Dynamic Amnesty Programs

Sam Kapon*

[Click here for the latest version]

6th May 2021

Abstract

A regulator faces a stream of agents each engaged in crime with stochastic returns. The regulator designs an amnesty program, committing to a time path of penalty reductions for criminals who self-report before they are detected. In an optimal time path, the intertemporal variation in the returns from crime can generate intertemporal variation in the generosity of amnesty. I construct an optimal time path and show that it exhibits amnesty cycles. Amnesty becomes increasingly generous over time until it hits a bound, at which point the cycle resets. Agents engaged in high return crime self-report at the end of each cycle, while agents engaged in low return crime self-report always.

Keywords— Dynamic Mechanism Design, Self-Reporting, Amnesty, Crime

1. Introduction

To stop ongoing crime, a regulator can offer preferable treatment to criminals who self-report. These *amnesty*, or *self-reporting programs*, appear in such diverse contexts as illegal gun ownership, collusion, desertion in war, tax evasion, espionage[1], civil conflict, and corruption. For instance, the U.S. Department of Justice operates a leniency program for self-reporters,

---

*New York University, sdk301@nyu.edu. I am indebted to Sylvain Chassang for his guidance and patience. I am grateful to Guillaume Fréchette and Dilip Abreu for their support. This paper has benefited from discussions with and comments from David Pearce, Erik Madsen, Debraj Ray, Paula Onuchic, Dmitry Sorokin, Mauricio Ribiero, Alessandro Lizzeri, Nikhil Vellodi, Andrew Schotter, Joshua Weiss, Stefan Bucher, Lester Chan, Ariel Rubinstein, Maher Said, Dhruva Bhaskar and Alistair Wright. I thank seminar participants at the 2020 Young Economist Symposium, Berkeley Haas, and various NYU seminars for their comments.

[1] Such as the amnesties offered to British informants by the Irish Republican Army in the 1980's.

which has become its "most important investigative tool for detecting cartel activity."[2] The Red Army's amnesty for military desertion in June 1919 induced the return of over 100,000 deserters.[3] Australia's gun buy-back of 1997 collected more than 650,000 weapons.[4] The Chieu Hoi program offered amnesty to defectors during the Vietnam war, enticing over 100,000.[5]

An extensive theoretical literature has investigated the use of self-reporting programs in one-shot regulation.[6] Less attention has been paid to the inter-temporal properties of these programs, which are often offered on a repeated, time-limited, basis. The Red Army's Central Anti-Desertion Commission operated repeated amnesty periods, interspersed with periods of harsh enforcement and a similar program has historically been applied to desertion in French militaries.[7] The Brazilian gun buyback program has been run four times since 2013.[8] The U.S. has operated a number of tax-related self-reporting programs, often on a repeated, time-limited, basis.[9] Other programs are offered continuously. For instance, the U.S. Department of Justice's cartel leniency program and the Mexican gun buyback are, and the Chieu Hoi program was, run continuously without explicit adjustment to the terms of self-reporting.

In this paper, I ask: how should the terms of self-reporting programs be designed *over time*? Should they be constant or fluctuate inter-temporally? I study a mechanism design problem in which criminal agents arrive at a time-homogeneous rate and their returns from crime are private, idiosyncratic and evolve over time. In particular, criminals can transition from a high return state of crime to a low return state of crime. A regulator commits to a time-path of penalty reductions for agents who self-report before they are detected that applies uniformly to all agents. The range of possible punishments is bounded and the agent may be exogenously detected, at which point the regulator applies the maximum punishment possible. An agent's only decision is when, if ever, to self-report.

The inter-temporal variation in the returns from crime drives the desirability of dynamic policies for the regulator in this environment. Because returns from crime can transition from high to low, static self-reporting programs are subject to exploitation. The intuition for this is best seen by comparing two extreme policies. The first is a *static policy*, offering the same terms for self-reporting at all times. This program lets the agent benefit both

[2] https://www.justice.gov/atr/leniency-program
[3] See Figes (1990).
[4] See Leigh and Neill (2010).
[5] See Wosepkat (1971)
[6] See for instance, Kaplow and Shavell (1994), Malik (1993) and Andreoni (1991) for early contributions.
[7] See Wright (2012) for desertion in the Red Army and Forrest et al. (1989) for desertion in French militaries.
[8] See Macinko et al. (2007).
[9] OECD (2015), Luitel and Sobel (2007)

from crime while his return is high and from self-reporting once his return is low. At the opposite extreme is a *one-time policy*, in which agents only have one chance to self-report for favorable treatment and are otherwise treated harshly, as if detected exogenously. Under the one-time policy, agents with high returns from crime choose to self-report rather than wait for their return to become low, knowing that by then the option to self-report will be gone. The one-time policy is therefore able to generate self-reporting by higher return agents than is the static policy.

The drawback of the one-time policy is that agents who arrive after the single reporting opportunity never self-report. The regulator must then balance two forces: (i) enticing contemporaneous agents to self-report by offering a future with less opportunity for self-reporting and (ii) enticing future agents to self-report by not completely shutting down these opportunities, as is done in the one-time policy. This trade-off is explored in the remainder of the paper.

The basic features of the model are motivated by the following observations. First, the returns from crime often accrue slowly over time. For instance, deserters value each moment away from their military posts, and cartels accrue profits from price-fixing slowly over time. Second, returns from crime are private, idiosyncratic and *change over time*. Military deserters face uncertain food and shelter availability, and an uncertain risk of being caught (Forrest et al., 1989). Illegal gun owners may leave crime (Willmer, 1971) or find themselves in need of the money from a gun buyback (Dreyfus et al., 2008). Cartels face fluctuating demand conditions, new entrants de-stabilize collusion, the risk of detection changes over time (Connor (2007), Gärtner (2014)), and these may be difficult to observe until long after the cartel has been detected, or ever. Third, in many settings of interest, crime has long-term, irreversible effects: a deserter cannot stop being a deserter without military permission, a change in tax payment can spark IRS scrutiny, and in general the evidence of crime is persistent. This irreversibility motivates the assumption that the only way to leave crime is to self-report to the regulator. Finally, amnesty typically takes the form of a reduction in penalties for any agent who self-reports at a given time, motivating the regulator's problem as a choice of a time path for amnesty that applies uniformly to all agents.[10]

The main result (Theorem 2) characterizes an optimal amnesty policy and shows that it takes a *cyclical* form; at regularly spaced times, the regulator offers the minimum self-reporting penalty available, and in between offers a decreasing schedule of penalties that eventually hits this minimum, after which the cycle resets. The decreasing schedule of penalties induces agents with a low return from crime to immediately report, and is chosen

[10] In Section 7, the features of the model are further discussed and interpreted.

so that these agents are indifferent between immediately reporting and waiting until the end of the cycle to report. Agents with high returns from crime report at the end of each cycle, when the self-reporting penalty is at its minimum. The frequency of cycles increases with the risk of detection, the penalty for detection, and the rate of transition from high to low return crime.

To arrive at this policy, I first provide a key lemma (Lemma 1); any regulatory policy can be transformed into another policy in which (*i*) high return agents exhibit the same reporting behavior as in the original policy, (*ii*) low return agents self-report immediately at all times and (*iii*) low return agents are indifferent between reporting immediately and behaving like high return agents. With this lemma, the regulator's problem can be solved in two steps: (*i*) optimize over policies in which agents operating high and low return crimes exhibit the same reporting behavior and the regulator experiences no loss from low return crime, then (*ii*) apply the construction in the lemma to the policy from the first step to recover an optimal policy for the general case in which the regulator's loss from agents who operate low return crime is non-zero.

I formulate the first step recursively, with the decision nodes as agents' reporting times and the regulator's state as the self-reporting penalty that she must offer at the contemporaneous decision node. At any decision node, the regulator chooses (*i*) the *delay* until the next decision node and (*ii*) the self-reporting penalty at that time i.e. the *generosity* of the next penalty. This latter choice becomes the state at the next decision node. The constraint of the problem is a one-shot incentive compatibility constraint: an agent operating high return crime should prefer to report immediately at the decision node than to delay reporting until the next decision node. If the constraint is satisfied for high return agents, then it is satisfied for low return agents, and so the constraint for the latter can be dropped.

The one-shot incentive compatibility constraint can be satisfied in many different ways. The regulator can combine a long delay with a generous amnesty at the next reporting time. Or, the regulator can combine a short delay with an ungenerous amnesty at the next reporting time. Each of these generate different values for the regulator. With the former, the regulator must suffer a long period of time in which criminals are allowed to accumulate. With the latter, the regulator must offer an ungenerous amnesty at the next decision node, which is indirectly costly because it enters the future incentive compatibility condition and constrains the policies available to the regulator. The crucial step of the main result (Theorem 2) is to show that, whatever the penalty the regulator must immediately deliver (the state), the optimal way to do it is by offering the next penalty to be the minimum self-reporting penalty, and the delay as long as necessary to satisfy the incentive compatibility condition. Repeatedly applying this optimal policy in the recursive problem generates the path for an optimal policy described above.

A *backloading* motive on the part of the regulator is the driving force behind the optimality of this policy. In particular, I show that the regulator and agent face linearly related payoffs between decision nodes, but *across* decision nodes, agents face not only time discounting r, as does the regulator, but also a risk of detection ρ, which acts like additional time discounting. This effectively makes the agents less patient than the regulator. Intuitively, this relationship means that the regulator prefers to incentivize reporting using the delay dimension rather the generosity dimension. In particular, the regulator always incentivizes high return agents to report using the same amnesty level — the minimum self-reporting penalty — and creates as much delay between these amnesties as is necessary to incentivize reporting.

In Section 6, I show that by allowing the arrival rate of agents to decay over time, the main result as well as the backloading intuition are further clarified. I prove the following results: if the rate of decay is *smaller* than the rate of detection, the policy of the main result remains optimal, but if the rate of decay is *larger* than the rate of detection, then a different policy will be optimal. In this latter case, when the rate of decay is larger than the rate of detection, an optimal policy instead takes a *front-loaded* form: the regulator offers an amnesty that becomes continuously less generous over time and, within finite time, makes an upward jump to some long-run level at which it remains fixed. Both high and low return agents report on the increasing portion, while only low return agent report at the long-run level afterwards. This front-loading of reporting by high return agents, relative to the optimal policy described in Theorem 2, reflects the front-loaded arrival of agents to the model.

In Section 5, I discuss a number of empirical settings in which dynamic self-reporting policies play a role. In the case of military desertion, I recount qualitative evidence from a case study of the Red Army's anti-desertion campaign in Karelia detailed in Wright (2012), among other sources, to argue that the dynamic policies observed were a result of deliberate design intended to induce fast self-reporting. I move to illegal gun ownership and consider how the design results of the model may be applied to improve amnesty and buyback programs. Last, I discuss the application of the model to voluntary disclosure and amnesty programs in tax collection.[11]

After reviewing the literature, I introduce the model in Section 2 and analyze it in Section 3. In Section 4, I present a number of comparative statics and discuss how investments

[11]In some settings, in particular tax collection, the perceived fairness of enforcement may lead to a moral obligation to comply that generates higher compliance than would be implied by enforcement strength and financial considerations alone. In such cases, amnesty may backfire and lead to a deterioration of this moral obligation. I discuss this further in Section 7.

in other features of the environment can act as complements to dynamic amnesty. In Section 5, I present applications of the model. In Section 6, I extend the model to allow the rate at which agents arrive to decay over time. The assumptions of the model and alternative modeling choices are discussed in Section 7. I conclude in Section 8.

Contribution. This paper makes two contributions. First, it proposes a novel mechanism through which intertemporal variation in amnesty may be optimal. In certain settings, such as the Red Army's anti-desertion campaign, qualitative evidence is provided that supports this mechanism as a driver of the decision to vary amnesty over time. In other settings, such as tax and gun amnesty, in which the intertemporal variation in amnesty is more naturally understood as a response to public pressure or short-term budget constraints, the model is used to highlight a possible benefit to such intertemporal variation and to propose potential policy improvements.

Second, the paper solves a novel dynamic mechanism design problem, in which randomly arriving agents have stochastic values for an interaction with a regulator and can choose to irreversibly end their interaction with the regulator at some cost (i.e. amnesty). The most closely related work is in the determination of an optimal pricing path for a durable goods monopolist facing a stream of randomly arriving buyers with stochastic values for the product, which I discuss below in the related literature section.

Literature. This paper is related to the theoretical literature on self-reporting programs and the dynamic mechanism design literature, in particular intertemporal price discrimination in the economics and operations research literatures.

The early work of Kaplow and Shavell (1994), Malik (1993), and Andreoni (1991) studied law enforcement and self-reporting behavior in one-shot settings. Much of the subsequent literature is concerned with one-shot self-reporting settings in which the optimal inter-temporal use of amnesties cannot be studied.

Nevertheless, the dynamic properties of self-reporting programs have received some attention in the theoretical and empirical literature, although no theory has been developed that accounts for time-variation in the returns from crime. For instance, Marchese and Cassone (2000) rationalizes repeated tax amnesties as a method of discriminating between tax payers who are ex-ante different. Wang et al. (2016) studies how a regulator should design remediation and inspection policies for environmental hazards that arrive randomly over time e.g. leaks. A firm has an option to delay repair of its environmental hazard and the paper focuses on the interaction between the inspection policy and penalties. As the paper shows, when the rate of inspection (which is like the rate of detection in this paper) cannot be chosen but is instead Poisson, optimal self-reporting programs are always static, unlike in this paper. I focus on the role that dynamic self-reporting programs play absent any control of inspection policies but in the presence of dynamic returns from crime. In this sense, the papers are complementary.[12]

This paper is related to work in dynamic mechanism design such as Battaglini (2005) and the work on intertemporal price discrimination by a durable goods monopolist, such as Conlisk et al. (1984), Deb (2014), Garrett (2016) and Araman and Fayad (2020). The most closely related work is Garrett (2016) who studies a durable goods monopolist choosing a price path for dynamically arriving agents with changing values for a product and finds that cyclical pricing is optimal. A fundamental difference between our papers is the limited penalties the regulator in this paper has at her disposal; the regulator cannot punish above a maximum level, and faces a bound on the self-reporting incentive she can offer.[13] This constraint makes the regulator's problem in this paper non-trivial (since otherwise she could set the maximum penalty and the penalty for self-reporting low enough to induce immediate reporting by agents) but precludes the use of techniques applied in Garrett (2016). Solving the model in this paper therefore requires a different approach that explicitly incorporates these limited penalties. Preferences in this paper also differ from those of the price discrimination setting, and this leads to different intuition underlying the optimal policy.[14]

This paper is not the first to note a link between amnesty and intertemporal price discrimination. Marchese and Cassone (2000) applies intuition and techniques from the literature on inter-temporal price discrimination to study a model of tax amnesty. Since values there are static, results are more closely connected to a monopolist discriminating between consumers with different but static values e.g. Conlisk et al. (1984).

