^D_\pi(a,1) - g(a)\\
&\forall\theta:\ u^A_{\pi'}(a,\theta) = u^A_\pi(a,\theta)\\
&\lambda\pi'(a) = \lambda\pi(a)
\end{aligned}
$$

As opposed to purely deterrent policies, these policies affect the attacker's payoffs in both states, therefore, they have a different impact on the arms race and on the intensity of attacks. In particular, we have:

Proposition 6. ***(Effect of downstream deterrence policies)*** *For any initial policy π and downstream deterrent policy π' such that $u^D_\pi(a^*(0),0) - u^D_\pi(a^*(1),1) > u^D_{\pi'}(a^*(0),0) - u^D_{\pi'}(a^*(1),1)$, we have:*
(i) A short term deterrence effect: $\forall\theta$, $a^*(\theta) > a^{*\prime}(\theta)$
(ii) Longer technology cycles: $t^* < t^{*\prime}$ (iii) Less investments in hiding technologies: $\alpha^\pi_0 > \alpha^{\pi'}_0$

Proposition 6 studies the effect of downstream policies which reduce the attacker's short-term gains from investing. In addition to (i) reducing the intensity of attacks these policies affect the arms race. In particular, (ii) they lead to longer technology cycles. This effect is due to per-period gains being lower, therefore, the attacker only invests if he could "enjoy" being undetectable for longer. This implies that the attacker's response has to be such that he reduces the defender's incentives to invest and a softer best response by the later. To achieve this, it has to be that he invests less in hiding technologies at the beginning of each cycle.

As an implication, these policies do not face the same type of constraints as the purely deterrent policies as they soften the arms race. Therefore, higher levels of deterrence can be achieved, nevertheless they can be more costly as they involve monitoring a wider area for instance.

D.Optimality of arms race policies: In order to determine the optimal policy, policy makers compare the optimal arms race policy to policies which implement entente equilibria. When the attacker's investment cost is high, or the defender's investment cost is low, the former invests in hiding technologies with lower probabilities under an arms race policy. This implies that in this situation, the cost of investments in arms race equilibria are lower. Formally:

Proposition 7. ***The optimal policy:*** *Policies which implement arms race are optimal if and only if:*
Given the attacker's cost of investment F^A, the defender's cost of investing is low: $F^D \leq F^D\bar{(}F^A)$ Given the defender's cost of investment F^D, the attacker's cost of investing is high: $F^A \leq F^A\bar{(}F^D)$

This is an implication of proposition 3 as lower investment costs imply that, given an initial belief ρ_0, the stopping belief ρ^* is increasing in F^D. As a consequence, lower investment costs for the defender make her invest more frequently for any initial belief. However, as the length of the cycle is determined by the attacker's incentives, initial investments have to be such that this length is constant. (vi) in proposition 2 implies that learning has to be slower in that case, which in turn is associated with higher initial beliefs. In conclusion, lower investments costs for the defender are associated with less investment in hiding technologies and, therefore, higher gains for the defender to engage in the arms race.Similarly, an increase in the cost of investment in hiding technologies leads to longer cycles which, in equilibrium, is associated with slower learning and less investment in these technologies.

In addition to the effect of the costs, optimality of the arms race policy is determined by both player's payoffs given states. This aspect is important in contexts such as smuggling where, the flow payoffs depend on the characteristics of the smuggled goods whereas smuggling technologies are not. We can show that:

Proposition 8. *Fix the costs of investments F^A and F^D and the optimal entente policy π'. The arms race policy π is not Pareto dominated only if, for X and $Y > 0$:*
Given $u_\pi^D(a^*(\theta), \theta)$: $u_\pi^A(a^*(0), 0) - u_\pi^A(a^*(1), 1) \leq X$.
Equivalently, $u_\pi^A(a^*(\theta), \theta)$: $u_\pi^D(a^*(1), 1) - u_\pi^D(a^*(0), 0) \geq X$.

Here, the optimal arms race policy is not dominated whenever the defender has low gains from investing. When this is the case, he only invests if the attacker waits for sufficiently long before making investment decisions. This later condition requires investment probabilities by the attacker to be low and as a result, this leads to less wasteful investments in hiding and detection technologies. Similarly, when the cost of having undetectable attacks is high for the defender, she invests frequently unless the attackers invests with sufficiently low probabilities.

This result leads to two predictions: First, arms race should be observed in environments where fraud is costly for instance. As weapons and terrorism are more costly for society that drug smuggling, this threat induces more aggressive monitoring policies and leads to an arms race. This arms race is beneficial as the gains from less attacks are higher that the cost of having frequent investments. On the other hand, when attackers have few gains from investing, the perspective of engaging in an arms race is not costly for society, thus, arms race is desirable.

5. Conclusion

Detection of fraud can be challenging and often depends on both the attack and the detection technologies. As these technologies are endogenously developed, the attacker and the defender often engage in an arms race where the latter faces uncertainty about the former's technology. I construct a model to study these interactions and the effect of policy interventions on the dynamics of attacks and investments .

