*The University neither endorses not condemns opinions expressed in this thesis.*

# THÈSE

**En vue de l'obtention du**

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par l'Université Toulouse 1 Capitole**


**Présentée et soutenue par**

**Alae BAHA**


Le 6 juillet 2022

# Essays in microeconomic theory

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
**TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par
**Renato GOMES et Carole HARITCHABALET**

Jury

**M. Daniel GARRETT,** Rapporteur
**Mme Sinem HIDIR,** Rapporteure
**M. Takuro YAMASHITA,** Examinateur
**M. Renato GOMES,** Directeur de thèse

# Essays in economic theory

Alae Baha

**Acknowledgements**

First, I would like to express my deepest gratitude to my thesis advisor Renato Gomes for his advice, feedback, patience, encouragement, and enthusiasm throughout the Ph.D. I am also thankful to my co-advisor Carole Haritchabalet, and to Takuro Yamashita and Daniel Garrett for their advice at different stages of the Ph.D. I am also grateful to members of the Theory and industrial organization groups at TSE, in particular Jacques Cremer, Bertrand Gobillard, Johannes Hörner, Bruno Julien, Patrick Rey and Anna Sanktjohanser for their encouragement during the workshops and insightful discussions during meetings. Finally, I am also thankful to Nour Meddahi for his efforts to provide a better research environment for Ph.D.s and to the administrative staff at TSE, and in particular Christel Gilabert, Elvire Jalran, and Ludmila Namolovan-Stephan for their help before and during the jobmarket and the Ph.D. defense.

I am also thankful to TSE students for their discussions about research and for making my Ph.D. an enjoyable experience. In particular, I want to take my office mates Amirreza Ahmadzadeh, Paul-Henri Moisson, and Stephan Waizmann, and my co-author Esteban Muñoz and Christopher Sandmann for all the research-related discussions around coffees, whiteboards, and meals. Finally, I wanted to thank my classmates Luise Eisfeld, Alipio Ferreira, Ivan Salles Fonseca, Antoine Jacquet, Gosia Majewska, Romain Mongeau, Thuy Nguyen, Vatsala Shreeti, Fernando Stipanicic, Matthieu Tarbé, Viridiana Torres, Philipp Wangner, Jun Yan and Li Yu as well as fellow basketball players Simon Arthur, Luca Benatti, Léa Bignon, Captain Lisa Botbol, Max Lesellier. Knowing all of you is an immense privilege.

I am also profoundly grateful to my parents and siblings for their encouragement and support.

**Résumé**

Cette thèse s'inscrit dans un programme de recherche qui étudie la conception de mécanismes optimaux dans des environnements où le processus de production, et donc l'intensité des asymétries d'information, est endogène. Il se compose de trois articles, chacun étudiant l'effet des choix de mécanismes ex-ante sur l'intensité des asymétries d'information ex-post ainsi que ses effets en terme d'optimalité du mécanisme choisi.

Le premier article se consacre à l'étude de l'approvisionnement optimal dans un environnement dans lequel les agents prennent des décisions d'investissement qui leur permettent de produire uniquement dans un état du monde inconnu ex-ante et a des applications dans l'approvisionnement en présence d'incertitude technologique telles que l'approvisionnement (ex-ante) en vaccins. Le deuxième article étudie les politiques de contrôle optimales dans un environnement dans lequel la capacité du moniteur à détecter la fraud/crime dépend à la fois de ses investissements passés et de ceux de l'agent et a des applications dans la cybersécurité, le trafic de drogue, le blanchiment d'argent, l'évasion fiscale et le dopage. Enfin, le troisième article étudie un problème de production dans lequel la productivité de l'agent dépend d'une allocation de ressources (ou de temps) non observable et trouve des applications en économie du travail.

**Abstract**

This dissertation is part of a research agenda which studies optimal mechanism design in environments in which the production process, an thus the intensity of the asymmetries of information, is endogenous. It consists of three papers, each which studies the effect of ex-ante mechanism choices on the ex-post intensity of the asymmetries of information and its implication in terms of optimal mechanisms.

The first paper studies the optimal procurement in an environment in which agents make investment decisions that allow them to produce only in ex-ante unknown state of the world and has applications in procurement under technology uncertainty such as the procurement of vaccines. The second paper studies the optimal monitoring policies in an environment in which the monitor's ability to detect misbehavior depends on both her and the agent's past investments and has applications in cyber security, drug smuggling, money laundering tax evasion and doping. Finally, the third paper studies a production problem in which the agent's productivity depends on an unobservable resource (or time) allocation and has applications in labor economics.

# Chapter 1

# Designing Contracts For Technology Procurement

**Abstract**

This paper studies the problem of optimal procurement under uncertainty about relevant production technology. A buyer chooses a symmetric procurement mechanism that depends on this technology and the relevant sellers. Sellers observe this mechanism and choose production technologies. In this setting, the choice of mechanisms shapes both the agent's rents and their technology adoption profiles. I show that the optimal mechanism induces mixing in the technology adoption by the least efficient agents' types whereas the most efficient ones adopt only the technology that is most likely to succeed. This mechanism can be implemented through first-price auctions in which case mixing allows the principal to benefit from both more efficient trade and a more aggressive bidding behavior.

## 1.1. Introduction:

In environments involving technology procurement, sellers often undertake specialization decisions to acquire the capacity to produce a good or fulfill a contract. This specialization can, for instance, take the form of technology choice, the design of product characteristics, the development of project-specific human capital, etc. A key challenge, inherent to these choices, is that the relevant specialization is often unknown at the moment of decision making due for instance to uncertainty about the future needs of the buyer, about the development of complementary products/technologies, or about the research path which might lead to a breakthrough.

From the buyer's perspective, endogenous specialization by the sellers adds another dimension to the mechanism design choice: Trade mechanisms shape not only the rent of the relevant sellers but also the levels of competition in each technology niche as well as information conditional on technology choices. This paper investigates these aspects and studies how the principal's choice of mechanisms impacts specialization by the sellers as well as the optimal symmetric mechanism in this type of environments.

In order to fix ideas, consider the recent development of Covid vaccines. In this example, companies had to choose a technology to develop a vaccine which could be M-RNA, whole microbe approach, or subunit approach. At the moment of choosing a technology, the type of vaccines that could succeed was unknown, however, this choice still affects the future levels of competition in the market for each potential relevant technology. In this case, by being more generous (in terms of payments) towards one technology, the buyers can induce more firms to adopt that specific technology at the expense of less competition in other technology markets. In this case, specialization is a key determinant of the optimal mechanism as it affects both the levels of competition and the risk of no supply of vaccines.

In order to study these aspects, this paper develops a model of a multi-market principal who has the ability to commit to future mechanisms. The principal faces $N$ ex-ante identical potential sellers who are characterized by their production costs, which are private information, and their endogenous technology choices. These choices allow sellers to produce in one of two exogenous and unknown technology states. In this game, the principal chooses (symmetric) technology recommendations and trade rules as a function of the technology state, agents' reports, and the realized technology adoption profile. After observing this choice, agents undertake simultaneous specialization decisions. Finally, the state is realized and publicly observed and the good is traded according to the chosen mechanism.

The key novelty of this paper is that mechanism choice for a given realization of the technology state has an impact on overall specialization decisions taken by the agents. This implies that the optimal mechanism should internalize cross-markets externalities that can take two forms:

- A competition externality resulting from some agents changing specialization decisions.
- An informational externality due to the technology choice being informative about the agent's cost.

These externalities imply that the cost of implementing separating mechanisms is partially determined by mechanism choice, therefore, the principal's problem incorporates the effect of his choice on the technology adoption profiles. The main result of the paper shows that the optimal mechanism is such that the most efficient subset of types adopts the technology which is the most likely to succeed, whereas the other types strictly mix between the two technology choices. Proposition 2 shows that any mechanism with such a structure can be implemented through first-price auctions with maximum and minimum bids. In this case, maximum bids play the same role as in standard models and exclude the least efficient types from trade. On the other hand, minimum bids in one technology market induce pooling at the bottom and divert more efficient types from this technology, therefore, they provide incentives for those types to adopt the technology which is the most likely to succeed.

Now, to provide intuition behind the optimality of mixing in an auction context, first consider a mechanism that allocates the least efficient types to technology "x" and the most efficient ones to the technology "y". Consider a mixing structure that keeps the probability of trade in each technology market constant (or equivalently, the probability of facing at least one seller constant). Adopting this structure provides two sources of gains to the principal: An efficiency effect and a competition effect.

These effects are driven by the fact that mixing induces first-order stochastic dominance conditional on adopting technology $x$. That is, conditional on adopting this technology, the probability that the agent's cost is lower than some given level is now higher. As an implication, the probability of trade with the least efficient types is lower under the mixing regime. On the other hand, the competition effect is due to the agents' information being more dispersed inducing any given agent's type in the mixing region to bid more aggressively in a first-price auction and reduces the expected price that the principal pays. Similar reasoning applies to all mixing types in technology y.

**Related literature:** This paper relates to the literature on competing mechanisms. Building on the seminal work of Mcafee(1993) (see also Auster and Gottardi (2019), Jehiel and Lamy (2018), Martimort and Stole (2002), and Pavan and Calzolari (2009)), this literature studies how principals, usually interpreted as sellers, compete through mechanisms in order to attract agents. The key insight from this literature is that the set of agents who participate in a given mechanism is endogenous and depends on choices made by all principals. I extend this analysis to an environment where single principal designs mechanisms for multiple markets, corresponding to different realizations of the technology state. In this context, cross-mechanism externalities are fully internalized and constitute therefore an important element for the analysis. Studying this aspect is a key part of my contribution.

A technical challenge which arises in this literature is that each principal's feasible mechanisms can depend on other principals' strategies leading to an infinite regress problem that makes the model intractable and the message space more complex. To deal with this issue, the literature assumes no cross-principals externalities in the sense that a unilateral deviation by a principal doesn't have an impact on the payoffs of agents participating in other principals' mechanisms nor on their incentive compatibility constraints. The infinite regress issue does not arise when a single principal is designing both mechanisms, therefore, and as these externalities are a key part of the analysis in this paper, I study properties of optimal mechanisms when there are cross externalities.

Further away from this paper, part of the literature focuses on the characteristics of the relevant message space for competing mechanism games (Epstein & Peters (1999), Peters & Szentes (2012), etc.) and the set of implementable outcome functions (Peters (2010)). I refer the interested reader to Peters (2014) for a complete review. In a more applied work, Yamashita (2010) characterized a message space implementing the optimal mechanism for the principal as well as the corresponding equilibrium in a common agency environment. In this paper, agents report their payoff types as well as mechanisms proposed by other principals which implies that in the equilibrium, every principal has full information about equilibrium mechanisms. In our environment, as the principal chooses mechanisms in all the technology markets, the message space is trivial and boils down to agents' payoff types.

Finally, the implementation result in this paper relates to the literature which studies endogenous entry in auctions (see Jehiel and Lamy (2015) Mcaffee and Mcmillan (1987), Levin and Smith) and competing auctions (See Ellison, Mobius and Fudenberg). As allowing for externalities in a competing mechanism setting makes the equilibrium number of bidder endogenous, I contribute to this literature by studying optimal auction design in an environment in which the outside option in each market is endogenous. A key distinction from Ellison and al. is that in my setting, the auction choice is endogenous and chosen optimally by the buyer.

The rest of the paper is organized as follows: In section 2, I develop a framework to tackle the problem of specialization, in section 3, I study the optimal mechanism and its implementability through first-price auctions, and in section 4 I conclude.

## 1.2. Framework:

A principal wants to acquire a good, which she values at $V > 0$. She faces N sellers indexed by $i = \{1, 2, ..., N\}$. Sellers' production costs are determined by their type $(\theta, \gamma)$ where $\theta_i \in \Theta$ is the seller's efficiency and $\gamma_i \in \{\gamma^1, \gamma^2\}$ is his choice of technology. Sellers have the ability to produce in a given state of the word $k \in \{1, 2\}$ only if their technology matches the state, that is: A given seller $i$ can produce in a state $k$ only if $\gamma_i = \gamma^k$. A seller of type $(\theta, \gamma)$ has therefore a production cost:

$$C_k(\theta, \gamma) = \begin{cases} \theta \text{ If } \gamma = \gamma^k \\ \infty \text{ Otherwise} \end{cases} \tag{1.1}$$

The state of nature is exogenous. Denote by $\rho_k$ the probability associated to state $k$ and assume, without loss of generality, that $\rho_1 \geq \rho_2$. Similarly, seller's efficiency type is exogenous and determined by a probability distribution function $f : \Theta \to \mathbb{R}^+$ and its associated cumulative distribution $F : \Theta \to [0, 1]$. I assume that types are independent and identically distributed and that $f(\theta)$ and $F(\theta)$ satisfy the standard regularity conditions: $f(\theta)$ continuous and $\forall \theta \in \Theta : f(\theta) > 0$ and that: $\theta + \frac{1-F(\theta)}{f(\theta)}$ is increasing in $\theta$. Finally, wlog I assume that $V \geq \bar{\theta}$.