2. The Model

A stream of criminal agents (he) must decide whether to continue to operate or apply for amnesty. A regulator (she) chooses and commits to a penalty policy that is relevant for an agent's decision. Calendar time is continuous, $t \in \mathbb{R}_+$.

[12]In the environmental hazard setting, the authors also show that it is without loss of generality to study mechanisms that induce immediate reporting and repair of the hazard. In the setting of this paper, the analogue of this is true only for low return agents.

[13]A natural lower bound on this incentive is that the regulator can at most offer not to punish a self-reporting agent at all, but the model allows any arbitrary bound.

[14]The regulator is concerned only with *fast* self-reporting by the agents. In the durable good monopoly setting, it would be as if the monopolist cared only that buyers purchased quickly, but not about the price.

2.1 The Agents

I present below the details of the agents' environment.

Arrival and Flow Gain. Infinitesimal agents arrive at constant flow rate normalized to 1. Each agent is endowed with an individual *flow gain* process that follows a standard two-state continuous-time Markov chains, denoted x_t, independent across agents, with state space $E = \{x^l, x^h\}$ such that $0 \leq x^l < x^h$.[15] For simplicity, state x^l is absorbing and agents transition from state x^h to x^l at Poisson rate λ. Upon arrival, agents are initialized in state x^h.[16] I index x_t by time since arrival so that x_0 is the initial state of an agent upon arrival.

Choice and Detection. An agent arriving at time t_0 chooses a $[0, \infty]$-valued stopping time with respect to the filtration generated by $(x_t)_{t \geq 0}$, denoted τ, to irreversibly stop. I will use the terms *stopping* and *reporting* interchangeably, so that if an agent stops at some time t, I will also say that the agent reports his crime at t. For any stopping time τ, the calendar time at which the agent stops is $t_0 + \tau$. Upon stopping at calendar time t, the agent pays a terminal penalty $p_t \in [\underline{p}, \overline{p}]$ and his flow gains stop accruing. An agent is randomly detected by the regulator at time $t_0 + \tau_\rho$, where τ_ρ is an individual specific exponentially distributed stopping time with rate parameter ρ, independent of $(x_t)_{t \geq 0}$ and across agents. If the agent is detected, he pays the maximum penalty $\overline{p}$ and his flow gains stop accruing.

Payoffs. To compute an agent's value from a stopping time, first let $w(x, t_0, t)$ denote the value of an agent who arrives at time t_0 in state x and delays reporting for a deterministic length of time t, net of the reporting penalty paid after delay t. This can be written,

$$w(x, t_0, t) := \mathbb{E}\left[\overbrace{\int_0^{t \wedge \tau_\rho} e^{-rs} x_s ds}^{(a)} - e^{-rt \wedge \tau_\rho} \overbrace{\mathbf{1}_{\tau_\rho < t} \overline{p}}^{(b)} \,\middle|\, x_0 = x, t_0 = t \right]$$

where the expectation is taken with respect to the distribution of τ_ρ and x_t. Note that λ does not explicitly appear, but rather controls the evolution of x_t. The first term, (a), represents the accrued flow gain until the minimum of (i) the time the agent chooses to stop and (ii) the time the regulator detects the agent. The second term, (b), is the penalty the agent receives when he is exogenously detected, $\overline{p}$, before choosing to stop. Notice that,

[15] As in Garrett (2016), I take the intuitive approach to aggregate the uncertainty at the individual level. An earlier version of the paper modeled arrival as a stochastic counting process, and the same results can be obtained there.

[16] As I detail in Section 7, this assumption can be relaxed to allow for a time-independent arrival distribution across states without affecting the results.

conditional on x_0, the expression is independent of t_0, since x_s is identically distributed for all agents conditional on x_0. So I can drop the dependence on t_0, and simply refer to the expression as $w(x,t)$.

Given this definition, an agent's expected payoff from stopping time τ when arriving at time t_0 in state x_0 under penalty policy $\mathbf{p} = (p_t)_{t\geq 0}$ is

$$W(x,t_0,\tau,\mathbf{p}) := \mathbb{E}\left[w(x,\tau) - e^{-r(\tau_\rho \wedge \tau)} \overbrace{\mathbf{1}_{\tau_\rho \geq \tau} p_{\tau+t_0}}^{(c)} \middle| x_0 = x \right]$$

where the expectation is taken with respect to the distribution of τ_ρ, x_t and τ. The last term, (c), is the penalty the agent pays when stopping before he is detected by the regulator. The independence of τ_ρ from τ and x_t implies that

$$w(x,t) = \mathbb{E}\left[\int_0^t e^{-(r+\rho)t} x_t dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho + r} \overline{p} \middle| x_0 = x \right]$$

and

$$W(x,t_0,\tau,\mathbf{p}) = \mathbb{E}\left[w(x,\tau) - e^{-(\rho+r)\tau} p_{\tau+t_0} \middle| x_0 = x \right]$$

Note that the agent effectively discounts at rate $\rho + r$, which I call the *effective discount rate.* The agent solves the problem,

$$W^*(x,t_0,\mathbf{p}) := \sup_{\tau \geq 0} W(x,t_0,\tau,\mathbf{p}) \qquad (\mathcal{A})$$

If a policy τ achieves value $W^*(x,t_0,\mathbf{p})$ it is called an *optimal stopping time* for the agent who arrives at time t_0 in state x. When it is clear, I will suppress the dependence of $W(x,t_0,\tau,\mathbf{p})$ on $\mathbf{p}$, denoting it by $W(x,t_0,\tau)$.

2.2 The Regulator

The regulator commits at time 0 to a *penalty policy* and an *obedient recommendation policy* i.e. a pair $(\mathbf{p},\mathbf{a}) := ((p_t)_{t\geq 0}, (a_t)_{t\geq 0})$ indexed by calendar time. The first component, $(p_t)_{t\geq 0}$, is a measurable function from $\mathbb{R}_+$ to $[\underline{p}, \overline{p}]$ with $\overline{p} \geq 0$.[17] The second component, $(a_t)_{t\geq 0}$, is any function with $a_t \in \{0,1\}^{\{x^l,x^h\}}$ satisfying the following conditions,

(1) $a^x(t) := a_t(x)$ is measurable for each x and

(2) $\tau^{\mathbf{a}} := \inf\{t - t_0 | t \geq t_0 \text{ and } a_t(x_t) = 1\}$ is an optimal stopping time for an agent arriving

[17] Note that I've placed no restrictions on $\underline{p}$. $\underline{p}$ can, for instance, be negative and represent a reward, as in the case of gun buybacks. Nevertheless, because p_t here is pure money burning from the perspective of the regulator, the most natural cases involve $\underline{p} \geq 0$.

at time t_0.

The stopping time defined in (2) is called the *stopping time induced by a* for an agent arriving at t_0.[18] Let $\overline{\mathcal{M}}$ be the set of policies $(\mathbf{p}, \mathbf{a})$ that satisfy only (1), and $\mathcal{M}$ the set of policies that satisfy both (1) and (2).

Payoffs. To calculate the regulator's payoff, I must calculate the characteristics of the population of agents. In particular, I need to know (i) the mass of arrived agents who have not already chosen to report or been detected and (ii) the distribution across states x^h and x^l for such agents. Given a penalty policy $\mathbf{p}$, the stopping times induced by the obedient recommendation policy $\mathbf{a}$ induce a pair of paths, $(\mu_t^h)_{t\geq 0}$ and $(\mu_t^l)_{t\geq 0}$, describing the mass of agents in states x^h and x^l at each time t, respectively. The regulator discounts the future at the same rate as the agent, r, and her payoff from a policy $(\mathbf{p}, \mathbf{a})$ is:

$$V(\mathbf{p}, \mathbf{a}) := -\int_0^\infty e^{-rt}\left(\mu_t^h + \alpha_l \mu_t^l\right) dt$$

where $\alpha_l \geq 0$. The regulator solves the problem,

$$V^* := \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}} V(\mathbf{p}, \mathbf{a}) \tag{$\mathcal{P}$}$$

A policy that achieves V^* is called *optimal*.[19]

3. Model Analysis

In this section, I compare dynamic policies to static policies and characterize an optimal policy. This is broken up into five steps:

(1) Static Policies. I define static policies as policies in which the terms of self-reporting are constant over time, and demonstrate their basic properties.

(2) Low State Screening Lemma. I show how to transform a policy into one which always induces immediate reporting by agents in the low state, under some conditions. This simplifies the subsequent analysis.

[18] Observe that the regulator is restricted to deterministic penalty policies (which I discuss in Section 7) This implies that a_t does not affect the agents' values, but rather (i) serves as a useful accounting device and (ii) allows the regulator to break ties in her favor.

[19] An optimal policy in the class is not necessarily the optimal mechanism in a general mechanism design approach, in which the regulator elicits reports from agents about their arrival time and returns from crime, and tailors self-reporting policies to these reports. I focus on this restricted class of policies, time paths for self-reporting penalties that apply uniformly to all agents, to remain as close as possible to the types of policies implemented in practice.

(3) *The Value of Dynamic Policies.* I characterize the set of model parameters for which a static policy is sub-optimal.

(4) *Recursive Representation.* Using the low state screening lemma proved earlier, I will write down and analyze a recursive representation of the problem.

(5) *Optimal Policy.* Using the recursive representation, I provide the main result of the paper, a characterization of an optimal policy.

3.1 Static Policies

A static policy is one which is constant over time; it induces no added inter-temporal reporting considerations for the agent.

Definition 1. *A* static policy $(\boldsymbol{p}, \boldsymbol{a})$ *is such that* $p_t = v$ *for all* t *for some* $v \in [\underline{p}, \overline{p}]$. *A* dynamic policy *is any policy which is not a static policy. Denote by* $\boldsymbol{p}^v$ *the penalty policy in which* $p_t = v$ *for all* t.

An agent's decision problem under such a policy takes a particularly simple form. Since the policy does not exhibit any inter-temporal variation, the agent's problem in each state for x_t is exactly the same at any time. It follows that the only stopping polices relevant for computing the agent's value are (i) never stop in either state, (ii) stop immediately in both the high and low states, and (iii) wait until the low state and then stop immediately.

The simplicity of the agent's decision problem under a static policy allows for a straightforward proof of the following proposition, which states that the optimal static policy is one which offers the agent $\underline{p}$.

Proposition 1. *An optimal static policy is* $(\boldsymbol{p}^{\underline{p}}, \boldsymbol{a})$ *where* $a_t(x) = a_{t+s}(x)$ *for each* $x \in \{x^h, x^l\}$ *and* $t, s \in \mathbb{R}_+$.

The proof is given in Appendix B but I provide here the intuition. Suppose that an agent considers reporting at time t or $t+s$. Lowering v has two effects from the agent's perspective: self-reporting at t becomes more profitable and so does self-reporting at $t+s$. However, the value to the agent from self-reporting at t increases by more than his value to self-reporting at $t+s$, because the increase at $t+s$ is scaled down by the agent's effective discount rate. Decreasing v as much as possible, i.e. to $\underline{p}$, is therefore optimal. Note that it is critical for the result that v, the penalty for self-reporting, does not enter the regulator's objective function explicitly.

3.2 Low Type Screening

In this section, I provide a lemma that allows me to focus on a sub-class of policies in the search for an optimal policy. I show that the regulator can induce reporting by low types at all times at no cost to the incentives of high type agents. Let

$$\tau^h := \inf\{s - t | s \geq t \text{ and } a_s(x^h) = 1\},$$

which is a stopping time for an agent arriving at t that follows the recommendation for high type agents.

Definition 2. *Let $\mathcal{L}$ be the set of policies $(\boldsymbol{p}, \boldsymbol{a}) \in \mathcal{M}$ such that*

1. $a_t(x^l) = 1$ *for all* t *and*
2. $W(x^l, t, 0) = W(x^l, t, \tau^h)$.

The first requirement is that an agent in the low state, x^l, chooses to report immediately. The second requirement is that an agent in the low state is indifferent between immediately reporting and instead following the high type's strategy as prescribed by $\mathbf{a}$.[20]

Let τ^∞ be the policy that never reports and,

$$\Delta_l := (-\underline{p}) - W(x^l, 0, \tau^\infty, \mathbf{p}^{\underline{p}}).$$

The value Δ_l is the difference in payoffs for an agent in the low state between immediate reporting for penalty $\underline{p}$ and never reporting i.e. τ^∞. Notice that this value is measured at time 0, but the value is the same for any arrival time t_0.

The following lemma states that, as long as Δ_l is positive, any policy can be transformed without loss for the regulator's value into a policy in $\mathcal{L}$. The lemma requires that Δ_l is positive so that an agent in state x^l is willing to self-report for the most generous self-reporting penalty. As I show in Theorem 1, if this condition fails, a static policy is optimal. This immediately implies that either a static policy is optimal or the search for an optimal policy can be restricted without loss of value for the regulator to $\mathcal{L}$.

Lemma 1 (Low Type Screening Lemma)**.** *Suppose* $0 \leq \Delta_l$. *Then, for any policy* $(\boldsymbol{p}, \boldsymbol{a}) \in \mathcal{M}$, *there is another policy* $(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{a}}) \in \mathcal{L}$ *s.t.* $a_t(x^h) = \tilde{a}_t(x^h)$.