When engaging in the arms race, the defender learns about her monitoring ability through detection and, as she fails to detect attacks, she becomes more pessimistic and invests in a novel detection technology. The attacker reacts to this investment by investing in a hiding technology with a strictly positive probability, leading to cyclical patterns in these environments. Both the length of these cycles and the investment in each cycle depend on the monitoring policy. More stringent defense policies such as higher penalties lead to less intense attacks when they are detectable at the cost of a potentially more intense arms race in the equilibrium, which creates a tradeoff for the policymaker.

I show that the arms race policies are not Pareto-dominated when the defender's investment cost is low, the attacker's investment cost is high.

When this is the case, the attacker invests with a lower probability in hiding technologies in equilibrium which implies that engaging in the arms race is less costly to the defender and for deterrence.

The model has a few implications in term of optimal design of monitoring policies: First, any two policies that lead to the same "short term deterrence" when technologies are fixed and that have the same cost can be ranked: The most informative policy (such as higher monitoring rather that more punishment) is better as it induces less investments in hiding technologies. More importantly, it highlights some effects of this design that are important for evaluating policies: As higher levels of monitoring lead to more technology adoption, using detect fraud as a proxy for realized fraud can be misleading and one has to also evaluate the impact of this policy on the adoption of hiding technologies.

The model also leads to some verifiable empirical predictions. In environments where detection is informative about the technology state, one should expect serial correlation in detections as one detection implies that in the following period the monitor is more likely to be able to detect fraud. A second prediction, is that harsher defense policies with either more monitoring or higher penalties can lead to a more intense arms race between the two players. These policies can therefore harm the defender by inducing more investments in hiding technologies, in which case, the average intensity of attacks can be higher.

Through this paper, I restricted attention to a game with only one attacker and one defender. A natural extension is to consider multiple attackers in order to study the effect of learning about a whole population on the dynamics. Similarly, in an environment with multiple defenders, incentives to free-ride on each other's learning can emerge and make maximal security more desirable.

References

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.

Bergemann, D. and Valimaki, J. (2006). Bandit problems.

Blundell, W., Gowrisankaran, G., and Langer, A. (2020). Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations. *American Economic Review*, 110(8):2558–85.

Board, S. and Meyer-ter Vehn, M. (2013). Reputation for quality. *Econometrica*, 81(6):2381–2462.

Board, S. and Meyer-ter Vehn, M. (2022). A reputational theory of firm dynamics. *American Economic Journal: Microeconomics*, 14(2):44–80.

Bustos, S., Pomeranz, D., Serrato, J. C. S., Vila-Belda, J., and Zucman, G. (2022). The race between tax enforcement and tax planning: Evidence from a natural experiment in chile. Technical report, National Bureau of Economic Research.

Cabaj, K., Caviglione, L., Mazurczyk, W., Wendzel, S., Woodward, A., and Zander, S. (2018). The new threats of information hiding: The road ahead. *IT professional*, 20(3):31–39.

Dilmé, F. (2019). Reputation building through costly adjustment. *Journal of Economic Theory*, 181:586–626.

Dilmé, F. and Garrett, D. F. (2019). Residual deterrence. *Journal of the European Economic Association*, 17(5):1654–1686.

Eeckhout, J., Persico, N., and Todd, P. (2005). A rational theory of random crackdowns. *manuscript, University of Pennsylvania.*

Gibson, M. (2019). Regulation-induced pollution substitution. *Review of Economics and Statistics*, 101(5):827–840.

Gonzalez-Navarro, M. (2013). Deterrence and geographical externalities in auto theft. *American Economic Journal: Applied Economics*, 5(4):92–110.

Halac, M. and Prat, A. (2016). Managerial attention and worker performance. *American Economic Review*, 106(10):3104–32.

Hörner, J. and Skrzypacz, A. (2017). Learning, experimentation and information design. *Advances in Economics and Econometrics*, 1:63–98.

Johnson, S. D., Guerette, R. T., and Bowers, K. (2014). Crime displacement: what we know, what we don't know, and what it means for crime reduction. *Journal of Experimental Criminology*, 10(4):549–571.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68.

Ladegaard, I. (2019). Crime displacement in digital drug markets. *International Journal of Drug Policy*, 63:113–121.

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, 121(3):1029–1061.

Marinovic, I. and Szydlowski, M. (2022). Monitoring with career concerns. *The RAND Journal of Economics*.

Polinsky, A. M. and Shavell, S. (2000). The economic theory of public enforcement of law. *Journal of economic literature*, 38(1):45–76.

Riley, K. J. (2005). Border control. *The McGraw-Hill Homeland Security Handbook*, pages 587–612.

Telle, K. (2013). Monitoring and enforcement of environmental regulations: Lessons from a natural field experiment in norway. *Journal of Public Economics*, 99:24–34.

Varas, F., Marinovic, I., and Skrzypacz, A. (2020). Random inspections and periodic reviews: Optimal dynamic monitoring. *The Review of Economic Studies*, 87(6):2893–2937.

Yang, D. (2008). Can enforcement backfire? crime displacement in the context of customs reform in the philippines. *The Review of Economics and Statistics*, 90(1):1–14.

6. Appendix

Proof for proposition 1:

Part 1: A policy is an entente policy if and only if $\frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - \frac{\underline{u}^A_\pi}{1-e^{-r\Delta}} > F^A$ is an entente policy.