The game can be interpreted as a situation in which a buyer is interested in acquiring a good and faces uncertainty about the relevant production technology at the moment of designing trade mechanisms. This uncertainty can for instance be due to exogenous technological breakthrough in complementary technologies, or to some unknown and relevant product characteristics for a principal facing uncertainty about her preferences etc. Similarly, sellers have to develop the product with limited information about which technology might lead to a success. Using information they have about the trade mechanism, sellers self-select by adopting different technologies which makes the set on "ex-post relevant sellers" endogenous. These choices affect future competition and the object of interest will be to study how to study the effect of mechanisms choice on (i) ex-post efficiency and (ii) the technology adoption profiles and levels of competition.

A recent example of such a situation is for instance the procurement of vaccines during the Covid pandemic: In this situation, state had to commit to trading rules prior to the development of vaccines and under uncertainty about the type of vaccines (M-RNA, whole-microbe approach or subunit approach) which might succeed. On the other hand, each vaccine producer has an a cost of manufacturing and distributing vaccines which depends on his efficiency and distribution network for instance.

**Mechanisms and payoffs:** Denote by $\vec{\gamma} \in \{\gamma_1, \gamma_2\}^N$ the realization of a technology adoption profile. An ex-post direct mechanism for the buyer is a pair $(X, T)$ where $X : \theta^N \times \{\gamma_1, \gamma_2\}^N \to [0, 1]^N$ is an allocation rule which determines the probabilities of trade with each seller given reports and the realization of the technology adoption profile. Similarly, $T : \theta^N \times \{\gamma_1, \gamma_2\}^N \to \mathbb{R}^{+N}$ is the transfer rule which determines payments to each agent.

An ex-ante direct mechanism $M$ is a tuple $(Y, (X, T)_{\vec{\gamma}})$ where $Y : \theta \to [0, 1]$ is a technology recommendation which determines the probability that type $\theta$ adopts technology 1 and $(X, T)_{\vec{\gamma}}$ is a choice of ex-post direct mechanism for each realized technology adoption profile. Denote by $f_k(\theta)$ the density function conditional on adopting technology $k$ and by $F_k(\theta)$ the associated CDF. A mechanism is feasible only if this density is continuously measurable.

Players have quasilinear utility functions, that is, for each realized trade decision and vector of expected transfers, seller $i$ gets an expected utility:

$$U_i = E[T_i] - x_i C_k(\theta_i, \gamma_i) \tag{1.2}$$

Where $E[T_i]$ is the expected transfer seller $i$ receives and $x_i$ is the probability he trades with the buyer. Similarly, the buyer obtains payoffs:

$$U^P = \sum_{i=1}^{N} x_i V - E[T_i] \tag{1.3}$$

**Information and timing:** The timing of the game is as follows:

- Stage 0: Nature draws sellers' efficiency types $\theta_i$
- Stage 1: The buyer chooses a mechanism $M$
- Stage 2: Each seller observe the mechanism $M$ and his efficiency type $\theta_i$ and chooses a technology $\gamma_i$
- Stage 3: Nature draws the technology state $k$
- Stage 4: All players observe $k$ and the technology adoption profile $\{\gamma_1, ..., \gamma_N\}$ and each seller either rejects or accepts participating in the mechanism
- Stage 5: Participating sellers report their types and the payoffs are realized according to the mechanism $M$.

We will study symmetric Perfect Bayesian Equilibria of the game.

### 1.2.1. Feasible mechanisms:

As standard in mechanism design, an ex-post (direct) mechanism is feasible if and only if it satisfies incentive compatibility (truth telling is a best response), individual rationality (all player-types are willing to participate in the mechanism), positivity (probabilities of trade with each type are non-negative) and unit-demand (the sum of these probabilities is weakly lower than one) constraints. The key difference when considering ex-ante feasible mechanisms is that as the technology adoption profile is an equilibrium object which means that (i) these constraints have to hold for each such a profile which is realized with a strictly positive probability and (ii) the technology adoption probabilities have to be best responses for all types. Formally, given $\gamma_i$ and the adoption profile of other sellers $\gamma_{-i}$, we denote by $U(\theta_i, \hat{\theta}_i, \gamma, \gamma_{-i})$ and by $x(\theta_i, \hat{\theta}_i, \gamma, \gamma_{-i})$ the expected utility and the probability of trade type $\theta_i$ obtains from reporting $\hat{\theta}_i$. We define:

**Definition 1.** ***Feasible ex-post mechanisms:*** *Set* $(\gamma_i, \gamma_{-i})$ *a realization of the technology adoption profile and* $p(\vec{\gamma})$ *its associated probability and define as* $\Theta_\gamma$ *the set of types who adopt technology* $\gamma$ *with a strictly positive probability in equilibrium. An ex-post mechanism is feasible if and only if for all* $(\gamma_i, \gamma_{-i})$ *such that* $p(\gamma_i, \gamma_{-i}) > 0$ *it satisfies:*

*Incentive Compatibility:*$\forall \gamma_i, \gamma_{-i}, \forall \theta, \hat{\theta} \in \Theta_\gamma$*:* $U(\theta_i, \theta_i, \gamma_i, \gamma_{-i}) \geq U(\theta_i, \hat{\theta}, \gamma_i, \gamma_{-i})$

*Individual Rationality:* $\forall \gamma_i, \gamma_{-i} \theta \in \Theta_\gamma$*:* $U(\theta_i, \theta_i, \gamma_i, \gamma_{-i}) \geq 0$

*Positivity:* $\forall \gamma_i, \gamma_{-i} \theta_i, \theta_\{ - i\} \in \Theta_\gamma$*:* $x(\theta_i, \theta_{-i}, \gamma_i, \gamma_{-i}) \geq 0$

*Unit demand:* $\forall \gamma, \theta \in \Theta_\gamma$*:* $\sum_{i=1}^N x(\theta_i, \theta_{-i}, \gamma_i, \gamma_{-i}) \leq 1$

Here, incentive compatibility and individual rationality are standard and ensure that each agent type participates in the mechanism and has a best response which consists of truthful reports in equilibrium. The only difference with standard models is that the set of possible deviations is restricted to types adopting the same technology. In this game, this set is determined endogenously by the chosen mechanisms, therefore, feasibility requires each agent's technology adoption probabilities, and the resulting $\Theta_\gamma$, to be a best response for all agents' types. This is captured in the set of feasible ex-ante mechanisms:

**Definition 2.** ***Feasible ex-ante mechanisms:*** *A mechanism is ex-ante feasible if:*
*Ex-post feasibility: For each* $k, (\gamma_i, \gamma_{-i})$ *such that* $P(\gamma_i, \gamma_{-i}) > 0$*: The continuation mechanism is ex-post feasible*
*Obedience constraint:* $\forall \gamma, \gamma', \theta \in \Theta_\gamma$: $E_{\theta_{-i},\gamma_{-i}}[U(\theta_i, \theta_i, \gamma, \gamma_{-i})] \geq \max_{\hat{\theta} \in \Theta_{\gamma'}} E_{\theta_{-i},\gamma_{-i}}[U(\theta_i, \hat{\theta}, \gamma', \gamma_{-i})]$

The obedience constraint determines technology choices by sellers and it ensures that, given the chosen ex-post mechanisms, sellers cannot strictly benefit from a double deviation by changing the technology choice and misreporting type. In order to make notation more smooth, we find it convenient to set $N_k$ as the realization of the number of sellers who adopted technology $k$. As we restrict attention to symmetric mechanisms, ex-post feasibility has to depend only the realized number of relevant sellers and not their identity, therefore, the following holds:

**Lemma 1.** *Any feasible symmetric ex-post direct mechanism has a payoff equivalent symmetric direct mechanism which depends only the realized state* $k$*, the number of sellers who adopted technology* $\gamma^k$ *and their reports.*

The proof of this lemma is straightforward and relies on the fact that the buyer never trades with sellers who adopted an irrelevant technology independently from their reports. For any possible such reports, the principal can propose stochastic mechanisms which assign the same probabilities on outcome as mechanisms who depend on irrelevant sellers. This lemma allows us to use a more compact notation and a mechanism we denote by $IC_{N_k,k}$, $IR_{k,N_k}$, $pos_{k,N_k}$ and $D_{k,n_k}$ the ex-post feasibility constraints in state $k$ given the number of relevant sellers $N_k$.

## 1.3. The optimal mechanism:

Unlike a standard mechanism design setting, in this problem, the set of relevant types in each state is endogenous. As a result, the properties of the ex-post probability distribution over types are unknown. Therefore, we begin by identifying necessary optimality conditions regarding the distribution of types given the technology choice.

**Proposition 1.** *A mechanism $M$ is optimal only if there exists $\theta^*$ such that:*

- *The principal trades with a strictly positive probability with all types*
- $\forall \theta \in [\theta^*, \bar{\theta}] : \ \gamma(\theta) = \alpha\gamma_1 + (1-\alpha)\gamma_2$
- $\forall \theta \in [\underline{\theta}, \theta^*] : \ \gamma(\theta) = \gamma_1$

*Where* $\alpha = 1 - \frac{\rho_2}{(\rho_1+\rho_2)(1-F(\theta^*))}$

**Proof:** (see appendix for a detailed proof)

**Sketch of the proof:** The proof works as follows: First, I define a relaxed version of the problem in which feasibility constraints are required to hold in expectation with respect to the technology profiles. First, note that as the original problem requires these constraints to hold for each realization of the technology profile, any payoffs which can be reached in the original problem can also be reached in the relaxed version of the problem. This makes the solution to the latter, $M'$, an upper bound on the principal's payoffs. The first step of the proof is to show that if the structure described in proposition 1 is satisfied in the mechanism $M'$, this later is implementable within the original problem.

The second part of the proof shows that the solution to the problem $M'$ is uniquely achievable using the structure described in proposition 1. This is shown by starting from an arbitrary feasible mechanism $M'$ and showing that there exists a payoff equivalent mechanism $M''$ which (i) leads to the same ex-ante probability of adopting a given technology by an agent and (ii) delivers the same probabilities of trade with each agent type and (iii) satisfies the properties in proposition 1. Finally, to complete the proof, we compute the solution to the relaxed problem, $M'$, and show that any mechanism which does not belong to the described class of mechanisms delivers strictly lower payoffs. In this case, the solution to $M'$ implements more efficient trade and lower probabilities of trade with the least efficient types while keeping the total probability of receiving the good constant.

In this class of mechanisms, introducing mixing has two effects: An efficiency effect and a rent effect. To provide intuition behind the efficiency effect, consider a benchmark mechanism in which the least efficient types adopt the least likely technology and more efficient ones adopt the most likely technology. Introducing mixing while keeping the ex-ante probabilities of technology adoption fixed allows the principal to reduce the probability of trade with the least efficient agents' types to the benefit of some intermediary types. This effect is due to mixing allowing the principal to relax some feasibility constraints, which allows reoptimizing over the allocation. Similar reasoning can be used as long as the constraints are not binding in both markets which happens only if the mixing is done with the probability $\alpha$ described above. This probability makes rent increase at the same speed in both markets.

The rent effect comes from the fact that the agent's rent increases at a speed that is equal to the probability of trade. Reducing the probability of trade with the lowest types implies that rent is increasing at a slower rate for those types. Compared to the benchmark mechanism, the principal can construct a suboptimal mechanism that belongs to the described class and which induces pooling for some intermediary types so that the rent is strictly lower for all types. Starting from this mechanism one can reoptimize and strictly increase payoffs.