The replacement penalty policy, $\tilde{\boldsymbol{p}}$, has $\tilde{p}_t = p_t$ for any t such that the high type reports. Otherwise, the replacement penalty policy has $\tilde{p}_t = -W(x^l, t, \tau^h, \mathbf{p})$. The high type still

[20] I abuse notation slightly and denote by 0 a stopping time that places probability 1 on immediate stopping upon arrival.

finds it optimal to report when $a_t(x^h) = 1$, since the value to any stopping time that stops with positive probability at t such that $\tilde{a}_t(x^h) = 0$ was available to him under the original policy as well. To see an example, consider the policy in the left panel of Figure 1. In this

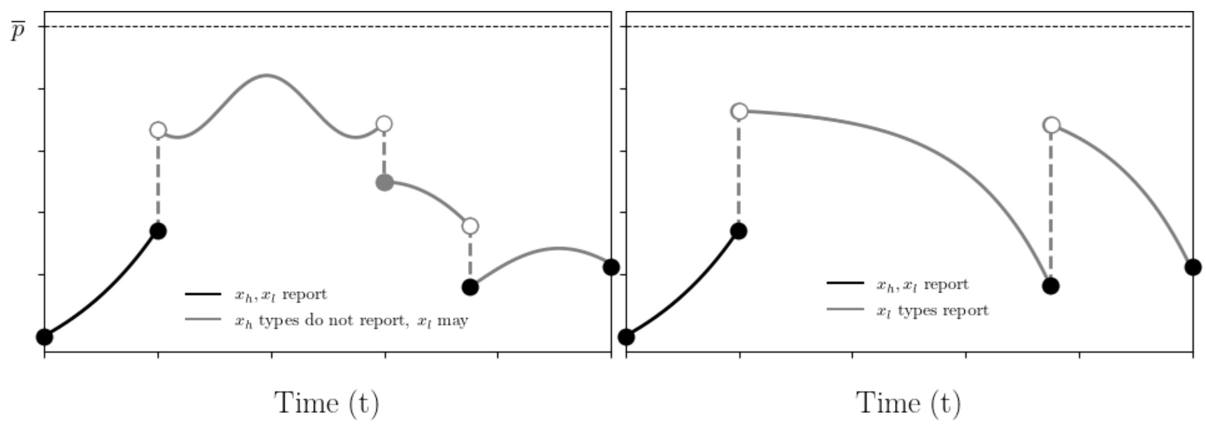

Figure 1: The left panel depicts the original policy. The right panel depicts the transformed policy, inducing the low type to report at all times without changing the reporting behavior of high types.

policy, the high and low types report on the black regions, while the low types may report elsewhere. There may also be times at which nobody reports.

To clarify the transformation that the lemma performs, I describe a two-step process (the first step of which is not pictured). Initialize $\tilde{\mathbf{a}} = \mathbf{a}$ and $\tilde{\mathbf{p}} = \mathbf{p}$. The first step of the transformation replaces $\tilde{p}_t$ on the gray region with $\overline{p}$ and replaces $\tilde{a}_t(x^l) = 0$ at such points. This only strengthens the incentives of the agents to report on the black regions, where $\tilde{a}_t(x^h) = a_t(x^h) = 1$, since this step (weakly) worsens any alternative to immediate reporting on the black region. The result of the second (and last) step of the transformation is depicted in the right panel of Figure 1, where I replace $\tilde{p}_t$ at any time at which $a_t(x^h) = 0$ with the negative of the low type's value from waiting until the next solid black point to report i.e. $-W(x^l, t, \tau^h, \mathbf{p})$. Replace $\tilde{a}_t(x^l) = 1$ at such points. Reporting at any such point delivers the agent a value which was available to him in the original policy and so the new recommendation, which sets $\tilde{a}_t(x^h) = 1$ exactly on the black region and $\tilde{a}_t(x^l) = 1$ everywhere, is an obedient recommendation policy. The penalty policy in the right panel of Figure 1, along with the corresponding recommendation policy, is then a member of $\mathcal{L}$, with high types incentivized to report on the black region, as in the original policy, and low types incentivized to report everywhere.

3.3 The Value of Dynamic Policies

With the previous lemma in place, I now answer the question: for what parameters can dynamic policies improve over static policies? The theorem below characterizes this set. Let $\theta := (\rho, r, \lambda, x^h, x^l, \overline{p}, \underline{p})$ denote an arbitrary parameterization. I will say that *dynamic policies strictly improve over static policies* under a particular parameterization if

$$V^* - \sup_{(\boldsymbol{p}^v, \boldsymbol{a}) \in \mathcal{M}} V(\boldsymbol{p}^v, \boldsymbol{a}) > 0.$$

Denote by Θ^* the set of parameters for which dynamic policies strictly improve over static policies.

Theorem 1. *The set of parameters θ for which dynamic policies strictly improve over static policies is non-empty and defined by the relation $(\rho + \lambda + r)\Delta_l \geq x^h - x^l > (\rho + r)\Delta_l$, i.e.*

$$\Theta^* = \left\{ \theta \middle| (\rho + r + \lambda)\Delta_l \geq x^h - x^l > (\rho + r)\Delta_l \right\}$$

Observe that when high types do not transition to the low state, i.e. $\lambda = 0$, the set is empty and dynamic policies do not improve over static policies.[21] The result is proved in Appendix D. I provide here the intuition in the simple case when the regulator puts a 0 weight on low type agents (i.e. $\alpha_l = 0$), the gains in the low state are 0 (i.e. $x^l = 0$) and the minimum penalty the regulator can offer is 0 (i.e. $\underline{p} = 0$).

Let $\tau^l := \inf\{t - t_0 | t \geq t_0 \text{ and } x_t = x^l\}$ which is the transition time from the high to the low state. Under the optimal static policy with $p_t = 0$ for all t and since $x^l = 0$, an optimal stopping time must stop weakly before τ^l. When p_t is constant, the only stopping times that the high type must consider to compute his value are τ^0, i.e. stop immediately, and τ^l, i.e. wait until entering the low state to stop.

Suppose first that $x^h < (\rho + r)\Delta_l$. Under the parametric assumptions, $(\rho + r)\Delta_l = \rho\overline{p}$ and $W(x^h, 0, \tau^l, \boldsymbol{p}^{\underline{p}}) = \frac{x^h - (\rho + r)\Delta_l}{\rho + r + \lambda}$. Then,

$$x^h < (\rho + r)\Delta_l \implies W(x^h, 0, \tau^l, \boldsymbol{p}^{\underline{p}}) < 0$$

which implies that the agent prefers to report immediately rather than delay reporting until τ^l. The static policy then induces reporting by all agents immediately upon arrival, which is the regulator's first best and cannot be improved.

Suppose instead that $x^h > (\rho + r + \lambda)\Delta_l$, which means that

$$\frac{x^h}{\rho + r + \lambda} > -W(x^l, 0, \tau^\infty, \mathbf{p}^{\underline{p}})$$

[21] This is not a consequence of the assumption that agents arrive in state x^h which, as I discuss in Section 7, can be generalized at no cost to the results and just minor cost to the notation.

The left-hand side is the total gain for an agent in the high state who chooses never to report, while the right-hand side is the total loss from detection for an agent who chooses never to report. The inequality implies that the total payoff from never reporting is strictly positive. Since in any penalty policy, $\mathbf{p}$, it must be that $p_t \geq \underline{p} = 0$, the inequality implies that no agent in the high state can ever be induced to report under any policy. Then, since $\mathbf{p}^{\underline{p}}$ ensures that low types report always, no other policy can increase the regulator's value and so dynamic policies are not necessary to achieve the regulator's optimal value.

Finally, suppose that $x^h \in \Big(\Delta_l(\rho+r), \Delta_l(\rho+r+\lambda)\Big]$. For any $T > 0$, define $\mathbf{p}^T := (p_t^T)_{t\geq 0}$ by

$$p_t^T := \begin{cases} 0 & \text{if } t = T \\ \overline{p} & \text{if } t \neq T \end{cases}$$

This is a *one-time* policy, that offers a single opportunity at time T to report for $p_T^T = \underline{p}$ and otherwise sets $p_t^T = \overline{p}$. Since $x^h \leq (\rho + r + \lambda)\Delta_l$, an agent places a negative value on the policy τ^∞, and so the agent would immediately report for penalty $p_t = 0$ if the only other available choice was τ^∞. By offering the one-time policy at T, the regulator effectively forces the agent to make this decision at time T; any policy other than τ^∞ is sub-optimal after T, since the agent gets no penalty reduction from self-reporting, so prefers to delay reporting as long as possible. As long as $T > 0$, the regulator is therefore able to induce reporting by a strictly positive mass of high types, while because $x^h > (\rho+r)\Delta_l$, no static policy induces reporting by high types. The low types only report at T, but this doesn't represent a loss for the regulator because $\alpha_l = 0$. So $\mathbf{p}^T$ strictly improves the regulator's value over any static policy.

Allowing for $x^l > 0$ and $\underline{p} \neq 0$ requires only slightly different algebra. Allowing for $\alpha_l > 0$ requires application of Lemma 1, but is otherwise the same.

3.4 Recursive Representation

A corollary of Theorem 1 is that if a dynamic policy is necessary to achieve the regulator's optimal value, that is $\theta \in \Theta^*$, then $0 < \Delta_l$. This is the condition required for the application of Lemma 1, and so when $\theta \in \Theta^*$, it is without loss for the regulator's optimal value to search for policies in which the low type agents report at all times. This means that the search for an optimal policy can be done in two steps:

(1) Solve the problem when $\alpha_l = 0$

(2) Apply Lemma 1 and the construction therein to the resulting policy to find an optimal

policy for $\alpha_l \geq 0$

The problem when $\alpha_l = 0$ admits a simplification because the regulator can constrain the search for an optimal policy to *pooling policies*, in which $a_t(x^h) = a_t(x^l)$. This is the outcome of two observations (i) the regulator can without loss of generality set the penalty to $\overline{p}$ whenever high types do not report, and (ii) low types find it optimal to report weakly earlier than high types. The regulator's problem is then reduced to choosing the times at which agents report, and the penalty at such times. Incentive constraints for the low type agent can be dropped since for any policy, the low type finds it optimal to stop weakly earlier than the high type. Incentives not to report when $p_t = \overline{p}$ are satisfied for all policies, and so these are dropped as well. These points are summarized in the following lemma,

Lemma 2. *When $\alpha_l = 0$ and $\theta \in \Theta^*$, then*

$$V^* = \begin{cases} \sup\limits_{(\boldsymbol{p},\boldsymbol{a})\in\overline{M}} \quad V(\boldsymbol{p},\boldsymbol{a}) \\ \text{subject to} \\ W(x^h, t, \tau^{\boldsymbol{a}}, \boldsymbol{p}) \geq W^*(x^h, t, \boldsymbol{p}) \text{ for each t} \\ p_t = \overline{p} \textit{ if } a_t(x^h) = 0 \\ a_t(x^h) = a_t(x^l) \text{ for each t} \end{cases}$$

Since $a_t(x^h) = a_t(x^l)$ for each t, I drop the dependence on x while I study the case $\alpha_l = 0$. To state the problem recursively, let

$$\underline{t}(t, \mathbf{a}) := \sup\{t - s | s \leq t \text{ and } a_s = 1\}$$

for any recommendation policy $(\mathbf{p}, \mathbf{a})$. Given t, this is the elapsed time since the last time agents reported. It is straightforward to show that the measure of high types at t, when agents choose the stopping time induced by $\mathbf{a}$, is $\mu_t^h = \frac{1-e^{-(\rho+\lambda)(t-\underline{t}(t,\mathbf{a}))}}{\rho+\lambda}$. Then, the regulator's value for a policy $(\mathbf{p}, \mathbf{a})$ is

$$V(\mathbf{p}, \mathbf{a}) = -\int_0^\infty e^{-rt} \frac{1 - e^{-(\rho+\lambda)(t-\underline{t}(t,\mathbf{a}))}}{\rho + \lambda} dt.$$

To describe the problem of an agent it is useful to define

$$\overline{t}(t, \mathbf{a}) := \inf\{s - t | s > t \text{ and } a_s = 1\}.$$

This is the delay until the *next* time, s, such that $a_s = 1$. Then, a *one-shot deviation* available to an agent is, rather than report at t with $a_t(x_t) = 1$, report instead at $t + \overline{t}(t, \mathbf{a})$. A consequence of obedience of $\mathbf{a}$ is that one-shot deviations are not profitable. That is, for any t such that $a_t(x_t) = 1$,

$$-p_t \geq w(x, \bar{t}(t, \mathbf{a})) - e^{-(\rho+r)\bar{t}(t,\mathbf{a})} p_{t+\bar{t}(t,\mathbf{a})}$$

where $-p_t$ is the penalty that the agent pays by self-reporting at time t.

The regulator's problem can be formulated recursively with, (i) *decision nodes* as the times at which agents report, (ii) the state as the penalty that must be offered at the decision node, p_t, (iii) choices as the *delay* until the next such time (the analogue of $\bar{t}(t, \mathbf{a})$) and the *penalty* at that time (the analogue of $p_{t+\bar{t}(t,\mathbf{a})}$) and (iv) the constraint as the one-shot incentive compatibility constraint defined above. The regulator's value, V^*, can then be computed by optimizing over the first calendar time at which agents report and the penalty at that time, followed by the solution to the recursive formulation of the regulators problem in which that penalty enters as the initial state.

The next lemma presents a recursive problem, and states that if a solution exists and is associated with a policy that ensures that the time between decision nodes is uniformly bounded below by some $\epsilon > 0$, then the solution to the recursive problem solves the regulator's problem.[22] Let $w_h(t) := w(x^h, t)$ and define

$$v(t) := \int_0^t e^{-rt} \frac{1 - e^{-(\rho+\lambda)t}}{\rho + \lambda}.$$

As I will show, $v(t)$ is the per-period loss of the regulator in the recursive representation of her problem. Let $P := [\underline{p}, -W(x_h, 0, \tau^\infty, \mathbf{p})]$, which is invariant to the policy $\mathbf{p}$ because the agent never self-reports under τ^∞. This set P is the domain on which the value function will be defined, and the upper bound is the maximum penalty that an agent would ever be willing to pay.

Lemma 3. *Suppose that $\alpha_l = 0$, $\theta \in \Theta^*$, and $\boldsymbol{V} : P \to \mathbb{R}$ satisfies*

$$\boldsymbol{V}(p) = \begin{cases} \sup\limits_{t \geq 0, p'} \quad -v(t) + e^{-rt}\, \boldsymbol{V}(p') \\ \text{subject to} \\ w_h(t) - e^{-(\rho+r)t} p' \leq -p \\ p' \in P \end{cases} \tag{R^*}$$

and there is some policy, $(t(p), p'(p))$ that achieves the value $\boldsymbol{V}(p)$, and has $\inf\limits_p t(p) > 0$.

[22] This ensures that the times at which agents are recommended to report, according to the solution to the recursive problem, contains no limit points.

Then,

$$V^* = \max_{t_0, p_0} \left\{ -v(t_0) + e^{-rt_0} \boldsymbol{V}(p_0) \right\}.$$

The approach follows Spear and Srivastava (1987), and the first constraint is referred to as the *promise-keeping constraint*.