Note first that an equilibrium without investment exists as, if $\forall t,\ \alpha_t = 0$, the defender's best response is to never invest. The attacker's best response to this strategy is to play according to the no investments benchmark. To show that, note first that as the attacker's action when attacks are not detectable have no impact on the continuation history, his attack intensity when it is the case is the same as the no investment benchmark. Therefore, his flow payoffs are $\bar{u}^A_\pi$ leading to payoffs $\frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - F^A$.[9] We also have:

$$\frac{\underline{u}^A_\pi}{1-e^{-r\Delta}} > \frac{\bar{u}^A_\pi}{1-e^{-r\Delta}}$$

Therefore, no investments is indeed an equilibrium under this policy. Now, I show that this is the unique equilibrium.

Case 1: Assume there exists a such that under policy π: $\bar{u}^A_\pi \leq u^A_\pi(a, 1)$, then the policy π is an entente policy

Proof: Set (σ^A, σ^D) any equilibrium strategies and define by $p(h^t)$ the probability of reaching the history h^t in equilibrium and by $p(h^{t+\Delta}|h^t)$ the distribution of the continuation histories.

For the sake of contradiction, assume that there exists an equilibrium in which the attacker invests and denote by H^1 the set of histories such that attacks are detectable and by H^0 its complementary set where attacks are not detectable. The attacker's payoffs can be written as:[10]

$$U^A_0 = \sum_{t=0}^{\infty} e^{-rt}\Big[\sum_{h^t} p(h^t) E_{a,\alpha}[u_\pi(a,\theta) - \alpha F^A | h^t]\Big]$$

That is, at each time t, he gets an expected utility which depends on the probability of reaching a history h^t times the instantaneous payoffs associated with his action. Note that as the defender's actions are part of the public history, their impact on the attacker's payoffs is taken into account through the history. We have:

[9] Note that here I am abstracting away from the posibility of investing several times as redundant investments are is wasteful and leads to payoffs strictly lower than $\frac{\bar{u}^A_\pi}{1-e^{-r\Delta}}$

[10] Here, the attacker's choice of attack's intensity when attacks are not detectable has no impact on the continuation history, therefore, when it is the case he will choice the same action as in the no investment benchmark

$$
\begin{aligned}
U_0^A &= \sum_{t=0}^{\infty} e^{-rt}\Big[\sum_{h^t} p(h^t)E_{a,\alpha}[u_\pi(a,\theta) - \alpha F^A|h^t]\Big] \\
&= \sum_{t=0}^{\infty} e^{-rt}\Big[\sum_{h^t\in H^1} p(h^t)E_{a,\alpha}[u_\pi(a,\theta) - \alpha F^A|h^t] + \sum_{h^t\in H^0} p(h^t)E_{a,\alpha}[u_\pi(a,\theta) - \alpha F^A|h^t]\Big] \\
&\leq \sum_{t=0}^{\infty} e^{-rt}\Big[\sum_{h^t\in H^1} p(h^t)E_{a,\alpha}[u_\pi^A(a,1) - \alpha F^A|h^t] + \sum_{h^t\in H^0} p(h^t)E_{a,\alpha}[u_\pi(a,\theta) - \alpha F^A|h^t]\Big] \\
&< \sum_{t=0}^{\infty} e^{-rt}\Big[\sum_{h^t\in H^1} p(h^t)E_{a,\alpha}[u_\pi^A(a,1)|h^t] + \sum_{h^t\in H^0} p(h^t)E_{a,\alpha}[u_\pi(a,\theta)|h^t]\Big] \\
&\leq \frac{\underline{u}_\pi^A}{1-e^{-r\Delta}}
\end{aligned}
$$

Here, the first inequality is obtained by using the assumption of some attack intensity delivering a higher utility ($\bar{u}_\pi^A \leq u_\pi^A(a,1)$). The second inequality uses $F^A > 0$ and finally, as these payoffs are reachable a strategy $\sigma^{A'}$ in which $\forall h^t : \alpha(h^t) = 0$, these payoffs are lower than the maximal payoffs that the attacker can get in the benchmark without investments. This, the attacker has a strictly profitable deviation: A contradiction.

Case 2: Assume that $\forall a$: $\bar{u}_\pi^A > u_\pi^A(a,1)$, then the policy π is an entente policy.

Proof: Similarly, assume that there exists an equilibrium in which the attacker invests with a strictly positive probability at some history h^τ at time τ. Define by $p(h^t)$ to be the probability of reaching the history h^t in the continuation game. The attacker invests at time τ implies that $\alpha_\tau = 1$ is one best response, therefore, we have:

$$\begin{aligned}
U^A(h^\tau) &= \sum_{t=0}^{\infty} e^{-r(t-\tau)}\Big[\sum_{h^t\in H^1} p(h^t)E_{a,\alpha}[u_\pi(a,\theta)-\alpha F^A|h^t] + \sum_{h^t\in H^0} p(h^t)E_{a,\alpha}[u_\pi(a,\theta)-\alpha F^A|h^t]\Big] \\
&= -F^A + \bar{u}^A_\pi + \sum_{t=1}^{\infty} e^{-r(t-\tau)}\Big[\sum_{h^t} p(h^t)E_{a,\alpha}[u_\pi(a,\theta)-\alpha F^A|h^t] \\
&\le \frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - F^A - \sum_{t=1}^{\infty} e^{-r(t-\tau)}\Big[\sum_{h^t} p(h^t)E_{a,\alpha}[\alpha F^A|h^t]\Big] \\
&\le \frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - F^A \\
&< \frac{\underline{u}^A_\pi}{1-e^{-r\Delta}}
\end{aligned}$$