### 1.3.1. Implementation through generalized auctions

In this subsection, I will discuss the implementation of the optimal mechanism through generalized auctions. More specifically, I will focus on the properties of the optimal auctions and the effect on the introduction of mixing on competition between bidders. To do so, I first show that a specific class of auctions can implement any candidate optimal mechanism:

**Proposition 2.** *Consider any mechanism belonging to the class described in proposition 1 and denote by $N_k$ the realized number of agents who adopted technology $\gamma^k$. This mechanism is implementable via posted prices when*

*$N_k = 1$ and a collection of generalized auctions with minimum and maximum bids for each technology market when $N_k > 1$. Moreover, auctions induce pooling for at least a subset of types if the cutoff type $\theta^* > \theta^{**}$ for some $\theta^{**}$.*

**Proof:** (see appendix)

**Discussion:** First note that as standard auctions and posted prices induce efficient trade, in this context they cannot implement the cutoff type is too high.More interestingly, using proposition 2, one can allows study the costs and the benefits from mixing in an auction setting. To do so, note first that compared to the benchmark in which agents make technology choices in pure strategies described above, mixing makes information, given the technology choice, more dispersed. As an implication, under the mixing regime, the density function given a technology choice is lower for all mixing types which makes conditional information more dispersed. [1]

A formal argument relies on first-order stochastic dominance: To provide intuition, consider $\theta^*$ and $\theta^{**}$ the cutoff type in case of mixing and under the benchmark respectively. First, in the For all seller types $\theta \in [\theta^{**}, \bar{\theta}]$: These types adopt technology 2 in the benchmark. Introducing mixing for this subset of types inducer a lower density function which implies that for each given type, the probability that he is facing a more efficient type is higher. These types will therefore bid more aggressively under the mixing regime which creates a competition effect. Moreover, as intermediary types also adopt this technology, the probability of trade with type $\theta^{**}$ is no longer equal to one (conditional on being a relevant seller) under mixing which induces more efficient trade.

On the other hand, introducing mixing in the interval $[\theta^*, \theta^* *]$ can be analyzed by comparing trade in state 1 under both mechanisms. Note first that for types $\theta \in [\underline{\theta}, \theta^*]$, as the distribution of types has the same local properties under both regimes, these types adopt the same bidding behavior and win the auction with the same probability under both regimes. This implies that an agent of type $\theta^*$ wins the auction with an ex-ante probability $\rho_1 F(\theta^*)$ under both regimes and obtains the same transfers. Finally, mixing increases both the rent and the probability of trade with types in the interval

[1] In this case, the benchmark mechanism can be implemented through minimum and maximum bids in each technology market.

$[\theta^{*}, \theta^{**}]$.

This last effect is due to the density in the interval $[\theta^{*}, \theta^{**}]$ being lower under mixing which implies that the probability of trade decreases at a slower rate under this regime. In order to assess the total effect of mixing, one has to compare these gains to the total cost of implementing the benchmark mechanism. This aspect was discussed in proposition 1.

Finally, in order to conclude this section, it is of interest to study the properties of these auctions:

**Corollary 1.** *Any mechanism with* $q_1 < q^*$ *is implementable with a minimum bid of* 0 *in state 1 and* $\underline{b} > 0$ *in state 2. Moreover,* $\underline{b}$ *is increasing in* $q_1$.

Here, as discussed above, the minimum bid allows the principal to implement pooling in state 2. Increasing this bid induces pooling for more types who are diverted away from the technology $\gamma_2$. As an outcome, by increasing the pooling region she increases the agents' incentives to adopt technology $\gamma^1$ and therefore increases $q_1$. An implication of this corollary is that optimal auctions discriminate between technologies by "forcing" efficient agents to adopt the most likely technology. This result provides a rationale behind the use of these auctions in practice.

## 1.4. Conclusion

This paper studies optimal mechanisms for a multi-market principal who faces uncertainty about the relevant technology. I show that the optimal mechanism belongs to a class of mechanisms in which the set of types adopting each technology is continuous. In this class, the least efficient agents strictly mix between the two available technologies whereas the most efficient ones adopt the most likely technology. Within this class, the virtual valuation satisfies the same properties as in standard mechanisms, however, implementation of a given cut-off type can induce pooling around the indifferent type.

This work highlights the benefits of mixing in this type of setting which can come from two sources: Efficiency gains and informational ones. Efficiency gains are due to mixing allowing to shift probabilities of trade from the least efficient to the intermediary types while keeping the total probability of trade in each market constant. On the other hand, informational gains come from the fact that the separation of types is less costly under mixing. This last effect is also related to a lower probability of trade with the least efficient types which makes the rent decrease at a slower rate with respect to cost (or equivalently, increase at a slower rate with respect to efficiency). Finally, the paper discusses implementation in an auction setting. In this setting, the cheaper cost of separation takes the form of more aggressive bidding behavior by the least efficient types.

## References:

Auster, S. and Gottardi, P., 2019. Competing mechanisms in markets for lemons. Theoretical Economics, 14(3), pp.927-970.

Epstein, L.G. and Peters, M., 1999. A revelation principle for competing mechanisms. Journal of Economic theory, 88(1), pp.119-160.

Albrecht, J., Gautier, P.A. and Vroman, S., 2014. Efficient entry in competing auctions. American Economic Review, 104(10), pp.3288-96.

Ellison, G., Fudenberg, D. and Möbius, M., 2004. Competing auctions. Journal of the European Economic Association, 2(1), pp.30-66.

Jehiel, P., & Lamy, L. (2015). On discrimination in auctions with endogenous entry. American Economic Review, 105(8), 2595-2643.

Jehiel, P. and Lamy, L., 2018. A mechanism design approach to the Tiebout hypothesis. Journal of Political Economy, 126(2), pp.735-760.

Levin, D. and Smith, J.L., 1994. Equilibrium in auctions with entry. The American Economic Review, pp.585-599.

McAfee, R.P., 1993. Mechanism design by competing sellers. Econometrica: Journal of the econometric society, pp.1281-1312.

Mcafee, & Mcmillan. Auctions with entry. Economics Letters, 1987, vol. 23, no 4, p. 343-347.

Pavan, A. and Calzolari, G., 2009. Sequential contracting with multiple principals. Journal of Economic Theory, 144(2), pp.503-531.

Peters, M. and Szentes, B., 2012. Definable and contractible contracts. Econometrica, 80(1), pp.363-411.

Peters, M., 2010. On the Revelation Principle and Reciprocal Mechanisms in Competing Mechanism Games. Discussion paper, Department of Economics, The University of British Columbia.

Peters, M., 2014. Competing mechanisms. Canadian Journal of Economics/Revue canadienne d'économique, 47(2), pp.373-397.

Ramey, G. (1996). D1 signaling equilibria with multiple signals and a continuum of types. Journal of Economic Theory, 69(2), 508-531.

Yamashita, T., 2010. Mechanism games with multiple principals and three or more agents. Econometrica, 78(2), pp.791-801.

## 1.5. Appendix

Notation: Throughout the appendix, we will use $f_k(\theta)$ and $F_k(\theta)$ to denote the density and the cumulative distribution function conditional on the adopted technology being $\gamma_k$.

### 1.5.1. Proposition 1:

First, let us define a relaxed version of the problem in which the incentive compatibility, unite demand constraints and obedience are satisfied in expectation only (rather than for all realizations of adoption profiles).

**Definition 3.** ***The relaxed problem*** *Set $M'$ to be the solution to maximizing $U^P$ subject to*

*Incentive Compatibility:* $\forall\gamma, \forall\theta, \hat{\theta} \in \Theta_\gamma$: $E[U(\theta_i, \theta_i, \gamma)] \geq U(\theta_i, \hat{\theta}, \gamma)$

*Individual Rationality:* $\forall\gamma, \theta \in \Theta_\gamma$: $U(\theta_i, \theta_i, \gamma) \geq 0$

*Positivity:* $\forall\gamma, \theta_i, \theta_\{ - i\} \in \Theta_\gamma$: $x(\theta_i, \theta_{-i}, \gamma) \geq 0$

*Unit demand:* $\forall\gamma, \theta \in \Theta_\gamma$: $\sum_{i=1}^{N} x(\theta_i, \theta_{-i}, \vec{\gamma}) \leq 1$

*Obedience constraint:* $\forall\gamma, \theta \in \Theta_\gamma$: $E[U(\theta_i, \theta_i, \gamma)] \geq \underset{\gamma' \neq \gamma_i}{max} \underset{\hat{\theta} \in \Theta_{\gamma'}}{max} E[U(\theta_i, \hat{\theta}, \gamma')]$

Note that these conditions are necessary for the mechanism to be feasible in the initial problem. Remarque: this situation would be the same as a game in which buyers report their types before observing the technology adoption profile.

**Step 1:** If the solution to the relaxed problem is such that $\exists\theta^*$ such that:

- $\forall\theta \in (\underline{\theta}, \theta^*)$: $\gamma(\theta) = \gamma_1$
- $\forall\theta \in (\theta^*, \bar{\theta})$: $\gamma(\theta) = \rho_1\gamma_1 + (1 - \rho_1)\gamma_2$

Then the mechanism $M'$ solves the original problem.

**Proof:**

Consider any solution to the relaxed problem. Incentive compatibility implies:

$$U(\theta, \theta, \gamma) = T_k(\theta) - x_k(\theta)\theta \geq T_k(\theta') - x_k(\theta')\theta = U(\theta, \theta', \gamma)$$

As standard in mechanism design, using $\theta' = \theta + \Delta$, and computing the limit as $\Delta \to 0$, incentive compatibility implies:

$$\frac{\delta U(\theta, \theta, \gamma)}{\delta \theta} = -x_k(\theta)$$

As $\forall \theta \in [\theta^*, \bar{\theta}] : \gamma(\theta) = \rho_1 \gamma^1 + (1 - \rho_1)\gamma^2$, the obedience constraint implies that the agent is indifferent between the two technologies. As an implication, we have $\forall \theta \in [\theta^*, \bar{\theta}] : E[U(\theta_i, \theta_i, \gamma_1)] \geq E[U(\theta, \theta, \gamma_2)]$. A necessary condition for this to hold is that $\forall \theta \in [\theta^*, \bar{\theta}] : x_1(\theta) = x_2(\theta)$.

Now, consider any mechanism $M'$ which solves the relaxed problem. A necessary condition is that, given $x(\theta^*)$, $x(\theta)$ is optimal. This implies that in the interval $[\theta^*, \bar{\theta}]$, the mechanism maximizes the buyers payoffs subject to $IR$, $IC$, $x(\theta) \leq x(\theta^*)$ and $E[x_1(\theta)] = E[x_2(\theta)] = x(\theta)$. This is equivalent to:

$$max \int_{\theta^*}^{\bar{\theta}} (x(\theta)v - T(\theta)] f(\theta)) d\theta$$

As standard in mechanism design problems, replacing transfers by their value $(R(\theta) + x(\theta)c(\theta))$, the problem becomes:

$$max \sum_{i=1}^{N} \int_{\theta^*}^{\bar{\theta}} x_k(\theta_i)(v - \theta_i - U(\bar{\theta}) - \int_{\theta}^{\bar{\theta}} x(s)ds) f_k(\theta) d\theta$$

subject to $0 \leq x_i(\theta) \leq x(\theta^*)$.

Using Fubini's to solve the double integral and setting $U(\bar{\theta}) = 0$ (otherwise, a profitable deviation would be to reduce all transfers by $U(\bar{\theta})$), this equation becomes:

$$max\sum_{i=1}^{N}\int_{\theta^*}^{\bar{\theta}} x_k(\theta_i)(v-\theta_i-\frac{1-F_k(\theta_i)}{f(\theta)})f_k(\theta)d\theta$$

As $\theta+\frac{1-F_k(\theta)}{f_k(\theta)}$ is increasing in $\theta$; pointwise maximization implies that $\exists\theta_1 \geq \theta^*$ such that $\forall\theta \in [\theta^*, \theta_1] : x_k(\theta) = x_k(\theta^*)$ and for all $\theta < \theta^*$ the mechanism is a separating mechanism. As an implication, the buyer trades with the seller with the lowest virtual valuation when $\theta \geq \theta_1$ and, if more than one buyer has a cost $\theta \in [\theta^*, \theta_1]$, the probabilities of trade are equal among those buyers. Using similar reasoning, we obtain that there exists $\theta_2\ leq\theta^*$ such that the buyer trades with the buyer with the lowest valuation if there exists at least one seller with $\theta_i < \theta_2$ and trades with the other sellers with equal probabilities otherwise.

Finally, to conclude our proof, given the definition of $\alpha$, separation induces the same ex-ante probabilities of trade in both states which implies that $M'$ is feasible in the initial problem by introducing pooling for all realizations of around $N_k$ in the relevant pooling region.

**Step 2:** Any feasible mechanism in the relaxed problem has a payoff equivalent mechanism such that for some $\theta^*$:

(i) $\forall\theta \in (\underline{\theta}, \theta^*)$: $\gamma(\theta) = \gamma_1$

(ii) $\forall\theta \in (\theta^*, \bar{\theta})$: $\gamma(\theta) = \alpha\gamma_1 + (1-\alpha)\gamma_2$

**Proof:** (To be completed)

**Step 3:** The solution to $M'$ is unique

**Proof:** To show this, consider two mechanisms, one of which does not satisfy the above properties. It is sufficient to show that the probabilities of trade describe in step 1 and only implementable via a unique $M'$.