To solve for a fixed point in (R^*), I conjecture an optimal policy and associated value function $\mathbf{V}(p)$, and then verify the conjecture. In particular, I conjecture that an optimal policy is $p'^{,*}(p) = \underline{p}$ and $t^*(p)$ is defined as the unique strictly positive solution to the equation

$$w_h(t^*(p)) - e^{-(\rho+r)t^*(p)} \underline{p} = -p \tag{I}$$

which is guaranteed to exist for each $p \in P$ whenever $\theta \in \Theta^*$. This $t^*(p)$ is a solution to the one-shot incentive compatibility constraint at equality: a high return agent is indifferent between immediately reporting for penalty p and reporting for penalty $\underline{p}$ after a deterministic amount of time $t^*(p)$.

I provide now a sketch of the proof of the verification of this policy as optimal. Recall that $\alpha_l = 0$. For the purposes of this sketch, I assume $x^l = \underline{p} = 0$ and I will describe the computations necessary for verification in the case in which the state is $p = \underline{p} = 0$.

Let $\mathbf{V}^*(p)$ be the value associated to the policy $(t^*(p), p'^{,*})$ and plug the definition of the conjectured optimal policy into (R^*) to get

$$0 = \begin{cases} \sup_{t,p'} \quad -v(t) + e^{-rt}\mathbf{V}^*(p') - \mathbf{V}^*(0) \\ \text{subject to} \\ w_h(t) - e^{-(\rho+r)t} w_h(t^*(p')) \leq 0 \\ p' \in P \end{cases} \tag{R_0^*}$$

It is straightforward to show that $t > t^*(0)$ cannot deliver an improvement, so I focus on choices $0 < t < t^*(0)$.[23] Integrate $v(t)$ to find that,

$$v(t) = \frac{1 - e^{-rt}}{r(\rho + \lambda)} - v^0(t)$$

where $v^0(t) := \frac{1 - e^{-(\rho+\lambda+r)t}}{(\rho+r+\lambda)(\rho+r)}$. After integrating $w_h(t)$ and plugging in the definition of $v^0(t)$,

$$w_h(t) = x^h(\rho + \lambda)v^0(t) - \frac{\rho\overline{p}}{\rho + r}(1 - e^{-(\rho+r)t})$$

Now observe that $w_h(t^*(0)) = 0$ implies that $\frac{\rho\overline{p}}{x^h(\rho+\lambda)(\rho+r)} = \frac{v^0(t^*(0))}{1-e^{-rt^*(0)}}$. Plug the equation for

[23] If $t = 0$ or $t = t^*(0)$, the objective is 0.

$w_h(t)$ along with this equality into the promise-keeping constraint and rearrange to see that the verification problem reduces to showing that,

$$0 = \begin{cases} \sup_{t,p'} & -v(t) - e^{-rt}v(t^*(p')) + (\frac{1-e^{-r(t+t^*(p'))}}{1-e^{-rt^*(0)}})v\left(t^*(0)\right) \\ \text{subject to} \\ v^0(t) + e^{-(\rho+r)t}v^0(t^*(p')) - (\frac{1-e^{-(\rho+r)(t+t^*(p'))}}{1-e^{-(\rho+r)t^*(0)}})v^0\left(t^*(0)\right) \leq 0 \\ p' \in P \end{cases} \tag{1}$$

Plug in the definition of $v(t)$, integrate and rearrange to find that this is equivalent to,

$$0 = \begin{cases} \sup_{t,p'} & v^0(t) + e^{-rt}v^0(t^*(p')) - (\frac{1-e^{-r(t+t^*(p'))}}{1-e^{-rt^*(0)}})v^0(t^*(0)) \\ \text{subject to} \\ v^0(t) + e^{-(\rho+r)t}v(t^*(p')) - (\frac{1-e^{-(\rho+r)(t+t^*(p'))}}{1-e^{-(\rho+r)t^*(0)}})v^0(t^*(0)) \leq 0 \\ p' \in P \end{cases} \tag{2}$$

The objective and constraint of this problem are closely related: the constraint is just a *distorted* version of the objective with an additional discount of ρ across reporting times, which accounts for the fact that an agent may be detected and punished at the maximum $\overline{p}$ before reaching the next reporting time.[24] Intuitively then, it appears natural that the conjecture makes use of the *delay* dimension rather than the *generosity* dimension to satisfy promise-keeping — rather than incentivizing reporting by making the next amnesty ungenerous, the next amnesty is kept as generous as people and is pushed as far into the future as necessary to satisfy promise-keeping. This is the *backloading* motive.

The verification problem can be rewritten

$$0 = \begin{cases} \sup_{t,p'} & h(r, t, t^*(p')) \\ \text{subject to} \\ h(\rho + r, t, t^*(p')) \leq 0 \\ p' \in P \end{cases} \tag{3}$$

where

$$h(a, t, t') = v^0(t) + e^{-at}v^0(t') - \left(\frac{1 - e^{-a(t+t')}}{1 - e^{-at^*(0)}}\right) v^0(t^*(0)).$$

To complete the verification, I show that $h(a, t, t',)$ satisfies a *single-crossing* property. In

[24]In Section 6, I show the left-hand side of the promise-keeping constraint is in fact the regulator's objective under the alternative assumption that agents arrive to the model at time-inhomogeneous rate $e^{-\rho t}$.

particular, I first define a set of conditions on t' that must hold for any $t \leq t^*(0)$, and show that whenever $0 < t < t^*$ and t' satisfies these conditions, then

$$h(a, t, t') \leq 0 \text{ for some } a \in [0, 1] \implies h(\tilde{a}, t, t') < 0 \text{ for any } \tilde{a} \in [0, a]$$

Plugging in $a = \rho + r$ and $\tilde{a} = r$ to this condition completes the verification. The following proposition summarizes the result.

Proposition 2. *Suppose that $\alpha_l = 0$ and $\theta \in \Theta^*$. Further suppose that $p'^{,*}(p) = \underline{p}$, $t^*(.)$ is the unique strictly positive solution to Equation (I), and $\boldsymbol{V}^*(p)$ is the associated value function. Then*

$$V^* = \max_{t_0, p_0} \left\{ -v(t_0) + e^{-rt_0} \boldsymbol{V}^*(p_0) \right\}.$$

The proof proceeds along the lines described above to show that Equation (R_0^*) is satisfied and then applies Lemma 3 after observing that $\inf_p t^*(p) \geq t^*(\underline{p}) > 0$ when $\theta \in \Theta^*$.

3.5 An Optimal Policy

In the previous section, I showed how to write the regulator's problem recursively when $\alpha_l = 0$. In the following result, I describe the optimal path to which this recursive problem leads, and combine the result with Lemma 1 to construct an optimal policy when $\alpha_l \geq 0$. Recall that $\bar{t}(t, \mathbf{a}) := \inf\{s - t | s > t \text{ and } a_s = 1\}$.

Theorem 2. *If $\theta \in \Theta^*$, an optimal policy, $(\boldsymbol{p}^*, \boldsymbol{a}^*)$, is:*

- $p^*_{nt^*(\underline{p})+t_0} = \underline{p}$ *for $n \in \mathbb{N}$ for some $t_0 \geq 0$*
- $a^*_t(x^h) = 1$ *if and only if $t \in \{t_0, nt^*(\underline{p})\}_{n \in \mathbb{N}}$*
- $p^*_t = e^{-(\rho+r)\bar{t}(t,\boldsymbol{a}^*)}\underline{p} + (1 - e^{-(\rho+r)\bar{t}(t,\boldsymbol{a}^*)})\frac{(\rho\overline{p} - x^l)}{\rho + r}$ *for all $t \notin \{t_0, nt^*\}_{n \in \mathbb{N}}$*
- $a^*_t(x^l) = 1$ *for all $t \geq 0$*

If $\theta \notin \Theta^*$, then an optimal penalty policy is $\boldsymbol{p}^* := (p^*_t)_{t \geq 0}$ with $p^*_t = \underline{p}$ for all t, and an optimal obedient recommendation is constant i.e. $a_t(x) = a_s(x)$ for each $t, s \geq 0$ and $x \in \{x^h, x^l\}$.

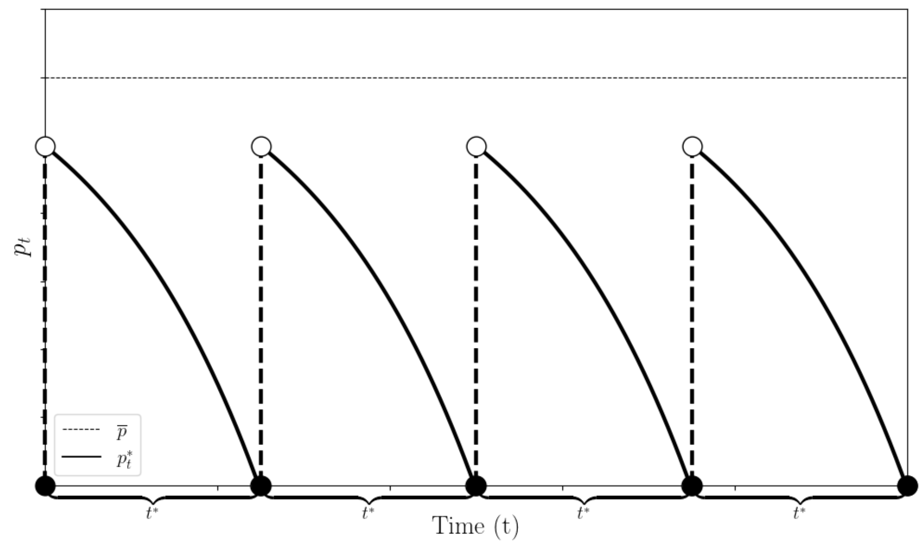

Figure 2: An optimal policy.

For $\theta \in \Theta^*$ and any $t \notin \{t_0, nt^*(\underline{p})\}_{n\in\mathbb{N}}$, the path of p_t guarantees that an agent in state x^l is indifferent between immediately reporting anywhere on this path and waiting until the next $t \in \{t_0, nt^*(\underline{p})\}_{n\in\mathbb{N}}$ to report. The t_0 in the theorem is an initial timing choice of the regulator, who is initially unburdened by incentives of prior agents. The optimal policy beyond t_0 is displayed in Figure 2.

In the recursive formulation of the problem in the previous section, I showed that the agent and regulator's value from a deviation from the proposed guess are closely connected. In particular, the agent's value is a distorted version of the regulator's value, applying an extra discount factor of ρ across the times at which high types report. In fact, it can be shown that the agent's value from a deviation is a linear function of the regulator's value from the deviation under the alternative assumption that agents arrive to the model at *time-inhomogeneous rate* $e^{-\rho t}$. Since incentive compatibility constraints would be unchanged under this alternative assumption, this suggests that if the regulator indeed faced a population arriving at time-inhomogeneous rate $e^{-\gamma t}$ for arbitrary $\gamma \geq 0$, a new result could be established: when $\gamma < \rho$, the optimal policy in Theorem 2 remains optimal, but if $\gamma > \rho$, a new optimal policy will prevail. This indeed turns out to be true: in Section 6, I show that when $\gamma < \rho$, the optimal policy remains optimal, and I characterize a new optimal policy when $\gamma > \rho$.

4. Comparative Statics

The frequency of cycles increases in the risk of detection (ρ), the maximum penalty ($\overline{p}$) and the rate of transition from high to low type crime (λ). So, when the regulator can invest to increase ρ or $\overline{p}$ ex-ante, the regulator can also increase the frequency with which high types self-report in the optimal policy. This comparative static highlights the complementarities between investment in features of the enforcement environment and the use of amnesty, in the presence of dynamic returns from crime.

While a more general analysis would allow the regulator to choose enforcement efforts dynamically and jointly with amnesty, the comparative static nevertheless provides useful insight. In some settings, the regulator may only be able to imperfectly affect the rate of detection; for instance, in the case of cartels or desertion during war, much of detection comes from third-parties reporting to the regulator. While the regulator can potentially affect the incentives of third-parties to report, this is a much less tightly controlled process than in the case of, for instance, environmental inspections where the regulator directly controls the rate of inspection.[25] Second, in some of the examples I discuss in the paper, policy-makers first implemented a blanket strengthening of enforcement and only then implemented an amnesty policy. One way to approximate this decision making process is by adding a pre-play stage to Section 2 in which the regulator chooses, at some cost, ρ and $\overline{p}$. The comparative statics tell us what would happen in such a model (up to the specification of a cost to raise ρ or $\overline{p}$).

5. Applications

I have shown that optimal policies exhibit cycles of amnesty. I first discuss an applications of the model to military desertion and detail a case of desertion amnesties during the Russian Civil War. Afterwards, I discuss the model's implications for tax amnesties and gun buybacks and amnesties.

5.1 Desertion

From June 1919-June 1920 alone, the Red Army's *Central Anti-Desertion Commission* recorded over 2.6 million deserters, nearly equal to the number of new recruits over the same period (Figes, 1990). During the Vietnam War, over 400,000 soldiers deserted. The war minister of Napoleonic Italy declared desertion as "the first and principal obstacle to the organization of the army of the Kingdom" (Grab, 1995).

[25] As in Wang et al. (2016).

Desertion amnesties are often offered during the course of war in an effort to entice return and have been applied extensively across history.[26] The Red Army created its anti-desertion commission in 1918 – it increased punishments, strengthened enforcement (for instance, dispatching armed groups to search for deserters) and implemented periodic amnesties to entice deserters back to their units (Wright, 2012). As noted in Figes (1990), "...the most successful means of combating desertion [in the Red Army] were the amnesty weeks." During and surrounding the Argentine War of Independence, the military engaged in "alternating carrot and stick", offering amnesties to deserters in December 1813, September 1815, and September 1821 (Slatta, 1980) . In Napoleonic Italy, "the government's repressive policy was mitigated by frequent amnesties designed to entice deserters and draft dodgers back to the army" (Grab, 1995). French Militaries in the 18th and 19th centuries offered periodic amnesties "interspersed with periods of severe repression, in an attempt to lure waverers back to their units" (Forrest et al., 1989).

It is not difficult to imagine how the difficulties that a deserter faces can change over time, and this provides the basis for the dynamic returns modeled in this paper. Forrest et al. (1989) provides, among other things, an account of desertion in France in the early 19th century. Deserters were often hungry unless they were lucky enough to receive help from local people. Snow could block passages through the mountains and the cold could be deadly. Deserters were "forced into the surrounding countryside [of their village], searching out caves and hiding-places that would offer protection until the forces of law and order had passed through." Under such circumstance, Forrest et al. (1989) remarks, "[I]t is hardly surprising that considerable numbers of deserters changed their minds."