Where the first inequality come from using $\bar{u}^A_\pi > u^A_\pi(a,1)$ and the last one is obtained by using $\frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - \frac{\underline{u}^A_\pi}{1-e^{-r\Delta}} > F^A$. The intuition here is that, as the attacker gets strictly higher flow payoffs when attacks are not detectable, he can do no better than keeping his technological advantage forever and get payoffs of $\frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - F^A$, however, these payoffs are lower than the ones he can secure by never investing.

Part 2: An equilibrium exists.

Proof: From part 1, we know that an equilibrium exists whenever $\frac{\bar{u}^A_\pi}{1-e^{-r\Delta}} - \frac{\underline{u}^A_\pi}{1-e^{-r\Delta}} > F^A$. When it is not the case, we show that there always exists a complete hiding equilibrium. By definition of the complete hiding equilibrium, the attacker always invests whenever $\theta = 1$. Therefore, the defender's payoffs are:

$$U^D_0 = \underset{\delta}{max} \frac{\underline{u}^D_\pi}{1-e^{-r\Delta}} - \sum_{t=0}^{\infty} e^{-r\Delta t}\delta_t F^D$$

This leads to $\delta_t = 0$ for all t. Now, we show that the attacker is in best response investing following each detection. As not investing at time 0 leads to a continuation game which is the same as the whole game G, we have:

$$U_0^A = \underset{a,\alpha}{max}(1-\alpha)\Big(u_\pi(a,1) + e^{-r\Delta}U_0^A\Big) + \alpha\Big(\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}} - F^A\Big)$$

This is a linear function of α and $\alpha < 0$ is a best response only if $\alpha = 0$ is also a best response, meaning that:

$$\begin{aligned} U_0^A &= \frac{u_\pi^A(a,1)}{1-e^{-r\Delta}} \\ &\le \frac{\underline{u}_\pi^A}{1-e^{-r\Delta}} \end{aligned}$$

Where the weak inequality comes from the optimality of the attacker's action in the no-investment benchmark. However, as $\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}} - \frac{\underline{u}_\pi^A}{1-e^{-r\Delta}} < F^A$, we have a contradiction and therefore, the unique best response for the attacker is $\alpha_0 = 1$ and we conclude that he is in best response and that whenever $\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}} - \frac{\underline{u}_\pi^A}{1-e^{-r\Delta}} < F^A$, a complete hiding equilibrium exists.

Part 3: If $\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}} - \frac{\underline{u}_\pi^A}{1-e^{-r\Delta}} < F^A$, any Markov Perfect Nash equilibrium under the policy π is either an arms race of a complete hiding equilibrium.

Proof: First note that in part 2, we showed that a complete hiding equilibrium always exists in this case. Now we will show that if there exists another other equilibrium, then this equilibrium is an arms race.

Step 1: In any Markov perfect equilibrium which is not a complete hiding equilibrium, the attacker invests with interior probability at the initial belief ρ_0.

Proof: Assume that the attacker invests with a probability 1 at the initial belief ($\alpha(\rho_0) = 1$). The best response for the defender is to never invest which means that this equilibrium is a complete hiding equilibrium: A contradiction.

Similarly, assume that $\alpha(\rho_0)) = 0$, then $\rho_0 = 1$ and for any time t such that $\rho_t = 1$, $\rho_{t+\Delta} = 1$, therefore the attacker never invests and gets payoffs of $\underline{U}_\pi^A$. The defender's best response is to never invest. A strictly profitable deviation for the attacker is to set $\alpha_0 = 1$ and get payoffs $\bar{U}_\pi^A - F^A > \underline{U}_\pi^A$: A

contradiction.
Therefore, in any equilibrium which is not a complete hiding equilibrium, the attacker invests with an interior probability at belief ρ_0.

Step 2: In any equilibrium, which is not a complete hiding equilibrium, where the attacker invests with probability 1 for some belief $\rho^* \neq \rho_0$, reached with strictly positive probability, we have $U^A_\pi(\rho^*) \geq U^A_\pi(\rho_0)$.

Proof: First assume that there exists a belief ρ^*, reacher with a strictly positive probability, such that the attacker invests with probability 1. ρ^* is reached with a strictly positive probability implies that not investing prior to reaching belief ρ^* is a best response for the attacker. Therefore, his payoffs at belief $\rho \geq \rho^*$ can be rewritten as follows:

$$U^A(\rho_t, 1) = u^A_\pi(a,1)\Delta + e^{-r\Delta}\Big[\lambda_\pi(a)U^A_\pi(\rho_0) + (1-\lambda_\pi(a))U^A(\rho_{t+\Delta},1)\Big] \quad (4)$$