Consider any type $\theta \in [\theta^*, \bar{\theta}]$: In this case it is sufficient to observe that $x(\theta)$ increases at a rate proportional to the local density. Now, assume that for some interval $[\theta_1, \bar{\theta}]$, the agents mix with probabilities $\alpha'(\theta) \neq \alpha$. This will lead these local densities to differ between markets. Therefore, either the mechanism induces inefficient trade if conditional densities or the technology choice in pure strategies, and the probabilities of trade decrease strictly slower under the mechanism $M'$. As an implication, for all types

$\theta \in (\theta_1, \bar{\theta})$, the probability of trade is strictly higher under the alternative mechanism, which is a contradiction. If the mechanism induces inefficient trade and similar probabilities, this mechanism induces higher transfers which contradicts efficiency.

# Chapter 2

# Dynamic monitoring of adaptive criminals

**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.

## 2.1. 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 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,

[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.

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.

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.

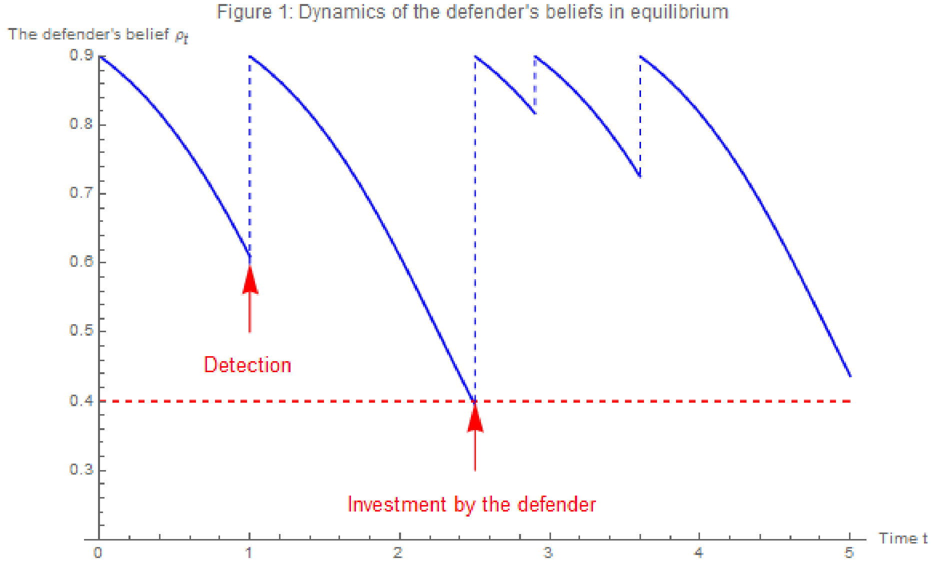
A policy implication of this result is that policy interventions such as raising fines or increasing the monitoring rates can lead players to engage in a technological arms race. This effect makes evaluating policies challenging as the global effect on deterrence depends on the cost of the arms race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

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




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.

### 2.1.1. 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.

## 2.2. The setting:

### 2.2.1. The model:

Consider a game where one defender (player $D$) and one attacker (player $A$) repeatedly interact at a fixed time interval $\Delta$. 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. In even periods ($t \in 2k\Delta$ with $n \in \mathbb{N}$), 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.

In odd periods ($t \in 2k\Delta + 1$), the attacker chooses the intensity of his attack $a_t \in [0, \bar{a}]$ which generates flow payoffs that depend on the monitoring policy $\pi$ and the technology state $\theta_t$. Under a policy $\pi$, 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 is the following:

- Stage 0: The state $\theta_{t-\Delta}$ is inherited from the past, and the defender updates her beliefs $\rho_t$ about it,

[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, and to Riley (2005) for an application to border control

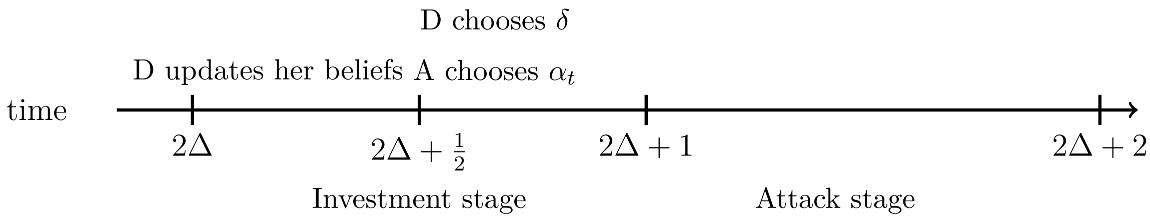



Figure 2.1: Preferred activity for principal and agent as a function of $\theta_L$ when types are not observed under $\sigma^P$.

- Stage 1: Both players simultaneously make investment decisions in hiding and detection technologies,
- Stage 2: The state $\theta_t$ 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.

**Law of motion of the defender's monitoring ability:** The monitoring ability $\theta_t$ 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 $\omega_t = 0$

[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 $\Delta$ goes to zero is equivalent to the equilibrium of the continuous-time version of the game.

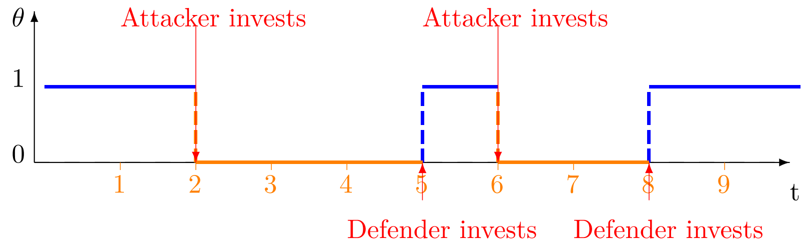


*Figure 1: The monitoring ability as a function of time and investments*

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 $\delta_t$ for each $t$ and history $h^{Dt}$. A (pure) Markov strategy for the defender consists on investment decisions $\delta_t$ as a function of her belief $\rho_t$ about $\theta_t$. More formally, a pure Markov strategy for the defender $\sigma^D$ is:

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

The attacker's private history at the beginning of time $t$, $h^{At}$, consists of the public history, investments $\alpha$, the intensity of the attack $a$ and the state $\theta$ up to (but not including) time $t$. A strategy for the attacker consists of a choice of the investment decision $\alpha_t$ and the intensity of attack for each $t$ and private history $h^{At}$. A Markov strategy for the attacker consists of investment decisions and intensity of the attack as a function of $(\rho_t, \theta_t)$. Formally, a pure Markov strategy for the attacker is a function:

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

**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:

$$U_t^D = E_{a,\theta,\delta}\left[\sum_{\tau=0}^{\infty} e^{-r(\tau-t)\Delta}\Big(u_\pi^D(a_\tau,\theta_\tau)\Delta - \delta_\tau F^D\Big)\right] \tag{2.1}$$

The defender's expected instantaneous payoffs at time $t$ in equation (2.1) 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:

$$U_t^A = E_{\alpha,\delta}\left[\sum_{\tau=t}^{\infty} e^{-r(\tau-t)}(u_\pi^A(a_\tau,\theta)\Delta - \alpha_\tau F^A\right] \tag{2.2}$$

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.

### 2.2.2. 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 $\theta$. 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:

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

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

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

**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

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

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.

**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

[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.

on the arms race. To study the effect of these policy measures, consider a starting policy $\pi$ 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 $\pi$ are:

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

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:

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

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 $\alpha$.[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.

[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.

## 2.3. 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 4.** ***(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 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 5.** ***(Arms race Equilibria)*** *An equilibrium is an arms race if for any private history $h_t^i$ 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: $\forall t$ : $\exists \tau_1, \tau_2 > t : E_{h^{\tau_1}}\left[\alpha_{\tau_1}\right], E_{h^{\tau_2}}\left[\delta_{\tau_2}\right] > 0$*

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 6.** ***(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 3.** ***(Types of equilibria)*** *For any policy $\pi$, 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 3 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 $\rho_t$ 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 $\Delta$ goes to zero, beliefs at time $t$ evolve according to:

$$\dot{\rho_\tau} = -\rho_\tau(1-\rho_\tau)\lambda_\pi(\hat{a}_\tau(1)) \tag{2.3}$$

This equation captures the fact that as the defender fails to detect attacks, her belief about her monitoring ability $\theta_t$ 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 $\tau$. Therefore, these intensities have to be determined in the equilibrium path.

**Proposition 4.** ***(The intensity of attacks)*** *In any Markov Perfect Equilibrium, the intensity of the attack is chosen myopically*

Proposition 4 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 $\tau$ 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 4 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.

## 2.4. 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.

### 2.4.1. The arms race equilibrium:

**Proposition 5.** ***(Arms race equilibria)*** *If* $u^A_\pi(a^*(0),0) - u^A_\pi(a^*(1),1) > (1-e^{-r\Delta})F^A$ *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* $\rho^*$ *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* $\rho_0$ *is such that* $t^* = t^A = t^D$

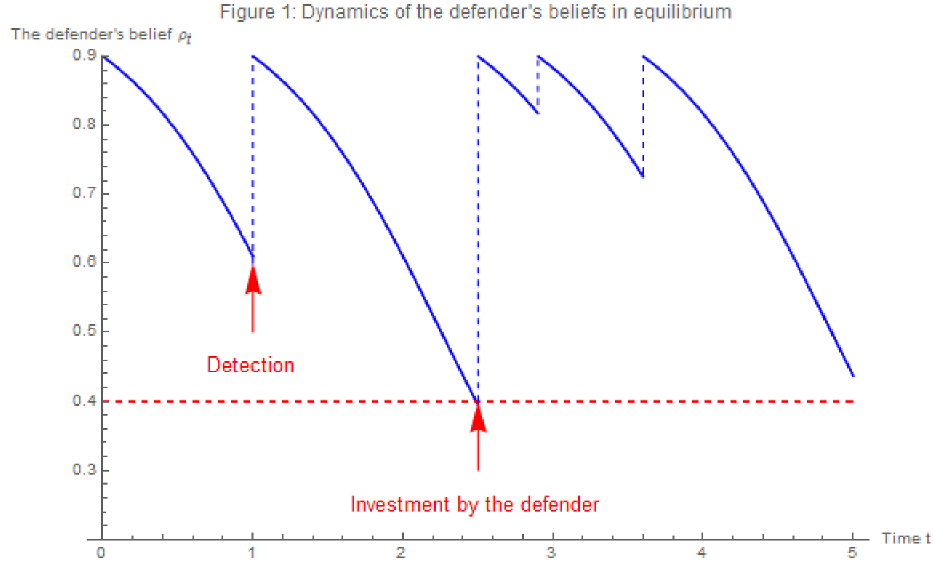
**Equilibrium cycles:** Proposition 2 allows us to describe the beliefs and technology cycles of this game (see figure 3). A cycle starts when both players receive an informative signal about the state of the monitoring ability due to detection of an attack (time t=1, t=2.9 and t=3.6 in figure 3) or 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 $\rho^*$ 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 $\rho_0$, 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*) $t^A = t^D$.

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




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 $\alpha_0$, the distance between the initial belief $\rho_0$ and the stopping belief $\rho^*$. 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 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.

### 2.4.2. 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: $u^A_\pi(0, \theta) = u^D_\pi(0, \theta) = \lambda_\pi(0) = 0$.

**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 $\pi$ and $\pi'$ which lead to an arms race.

**Definition 7.** *Fix $a^*(1)$ and $a^{*\prime}(1)$ the equilibrium intensities of detectable attacks under policies $\pi$ and $\pi'$.The policy $\pi$ is more informative than the policy $\pi'$ 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 6.** ***(Pareto dominance of more informative policies)*** *Consider any two policies $\pi$ and $\pi'$ 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 $\pi$ the most informative policy. We have:*
*(i) The policy $\pi$ induces less investments in R&D by the attacker: $\alpha_0^\pi < \alpha_0^{\pi'}$*
*(ii) The policy $\pi$ Pareto-dominates the policy $\pi'$*

Proposition 6 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 $\rho_0$, 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 $\pi$ to Pareto dominate the policy $\pi'$. [8]

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 $\pi$ is characterized by:

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

All policies $\pi$ and $\pi'$ 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 $\pi$ Pareto dominates $\pi'$. 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

[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.

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 $\pi$ to be an initial policy and consider a policy $\pi'$.