5.1.1 The Red Army and the Anti-Desertion Commission

Wright (2012) offers an account of the anti-desertion effort in the Red Army during the years 1918-1920, along with a detailed case study of the anti-desertion experience in Karelia. As noted in the case study, historians have deemed material shortages – 'uniforms, linen, tea, tobacco, and soap' – a primary reason for mass desertion during the Russian Civil War. Other important factors were the intensity of fighting, proximity to the White Army, and seasonality.[27] In response to the mass desertion problem, the Red Army created the Central Anti-Desertion Commission in December 1918. In June 1919, after an organizational period, the military introduced the use of periodic amnesties. During the months June to October 1919, multiple time-limited amnesty periods were offered, alongside harsh repression. The amnesties allowed deserter to reenter the military with no repercussions.

[26] In contrast to desertion amnesties offered after a war, as a method of reconciliation and forgiveness.
[27] For instance, the harvest season led soldiers to return home to sow their fields.

The model provides one lens through which to view the use of repeated, time-limited amnesties — by repeatedly offering an amnesty for only a short window, the anti-desertion program balanced two issues: (i) that deserters would not report under a permanently offered amnesty unless their conditions became unbearable and (ii) that offering the program just once would ignore many deserters who would eventually be willing to re-enter the ranks. The application of amnesties, in this form, appeared to be a deliberate choice, rather than indecision – as noted in Rendle (2014), "One contemporary later argued that amnesties could have a significant impact as long as they were introduced when deserters were receptive, were not too frequent, and were applied alongside repression." Newspaper ads stressed the time-limited nature of the amnesty, in order to encourage deserters' return. The following is an extract of a newspaper publication described in Wright (2012): 'Deserters, townspeople! Today is the last day to appear before the commission [anti-desertion]. Hurry; present yourself today as tomorrow will be too late!'.

It is instructive to consider the amnesties in the broader context in which they were offered. Aside from the general idiosyncratic variation in a deserter's plight as described earlier, some of the most important time-varying factors were the advances of the White army and the harvest season. In particular, deserters from the Red Army often returned at the end of their harvests, which is responsible for the success of some amnesties (Wright, 2012). A theory based on a public end to the harvest season would be able to account for annual amnesties, but even at a relatively small regional level, amnesties were more frequent — Wright (2012) describes a number of amnesty weeks separated by periods with no amnesty during the June-October 1919 period in the Karelia region. When instead the harvest timing is idiosyncratic (based on crop, sub-region, etc...), then the harvest season is less predictable and the theory presented in this paper applies.

Undoubtedly, a complete analysis of the amnesty-granting decision requires understanding the relationship between society, the military, and its personnel. As discussed in Wright (2012), the Red Army's overall decision to apply amnesty can be seen in part as a way of striking a balance between repression and restraint, in a bid to win the support of the peasantry. Nevertheless, this paper develops a formal model with a force, echoed in qualitative evidence from the period, that drives towards the intermittent application of amnesties often observed as a response to the uncertainty and variation in the life of a deserter.

5.2 Gun Amnesties and Buybacks

A typical gun amnesty program commits to a 'no-questions' asked acceptance of illegally owned firearms, freeing participants from the risks of illegal gun ownership.[28] Buy-back programs go one step further, offering to pay for each firearm surrendered. During the Argentinian buyback of 2007-08, the government collected more than 100,000 weapons (Lenis et al., 2010). During the Brazilian buy-backs of 2003, 2009 and 2011, the government collected more than 1 million weapons. When operated on a small scale, the evidence, especially in the U.S., points to the lack of any effect of gun buybacks on gun violence (Plotkin, 1996). On a large scale, however, these programs can potentially be effective (Lenis et al. (2010), Macinko et al. (2007)), especially when coupled with changes to the enforcement environment.

The inter-temporal properties of these programs vary considerably. Brazil, for instance, has operated a temporary buyback program four times since 2003. Sweden has operated three temporary buyback programs since 1993. Mexico operates a permanent buyback program. Tasmania operates a permanent amnesty program and all of Australia will begin to do so in 2021. In many cases, short-term gun buybacks are operated when public support is strong (e.g. after a tragedy) or when private funding is available (Plotkin, 1996).

The model in this paper explores one reason why the intermittent nature of some programs can be an advantage and how one can improve the design of programs that are offered continuously, such as Mexico's gun buyback program. When the *option value* of participating in a gun-amnesty or buy-back is a first-order concern, the optimal amnesty program has both a permanent and temporary component – in a stylized setting, I have shown that an optimal policy induces self-reporting by agents with low returns from gun ownership at all times, but induces self-reporting by agents with high returns from gun ownership only intermittently. When instead this option value is not first-order, a static program is optimal. When illegal guns returned in amnesties/buy-backs have come into the owner's possession innocuously – for example, through inheritance – the option value of amnesty is irrelevant and disposing the gun as soon as possible is, to a first-order approximation, the owner's only objective. On the other hand, if the value of owning an illegal gun is derived from its operation by the owner, for safety, recreational or criminal reasons, then the owner has a more complicated objective: he would like to take advantage of an amnesty or buy-back when the weapon is no longer useful to him, but not before.[29] By offering only a limited-time

[28] The exact content of 'no-questions' asked varies from program to program.

[29] However, it must be noted that this change in value cannot come from a malfunction in the gun itself. As shown in Mullin (2001), such a change in value will lead gun owners to turn their gun in during a buy-back only to turn around and use the money to buy a new gun.

buy-back, the regulator can entice a gun owner to self-report faster than he would under a permanent buy-back program.

Whether the optimal policy takes a dynamic form and represents a significant improvement over the optimal static policy depends on parameters of the environment. While some parameters like the detection rate and penalties can be estimated from available data, the speed at which people transition from high to low value gun ownership cannot be. Many gun amnesty and buyback programs are accompanied with anonymous surveys of participants (McGuire et al., 2011). One way to estimate these parameters is to add two questions which are often left out of these surveys. The first is "how long have you owned your gun?" Answers to this question provide information on λ, the rate of transition from high to low value gun ownership. This can then inform the frequency and form of the optimal policy. The second is "if you owned it during the last buyback, why didn't you turn it in then?". Answers to this question can provide direct evidence on the motives of the participants. For instance, some may have not known about the program, or may have been otherwise misinformed about its conditions. Failing to account for such delay would overstate the value of a dynamic policy.

5.3 Voluntary Disclosure and Tax Amnesty Programs

Ptolemy V implemented the first recorded tax amnesty, circa 200 BC. Modern tax authorities have repeatedly implemented tax amnesties and voluntary disclosure programs including in the United States, Germany, Italy, India, the Phillipines, and Spain. Since 1980, more than 40 U.S. states have implemented a tax amnesty and 20 have implemented three or more. The use of tax amnesty is controversial, despite its prevalence (Le Borgne and Baer, 2008). Indeed, OECD (2015) states that tax amnesty programs, which it defines as programs that offer a reduction of the original tax amount, are "unlikely to deliver benefits that exceed their cost" (one of the reasons for which may be fairness considerations, which I discuss in Section 7). On the other hand, voluntary disclosure programs which offer reductions of penalties and interest and protection from prosecution can provide substantial benefits (OECD, 2015). In the model, such a constraint is best implemented by imposing $\underline{p} > 0$, representing the negative long-run effects on compliance and morale of programs which are too generous to evaders.

Le Borgne and Baer (2008) discusses the two main motivations for implementing a tax amnesty or voluntary disclosure program: "The two primary reasons for introducing tax amnesties are (i) to raise revenue in the short-term, and/or (ii) to increase compliance (e.g., by encouraging taxpayers to declare and pay previously undeclared tax, file tax returns, or register to pay taxes, so as to increase revenue and horizontal equity in the medium term)."

It it often argued, especially recently, that tax amnesties have been implemented with an eye to (i).[30] The model I present in this paper shows that despite this focus, the intermittent nature of these programs, as compared to a permanent policy, is also valuable from the perspective of (ii), increasing long-term compliance, when the value to tax evasion changes over time.

A basic question relevant for examining tax amnesties and voluntarily disclosure programs is, why do people apply? In the model presented, the change in profits from tax evasion leads evaders to self-report. Although direct evidence regarding motivation is not widely available, one source of evidence on this question comes from Ritsema et al. (2003), who implemented a survey of participants in the 2003 Arkansas tax amnesty program. The authors find that income, ease of evasion and inability to pay were three important determinants in the decision to evade taxes. These are factors which vary over time and therefore the model applies to tax settings in which these are first-order factors in deciding to apply for amnesty or to voluntarily disclose evasion.

As detailed in OECD (2015), there are many examples of both permanent and repeated, temporary, tax amnesties and disclosure programs. Within the literature on tax evasion, the use of repeated, temporary amnesties has been a subject of some theoretical investigation. Marchese and Cassone (2000) rationalizes repeated tax amnesties as the tax authority price discriminating between ex-ante honest heterogeneous taxpayers. Although the model in the present paper abstracts from important features of tax evasion and collection (importantly, that the regulator does not care about collected penalties), it offers a new take on the relative value of permanent versus repeated, temporary programs, focusing on how such programs reinstate those who have already decided to evade i.e. (ii) in the taxonomy of Le Borgne and Baer (2008). In this context, when the value to evasion is persistent and changes over time, it is sub-optimal to offer a static program and a cyclical program can provide stronger incentives for agents to self-report.

Andreoni (1991), in a one-shot setting, argues that a permanent partial voluntary disclosure program can be valuable as it provides agents with insurance for negative income shocks. One of the purposes of that paper is to rationalize some permanent programs observed in the United States and Canada. From a design perspective, the intuition presented in this paper suggests that such a program can be further improved when agents' values for tax evasion are imperfectly persistent, by offering only occasional opportunities for voluntary disclosure.

[30]See Luitel and Tosun (2014).

6. Generalizing the Arrival Distribution

In this section, I suppose that, rather than a time-homogeneous arrival of agents, the regulator faces a stream of agents arriving at time-inhomogeneous rate $e^{-\gamma t}$ for some $\gamma \in [0, \infty)$. The model studied in the Section 3 corresponds to $\gamma = 0$, while $\gamma > 0$ corresponds to a setting in which the distribution of agents is skewed towards time 0. I show that when $\gamma < \rho$, the main theorem of Section 3 still holds; an optimal policy consists of amnesty cycles that take the form described in Theorem 2. When instead $\gamma > \rho$, a new optimal policy can be described as follows: after an initialization period like that in Theorem 2, the regulator offers an interval with an *increasing* self-reporting penalty, and after this interval offers a fixed penalty forever.

I operate in this section under the assumption that $\underline{p} = x^l = 0$, but this is only for simplicity and all of the results generalize.

Assumption 1. $\underline{p} = x^l = 0$

Let $V_\gamma(\mathbf{p}, \mathbf{a})$ denote the regulator's value from a policy $(\mathbf{p}, \mathbf{a})$ when the arrival rate of agents is $e^{-\gamma t}$ for $\gamma \in [0, \infty)$. Then, as in Section 2, the regulator solves

$$V_\gamma^* = \sup_{(\mathbf{p}, \mathbf{a})} V_\gamma(\mathbf{p}, \mathbf{a}).$$

The steps for proving Theorem 2 apply with little adjustment to V_γ^*, as long as $\gamma < \rho$.

Proposition 3. *Suppose $\gamma < \rho$. Then, the optimal policy in Theorem 2 remains optimal.*

When $\gamma < \rho$, the arrival rate of agents is still relatively steady over time, and the fact that agents arrive more quickly near time 0 is not enough to overcome the backloading motive that leads to the cyclical policy. The proof is given in Appendix F.3.

This is no longer true when $\gamma > \rho$. In this case, the arrival of agents is front-loaded and the policy described in Theorem 2 does not deliver the regulator's optimal value. After the choice of the first reporting time, the optimal policy takes the following form:

(i) an interval with an *increasing* self-reporting penalty, on which all types report,

(ii) an upward jump at the end of this interval and afterwards

(iii) a constant self-reporting penalty, with only low types reporting.

The proposition below states the form of the optimal policy. When $\theta \notin \Theta^*$, a static policy is again optimal, so I restrict the proposition to the case $\theta \in \Theta^*$. Let

$$t^I := ln\left(\frac{x^h - \frac{gx^h}{f})}{x^h - \rho\overline{p}}\right)\frac{1}{\rho + r}$$

which will turn out to be the length of the interval in which self-reporting penalties are increasing.

Proposition 4. *Suppose $\gamma > \rho$ and $\theta \in \Theta^*$. Then, there exists t_0 such that an optimal policy, $(\boldsymbol{p}, \boldsymbol{a}) = ((p_t^*), (a_t^*))_{t \geq 0}$, is:*

- $p_t^* = e^{-(\rho+r)(t_0-t)}\underline{p} + (1 - e^{-(\rho+r)(t_0-t)})\frac{(\rho\overline{p})}{\rho+r}$ *for* $t < t_0$
- $p_t^* = (e^{(\rho+r)(t-t_0)} - 1)\frac{x^h - \rho\overline{p}}{\rho+r}$ if $t_0 \leq t \leq t_0 + t^I$ *and*
- $p_t^* = \frac{\rho\overline{p}}{\rho+r}$ *for* $t \geq t_0 + t^I$.
- $a_t^*(x^h) = 1$ *if and only if* $t_0 \leq t \leq t_0 + t^I$
- $a_t^*(x^l) = 1$ *for all* t

The result is proved in Appendix F.3. An example of the optimal policy in Proposition 4 beyond t_0 is depicted in Figure 3. As in Theorem 2, the existence of t_0 is a result of the fact that the regulator has no prior incentive constraints to satisfy until the initial amnesty offer.