Note first that in any such equilibrium, $\frac{\bar{u}^{A}{}_{\pi}}{1-e^{-r\Delta}} - F^A \geq U^A_\pi(\rho_0)$. Indeed, assume not and denote by t^* the time at which belief ρ^* is reached if there is no detection and by $P_t = \Pi^t_{\tau=0}\lambda(a(\rho_t))$ the probability of reaching each time t. We have:

$$\begin{aligned}
U^A_\pi(\rho_0) &= \frac{1}{1-\sum_{\tau=0}^{t^*-\Delta} P_\tau}\Big[\Big[\sum_{t=0}^{t=t^*-\Delta} e^{-rt}(P_t u(a(\rho_t)))\Big] + P_{t^*}e^{-rt^*}U^A(\rho^*)\Big] \\
&< \frac{1}{1-\sum_{\tau=0}^{t^*-\Delta} P_\tau}\Big[\Big[\sum_{t=0}^{t=t^*-\Delta} e^{-rt}(P_t u(a(\rho_t)))\Big] + e^{-rt^*}U^A_0\Big] \\
&\leq \underset{a_t}{max}\frac{1}{1-\sum_{\tau=0}^{t^*-\Delta} P_\tau}\Big[\Big[\sum_{t=0}^{t=t^*-\Delta} e^{-rt}(P_t u(a(\rho_t)))\Big] + e^{-rt^*}U^A_0\Big] \qquad = \frac{\underline{u}^A_\pi}{1-e^{-r\Delta}}
\end{aligned}$$

The first equation is just a rewriting of the attacker's payoff function as being the sum over t of the utilities he gets once reaching beliefs ρ_t times the probability of reaching belief ρ_t which is P_t. And the term $frac11 - \sum_{\tau=0}^{t^*-\Delta} P_\tau$ comes from the fact that conditional on detection or investment by the defender, the game reboots and the continuation payoffs are U^A_0. The first inequality is due to assuming that the attacker's payoffs are lower at belief

ρ^*. This means that the attacker can get higher payoffs if he could reboot the game and move back to belief ρ_0 whenever belief ρ^* is reached. The maximal payoffs he could get in that case are reached without investing and are therefore weakly lower than $\frac{u^A_\pi}{1-e^{-r\Delta}}$. However, as never investing allows him for secure at least $\frac{u^A_\pi}{1-e^{-r\Delta}}$, this implies that he has a strictly profitable deviation: A contradiction. Therefore, $U^A_\pi(\rho^*) \geq U^A_0$.

Step 3: In any equilibrium, which is not a complete hiding equilibrium, where the attacker invests with probability 1 for some belief $\rho^* \neq \rho_0$, reached with strictly positive probability, we have $U^A_\pi(\rho^*) \geq U^A_\pi(\rho)$ for all $\rho \geq \rho^*$.

Proof: The proof is similar to step 2. Assume not and that there exists a belief ρ', we can construct a strategy which is feasible in which for each $t \geq t^*$, the attacker plays a mixed strategy which follows the same distribution of actions as the one whhich follows time $t(\rho')$. This strategy allows reaching strictly higher payoffs. This strategy is itself weakly dominated by the strategy of never investing and playing the optimal short term action which is a contradiction.

Step 4: In any equilibrium, which is not a complete hiding equilibrium, where the attacker invests with probability 1 for some belief $\rho^* \neq \rho_0$, reached with strictly positive probability, the defender invests with a strictly positive and interior probability at belief $\rho_{t^*-\Delta}$.

Proof: Consider time $t^* - \Delta$. Assume first that the attacker invests with probability 1 at belief $\rho_{t-\Delta}$, the belief ρ^* is never reached in equilibrium: A contradiction. Similarly, Assume first that the attacker invests with probability 0 at belief $\rho_{t-\Delta}$, the attacker's payoffs are:

$$U^A_\pi(\rho_{t^*-\Delta}) = \max_a u(a) - e^{-r\Delta}\Big[\lambda(a)U^A_\pi(\rho_0) + (1-\lambda(a))U^A(\rho_{t^*}, 1)\Big]$$

We have $\forall a$:

$$
\begin{aligned}
&u^A_\pi(a,1) + e^{-r\Delta}\Big[\lambda_\pi(a)U^A_\pi(\rho_0) + (1-\lambda_\pi(a))U^A(\rho_{t^*},1)\Big] \\
&\le \underline{u}^A_\pi + e^{-r\Delta}\Big[\lambda_\pi(a)U^A_\pi(\rho_0) + (1-\lambda_\pi(a))U^A(\rho_{t^*},1)\Big] \\
&< \bar{u}^A_\pi + (1-e^{-r\Delta})F^A + e^{-r\Delta}\Big[\lambda_\pi(a)U^A_\pi(\rho_0) + (1-\lambda_\pi(a))U^A(\rho_{t^*},1)\Big] \le \bar{u}^A_\pi + (1-e^{-r\Delta})F^A + e^{-r\Delta}
\end{aligned}
$$

Investing at time $t-\Delta$ provides payoffs of $\bar{u}^A_\pi + (1-e^{-r\Delta})F^A + e^{-r\Delta}U^A(\rho_{t^*},1)$ which implies that the attacker has a strictly profitable deviation and contradicts the definition of ρ^*. Here the first inequality comes from optimality of $\underline{u}^A_\pi$ in the stationary technology benchmark and the second one is due to $\underline{U}^A_\pi < \bar{U}^A_\pi - F^A$. This implies that in any such an equilibrium, the defender invests with a strictly positive and interior probability at time $t^* - \Delta$ which concludes this proof.