**Definition 8.** *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 7.** ***(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* $\pi'$
*(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 $\rho_0$. 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 $\pi'$ 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 $\theta$ 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 9.** *A policy intervention $\pi'$ 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 8.** ***(Effect of downstream deterrence policies)*** *For any initial policy $\pi$ and downstream deterrent policy $\pi'$ 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: $\alpha_0^{\pi} > \alpha_0^{\pi'}$*

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 9.** ***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 $\rho_0$, the stopping belief $\rho^*$ 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 10.** *Fix the costs of investments $F^A$ and $F^D$ and the optimal entente policy $\pi'$. The arms race policy $\pi$ is not Pareto dominated only if, for $X$ and $Y > 0$:*
*Given $u^D_\pi(a^*(\theta), \theta)$: $u^A_\pi(a^*(0), 0) - u^A_\pi(a^*(1), 1) \leq X$.*
*Equivalently, $u^A_\pi(a^*(\theta), \theta)$: $u^D_\pi(a^*(1), 1) - u^D_\pi(a^*(0), 0) \geq X$.*

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.

## 2.5. 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.

## 2.6. Appendix

***Proof for proposition 1:***

**Part 1:** A policy is an entente policy if and only if $\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}} - \frac{\underline{u}_\pi^A}{1-e^{-r\Delta}} > F^A$ 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 $\bar{u}_\pi^A$ leading to payoffs $\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}} - F^A$.[9] We also have:

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

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 $\pi$: $\bar{u}_\pi^A \leq u_\pi^A(a,1)$, then the policy $\pi$ is an entente policy

**Proof:** Set $(\sigma^A, \sigma^D)$ 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

[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 $\frac{\bar{u}_\pi^A}{1-e^{-r\Delta}}$

not detectable. The attacker's payoffs can be written as:[10]

$$U_0^A = \sum_{t=0}^{\infty} e^{-rt}\Big[\sum_{h^t} p(h^t) E_{a,\alpha}[u_\pi(a,\theta) - \alpha F^A | h^t]\Big]$$

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:

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

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 $\pi$ is an entente policy.

**Proof:** Similarly, assume that there exists an equilibrium in which the

[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

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

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

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:

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

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 $\alpha$ and $\alpha < 0$ is a best response only if $\alpha = 0$ is also a best response, meaning that:

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

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 $\pi$ 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 $\rho_0$.

**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 $\bar{U}^A_\pi - F^A > \underline{U}^A_\pi$: A contradiction.
Therefore, in any equilibrium which is not a complete hiding equilibrium, the attacker invests with an interior probability at belief $\rho_0$.

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

$$U^A(\rho_t, 1) = u^A_\pi(a, 1)\Delta + e^{-r\Delta}\Big[\lambda_\pi(a)U^A_\pi(\rho_0) + (1 - \lambda_\pi(a))U^A(\rho_{t+\Delta}, 1)\Big] \quad (2.4)$$

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 $\rho^*$ 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:

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

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 $\rho_t$ times the probability of reaching belief $\rho_t$ 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 $\rho^*$. This means that the attacker can get higher payoffs if he could reboot the game and move back to belief $\rho_0$ whenever belief $\rho^*$ is reached. The maximal payoffs he could get in that case are reached without investing and are therefore weakly lower than $\frac{u_\pi^A}{1-e^{-r\Delta}}$. However, as never investing allows him for secure at least $\frac{u_\pi^A}{1-e^{-r\Delta}}$, this implies that he has a strictly profitable deviation: A contradiction. Therefore, $U_\pi^A(\rho^*) \geq U_0^A$.

**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 $U_\pi^A(\rho^*) \geq U_\pi^A(\rho)$ for all $\rho \geq \rho^*$.

**Proof:** The proof is similar to step 2. Assume not and that there exists a belief $\rho'$, 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 $\rho^*$ 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:

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

We have $\forall a$:

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

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 $\rho^*$. 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 2.4 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:

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

This inequality can be rewritten as:

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

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

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

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:

$$U^D(\rho) = \rho\bar{u}_\pi^D + (1-\rho)\underline{u}_\pi^D + e^{-r\Delta}U_\pi^D(\rho_{t+\Delta}) \tag{2.5}$$

This function is strictly increasing in $\rho$ 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.
On the other hand, if $\bar{u}^D_\pi < \underline{u}^D_\pi$, 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 $\pi$ and $\pi'$ two strongly short-term-equivalent policies. Consider any equilibrium investment probabilities under policy $\pi$: $\alpha(\rho)$ 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 $\pi$ and $\pi'$. 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, $\alpha_\pi$ and $\delta_\pi$ are also best responses against each other under policy $\pi'$ which concludes the proof.

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

***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 $\rho_0$ and from step 6 the attacker's action is the myopic action.
Now, the defender's payoffs are:

$$U^D(\rho) = \delta U^D + (1-\delta)\Big[\rho \bar{u}^D_\pi + (1-\rho)\underline{u}^D_\pi + e^{-r\Delta}U^D_\pi(\rho_{t+\Delta})\Big] \tag{2.6}$$

For any belief $\rho^*$ such that the defender invests,

Now, denote by $\sigma'$ 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, $\sigma'$ 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:

$$U^D_t = -h(\rho a^* + (1-\rho))\Delta + \max_{\delta} \; e^{-r\Delta}\Big[\delta(U^D_0 - F^D) + (1-\delta)E[U^D_{t+\Delta}]\Big] \tag{2.7}$$

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 2.8 is monotonically decreasing in $\rho$, 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 $\Delta$ 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:

$$U_t^D = -h(\rho a^* + (1 - \rho))\Delta + \max_{\delta} e^{-r\Delta}E[U_{t+\Delta}^D]\Big] \tag{2.8}$$

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 2.8 as:[12]

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

The general solution to this differential equation is:

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

[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, 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 $\rho^*$ and $U_0^D$:

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

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

$$
\begin{aligned}
U(\rho^*) &= \frac{1}{r+a^*m\rho^*}\Big[-h(\rho^*a^* + (1-\rho^*)) + a^*m\rho^*U_0^D\Big] \\
&= U_0^D - \frac{1}{r+a^*m\rho^*}\Big[h(\rho^*a^* + (1-\rho^*)) + rU_0^D\Big]
\end{aligned}
$$

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 $\rho_0$, we show that for all $\rho_0$, 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.

# Chapter 3

# Ressource allocation in the presence of moral hazard and endogenous adverse selection

**Abstract**

A principal wants to develop a new product by delegating its production to an agent. Production is dichotomic and stochastic. The agent allocates resources between a task that yields direct production and a task that increases his productivity. Increasing productivity makes effort more costly. We show that when the resource allocation is non-observable, the agent's final productivity in the contract proposed by the principal is lower than the optimal one. In this setting, raising bonuses encourages both effort and increases in productivity, as a result, compared to a benchmark in which the allocation is observable, the principal has incentives to reduce the bonus due to the agent being less productive and incentives to increase the bonus to encourage him to increase his productivity. The main result of our paper shows that, when both the initial productivity and the cost of increasing productivity are small, this leads to higher bonuses than the full observability benchmark.

## 3.1. Introduction

One of the key drivers of a firm's productivity is the choice of how to allocate resources between different activities. This allocation is often decided or recommended by workers who have superior information about either the firm's/project needs or an ability to monitor how resources are being used. For instance, workers decide the share of time they spend exerting productive effort and the one they spend on on-the-job learning to increase their productivity in an unobservable way. Similarly, in the context of product development, teams of workers carry the R&D programs and decides how to allocate their time and effort between market studies to improve knowledge about consumer's preferences and product development. In both examples, the agent's final productivity is both endogenous and unobservable. However, the source of the final asymmetry of information is different: In the first example, the resource allocation is taken under moral hazard as the top management and the owners lack the ability to monitor the worker's learning. In the second example, there is adverse selection as they do not observe part of the relevant information for decision making.

Once the resource allocation is determined, firms face the standard moral hazard which is inherent to risky production environments. Resource allocation shapes the intensity of this problem because it determines both the productivity of effort and its cost. This situation can create some tensions between the firms and their workers as both players share the benefits of success, whereas the workers privately supports any potential increases in the cost of effort. These tensions lead to misalignment of incentives, which, combined with the asymmetries of information in the resource allocation problem can lead to contractual distortions. This paper contributes to the literature by studying the source of misalignment of incentives in terms of resource allocation and how the nature of the asymmetries of information shapes the contractual distortions.

In order to fix ideas, consider, for instance, a company that wants to develop a new product: the agent carries out "exploratory tasks" such as market analysis, studying the consumer's needs, identifying potential targets, studying the competition, etc. that aim at identifying the components of the product that should be improved. Following this exploratory phase, the agent performs "implementation tasks" to develop the product and introduce it to the market. Launching the new product successfully depends on the combination of both tasks. However, the principal only observes the success or failure without observing how time was allocated between tasks. In this situation, exploratory tasks help at making implementation effort better targeted, and they can be interpreted as tasks that improve the project-specific productivity of the agent which are shared between the two players; however, when the tasks allocation is not observable, the cost is supported by the agent only. In other words, upon observing a failure/success of production, the principal cannot observe if such was the aftermath of insufficient/sufficient R&D.

In this paper, we study these questions in a principal-agent model where one agent takes charge of a risky project and has two tasks to perform. The first task is a costly choice of effort that increases the probability of success of the project, whereas the second task is an exploration task that has an impact on the agent's productivity and increases the likelihood of success for each given level of effort. The principal either observes the agent's initial productivity (subsection 3.1) or the chosen resource allocation (subsection 3.2) and designs a contract that consists of transfers given the success or failure of the projects and the observed variable. Finally, then the agent chooses how to allocate a perfectly divisible unit of time between the two tasks and the amount of productive effort he exerts during the designated time.

When the resource allocation is not observable, for any level of bonus, the agent exerts under-allocates resources to increasing his productivity compared to the principal's preferred allocation (proposition 1). The intuition behind this result is that both players benefit from the total increase in the probability of success while potential increases in the total cost are privately supported by the agent. More specifically, the principal's favorite allocation is the one which maximizes the probability of success given the level of bonus, however, as increasing the productivity affects the cost of infra-marginal unites of effort, the agent's cost increases. This later effect is not taken into account by the principal.

Due to this misalignment of incentives, increasing the levels of bonuses have two effects on the agent's: A "performance effect" which makes him, given the resource allocation, increase his effort and generate higher probabilities of success. This effect is the standard one in the moral hazard literature. The second effect is a "productivity effect" which leads the agent to have incentives which are closer to the principal's interest and allocate more resources to increasing the productivity. This implies that empirical measure of bonuses reflect both the agent's incentives to exert effort and their incentives to allocate time to increasing their productivity. Interestingly, in applications such as on-the-job learning, while the first effect is specific to the period in which bonuses are high, the second effect can be impacted by promised future bonus, as workers can have incentives to smooth their learning.

In the baseline model, proposition 2 compares bonuses under unobservable allocations to ones in a benchmark in which the principal observes both the initial productivity and the resource allocation. Compared to the benchmark, the principal has incentive to reduce the bonus given that the agent has a lower productivity and an incentive to increase the bonus as higher bonuses increase the agent's incentives to increase his productivity. The total effect depends on which of the two forces dominate and we show that, when the effect of increasing the productivity on the cost of the infra-marginal units is low, the second effect dominates and the agent receives higher bonuses. On the other hand, when the cost effect is high, providing incentive to the agent to learn becomes more costly and the principal prefers reducing the bonus.

This result can be reinterpreted through the lens of the worker's ability to learn. When this ability is high, the worker is a fast learner and small increases in the bonus create high incentives to improve his productivity. This situation creates high incentives to the principal to increase the bonus and lead to higher bonuses. On the other hand, when facing slow learners, the principal cannot benefit much from the learning effect, therefore, she has incentives to reduce the bonus as she faces less productive types compared to the benchmark. Extending the results to settings in which there is asymmetric information about the "learning type" can be interesting and will be left to future work.

Finally, our last set of results analyse a setting in which the initial productivity is the agent's private information, however, the allocation is observable to the principal. This corresponds to environments such as allocating workers to tasks in which the manager has superior information about the project's needs but not about the role of each worker. In this situation, we show that when higher allocations are sufficiently costly, the principal can ignore the moral hazard problem and propose the same contracts as when both the allocation and the type are observable. This result is due to the fact that low types are too unproductive to accept a contract with low ex-post productivity whereas high types face a sufficiently high cost of mimicking a wasteful allocation.

**Related literature.** This paper contributes to the literature that studies multitasking (see Dewatripont et al. (1999) for a survey and Holmstrom and Milgrom (1991) for a seminal work). In that vein, the closest contribution is found in the work of Mukherjee and Vasconcelos (2011), who study optimal job design in a multitasking environment when the firms use implicit contracts. Crucially, in their setting, the two tasks available to the agents are assumed to be independent, as opposed to our canonical setting, in which one of the tasks determines the productivity of the other one.

Our work also relates to he literature that studies simultaneous adverse selection and moral hazard. In particular, Gottlieb and Moreira (2017) show that in this environment, under a multiplicative separability condition, the optimal mechanism offers a single contract. Guesnerie et al. (1989) show that in most cases, the moral hazard aspect does not entail welfare losses compared to the pure adverse selection case. Other notable contributions in that direction were made by Sung (2005), Ma (1991), and Ollier and Thomas (2013).