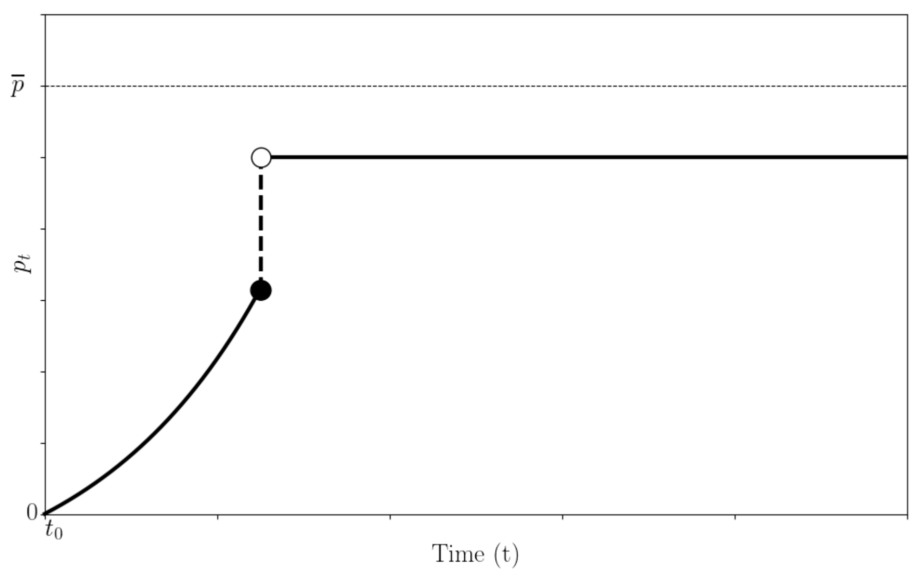

Figure 3: An Example of the Optimal Policy in Proposition 4

7. Interpretation and Extensions

In this section, I provide interpretations of the assumptions in the model. I also detail some extensions and alternative modeling choices, as well as the model's limitations.

Arrival Time. In practice, amnesty programs can be valuable when the enforcement environment is too weak to deter all crime. In light of this, arrival by the agent to the model can be interpreted in two ways.

1. First, the time the criminals begins committing the crime is perfectly observed (as in desertion during war), but the crime is initiated in a state in which the returns are too high to be deterred and arrival to the model is the time that returns reach a level at which the regulator can induce self-reporting. If returns are private, arrival time to the model is then naturally viewed as private information of the criminal.

   Consider, for instance, the case of military desertion. A military will know, within a few hours, that a soldier has deserted. Therefore, the military can condition on desertion time when designing the amnesty program. For instance, the military could say that anyone who returns within two days will not be labeled a deserter and will be punished only mildly. Such programs could entice deserters who quickly decided they made a mistake, but not those who made calculated decisions and chose to remain deserters for longer periods.[31] In the Russian Civil War, the Red Army's desertion amnesties did not restrict the application of amnesty by desertion date. In other desertion settings, there have been such restrictions, but they are rare and often extend years in the past (such as the Sri Lankan desertion amnesty of 2008, which extended to all who deserted after 2005).

   To formalize this interpretation of the model, (i) suppose the regulator observes arrival time and chooses a time path for amnesty with this arrival time as time 0 and (ii) include a third state, $x^{hh} > x^h$, in which the agent begins his crime. The state x^{hh} is set so high that the agent would never self-report, reflecting the weak enforcement environment. At some Poisson rate, γ, he transitions from x^{hh} to x^h and never returns to x^{hh}. While the regulator is allowed to condition her amnesty path on the time at which the agent begins crime, she does not observe the time of transition from x^{hh} to x^h. This version of the model is formally equivalent to the model described in Section 6, and so the results from that setting can be applied here as well. An alternative to this is appropriate when the regulator may be unsure of the precise arrival time of an agent, but has some information. In particular, one can suppose the regulator faces a single agent who arrives at some exponentially distributed time from 0, with parameter γ. This version of the model is technically equivalent to the version in Section 6.

2. Second, in other cases, it may be more natural to suppose that the regulator observes

[31]For those, enforcement is too weak and offering an immediate amnesty is like the military asking them not to desert, which it already does.

neither the private return nor the actual time at which the crime begins. This is the natural interpretation in, for instance, the gun amnesty and buyback case. In this case, one can introduce a state x^{hh} as above which transitions to x^h at some rate γ. In this case, however, the regulator does not observe the agents' arrival time to the model, and so the analysis is identical to that in Section 3.

Each of these modeling approaches accord with the view that enforcement is too weak to deter crime ex ante and so must focus on detection ex post through self-reporting. Nevertheless, deterrence is a first-order concern in most settings, and so I discuss it in greater depth below.

Deterrence. The model ignores the decision to become a criminal – criminals arrive criminals. A version of the model that introduces a third, very high state in which the agent arrives and chooses whether to become a criminal leaves the results unchanged. A richer model would have multiple states in which an agent can arrive to the model and decide whether to become a criminal; agents arriving in the highest states cannot be deterred from crime, while agents arriving in the intermediate and low states can be. As agents transition from the high states down into the intermediate and low states, they engage with self-reporting programs. In this model, the optimal policy will trade off *deterrence* with *ex-post detection* from self-reporting. When the main motivation is deterrence, self-reporting programs will be counter-productive. This can happen when the distribution of arriving values is concentrated below the level deterred by shutting down self-reporting programs. When the main motivation is ex-post detection, then self-reporting programs like the ones studied here will be useful. This can happen when the distribution of entering values is concentrated above the level deterred by shutting down self-reporting programs.

When the time at which the crime is initially committed can be observed and conditioned on (as in the discussion above on *Arrival Time*), then one simple way to deal with this is for the regulator to commit to a large t_0, i.e. a long initial time without any amnesty, after which she can implement the optimal amnesty policy already described. If instead the time at which the crime is initially committed is not used in the description of the optimal policy, either because it is unknown or is too difficult to implement, the deterrence issue is more complicated. I investigate this case further in Appendix A.2, and show that the insights developed here can be partially extended to a model that allows the regulator to express a deterrence motive.

Exogenous Detection Penalty. The penalty the agent receives when exogenously detected is $\overline{p}$. In this setting, this is without loss generality because the regulator does not place any value on collected penalties for their own sake.

Deterministic Policies. The regulator's choice is restricted to *deterministic policies*. In Appendix A.1, I provide an extension of the main result to a restricted class of Poisson random policies. Nevertheless, I do not rule the possibility that general random policies improve the regulator's value.

Initial Distribution of Values. Arriving agents are initialized in state x^h. Relaxing this to allow for a time-independent distribution of arriving values across the high and low states does not change the results. As I have shown, an optimal policy induces low value agents to report either always or never. Allowing for a time-independent distribution of arriving values then just scales the regulator's problem by a constant factor, leaving the optimal policy unchanged.

Uncontrollable Rate of Detection. In the model, the regulator does not control the rate of detection. In Wang et al. (2016), a self-reporting problem is studied in the context of environmental regulation in which the regulator controls the rate of detection. In that setting, in the absence of dynamic values, the paper shows that manipulating the risk of detection can amplify the value to dynamically adjusted self-reporting penalties. The model presented in this paper then fulfills a role complementary to Wang et al. (2016) by showing how self-reporting programs should behave in the absence of inspection control but in the presence of dynamic values to crime. It addresses the design of self-reporting programs in empirical settings in which the rate of inspection is indeed more difficult to control, unlike the Environmental Protection Agency (EPA) audit setting considered in Wang et al. (2016). For instance, in the case of price-fixing cartels, much of detection comes from buyer complaints which the anti-trust authorities do not directly control (Harrington, 2005). In the case of the anti-desertion campaigns in the Red Army, enforcement was locally delegated but deserters could be caught anywhere or discovered by people other than those tasked with explicit enforcement (Wright, 2012). In general, at least some detection typically comes from third-party reporting which the regulator cannot directly control.

Absorbing Low State. The model does not allow for the possibility that agents in the low state transition back to the high state. This assumption is made for tractability. Lemma 1 generalizes to this case and so a policy similar to the policy in Theorem 2 is approximately optimal when transitions back to the high state are infrequent, with the loss relative to optimality shrinking with the size of the transition.

Collected Penalties. The regulator's objective function does not include collected penalties, treating p_t and $\overline{p}$ as pure money burning.

In the case of tax amnesty, where one of the main motivations is short-term revenue, this issue is salient.[32] One way to incorporate revenue considerations is to generalize the regulator's objective function to be a weighted combination of the loss from tax evasion (x_t) and the profit from penalties. When the weight on collected penalties is equal to the weight on the loss from operation, the game between regulator and agents is zero-sum: in this case the regulator minimizes the agents' value, which is achieved by never granting a penalty reduction i.e. $p_t = \overline{p}$ for all t. When instead the regulator places a *lower* weight on the profit from penalties, there is scope for self-reporting to benefit the regulator. When the penalty represents prison time, a non-positive weight on penalties is apt. When penalties are financial, a lower weight represents the cost of collecting penalties and proving guilt, which is administratively expensive (Franzoni et al., 1996). Although the proof of Theorem 2 does not generalize to this setting, the main force at work remains in tact.

In the case of desertion, one interpretation of collected penalties is prison time. Militaries have found ways of preserving manpower while still imposing punishment, such as random punishment (Becker (1968), Chen (2017)), postponing prison sentences until after a war, relegating deserters to the worst duties, organizing penal battalions, and others. Nevertheless, a natural variation of the model would introduce a *loss* from collecting penalties, since by imprisoning a deserter, the military loses out on manpower. In this case, the incentive to offer self-reporting programs is even greater, since they give the regulator a way to avoid punishment and preserve manpower. The fundamental force of the paper is therefore strengthened in this context. The inclusion of this force complicates the analysis and how it affects the results is left as an open question. A more complex model may study how the government uses self-reporting incentives as a tool to speed up reporting and save on enforcement costs.

In other cases, it is more natural to ignore the profits from collected penalties. In cartels, for instance, the penalties may be viewed as pure transfers, while anti-competitive behavior represents a dead weight loss (Motta and Polo, 2003). Valuing penalties from illegal gun ownership may be accommodated in a way similar to tax evasion, with different weights on penalties and behavior. But, gun ownership generates externalities through misuse, theft and unregulated sale (Cook and Ludwig, 2006); these are not internalized by the gun owner in weak enforcement environments by definition. In this case, the weight the regulator places on collected penalties is small. Gun ownership penalties also involve prison time, which should not be treated positively in the regulator's objective function.

[32]See Le Borgne and Baer (2008) for a discussion of this issue.

Quitting. One of the real-life features motivating the model is that certain aspects of crime are irreversible, without regulatory approval, like desertion. Nevertheless, it is interesting to think about a case in which an agent is given an option to "quit" without self-reporting. This may be especially important in cases like gun amnesties and buybacks, where it seems especially easy to dispose of or hide an illegal gun when it is not in use and remove the evidence of its existence. When quitting is not free (because it is still risky to dispose of an illegal gun or hide it in a home where it may be mishandled), then the results in the paper continue to apply, except the regulator is limited in how high a penalty she can entice a criminal to accept. When quitting is free, there is no role for amnesty when $\underline{p} \geq 0$ because an agent always weakly prefers to quit rather than self-report. When instead $\underline{p} < 0$, i.e. a buyback is feasible, the basic forces remain.

Ethics. In some contexts with weak enforcement, the regulator may still enjoy a high rate of compliance. One reason posited for this in the case of tax compliance is ethics. Alm and Torgler (2011) argue that it is puzzling that so few citizens cheat on their taxes, given the relatively small risk of detection and penalties upon detection. The paper argues that ethics partially explains this high rate of compliance. Citizens comply because they perceive the tax system as fair, and the government as an institution that upholds this fairness. Generous tax amnesties may erode this sense of fairness and lead citizens who feel they have been unfairly treated (because they paid their taxes on time) to cheat on taxes in the future. In this way, a tax amnesty could create lasting damage to compliance.

Similar ethical forces could arise in other contexts as well. The use of amnesty may be particularly damaging in situations when enforcement is weak but compliance is high. Amnesties as described in this paper may then best be used in situations where fairness considerations are less relevant, or when a regulator can determine a minimum penalty level for amnesty that will not erode compliance through this channel.

Uniqueness. The policy described in Theorem 2 (and Proposition 3) when $\theta \in \Theta^*$ and $\alpha_l > 0$ is the unique optimal policy among all policies $(\mathbf{p}, \mathbf{a})$ in which $\{t|a_t(x^h) = 1\}$ has at least one isolated point.

The policy described in Proposition 4 when $\theta \in \Theta$ and $\alpha_l > 0$ is the unique optimal policy among all policies $(\mathbf{p}, \mathbf{a})$ in which any limit point of $\{t|a_t(x^h) = 1\}$ is a member of an interval.

These imply that the policy in Proposition 4 represents a strict improvement over the policy in Proposition 3 when $\gamma > \rho$, but the proof technique does not rule out the possibility that the policy in Proposition 3 and the policy in Proposition 4 deliver the same value when $\gamma < \rho$. Nevertheless, numerical simulation suggests that, when $\gamma < \rho$, the policy of Proposition 4 is always strictly worse than that of Proposition 3.

8. Conclusion

In this paper, I studied the problem of a regulator who designs amnesty programs to induce self-reporting of crime. I show (Theorem 1) that when the returns from crime can change over time ($\lambda > 0$), the generosity of an optimal amnesty program may change over time as well. In such cases, Theorem 2 establishes that the optimal policy is cyclical and describes its form. Except for an initial period, the minimum possible penalty ($\underline{p}$) for reporting is offered at evenly spaced points in time. In between such times, a decreasing schedule of penalties is offered. Agents with a high return from crime report only at the end of each cycle while those with a low return from crime report at all times during the cycle. A *backloading* motive on the part of the regulator drives the optimality of this form of amnesty. Agents discount more than the regulator across times at which high type agents are recommended to report, in particular they add an extra ρ to the regulator's discount factor across such time. The regulator therefore finds it optimal to incentivize reporting by high type agents at any time by offering the next amnesty to be the minimum self-reporting penalty, $\underline{p}$, as far into the future as is necessary to satisfy incentive constraints.

These results are generalized to a model that allows the arrival rate of agents to be time-inhomogeneous. Agents are assumed to arrive to the model at rate $e^{-\gamma t}$. When $\gamma < \rho$, so that arrival is sufficiently *slow*, then the main result described above continues to hold. When instead $\gamma > \rho$, a new optimal policy is characterized, which *front-loads* reporting by high return agents, which is a reflection of the front-loaded arrival of agents.

There are a number of avenues for future work. First, it would be useful to study a version of the problem in which the regulator can control, at some cost, the rate of detection. Second, new insights might result from incorporating political economy constraints into the model. In particular, the regulator may not be able to fully commit to a policy because she is occasionally replaced by a new regulator, wiping away previous commitments, as in a case of tax amnesty where the government is replaced every few years. Third, it would be interesting to further examine the deterrence margin beyond what has been discussed in Section 7. For instance, when the regulator cannot condition her policy on the time at which crime begins, randomization can be useful — by randomizing the timing of amnesty, agents cannot take advantage by initiating crime at times close to attractive amnesties. In that case, how should the regulator randomize amnesty?

References

Abreu, D. and H. Matsushima (1992): "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica: Journal of the Econometric Society*, 993–1008.