Step 5: There exists no equilibrium which is not a complete hiding equilibrium and in which the attacker invests with probability 1 at some belief.

Proof: Note consider equation 4 and assume that given some continuation payoffs $U^A_\pi(\rho_0)$ and $U^A(\rho_{t+\Delta},1)$, for two attack intensities a and a' with $a > a'$ an attack intensity a provides higher payoffs the ones given by a'. We have:

$$
\begin{aligned}
&u^A_\pi(a) + e^{-r\Delta}U^A(\rho_{t+\Delta},1) + e^{-r\Delta}\lambda(a)\Big[U^A_\pi(\rho_0) - U^A(\rho_{t+\Delta},1)\Big] \\
&\ge u^A_\pi(a') + e^{-r\Delta}U^A(\rho_{t+\Delta},1) + e^{-r\Delta}\lambda(a')\Big[U^A_\pi(\rho_0) - U^A(\rho_{t+\Delta},1)\Big]
\end{aligned}
$$

This inequality can be rewritten as:

$$
u^A_\pi(a) - u^A_\pi(a') + \Big(\lambda(a) - \lambda(a')\Big)e^{-r\Delta}\lambda(a)\Big[U^A_\pi(\rho_0) - U^A(\rho_{t+\Delta},1)\Big] \ge 0
$$

As $\lambda(a)$ in increasing in a, this implies that for all belief ρ with the associated time $\tau(\rho)$ such that $\Big[U^A_\pi(\rho_0) - U^A(\rho_{\tau+\Delta},1)\Big] \ge \Big[U^A_\pi(\rho_0) - U^A(\rho_{t+\Delta},1)\Big]$, we have:

$$
u^A_\pi(a) - u^A_\pi(a') + \Big(\lambda(a) - \lambda(a')\Big)e^{-r\Delta}\lambda(a)\Big[U^A_\pi(\rho_0) - U^A(\rho_{\tau+\Delta},1)\Big] \ge 0
$$

This implies that attacker would again prefer the higher action. We also have from step 3 that $\forall \rho : U_{\pi}^{A}(\rho) < U_{\pi}^{A}(\rho^{*})$. This implies that the lowest attack intensity is played at time $t^{*} - \Delta$.
From the defender's perspective, as she invests at time $t^{*} - \Delta$ and not after t^{*}, we have: $U_{\pi}^{D}(\rho_{t^{*}}) \geq U_{\pi}^{D}(\rho_{t^{*}-\Delta})$. This implies:

$$\begin{aligned}&\rho u_{\pi}^{D}(a(\rho), 1) + (1-\rho)\underline{u}_{\pi}^{D} + e^{-r\Delta} U_{\pi}^{D}(\rho^{*})\\ \leq &\underline{u}_{\pi}^{D} + e^{-r\Delta} U_{\pi}^{D}(\rho^{*})\\ &\iff u_{\pi}^{D}(a(\rho), 1) \leq u_{\pi}^{D}(0)\end{aligned}$$

Not, as $\forall \rho : \; a(\rho) > a(\rho_{t^{*}-\Delta})$, we have:

$$\forall \rho : \; u_{\pi}^{D}(a(\rho), 1) \leq u_{\pi}^{D}(0)$$

This implies that $\forall \rho < \rho^{*} : \; U_{\pi}^{D}(\rho) \leq U_{\pi}^{D}(\rho_{t^{*}-\Delta})$. This implies that $U_{\pi}^{D}(\rho_{0}) - F^{D} < U_{\pi}^{D}(\rho_{t^{*}-\Delta})$ which contradicts investing being a best response for the defender. Therefore, the unique equilibrium such that the attacker invests with probability 1 at some belief is the complete hiding equilibrium.
Step 6: Any equilibrium in which not investing is a best response for the attacker for all beliefs is an arms race equilibrium.
Proof: As never investing is a best response for the attacker, his payoffs are $\underline{U}_{\pi}^{A}$ and, by optimality of $a^{*}(1)$, the attack intensity when attacks are detectable is the same as the no investment benchmark, therefore, for any belief, the defender's payoffs when she does not invest can be rewritten as:

$$U^{D}(\rho) = \rho \bar{u}_{\pi}^{D} + (1-\rho)\underline{u}_{\pi}^{D} + e^{-r\Delta} U_{\pi}^{D}(\rho_{t+\Delta}) \tag{5}$$

This function is strictly increasing in ρ whenever $\bar{u}_{\pi}^{D} > \underline{u}_{\pi}^{D}$ (See Keller, Rady and Cripps (2005).),[11], therefore, if there exists a belief such that the defender invests with a strictly positive probability, she also invests with probability 1 for all lower beliefs which implies that there exists no history such that she stops investing and this equilibrium is an arms race.

[11] A more explicit computation is provided when proofing step 1 of proposition 3

On the other hand, if $\bar{u}_{\pi}^{D} < \underline{u}_{\pi}^{D}$, the defender never invests and the unique equilibrium is a complete hiding. I conclude that any equilibrium is therefore either a complete hiding or an arms race equilibrium.