In terms of economic applications, the model adds to the literature that studies human capital formation in firms. The seminal contribution is found in Becker (1964), which has been extensively studied in the literature on

Labour Economics (see Leuven (2005) for a review). Related to our problem: Acemoglu and Pischke (1998) study a model in which the superior information of the current employer regarding its employees' abilities relative to other firms creates ex-post monopsony power, and encourages this employer to provide and pay for training, even if these skills are general, and Schlicht (1996) presents a model of moral hazard in which the trainee can form an opinion about the amount of on-the-job training only after training is completed, this creates a possibility for the firm to offer less training and make extra profit.

## 3.2. The model

**The production problem**. Consider a model where one principal (she) interacts with one agent (he) and can commit to a contingent wage schedule for each outcome. The principal wants to produce a good which gives her a value $v \in [0, 1]$. The agent has a publicly known initial productivity $\theta_0 \in \{\theta_L, \theta_H\}$ with respective probabilities $p$ and $1 - p$, with $p \in [0, 1]$. We set $0 < \theta_L < \theta_H$ He has two decision variables: a choice of total effort (henceforth, just "effort") $e \in [0, 1]$ and choice of resource allocation $\alpha \in [0, 1]$. Resources can be allocated to one of two tasks: A share $\alpha$ is allocated by the agent to a task that increases his productivity (e.g learning, investments, etc.), and a share $1 - \alpha$ is allocated to production that directly generates output.

Exploration is assumed to be costless, however, it affects the cost of exerting. Set $C(\alpha, e)$ to be the cost of exerting effort $e$ and allocating a share of resources $\alpha$ to increasing productivity. We set the cost function to be quadratic in effort $e$ and allocation $\alpha$:

$$C(e, \alpha) = \gamma \alpha^2 e + \frac{e^2}{2}. \tag{3.1}$$

Here, $\gamma > 0$ captures the intensity of complementarity between the resource allocation and the effort in the agent's cost. Production is stochastic and an effort-resource allocation scheme $(\alpha, e)$ leads to success with probability $\phi(\alpha, e, \theta_0)$. we assume that in addition to its effect on costs, resource allocation affects the probability of success in production and set:

$$\phi(\alpha, e, \theta_0) = max\{1, (\theta_0 + \alpha)\}e. \tag{3.2}$$

We interpret the problem as the one of a firm delegating product development to the agent. The agent has an initial knowledge $\theta_0$ about the product components which are relevant to the consumers. He decides the proportion of time/workers $\alpha$ to dedicate to market analysis to increase this knowledge. The rest of his time/workers is dedicated to developing the product. A trade-off arises as exerting a given amount of effort in less time or with fewer resources increases the cost for the agent. The agent's choice of learning trades off the **encouragement effect** due to the agent being more productive with the **discouragement effect** due to effort being more costly. However, as the benefits from a higher probability of success are shared between the principal and the agent while the cost is privately supported by the latter, disagreement regarding resource allocation can emerge. The objective of this paper is to study when players "disagree" in terms of resource allocation as well as the contractual implications of this disagreement.

**Observability of $\alpha$.** Both the agent's initial productivity $\theta_0$ and her choice of effort $e$ are unobservable to the principal. However, she can observes a signal $\hat{\alpha} \in \hat{A}$ about the resource allocation $\alpha$. We will study both the extreme cases of perfect observability ($\hat{\alpha} = \alpha$) and perfect non-observability ($\hat{\alpha}$ independent from $\alpha$).

**The payoffs**. We assume both players to be risk-neutral and denote by $s \in \{0, 1\}$ a variable which takes a value $s = 1$ when production succeeds and $s = 0$ otherwise. Finally, we denote by $T \in \mathbb{R}$ any realized transfer from the principal to the agent. For any given outcome $s$ and transfer $T$, the principal gets payoffs:

$$U^P = sv - T.$$

Similarly, the agent obtains payoffs:

$$U^A = T - C(e, \alpha).$$

**The mechanism design problem**. Without loss of generality, we restrict attention to direct mechanisms and normalize the agent's outside option to 0. A direct mechanism in this environment is a choice of transfers $T$ as a function of the agent's report about his type $\theta$, outcomes, and the signal $\hat{\alpha}$ :

$$T : \Theta \times s \times [0, 1] \to R$$

We assume that the agent has limited liability. Therefore, a feasible allocation satisfies, in addition to this constraint, the standard incentive compatibility and participation constraints. Denote by $\alpha(\theta)$ the process allocated to type $\theta$, a mechanism is feasible if and only if for all $\theta, \hat{\theta}, \alpha$, we have:

$$\underset{\alpha,e}{max} E_{s,\hat{\alpha}}\left[T(\hat{\alpha}, s, \theta) - C(\alpha, e, \theta)\right] \geq \underset{\alpha,e}{max} E_{s,\hat{\alpha}}\left[T(\hat{\alpha}, s, \hat{\theta}) - c(\alpha, e, \theta)\right] \quad (3.3)$$

$$\underset{\alpha,e}{max} E_{s,\hat{\alpha}}\left[T(\hat{\alpha}, s, \theta) - C(\alpha, e, \theta)\right] \geq 0 \quad (3.4)$$

$$\forall s : T(\hat{\alpha}, s, \theta) \geq 0. \quad (3.5)$$

Where 3.3, 3.4 and 3.5 are the standard incentive compatibility, participation and limited liability constraints respectively As the principal observes only reports, $\hat{\alpha}$ and realizations of production, by standard arguments, we obtain the following lemma that asserts the generalizability of linear contracts for our setting.

**Lemma 2.** ***Linear contracts*** *Any feasible mechanism has a payoff equivalent mechanism which consists of a fixed fee $F(\theta, \hat{\alpha})$ and a bonus $b(\theta, \hat{\alpha})$ in case of success.*

From here on, we will restrict attention to linear contracts and denote by

$(F, b)$ the mechanism which consists of a fixed fee $F$ and a bonus $b$.

**The timing**. The timing of the game is as follows.

**Stage 1:** nature draws the agent's type from a distribution with a CDF $F$. This type is privately observed by the agent.

**Stage 2:** the principal offers a menu of contracts $(F, b)_\theta$

**Stage 3:** the agent reports her type and chooses a contract.

**Stage 4:** the agent chooses a production process and effort level $e$ and the principal observes $\hat{\alpha}$.

**Stage 5:** the production outcome is realized and the agent is paid according to the chosen mechanism.

## 3.3. The equilibrium

In this section, we identify the type of misalignment of incentives in resource allocation that arises between the principal and the agent. For this purpose, we find it convenient to compare the equilibrium allocation to a benchmark in which $\theta_0$ and $\alpha$ are publicly observable and which implements, in equilibrium, the principal's preferred allocation. We redefine the problem as one in which the agent directly chooses the probability of success $\phi$ and the resource allocation $\alpha$. Denote by $\mathcal{C}(\alpha, \phi, \theta)$ the cost of this probability of success. Hence:

$$\mathcal{C}(\alpha, \phi, \theta_0) = \frac{\gamma\alpha^2}{\theta_0 + \alpha}\phi + \frac{1}{(\theta_0 + \alpha)^2}\frac{\phi^2}{2}.$$

Given the contract proposed to each type $\hat{\theta}$, the agent maximises:

$$\underset{\alpha, \hat{\theta}, \phi}{max} F(\hat{\theta} + \phi b(\hat{\theta}) - \mathcal{C}(\alpha, \phi, \theta). \tag{3.6}$$

Conditional on the report $\hat{\theta}$ and the choice of allocation $\alpha$, the probability of success $\phi^*(\theta, \theta_0, \alpha)$ that solves the above maximization problem is given by:

$$\phi^*(\theta, \theta_0, \alpha) = max\{0, (b(\theta_0 + \alpha) - \gamma\alpha^2)(\theta_0 + \alpha)\} \tag{3.7}$$

If the principal could observe both $\theta_0$ and the allocation $\alpha$, the problem boils down to a standard moral hazard problem in which $\alpha$ maximizes 3.7 and in which the principal's objective is to provide optimal incentives for the agent to exert effort. This choice of allocation is meant to reduce the intensity of moral hazard and make it easier to provide incentives to the agent whose marginal cost is lower. An incentive problem arises when the agent chooses $\alpha$ in an unobservable way. This can be explained in two ways: (i) whereas the principal makes his choice to "soften" the moral hazard problem related to effort, the agent does not have such concerns moreover, (ii) the benefits derived from the agent being more productive (lower marginal cost) are shared, whereas the cost of such gains is supported only by the agent who has to produce with less available resources.

On the other hand, when $\alpha$ is observable but the agent who has private information about her initial productivity has incentives to over-report his type so that he can induce fewer resources to be allocated to the support task. The objective of this section is to analyze these effects and the contractual distortions which can emerge depending on whether the agent's private information about her final productivity $(\theta_0 + \alpha)$ results from and ex-ante private information about $\theta_0$, a moral hazard problem due to unobservable allocation $\alpha$ or both.

### 3.3.1. Benchmark: Observable productivity and allocation

When both the agent's productivity and his chosen allocation are observable, the principal can implement her preferred allocation and the problem boils down to a standard moral hazard on effort. In this case, the optimal contract can be implemented through a bonus and a choice of allocation which solves:

**Proposition 11.** *When both the productivity and the allocation are observable, the optimal mechanism implements an allocation* $\alpha^B = min\{\frac{v}{2\gamma}, 1\}$ *and payments in case of success are* $b^B = \frac{v}{2} + \frac{\gamma(\alpha^B)^2}{2(\theta+\alpha^B)}$

First, note that this allocation proposed different bonuses to different types. This is due to learning changing both the marginal and the total cost for the agent which implies that the problem that the principal faces changes with the allocation. When the agent's initial productivity is maximal, the problem is a standard moral hazard problem and the bonus is $b = \frac{v}{2}$.

For types such that the principal implements an allocation which increases the productivity, the bonus is increasing in the allocation which means that, in equilibrium, both the benefits, in terms of a higher probability of success, and the cost of a higher allocation are shared between the two players. However, this relies on observability of both $\alpha$ and $\theta$. In many economic settings, this is not guaranteed, and the rest of this section will be dedicated to studying the contractual distortions which can emerge due to asymmetric information.

### 3.3.2. No ex-ante private information setting:

In economic settings like product development, or more generally, ones involving vertical relations in firms, it is often the case that initial productivity is common knowledge as the new "project" is commonly developed by the principal and the agent before the later benefits from delegation of decision making. This situation makes the ex-post asymmetries of information fully depend on the choice of resource allocation. We first study this type of environments and identify the type of misalignment of incentives in this allocation that can emerge and its contractual implications.

Formally, we consider a situation in which the agent's type in known to the principal (equivalently, we set $p \in \{0, 1\}$). By standard arguments, in this case, the optimal contract is equivalent to a choice of bonus in case of success and for any bonus $b \in (0, v)$, we have:

**Proposition 12.** ***Misalignment of incentive***

*For any level of bonus $b \in (0, v)$, there exists an initial productivity level $\theta^*$ such that (i) the agent under-allocates resources to increasing productivity for all $\theta < \theta^*$ and (ii) the agent reaches maximal ex-post productivity otherwise: for all $\theta \geq \theta^*$: $(\theta_0 + \alpha = 1)$.*

**Proof:** This result can be show using the agent's first order condition with respect of the probability of success (equation 3.7). Plugging this probability into his maximisation problem and solving for the equilibrium choice of allocation $\alpha$ leads to:

$$\alpha^A = max\{1 - \theta_0, \frac{b}{2\gamma}\} \tag{3.8}$$

Now, in order to compare this value with the principal's preferred allocation, note first that the principal's problem is linear in the probability of success $\phi$ and that for any bonus $b \in (0, v)$, it is strictly increasing in this probability. Therefore, the principal's preferred allocation is the one which maximizes the probability of success $\phi$ given the bonus $b$. Differentiating equation 3.7 with respect to $\alpha$ leads to:

$$\frac{\partial \phi}{\partial \alpha} = 2(b - \gamma\alpha)(\theta_0 + \alpha) - \gamma\alpha^2. \tag{3.9}$$

This equation can be rewritten as:

$$\frac{\partial \phi}{\partial \alpha} = 2(b - 2\gamma\alpha)(\theta_0 + \alpha) + \gamma\alpha^2 + 2\gamma\alpha\theta_0. \tag{3.10}$$

Evaluated at any $\alpha \leq \alpha^A$, this derivative is strictly positive which implies that the principal strictly prefers allocating more resources to the support task. Finally, this implies that whenever the agent's ex-post productivity is not maximal, we have $\alpha \in [0, 1 - \theta]$ which concludes the proof.