Alm, J. and B. Torgler (2011): "Do ethics matter? Tax compliance and morality," *Journal of Business Ethics*, 101, 635–651.

Andreoni, J. (1991): "The Desirability of a Permanent Tax Amnesty," *Journal of Public Economics*, 45, 143–159.

Araman, V. F. and B. Fayad (2020): "Intertemporal price discrimination with time-varying valuations," *Operations Research*.

Basu, K., K. Basu, and T. Cordella (2016): "Asymmetric punishment as an instrument of corruption control," *Journal of public economic theory*, 18, 831–856.

Battaglini, M. (2005): "Long-term contracting with Markovian consumers," *American Economic Review*, 95, 637–658.

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

Boise, C. M. (2006): "Breaking Open Offshore Piggybanks: Deferral and the Utility of Amnesty," *Geo. Mason L. Rev.*, 14, 667.

Chen, D. L. (2017): "The deterrent effect of the death penalty? Evidence from British commutations during World War I," *Evidence from British Commutations During World War I (Feb 26, 2017)*.

Conlisk, J., E. Gerstner, and J. Sobel (1984): "Cyclic Pricing by a Durable Goods Monopolist," *The Quarterly Journal of Economics*, 99, 489–505.

Connor, J. M. (2007): *Global price fixing*, Springer.

Cook, P. J. and J. Ludwig (2006): "The Social Costs of Gun Ownership," *Journal of Public Economics*, 90, 379–391.

Deb, R. (2014): "Intertemporal Price Discrimination with Stochastic Values," *Unpublished manuscript, University of Toronto*.

Dreyfus, P., L. E. Guedes, B. Lessing, A. R. Bandeira, M. de Sousa Nascimento, and P. S. Rivero (2008): "Small Arms in Rio de Janeiro," *The Guns, the Buypack, and the Victims*.

Figes, O. (1990): "The Red Army and Mass Mobilization during the Russian Civil War 1918-1920," *Past & Present*, 168–211.

Forrest, A. I., A. J. F. Forrest, et al. (1989): *Conscripts and Deserters: The Army and French Society During the Revolution and Empire*, Oxford University Press on Demand.

Franzoni, L. et al. (1996): "Punishment and Grace: on the Economics of Permanent Amnesties," Tech. rep.

Garrett, D. F. (2016): "Intertemporal Price Discrimination: Dynamic Arrivals and Changing Values," *American Economic Review*, 106, 3275–99.

Gärtner, D. L. (2014): "Corporate Leniency in a Dynamic World: The Preemptive Push of an Uncertain Future," *Available at SSRN 2340973*.

Grab, A. (1995): "State power, brigandage and rural resistance in Napoleonic Italy," *European History Quarterly*, 25, 39–70.

Harrington, J. E. (2005): "Detecting cartels," Tech. rep., working paper.

——— (2008): "Optimal corporate leniency programs," *The Journal of Industrial Economics*, 56, 215–246.

——— (2013): "Corporate Leniency Programs when Firms have Private Information: The Push of Prosecution and the Pull of Pre-emption," *The Journal of Industrial Economics*, 61, 1–27.

Kaplow, L. and S. Shavell (1994): "Optimal Law Enforcement with Self-Reporting of Behavior," *Journal of Political Economy*, 102, 583–606.

Krasikov, I., R. Lamba, and T. Mettral (2019): "Implications of unequal discounting in dynamic contracting," *Available at SSRN 3142932*.

Landeo, C. M. and K. E. Spier (2020): "Optimal Law Enforcement with Ordered Leniency," *The Journal of Law and Economics*, 63, 71–111.

Le Borgne, E. and K. Baer (2008): *Tax Amnesties: Theory, Trends, and Some Alternatives*, International Monetary Fund.

Leigh, A. and C. Neill (2010): “Do Gun Buybacks Save Lives? Evidence from Panel Data,” *American Law and Economics Review*, 12, 509–557.

Lenis, D., L. Ronconi, and E. Schargrodsky (2010): “The Effect of the Argentine Gun Buy-back Program on Crime and Violence,” in *Universidad Torcuato Di Tella. Argentina Retrieved from http://socrates. berkeley. edu/˜ raphael/IGERT/Workshop/PEVAF_Se ptember_27_2010. pdf*.

Luitel, H. S. and R. S. Sobel (2007): “The Revenue Impact of Repeated Tax Amnesties,” *Public budgeting & finance*, 27, 19–38.

Luitel, H. S. and M. S. Tosun (2014): “A reexamination of state fiscal health and amnesty enactment,” *International Tax and Public Finance*, 21, 874–893.

Macinko, J. et al. (2007): “Reducing Firearm Injury: Lessons from Brazil.” *LDI issue Brief*, 12, 1–4.

Malik, A. S. (1993): “Self-reporting and the Design of Policies for Regulating Stochastic Pollution,” *Journal of Environmental Economics and Management*, 24, 241–257.

Marchese, C. and A. Cassone (2000): “Tax Amnesty as Price-Discriminating Behavior by a Monopolistic Government,” *European Journal of Law and Economics*, 9, 21–32.

Marvão, C. and G. Spagnolo (2018): “Cartels and Leniency: Taking Stock of What We Learnt,” in *Handbook of Game Theory and Industrial Organization, Volume II*, Edward Elgar Publishing.

Maskin, E. (1999): “Nash Equilibrium and Welfare Optimality,” *The Review of Economic Studies*, 66, 23–38.

McGuire, M., M. Manno, A. Rook, L. Maranda, E. Renaud, A. DeRoss, and M. Hirsh (2011): “Goods for guns—the use of a gun buyback as an injury prevention/-community education tool,” *Journal of Trauma and Acute Care Surgery*, 71, S537–S540.

Motchenkova, E. (2004): “Effects of Leniency Programs on Cartel Stability,” *CENTER discussion paper*.

Motta, M. and M. Polo (2003): “Leniency programs and cartel prosecution,” *International journal of industrial organization*, 21, 347–379.

Mullin, W. P. (2001): “Will Gun Buyback Programs Increase the Quantity of Guns?” *International Review of Law and Economics*, 21, 87–102.

OECD (2015): *"Update on Voluntary Disclosure Programmes: A pathway to tax compliance"*.

Plotkin, M. R. (1996): in *Under Fire: Gun Buy-Backs, Exchanges and Amnesty Programs*, Police Executive Research Forum Washington, DC.

Rendle, M. (2014): "Mercy Amid Terror? The Role of Amnesties During Russia's Civil War," *Slavonic & East European Review*, 92, 449–478.

Ritsema, C. M., D. W. Thomas, and G. D. Ferrier (2003): "Economic and Behavioral Determinants of Tax Compliance: Evidence from the 1997 Arkansas Tax Penalty Amnesty Program," in *IRS research Conference*, Citeseer.

Shiryaev, A. N. (2007): *Optimal Stopping Rules*, vol. 8, Springer Science & Business Media.

Spagnolo, G. (2000): "Optimal leniency programs," *FEEM Working Paper*.

Spear, S. E. and S. Srivastava (1987): "On Repeated Moral Hazard with Discounting," *The Review of Economic Studies*, 54, 599–617.

Touzi, N. and N. Vieille (2002): "Continuous-time Dynkin games with mixed strategies," *SIAM Journal on Control and Optimization*, 41, 1073–1088.

Wang, S., P. Sun, and F. de Véricourt (2016): "Inducing Environmental Disclosures: A Dynamic Mechanism Design Approach," *Operations Research*, 64, 371–389.

Willmer, M. (1971): "On the Estimation of the Number of Persons Entering and Leaving Crime Annually," *Journal of Research in Crime and Delinquency*, 8, 175–181.

Wosepkat, J. L. (1971): "Repatriation and the Chieu Hoi Amnesty Approach in Vietnam: Consequences and Prospects," *The International Lawyer*, 637–651.

Wright, A. S. (2012): "Stemming the Flow: The Red Army Anti-Desertion Campaign in Soviet Karelia (1919)," *Revolutionary Russia*, 25, 141–162.

Zhou, J. and D. L. Gärtner (2012): "Delays in Leniency Application: Is There Really a Race to the Enforcer's Door?" .

Appendix

A. Extensions

In this section I discuss a number of extensions to the baseline model.

A.1 Poisson Randomization

Until now, I have restricted myself to the search for optimal deterministic policies. Although I don't characterize the optimal policy for general random policies, I expand the model to allow for a limited class of random policies and show that the deterministic optimal policy remains optimal. First, extend V linearly to random policies.

I prove the result for the case of $\alpha_l = x_l = \underline{p} = 0$, but it extends readily to the general case. Consider then the following class of policies:

Definition 3. *A randomized policy $(\boldsymbol{p}, \boldsymbol{a})$ is called a γ-Poisson policy if there exists t_0 and a sequence of random variables $(t_i)_{i\geq 0}$ s.t.*

- *$t_{i+1} - t_i$ independent of t_i and exponentially distributed with rate parameter γ*
- *$p_{t_i} = 0$ for $i \in \mathbb{N}$ and $p_t = \overline{p}$ otherwise*
- *$a_t(x^h) = 1$ if and only if $t \in \{t_0, t_i\}_{i\in\mathbb{N}}$*

The set of Poisson policies is denoted Γ.

These policies feature inter-arrival times of complete amnesty that are exponentially distributed with mean $\frac{1}{\gamma}$.[33] I restrict to the setting in which $x^l = \alpha_1 = 0$ and argue that the policy in Theorem 2 remains optimal when allowing the regulator to choose from Γ. Let $\mathcal{M}^\Gamma := \mathcal{M} \cup \Gamma$.

$$V^{\Gamma,*} := \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^\Gamma} V(\mathbf{p}, \mathbf{a}) \qquad (\mathcal{P}^\Gamma)$$

i.e. the expanded regulator's problem allowing for policies in Γ (with some abuse of notation since $\mathbf{p}$ and $\mathbf{a}$ are now random variables).

Theorem A.1. *Suppose $\underline{p} = x^l = \alpha_1 = 0$. Then*

$$V^* = V^\Gamma > V(\boldsymbol{p}, \boldsymbol{a})$$

[33]Notice that the choice of t_0 is as in problem ($\mathcal{P}$) to ensure that this initial choice doesn't drive any differences between the two problems.

for any $(\boldsymbol{p}, \boldsymbol{a}) \in \Gamma$.

The proof is given in Online Appendix F.2. Random mechanisms in Γ which are incentive compatible for the agent require putting substantial probability on long periods without amnesty, relative to the deterministic mechanism with inter-arrival times t^*. For the same reason as in Section 3, this sampling is more costly for the regulator than the agent.

A.2 Deterrence

In this section, I allows agents to decide whether to begin committing crime, maintaining the assumption that the regulator cannot observe the time at which crime is committed.

Model. To formally allow for a decision to enter crime, I study exactly the model of Section 2 but introduce a third state for values, x^{hh}, which is higher than x^h. This state x^{hh} is so high that agents can be neither induced to self-report nor deterred from entering using any policy $(p_t, a_t)_{t \geq 0}$. Upon arrival to the model, agents make a once-and-for-all decision whether to begin committing crime. It is also important now to allow for arrival in both states x^h and x^{hh} (as in the baseline model of Section 2, allowing for arrival in state x^l does not change results). To this end, let μ^0 be a time independent arrival distribution across states $\{x^{hh}, x^h\}$.[34]

For simplicity, I study a case in which $x^l = 0$ and $\underline{p} = 0$.

Assumption 2. $\underline{p} = x^l = 0$.

Agents can transition only from x^{hh} to x^h or from x^h to x^l. As before, transitions from x^h to x^l occur at rate λ and transitions from x^{hh} to x^h occur at rate λ_{hh}. I assume the following on the new features of the model,

Assumption 3. $x^{hh} > \frac{(\rho+r+\lambda_{hh})}{\rho+r}\rho\overline{p}$,

Assumption 4. $x^h \in (\rho\overline{p}, (\rho + r + \lambda)\frac{\rho}{\rho+r}\overline{p})$

The assumption $x^{hh} > \frac{(\rho+r+\lambda_{hh})}{\rho+r}\rho\overline{p}$, guarantees that agents arriving in state x^{hh} always enter. If $x^h \notin (\rho\overline{p}, (\rho + r + \lambda)\frac{\rho}{\rho+r}\overline{p})$, then setting $p_t = 0$ is optimal for the same reasons as in Theorem 2.

[34] Note that if μ^0 put probability 1 on x^h, or equivalently $x^{hh} = x^h$, then the optimal policy would be static. If $p_t = \overline{p}$, then giving agents the option to not begin committing crime in the first place is like offering a once-and-for-all amnesty. If this does not deter x^h types from entering, then no self-reporting policy can induce self-reporting by x^h types after they have entered. So, either the static policy $p_t = \overline{p}$ for all t deters both x^h and x^l types from entering, or it does not and then the static policy $p_t = \underline{p}$ for all t is optimal — this policy at least induces immediate reporting by low types.

Analysis. It is immediate to see that if μ^0 puts probability 1 on x^{hh}, the analysis of the model proceeds unchanged by replacing arrival with arrival *and* transition to state x^h in the formulation of Section 2.

A more difficult but realistic setting allows μ^0 to put positive probability on both x^{hh} and x^h. In designing her policy, the regulator must take account of a natural trade-off between (i) using self-reporting to entice agents who entered in state x^{hh} but have transitioned to state x^h and (ii) shutting down self-reporting to ensure that state x^h agents do not enter. As before, it is possible to induce reporting by x^l agents everywhere, so they do not pose any new difficulty in this environment. While I do not solve for the general optimal policy, I argue that a version of the cyclical policy in Theorem 2 strictly improves the regulator's value over static policies.

The assumptions made on x^{hh} and λ_{hh} guarantee that:

- $p_t = \overline{p}$ for all t deters agents in state x^h from entering but not agents in state x^{hh} (first and second assumptions)
- agents in state x^h can be induced to self-report, but not with a static policy (second assumption)

As long as both $\mu^0(x^{hh}) > 0$ and $\mu^0(x^h) > 0$, a positive mass of x^{hh} agents enter and eventually transition to state x^h. In this case, the trade-off between deterrence and ex post detection comes to the fore.

Consider now a static level, v, such that if $p_t = v$ for all t, agents who arrive in state x^h choose not to enter. Since $x^h \in (\rho\overline{p}, (\rho + r + \lambda)\frac{\rho\overline{p}}{\rho+r})$, there exists such a $v \in [\underline{p}, \overline{p}]$. Let $\underline{v}$ be the smallest such static v. Then, as long as $p_t \geq \underline{v}$, agents in state x^h will never choose to enter.