Proof for proposition 2:

Note that in both the complete hiding and the entente equilibrium, this equivalence is trivial as the state is fixed for all the duration of the game and (i) both players get similar short term payoffs to the non investment benchmark and (ii) the defender never invests in equilibrium and (iii) the attacker's either never invests if it is an entente policy or invests with probability 1 and time 0.
Now, assume that the equilibrium is an arms race equilibrium, and denote by π and π' two strongly short-term-equivalent policies. Consider any equilibrium investment probabilities under policy π: $\alpha(\rho)$ and $\delta(\rho)$. At reach time t, each player's instantaneous payoffs depend only on the state, and are equal under both policies. Moreover, as the two policies are strongly short-term-equivalent policies and using the fact that the attacker plays the myopic attack intensities in any arms race, this implies that playing any investment probability distribution generate the same probability distributions over investments deliver the same payoffs under policies π and π'. Finally, as the arrival rate of detections is the same under both policies ($\lambda_{\pi}(a_{\pi}^{*}) = \lambda_{\pi'}(a_{\pi'}^{*})$), the expected state given any history is the same under both policies (equivalently, the law of motion of beliefs given investment probabilities is the same under both policies), therefore, the expected gains from investing for player i player j plays a given distribution over investments is the same over both policies, therefore, α_{π} and δ_{π} are also best responses against each other under policy π' which concludes the proof.

Proof for proposition 3:

Step 1: In any arms race equilibrium, investments strategies are in cutoff strategies with:

$$\alpha(\rho) = \begin{cases} \in (0,1) \text{ if } \rho = \rho_0 \\ 0 \text{ Otherwise} \end{cases} .$$

$$\delta(\rho) = \begin{cases} 1 \text{ if } \rho \leq \rho^* \\ 0 \text{ Otherwise} \end{cases} .$$

Proof: First, note that by proposition 1 step 5, in any arms race equilibrium, the attacker invests with interior probability at the initial belief ρ_0 and from step 6 the attacker's action is the myopic action.
Now, the defender's payoffs are:

$$U^D(\rho) = \delta U^D + (1-\delta)\Big[\rho \bar{u}^D_\pi + (1-\rho)\underline{u}^D_\pi + e^{-r\Delta}U^D_\pi(\rho_{t+\Delta})\Big] \qquad (6)$$

For any belief ρ^* such that the defender invests,

Now, denote by σ' the strategy which consists of playing $a'(\rho) = a(\rho)$ and $\alpha'(\rho) = 0$ the strategy which consists of never investing and playing the same attack intensity as under the equilibrium strategy. From step 4, σ' is a best response as not investing is a best response for all beliefs under the equilibrium strategy. This strategy delivers payoffs $U'(\rho)$ such that:

$$\forall \rho: \; U'(\rho) \leq \frac{u(a^*) - ma^*S}{1 - e^{-r\Delta}}$$

Where a^* solves $(u')^{-1}(a^*) = mS$. This implies that $U^A_0 = \frac{u(a^*)-ma^*S}{1-e^{-r\Delta}}$. Moreover, as the upper bound on the right hand side is uniquely reached through a stationary attack intensity a^*, we obtain $\forall \rho: \; a(\rho, 1) = a^*$.
Now, the defender's value function at time t can be rewritten as:

$$U^D_t = -h(\rho a^* + (1-\rho))\Delta + \max_\delta \, e^{-r\Delta}\Big[\delta(U^D_0 - F^D) + (1-\delta)E[U^D_{t+\Delta}]\Big] \quad (7)$$

This problem is analogical to the one player version of Keller et. al (2005). Investing ($\delta_t = 1$) delivers payoffs that are independent from period t's state, therefore, it plays a similar role as pulling the safe arm, whereas not investing means that the defender continues experimenting and gets payoffs that depend on period t's: Investments play therefore the same role as pulling the risky arm in K.R.C. The payoff function in 8 is monotonically decreasing in ρ, therefore, the defender's strategy is a cutoff strategy (as $\forall \rho, \rho'$ with

$\rho' < \rho : U^D(\rho) < U^D(\rho_0) - F^D \implies U^D(\rho') < U^D(\rho_0) - F^D)$. Therefore, as Δ goes to zero, in any arms race equilibrium: $\delta(\rho) = \begin{cases} 1 \text{ if } \rho \leq \rho^* \\ 0 \text{ Otherwise} \end{cases}$.

Finally, the difference between the attacker's payoffs as a function of states is

$$U^A(\rho, 0) - U^A(\rho, 1) = \sum_{\tau = t(\rho)}^{t^*} e^{-r(\tau - t(\rho)}(u(1) - u(a^*))\Delta$$

This function is strictly decreasing in $t(\rho)$, therefore, either the attacker never invests or invests at the beginning of the cycle which concludes our proof.

Part 2: Equilibrium characterization:

Step 1: The equilibrium attack intensities are: $a(\rho, \theta) = \begin{cases} 1 \text{ if } \theta = 1 \\ a^* \text{ Otherwise} \end{cases}$

Proof: See step 4 in part 1.