**Economic intuitions and contractual implications**. The intuition behind this result is that in this type of environment, bonuses are a way to share benefits from success between the principal and the agent, however, whereas (i) both players gain from higher probabilities of success, (ii) only the agent pays for the cost. More specifically, (i) implies that the principal's preferred allocation is the one for which the probability of success is the highest. As a result, she prefers more resources to be allocated to increasing productivity whenever this reduces the cost of the marginal unit. However, as this higher allocation leads also to an increase in the cost of infra-marginal units, this might not be in the interest of the agent.

This later effect is captured by the fact that the cost function is, given $\alpha$, supermodular in $(\alpha, e)$ for low probabilities of success and submodular for high probabilities. This implies that the cost of the first units of $\phi$ strictly increases in $\alpha$. This later effect is not taken into account by the principal when he observes both the type and the allocation.

In terms of contracting, as opposed to a benchmark in which both $\alpha$ and $\theta_0$ are observable, changing bonuses has an impact not only on effort but also on the resource allocation, therefore, to assess the contractual implications of this misalignment of incentives, first not that it is straightforward from equation 3.8 that:

**Corollary 2.** *The resources allocated to the support task $\alpha$ are either maximal or strictly increasing in $b$ and strictly decreasing in $\gamma$.*

As a result, increasing bonuses has two effects on the agent's incentives: First, for any resource allocation, it provides incentives to provide higher levels of effort: This effect is the standard one in moral hazard settings. The second effect of higher bonuses is that they affect incentives to increase productivity: This later effect is absent in the case of pure moral hazard benchmark, we have:

**Proposition 13.** ***Optimal contract*** *There exists $\theta^*$ such that, compared to the observable type and allocation benchmark:*
*The bonus is strictly lower than the one in the benchmark for all $\theta > \theta^*$.*
*If $\theta < \theta^*$, we have:*

- *If $\gamma < \gamma^*$: The bonus is strictly* ***higher*** *than the one in the benchmark whenever the latter does not implement maximal ex-post productivity*

- *If $\gamma > \gamma^*$: The bonus is strictly* ***lower*** *than the one in the benchmark whenever the latter does not implement maximal ex-post productivity*

**Proof:** (See appendix)

The intuition behind this result is that, as opposed to the benchmark case, the fact that the agent decides the resource allocation has two effects on the principal's incentive to increase the bonus: First, because the agent is, in equilibrium, less productive, the marginal effect of higher bonuses on the agent's effort is smaller which reduces the principal's incentives to provide high bonuses. The second effect is related to the result in proposition 1 as higher bonuses not only lead to an increase in effort but also lead the agent to increase his productivity. The total effect of increasing the bonus is ambiguous and depends on which of the two effects dominate.

Proposition 2 links these effects to the complementarity between learning and effort in the cost function. When learning leads to a high increase in the cost (high $\gamma$), higher bonuses provide very small incentives to increase the productivity which makes the principal's optimal bonus smaller than in the benchmark case: In this situation, the effect of lower productivity dominates. In the opposite case, when $\gamma$ is small, the agent react to high bonuses by a big increase in his productivity, therefore, bonuses end up being higher than in the benchmark case in order to provide incentives to increase productivity.

Finally, the initial productivity determines the cost of providing additional incentives to the agent. When $\theta$ is sufficiently high, it becomes more costly to provide incentives to increase productivity as it requires higher rewards for all infra-marginal effort units (and probabilities of success), therefore, in this case, the principal always reduces bonuses compared to the benchmark.

**Discussion and empirical implications**. The contractual distortion discussed above has at least three empirical implications. First, it predicts that for similar tasks, the moral hazard is more intense when it is accompanied by unobservable decision making: This is the main driver of higher bonuses. As a result, when comparing bonuses for similar jobs, it predicts that industries or firms in which management cannot assess the quality of decision making and can only evaluate results should be the ones in which the bonuses are higher. This aspect can be relevant when comparing firms in which managers/ supervisors have either the same or different backgrounds from the decision-maker as in this situation, not only outcomes but also decisions can be observed.

A second implication is related to the evolution of human capital in industries. As on-the-job learning can be interpreted as allocating resources (or time) between different tasks, insights from this section can be used to analyze how its incentives evolve inside organizations. As the model predicts that bonuses lead incentives to learn to increase, one should expect that higher bonuses affect not only production (*a performance effect*) but lead workers to dedicate more time to increase their productivity (*a productivity effect*). These effects are especially relevant when studying the effect of future increases or short term increases in bonuses.

Finally, as the resource allocation is affected by the bonus, the model suggests that observed bonuses reflect not only the riskiness and the intensity of the moral hazard problem given technologies and resource allocations or other determinants of workers' productivity but also the fact that these bonuses can be increased to shape this productivity.

### 3.3.3. An adverse selection setting

In many environments, investments in increasing productivity can be publicly observed. Firms observe whether workers participated and validated a training program, stakeholders observe the amount of investments which was allocated to increasing productivity etc. In these environments, the resource allocation can be observed, however, the principal lakes information about the initial productivity which means that assigning the right resource allocation to each agent type can be challenging and requires understanding the inherent adverse selection problem.

To study this interactions, we set the signal about the allocation to be $\hat{\alpha} = \alpha$, meaning that the allocation is publicly observable, and set $p \in (0,1)$. To simplify the analysis, we first provide the necessary optimality and feasibility conditions for any contract and we assume that $\theta_H = 1$. This assumption is not necessary for our results, however, it simplifies the analysis and makes comparative statics easier to derive.

First, note that as $\alpha$ is observable, the optimal contract is similar to the one in which a mechanism is a choice of $\alpha$, a bonus $b$ and a fixed fee $F$ as a function of the agent's type. Denotes by $(\alpha_\theta, F_\theta, b_\theta)$ any such a contract. We have:

**Proposition 14.** ***Implementability of the benchmark allocation:*** *The benchmark allocation is implementable for each type if and only if* $\theta_L \geq \frac{\gamma - v}{2\gamma}$

**Proof:** (see appendix)

Proposition 3 allows identifying parameters such that the adverse selection can be ignored. In order to provide intuition, we find it necessary to first describe the benchmark allocation. If both $\alpha$ and $\theta$ are observable, the principal's optimal allocation and bonus satisfy $(b_L, b_H, \alpha_L) = (\frac{v}{2} + \frac{\gamma \alpha_L^2}{2(\theta+\alpha}, \frac{v}{2}, \frac{v}{2\gamma})$.

First, note that when the principal proposes the benchmark allocations, the high type's trade-off is between higher bonuses provided to low types and the cost of a wasteful allocation. First, the difference in bonuses is decreasing in $\gamma$ which means that higher costs of increasing productivity makes the principal implement a lower $\alpha$ and therefore, provide a lower bonus to the low type. This makes the high type less keen to mimic. Moreover, as $\gamma$ increases, it becomes more costly for the high type to mimic the low type due to the increase in his cost of effort.

These effects are captured by the fact that the high type's incentive constraint is easier to satisfy when raising productivity is costly. Now, fixing the cost of increasing the productivity $\gamma$, the bonus provided to the low type is increasing in his initial productivity. This implies that if $\theta_L$ is too low, the benchmark allocation provides a bonus which is high enough for the high type's incentive constraint to be violated. Similar effects drive type $\theta_L$'s incentives with the exception that, while both types share the same cost effect of higher allocation, only the low type gains in productivity. This aspect makes his incentive constraint easier to satisfy.

**Discussion and extension to unobservable initial productivity and hidden action:** As the misalignment of incentives in terms of resource allocation is independent of the observability of types, the results provided in the unobservable allocation setting extend naturally to a setting in which types are nonobservable. When the principal does not observe types, nor allocation, she can only increase bonuses in case of success.

Now, these bonuses can be increased in order to provide incentives to the agent of type $\theta_L$ to implement a higher allocation, however, as opposed to proposition 2, another cost appears in the principal's objective function which related to the fact that increasing the bonus leads to a sub-optimal contract also for the high type. This effect pushed bonuses (and allocation) down compared to the no-ex-ante asymmetric information setting. This effect will be stronger if the agent is more likely to be of a high type, therefore, one can expect changes in distributions towards a higher frequency of high types to lead to lower bonuses and less learning both because the optimal bonuses for these types are lower and because the effect on allocations is more limited.

## 3.4. Conclusion

In this paper, we study the problem of resource allocation in organizations. An agent makes, in addition to his unobservable choice of effort, a choice of resource allocation which affects his productivity: A higher allocation raises both productivity of effort and its cost. This situation creates misalignment of incentives between the two players and we show that the agent's ex-post productivity is either maximal (when the initial productivity is sufficiently high) or sub-optimal from the principal's perspective. This result is due to the principal preferring allocations which reduce the marginal cost of effort evaluated at the agent's optimum whereas the agent takes into account the effect on infra-marginal units of efforts.

This misalignment of incentives leads to contractual distortions which depend qualitatively on the cost of increasing the allocation. When this cost is high, the principal provides a bonus which is lower than the one in a benchmark in which the allocation is public knowledge. When the cost is low, increasing the bonus induces the agent to invest more in his productivity as he exerts a higher effort which leads to higher equilibrium bonuses. This later aspect has empirical implications in labor economics: First, bonuses are not only reflecting the (marginal) intensity moral hazard problem given the productivity of workers but also the fact that their private decisions such as learning affect this intensity. Moreover, this implies that a short term increase in bonuses can have a long term effect on output. This prediction is due to the fact that, in addition to their higher effort, workers can spend more time learning and, in a dynamic setting, have higher future output.

## 3.5. Appendix:

**Proof of proposition 2:**

The principal's problem is:

$$\max_{b} \phi(b)(v-b).$$

When both the initial productivity and the resource allocation are observable, the optimal contract is similar to a case in which the principal directly chooses $\alpha$, the first-order condition of his maximization problem with respect to the bonus is:

$$\frac{\partial \phi}{\partial b}(v-b) - \phi = 0$$

However, when $\alpha$ is non-observable, bonuses affect both the resource allocation and the choice of effort. In this case, the optimal bonus satisfies:

$$(\frac{\partial \phi}{\partial b} + \frac{\partial \phi}{\partial \alpha}\frac{\partial \alpha}{\partial b})(v-b) - \phi.$$

To proof the proposition, it is sufficient to show that for all bonuses which are lower than the optimal bonus in the benchmark case, the payoffs of the principal are strictly increasing in $b$ or low $\gamma$ and strictly decreasing otherwise. We do that in three steps: First, we show that the first order condition is a necessary and sufficient optimality condition, then, we show that for all bonuses which are lower than the argmax of equation 3.5, the difference between the first order derivatives with respect to bonuses of the principal's objective function and the one in the benchmark case is monotone. Finally, we conclude the proof by showing that the sign of this derivative is positive for low $\gamma$ and negative otherwise.

**Step 1: The first order condition is a necessary and sufficient optimality condition:**

Note first that the third order derivative of the principal's objective function with respect to $b$ can be written as:

$$\frac{d^3U^P}{db^3} = \frac{3v - 12b - 18\gamma\theta_0}{4\gamma^2}$$

This equation is positive when $b < \frac{3v-18\gamma\theta_0}{12}$ and negative otherwise. This implies that the second order derivative is concave and single peaked when $b = \frac{v-6\gamma\theta_0}{4}$. This implies that the second order derivative reaches its maximum when $b \leq v$ if $v \geq \frac{v-6\gamma\theta_0}{4}$ and when $b \geq v$ if $v \leq \frac{v-6\gamma\theta_0}{4}$. Using the fact that any feasible and optimal bonus is such that $b \in [0, 1]$, it is straightforward to show that the derivative when $b = v$ is sufficient for showing the result we want to show.

**Case 1:** $v \leq \frac{v-6\gamma\theta_0}{4}$

In this case, the second order derivative is increasing in the bonus $b$ for all $b \in [0, v]$. Therefore, in this interval, it is maximized at $b = v$ and the maximum is:

$$\frac{\partial^2U^P}{\partial b^2}(1) = -\frac{3v^2 + 12\gamma\theta v + 8\gamma^2\theta^2}{4\gamma^2} < 0$$

This implies that $\forall b < v$, we have: $\frac{\partial^2U^P}{\partial b^2} < 0$ and the principal's objective function is strictly concave in $b$ for any $b \in [0, 1]$ and the first order condition is a sufficient optimality condition.

**Case 2:** $v \geq \frac{v-6\gamma\theta_0}{4}$

In this case, the second order derivative is decreasing in the bonus $b$. Therefore, in the interval $[0, v]$, it is maximized at $b = 0$. We can identify two cases: First, for parameters such that this derivative is negative, we conclude that the objective function is strictly concave and the first order condition is a sufficient optimality condition.

In the second case, if $\frac{\partial^2U^P}{\partial b^2}(0) > 0$, we know that this derivative is continuous, therefore, using the fact that $\frac{\partial^2U^P}{\partial b^2}(1) < 0$, we know that there exists $b^* \in [0, v]$ such that the $\frac{d^2U^P}{db^2} > 0$ when $b < b^*$ and $\frac{d^2U^P}{db^2} < 0$ otherwise. This implies that the principal's objective function is increasing and convex in $b$ when $b < b^*$ and concave in the interval $b > b^*$. As a result, using the fact that the first order derivative is positive at $b = 0$, we know that $\forall b < b^*$, $\frac{\partial U^P}{\partial b} > 0$. This implies that in this case, the optimal bonus is $b \geq b^*$ which implies that either the optimal bonus is $b = v$ or the first order condition is sufficient.

**Necessity:** Now, using the fact that the first order derivative is strictly positive when $b = 0$ and strictly negative when $b = v$, we know that the optimal bonus is interior and that the first order condition is a necessary optimality condition which concludes our proof.

Now, denote by $b^B$ and $\alpha^B$ the optimal bonus and allocation in the benchmark case.

**Step 2: $\forall b \leq b^B$, the difference between the first order derivatives in the principal's problem and the benchmark is monotone in $\gamma$**

First, from the first order condition of the principal's problem in the benchmark, we have: for any $\alpha$:

$$b^B = \frac{v}{2} + \frac{(\alpha)^2\gamma}{2(\theta_0 + \alpha)}$$

Plugging this value in the allocation problem, we obtain:

$$\alpha^B = \frac{v}{2\gamma}$$

Now, we will show that $\forall b \in [0, b^B]$ the result in step 2 holds.

The difference between the derivatives of the principal's problem and the benchmark problem is:

$$\Delta = (v-b)((\frac{\partial \phi}{\partial b}(b, \alpha(b)) + \frac{\partial \phi}{\partial \alpha}\frac{\partial \alpha}{\partial b}(b, \alpha(b))(v-b) - \frac{\partial \phi}{\partial b}(b, \alpha^B)) - \phi(b, \alpha(b)) + \phi(b, \alpha^B) \tag{3.11}$$

Differentiating with respect to $\gamma$, we obtain:

$$\begin{aligned}\frac{\partial \Delta}{\partial \gamma} =& \frac{4b^3 - 3b^2 v + 3\gamma\theta_0 b(3b - 2v) - 4\gamma^3((\alpha^B)^2\theta_0 + (\alpha^B)^3)}{4\gamma^3} \\ &\leq \frac{4b^2 b^B - 3b^2 v + 3\gamma\theta_0 b(3b^B - 2v) - 4\gamma^3(\alpha^B)^2(\theta_0 + (\alpha^B)}{4\gamma^3} \\ &= \frac{b^2(4b^B - 3v) + 3\gamma\theta_0 b(3b^B - 2v) - 4\gamma^3(\alpha^B)^2(\theta_0 + \alpha^B)}{4\gamma^3}\end{aligned}$$

Where the inequality in the second line comes from the fact that $b \in [0, b^B]$. Now, replacing $b^B$ by its value we obtain:

$$\begin{aligned}\frac{\partial \Delta}{\partial \gamma} &\leq \frac{b^2(2\frac{\gamma(\alpha^B)^2}{\theta_0+\alpha} - v) + 3\gamma\theta_0 b(\frac{3(\gamma(\alpha^B)^2}{2(\theta+\alpha^B)} - \frac{v}{2}) - 4\gamma^3(\alpha^B)^2(\theta_0 + \alpha^B)}{4\gamma^3} \\ &\leq \frac{b^2(2\gamma\alpha^B - v) + 3\gamma\theta_0 b(\frac{3\gamma\alpha^B}{2} - \frac{v}{2}) - 4\gamma^3(\alpha^B)^2(\theta_0 + \alpha^B)}{4\gamma^3} \\ &\leq \frac{3\gamma\theta_0 b\frac{v}{4} - \gamma v^2(\theta_0 + \frac{v}{2\gamma})}{4\gamma^3} \\ &< \frac{-\gamma\theta_0\frac{v^2}{4} - \frac{v^3}{2}}{4\gamma^3} \\ &< 0\end{aligned}$$

The second inequality is obtained using $\theta \geq 0$ and the third inequality is obtained by replacing $\alpha^B$ by its value. Finally, the fourth inequality uses $b < v$ and and rearranging the terms.

**Step 3: Conclusion** From step 1, we know that the first order condition is both a necessary and sufficient optimality condition. This implies that, for $b^P$ the solution to the principal's problem, the derivative of her objective function with respect to bonuses is positive for lower bonuses ($\forall b < b^P : \frac{\partial U^P}{\partial b} > 0$) and negative otherwise ($\forall b > b^P : \frac{\partial U^P}{\partial b} < 0$). Therefore, to proof the proposition, it is sufficient to show that for $b = b^B$, the sign of the derivative of the principal's problem is positive for low $\gamma$ and negative otherwise.

Step 2 shows monotonicity of the difference between the two derivatives with respect to $\gamma$. This allows us to conclude that either for all $\gamma$ the principal always prefers higher bonuses compared to the benchmark, always prefers lower bonuses or there is a cutoff $\gamma^*$ such that $\forall \gamma < \gamma^*$ she prefers higher bonuses and $\forall\gamma > \gamma^*$ she prefers lower ones.

Now, note that learning is maximal under the the benchmark bonus if and only if $b^B \leq 2\gamma(1-\theta_0)$. Therefore, we will restrict attention to $\gamma \geq \frac{b^B}{2(1-\theta_0)}$. Now, evaluating the derivative of the principal's problem at $b = b^B$, and computing its limit as $\gamma$ goes to infinity we obtain:

$$\lim_{\gamma\to\infty} \frac{\partial U^P}{\partial b}(b^B) = \theta_0^2(v - 2b^B) < 0$$

Similarly, when $\lim_{\gamma \to \frac{b^B}{2(1-\theta)}^+}$, this derivative becomes:[1]

$$\lim_{\gamma \to \frac{b^B}{2(1-\theta)}^+} \frac{\partial U^P}{\partial b}(b^B) = \frac{1}{4\gamma^2}((v-b^B)(b^B)^2(3+\frac{6\theta}{1-\theta}+\frac{2\theta^2}{(1-\theta)^2})-(b^B)^3(1+3\frac{\theta}{1-\theta}+\frac{2\theta^2}{(1-\theta)^2})$$

Simple rearrangement of the terms leads to:

$$\begin{aligned}
4\gamma^2\frac{1-\theta}{(b^B)^2}\frac{\partial U^P}{\partial b}(b^B) &= (v-b^H)(3+3\theta+\frac{2\theta^2}{(1-\theta)}) - (b^B)(1+2\theta+\frac{2\theta^2}{(1-\theta)}) \\
&= (v-b^B)(3+\theta+2\theta(1+\frac{\theta}{(1-\theta)})) - (b^H)(1+2\theta(1+\frac{\theta}{(1-\theta)})) \\
&= (v-b^B)(3+\theta+\frac{2\theta}{(1-\theta)}) - (b^B)(1+\frac{2\theta}{(1-\theta)}) \\
&= (v-b^B)(2+\theta+\frac{1+\theta}{(1-\theta)}) - b^B\frac{1+\theta}{(1-\theta)} \\
&= v(2+\theta) + (v-2b^B)\frac{1+\theta}{(1-\theta)} \\
&= \theta v + \frac{1+\theta}{(1-\theta)}(v-2b^B+v(1-\theta)) \\
&= \theta v(1-\frac{1+\theta}{(1-\theta)}) + \frac{1+\theta}{(1-\theta)}(2v-2b^B) \\
&= \frac{2}{1-\theta}(-\theta^2 v + (1+\theta)(v-b^B))
\end{aligned}$$

This function is monotone in $\theta$. For $\theta = 0$, it is strictly positive and the

[1] We take the limit as the function has a kink at $\frac{b^B}{2(1-\theta)}$ given that bellow that level, the agent exerts maximal learning

principal provides higher bonuses that in the benchmark (using the fact that $(v-b^B)>0$ and for $\theta=1$, the function is negative (using $(v-b^B))<\frac{v}{2}<\frac{1}{2}$), therefore, the bonus is strictly lower than the benchmark which concludes the proof.

### 3.5.1. Proof of proposition 3:

This result is straightforward: As the more productive type has maximal productivity, he dedicates full resources to production when types are observable and obtains payoffs: $U^A(\theta_H,\theta_H)=\frac{v^2}{8}$.

On the other hand, the low type's cost function as well as the bonus change with the allocation and the bonus satisfies $b^B=\frac{v}{2}+\frac{\gamma\alpha^2}{2(\theta_L+\alpha)}$. As this bonus is higher than the one type $\theta_H$ obtains, the later trades-off higher bonus with more costly allocation due to an allocation which increases the cost without increasing his productivity. In this case, he exerts effort $e=b_L-\gamma\alpha_L^2$ and his incentive compatibility constraint is satisfied when:

$$\frac{v^2}{8}\geq\frac{(b_L-\gamma\alpha_L^2)^2}{2}=(\frac{v}{2}+\frac{\gamma\alpha_L^2}{2(\theta_L+\alpha_L)}-\gamma\alpha_L^2)^2$$

When $\theta_L+\frac{v}{2\gamma}\leq 1$, the optimal allocation for the principal is $\alpha_L=\frac{v}{2\gamma}$. In the opposite case, the allocation is $\alpha_L=1-\theta_L$. It is straightforward to verify that in both the high type's incentive compatibility constraint is satisfied. Note that for the only if statement, that we have:

$$\theta_L+\frac{v}{2\gamma}>1\implies\theta_L>1-\frac{v}{2\gamma}>\frac{\gamma-v}{2\gamma}$$

Now we derive the incentive compatibility constraint for the type $\theta_L$. $\theta_L$ prefers the contract $(F_L,b_L,\alpha_L)$ if and only if:

$$\frac{(b_L(\theta_L+\alpha_L)-\gamma(\alpha_L)^2)^2}{2}\geq\frac{(\theta_L v)^2}{8}$$

Rearranging the terms, and using $b_L=\frac{v}{2}+\frac{\gamma\alpha_L^2}{2(\theta_L+\alpha_L)}$; this constraint is equivalent to:

$$\frac{v(\theta_L + \alpha_L)}{2} - \frac{\gamma(\alpha_L)^2}{2} \geq \frac{\theta_L v}{2}$$
$$\iff v - \gamma\alpha_L \geq 0$$

Either (i) $\frac{v}{2\gamma} + \theta_L \leq 1$ and we have $\alpha_L = \frac{v}{2\gamma}$ and the incentive constraint is verified or (ii) $\alpha_L = 1 - \theta_L$ in which case the condition boils down to $\theta_L \geq 1 - \frac{v}{\gamma} > \frac{1}{2} - \frac{v}{\gamma}$

# Bibliography


Acemoglu, D. and Pischke, J.-S. (1998). Why Do Firms Train? Theory and Evidence*. *The Quarterly Journal of Economics*, 113(1):79–119.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

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.

Dewatripont, M., Jewitt, I., and Tirole, J. (1999). The economics of career concerns, part i: Comparing information structures. *The Review of Economic Studies*, 66(1):183–198.

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.

Gottlieb, D. and Moreira, H. (2017). Simple contracts with adverse selection and moral hazard.

Guesnerie, R., Picard, P., and Rey, P. (1989). Adverse selection and moral hazard with risk neutral agents. *European Economic Review*, 33(4):807 – 823.

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

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7:24.

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.

Leuven, E. (2005). The economics of private sector training: A survey of the literature. *Journal of Economic Surveys*, 19(1):91–111.

Ma, C.-t. A. (1991). Adverse selection in dynamic moral hazard. *The Quarterly Journal of Economics*, 106(1):255–275.

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

Mukherjee, A. and Vasconcelos, L. (2011). Optimal job design in the presence of implicit contracts. *The RAND Journal of Economics*, 42(1):44–69.

Ollier, S. and Thomas, L. (2013). Ex post participation constraint in a principal–agent model with adverse selection and moral hazard. *Journal of Economic Theory*, 148(6):2383–2403.

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.

Schlicht, E. (1996). Endogenous on-the-job training with moral hazard. *Labour Economics*, 3(1):81 – 92.

Sung, J. (2005). Optimal contracts under adverse selection and moral hazard: a continuous-time approach. *The Review of Financial Studies*, 18(3):1021–1073.

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.