Step 2: The length of the cycle is $t^A = \frac{1}{r} ln(1 + \frac{rF^A}{u(1) - u(a^*) + rmS - rF^A})$

Proof: From part 1, we have $\alpha(\rho_0) \in (0, 1)$. Denote by t^A the period at which the defender invests. For the attacker to be in best response, it has to be that:

$$\sum_{\tau=0}^{t^*} e^{-r\tau} u(a^*)\Delta = -F^A + \sum_{\tau=0}^{t^*} e^{-r\tau} u(1)\Delta$$

$$\iff \sum_{\tau=0}^{t^*} e^{-r\tau}(u(1) - u(a^*))\Delta = F^A$$

$$\iff F^A = \Big[u(1) - u(a^*)\Big]\Delta \frac{1 - e^{-rt^*}}{1 - e^{-r\Delta}}$$

$$\iff e^{-rt^*} = 1 - \frac{F^A(1 - e^{-r\Delta})}{(u(1) - u(a^*))\Delta}$$

$$\iff t^* = \frac{1}{r} ln\Big(\frac{(u(1) - u(a^*))\Delta}{(u(1) - u(a^*))\Delta - F^A(1 - e^{-r\Delta})}\Big) = \frac{1}{r} ln\Big(1 - \frac{F^A(1 - e^{-r\Delta})}{(u(1) - u(a^*))\Delta - F^A(1 - e^{-r\Delta})}\Big)$$

Step 2: The defender's stopping belief satisfies:

Proof: Consider any belief $\rho \in (\rho^*, \rho_0)$. In this belief, the defender does not invest and her value function evolves according to:

$$U_t^D = -h(\rho a^* + (1-\rho))\Delta + \underset{\delta}{max}\ e^{-r\Delta}E[U_{t+\Delta}^D]\Big] \tag{8}$$

Using $1 - r\Delta$ as a limit for $e^{-r\Delta}$ when delta goes to 0, I follow Keller et al., I rewrite the value function in equation 8 as:[12]

$$rU^D(\rho) = -h(\rho a^* + (1-\rho)) + am\rho\Big[U_0 - U^D(\rho) - (1-\rho)U^{D'}(\rho)\Big]$$

The general solution to this differential equation is:

$$U^D(\rho) = -\frac{h}{r} + \frac{\rho}{r + a^*m}(h(1-a^*) + a^*mU_0^D) + C(1-\rho)(\frac{\rho}{1-\rho})^{-\frac{r}{a^*m}}$$

Finally, using value matching ($U_0^D = U^D(\rho^*) + F^A$), we solve for C and have:

$$\begin{aligned} C =& \frac{1}{1-\rho^*}\Big(\frac{\rho^*}{1-\rho^*}\Big)^{\frac{r}{a^*m}}\Big[\frac{h}{r} + U_0^D - F^D\Big] \\ &- \Big(\frac{\rho^*}{1-\rho^*}\Big)^{1+\frac{r}{a^*m}}\frac{1}{r+am}\Big(h(1-a^*) + a^*mU_0^D\Big) \\ &= \frac{1}{1-\rho^*}\Big(\frac{\rho^*}{1-\rho^*}\Big)^{\frac{r}{a^*m}} \\ &\Big[\Big(\frac{h}{r} + U_0^D - F^D\Big) - \rho^*\Big(h(1-a^*) + a^*mU_0^D\Big)\Big] \end{aligned}$$

We finally get, given ρ^* and U_0^D:

$$\begin{aligned} U^D(\rho) =& -\frac{h}{r} + \frac{\rho}{r+a^*m}(h(1-a^*) + a^*mU_0^D) \\ &+ \frac{1-\rho}{1-\rho^*}\Big(\frac{\frac{\rho^*}{1-\rho^*}}{\frac{\rho}{1-\rho}}\Big)^{\frac{r}{a^*m}} \\ &\Big[\Big(\frac{h}{r} + U_0^D - F^D\Big) - \rho^*\Big(h(1-a^*) + a^*mU_0^D\Big)\Big] \end{aligned}$$

[12] as step 2 boils down to an adaptation of the cooperative problem in Keller et al., some parts of the proof will be skipped and I refer the interested reader to that paper for a more detailed proof

Finally, we use U_0^D as being a fixed point for this equation and smooth pasting, we have:

$$\begin{aligned} U(\rho^*) &= \frac{1}{r + a^* m \rho^*}\Big[- h(\rho^* a^* + (1 - \rho^*)) + a^* m \rho^* U_0^D \Big] \\ &= U_0^D - \frac{1}{r + a^* m \rho^*}\Big[h(\rho^* a^* + (1 - \rho^*)) + r U_0^D \Big] \end{aligned}$$

Finally, using simple algebra we derive $\rho^*(\rho_0)$ described in proposition 2.

Part 3: Existence

Proof: (to be completed) The proof is structured as follows: First, I assume that there exists some F^D such that an arms race equilibrium exists for some t^A.

Step 1: We show that for all $t^{A'} > t^A$, an arms race equilibrium exists, therefore, the set of lengths of the cycle that can be supported as an arms race equibrlium is compact.

Step 2: using the fact that the defender's payoffs are continuous and increasing in ρ_0, we show that for all ρ_0, and all $F^{D'} = F^D \; U_0^D - F^D > 0$, $U_0^D - F^D > 0 \implies U_0^D - F^{D'} > 0$, therefore, the set of lengths of the cycle that are supported under $F^{D'}$ is bigger which concludes the proof.