Under the requirement $p_t \geq v$, the regulator's problem reduces to exactly the problem studied in Section 2 with $\underline{p} = \underline{v}$ and arrival replaced by arrival in state x^{hh} *and* transition from x^{hh} to x_h. Given our assumptions, the parameterization θ is an element of Θ^*. Then applying Theorem 2 leads to an optimal regulatory policy that takes the cyclical form described in Theorem 2. Figure 4 depicts such a policy.

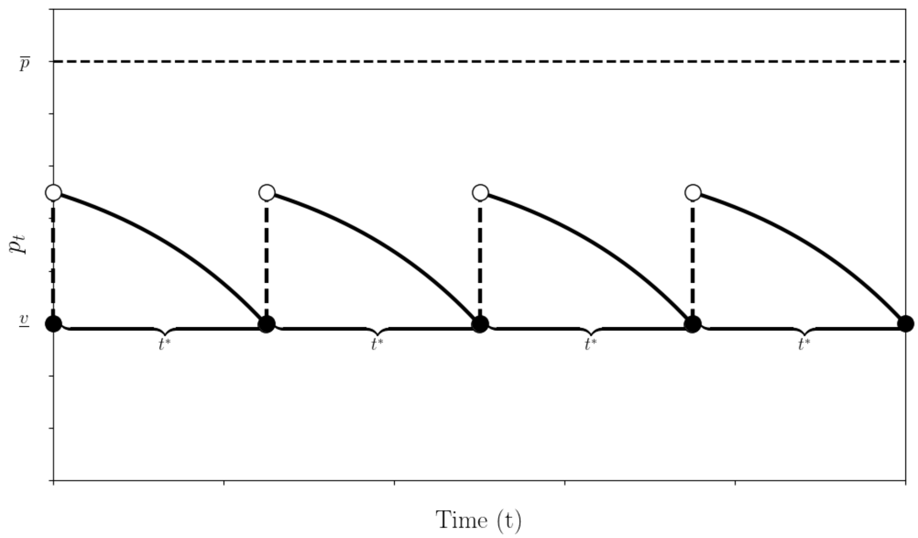

Figure 4: A optimal policy when $p_t \geq \underline{v}$.

Randomization. While this policy does improve the regulator's value relative to static policies, it is not in general optimal. At an intuitive level, randomization appears critical to achieve the optimal value because by randomizing, the regulator relaxes her deterrence constraint. For instance, in the cyclical policy, agents arriving in state x^h at the very beginning of a cycle have a much lower value to entering than do agents x^h arriving in the middle or end of a cycle. Randomizing allows the regulator to spread this deterrence more uniformly over the cycle.

It is straightforward to find numerical examples that can achieve the same level of deterrence as in the cyclical policy above — agents in states x^h and x^l never enter — but improve the regulator's value. For instance, consider the Poisson random policies studied in Appendix A.1. Any such policy that satisfies the agent's incentive compatibility condition *also* deters agents in state x^h from entering, because the arrival of an amnesty is i.i.d. over time. The description of an optimal policy in this setting with fully general random policies is left open.

B. Proof of Proposition 1

Proof of Proposition 1: Because $\mathbf{p}^v$ is constant and hence continuous, Theorem 3 in Shiryaev (2007) can be applied to show that there exists some $D \subset \{x^h, x^l\}$ such that[35]

[35] Application of the theorem in Shiryaev (2007) requires a re-casting of the stopping problem presented here. In particular, the state space must be expanded to account for the accumulating value. This formulation is omitted.

1. $\tau_v^* := \inf\{t \geq t_0 | x_t \in D\}$,
2. If τ is any other optimal stopping time for the agent, then $\mathbb{P}(\tau_v^* \leq \tau) = 1$.

It is therefore without loss of generality for the regulator to restrict to recommendation policies $\mathbf{a}$ such that $a_t(x) = a_s(x)$ for all $t, s \geq 0$ and $x \in \{x^h, x^l\}$, since these induce all stopping times of the form τ_v^*.

To prove the lemma, it is thus sufficient to argue that $\tau_{\underline{p}}^* \leq \tau_v^*$ for all $v \geq \underline{p}$. Given the characterization of τ_v^* described above, the only possibilities are

1. $\tau_v^* = \tau^{*,0} := 0$
2. $\tau_v^* = \tau^{*,\infty} := \infty$
3. $\tau_v^* = \tau^{*,l} := \inf\{t | x_t = x^l\}$

The agent's value for $\tau^{*,\infty}$ is independent of v. The agent's values for $\tau^{*,0}$ and $\tau^{*,l}$ are

$$\mathbb{E}\left[W(x, t, \tau^{*,0}; \mathbf{p}^v)\right] = -v$$
$$\mathbb{E}\left[W(x, t, \tau^{*,l}; \mathbf{p}^v)\right] = \mathbf{1}_{x=x^h}\left(\frac{x^h}{\rho + r + \lambda} + \frac{x^l - \rho\overline{p}}{\rho + r} - \frac{v\lambda}{\rho + r + \lambda}\right) + \mathbf{1}_{x=x^l}(-v)$$

To conclude that $\tau_{\underline{p}}^* \leq \tau_v^*$, observe that decreasing v increases $\mathbb{E}\left[W(x, t, \tau^{*,0}; \mathbf{p}^v)\right]$ by weakly more than $\mathbb{E}\left[W(x, t, \tau^{*,l}; \mathbf{p}^v)\right]$. Similarly, decreasing v weakly increases $\mathbb{E}\left[W(x, t, \tau^{*,l}; \mathbf{p}^v)\right]$ but has no effect on $\mathbb{E}\left[W(x, t, \tau^{*,\infty}; \mathbf{p}^v)\right]$. Decreasing v can therefore only induce a switch from $\tau^{*,\infty}$ to one of the other two, or from $\tau^{*,l}$ to $\tau^{*,0}$. The conclusion follows. $\square$

C. Proof of Lemma 1

Proof of Lemma 1: First, given $(\mathbf{p}, \mathbf{a})$, consider an alternative policy $(\hat{\mathbf{p}}, \hat{\mathbf{a}})$ s.t.

- $\hat{p}_t = p_t$ if $a_t(x^h) = 1$
- $\hat{p} = \overline{p}$ if $a_t(x^h) = 0$
- $\hat{a}_t(x^h) = \hat{a}_t(x^l) = 1$ if $a_t(x^h) = 1$
- $\hat{a}_t(x^h) = a_t(x^l) = 0$ if $a_t(x^h) = 0$

Observe that $\hat{a}$ is obedient and since $\mathbf{a}, \mathbf{p}$ are measurable, then so are $\hat{\mathbf{a}}$ and $\hat{\mathbf{p}}$, so that $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}$. I now generate a new policy, $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$, such that $\hat{a}_t(x^h) = \tilde{a}_t(x^h)$ and $\tilde{a}_t(x^l) = 1$ for all t. The definition of W^* and the assumption that $0 \leq \Delta_l = \rho\overline{p} - (\rho + r)\underline{p} - x_l$ imply that $W^*(x^l, t; \hat{p}) \in [-\overline{p}, -\underline{p}]$ *for all* t. The penalty process $\tilde{\mathbf{p}}$ is defined by

- $\tilde{p}_t = -W^*(x^l, t; \hat{p})$ if $\hat{a}(x^h, t) = 0$

- $\tilde{p}_t = p_t$ if $\hat{a}(x^h, t) = 1$

I argue now that $\tilde{\mathbf{p}}$ is measurable. Let $\underline{T}^h(t) := \inf \left\{ s \in \{s | \tilde{a}_s(x_t) = 1\} \cap [t, \infty) \right\}$. Since $\tilde{\mathbf{a}}$ is measurable so is $\underline{T}^h(t)$. Then, observe that

$$W^*(x^l, t; \hat{\mathbf{p}}) = \frac{x^l - \rho\overline{p}}{\rho + r}(1 - e^{-(\rho+r)(\underline{T}^h(t)-t)}) - e^{-(\rho+r)(\underline{T}^h(t)-t)} p_{\underline{T}^h(t)}$$

which is measurable since $\underline{T}^h(t)$ is. Consider any stopping policy τ under this this new policy. Then,

$$W(x, t, \tau; \tilde{\mathbf{p}}) = \mathbb{E}\left[\int_0^{\tau \wedge \tau_\rho} e^{-r(s-t)} x_s ds - e^{-r(\tau \wedge \tau_\rho - t)} \left(\mathbf{1}_{\tau_\rho \leq \tau} \overline{p} + \mathbf{1}_{\tau < \tau_\rho} \left(\mathbf{1}_{\hat{a}(x^h,t)=1} p_\tau + \mathbf{1}_{\hat{a}(x^h,t)=0} W^*(x^l, \tau; \tilde{p}) \right) \right) \right]$$

Now, let $\sigma := \tau \mathbf{1}_{A_\tau(x^h)=0} + \infty(1 - \mathbf{1}_{a_\tau(x^h)=0})$. Then, this expression can be written:

$$\begin{aligned} W(x, t, \tau; \tilde{\mathbf{p}}) = & \quad \mathbb{E}\left[\int_0^{\tau \wedge \tau_\rho \wedge \sigma} e^{-r(s-t)} x_s ds - e^{-r(\tau \wedge \tau_\rho \wedge \sigma - t)} \left(\mathbf{1}_{\tau_\rho \leq \tau \wedge \sigma} \overline{p} + \mathbf{1}_{\tau < \tau_\rho \wedge \sigma} p_\tau + \mathbf{1}_{\sigma \leq \tau_\rho \wedge \tau} W^*(x^l, \sigma; \tilde{p}) \right) \right] \\ \leq & \quad \mathbb{E}\left[\int_0^{\tau \wedge \tau_\rho \wedge \sigma} e^{-r(s-t)} x_s ds - e^{-r(\tau \wedge \tau_\rho \wedge \sigma - t)} \left(\mathbf{1}_{\tau_\rho \leq \tau \wedge \sigma} \overline{p} + \mathbf{1}_{\tau < \tau_\rho \wedge \sigma} p_\tau + \mathbf{1}_{\sigma \leq \tau_\rho \wedge \tau} W^*(x_\sigma, \sigma; \tilde{p}) \right) \right] \end{aligned}$$

where the second line is a result of the fact that $x^l \leq x_\sigma$. But then this implies that

$$W(x, t, \tau; \tilde{\mathbf{p}}) \leq W^*(x, t; \hat{\mathbf{p}})$$

Conversely, by using the strategy $\tau := \{t | a_t(x^h) = 1\}$, an agent guarantees himself $W^*(x_i, t; \hat{\mathbf{p}})$, and so I conclude that $W^*(x_i, t; \hat{\mathbf{p}}) = W^*(x_i, t; \tilde{\mathbf{p}})$. Then by the definition of $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$, $\tau := \inf\{t | \tilde{a}_t(x_t) = 1\}$ is optimal so $\tilde{\mathbf{a}}$ is an obedient recommendation strategy. This completes the proof. $\square$

D. Proof of Theorem 1

It is straightforward to compute

$$(\rho + r)\Delta_l = \rho\overline{p} - (\rho + r)\underline{p} - x^l$$

and

$$(\rho + r + \lambda)\Delta_l = \frac{\rho + r + \lambda}{\rho} \left(\rho\overline{p} - (\rho + r)\underline{p} - x^l \right)$$

So I will prove the result with these expressions in the place of Δ_l.

I first prove an intermediate result, for the case in which $x^l = \alpha_l = 0$.

Proposition D.1. *Suppose $x^l = \alpha_l = 0$. Then,*

$$\Theta^* = \left\{(\rho, r, \lambda, x^h, 0, \overline{p}, \underline{p}) \middle| (\rho + r + \lambda)\frac{\rho\overline{p} - (\rho + r)\underline{p}}{\rho + r} \geq x^h > \rho\overline{p} - (\rho + r)\underline{p}\right\}$$

Proof. By Proposition 1, it is without loss of generality to suppose that, for static penalty policies $\mathbf{p}^v$, the regulator's recommendation is constant i.e. $A_t(x) = A_s(x)$ for each $t, s \geq 0$ and $x \in \{x^h, x^l\}$. An agent's stopping times is then only one of the following three,

1. $\tau^{*,0} := 0$
2. $\tau^{*,\infty} := \infty$
3. $\tau^{*,l} := \inf\{t | x_t = x^l\}$

Observe that, for the case of x_t considered here, the regulator is indifferent between policies $\tau^{*,\infty}$ and $\tau^{*,l}$. By Proposition 1 once more, I need only show a strict improvement over the static penalty policy $\mathbf{p}^{\underline{p}}$ if I want to demonstrate that static policies can be strictly improved.

I first show that $\Theta^* \subseteq \{(\rho, r, \lambda, x^h, 0, \overline{p}, \underline{p}) | (\rho + r + \lambda)(\frac{\rho\overline{p} - (\rho + r)\underline{p}}{\rho + r}) \geq x^h > \rho\overline{p} - (\rho + r)\underline{p}\}$. Fix an arbitrary set of parameters, $\theta \in \Theta^*$. To show this, I compute the value to the agent under $\mathbf{p}^{\underline{p}}$ for any of his three possible optimal stopping times of the agent. The values under penalty policy $\mathbf{p}^{\underline{p}}$ for an agent arriving at t_0 in state x are

$$\begin{aligned}
W(x, \tau^{*,l}, t_0; \mathbf{p}^{\underline{p}}) &= \mathbb{E}_{x_0 = x}\left[\int_0^{\tau^{*,l}} x_t e^{-(\rho+r)t} - \frac{\rho}{\rho + r}\overline{p}(1 - e^{-(\rho+r)\tau^{*,l}}) - e^{-(\rho+r)\tau^{*,l}}\underline{p}\right] \\
&= \mathbf{1}_{x = x^h}\left(\frac{x^h - \rho\overline{p} - \lambda\underline{p}}{\rho + r + \lambda}\right) + \mathbf{1}_{x = x^l}(-\underline{p}) \\
W(x, \tau^{*,0}, t_0; \mathbf{p}^{\underline{p}}) &= -\underline{p} \\
W(x, \tau^{*,\infty}, t_0; \mathbf{p}^{\underline{p}}) &= \mathbf{1}_{x = x^h}\left(\frac{x^h}{\rho + r + \lambda}\right) - \frac{\rho\overline{p}}{\rho + r}
\end{aligned}$$

Suppose first that $x^h - (\rho\overline{p} - (\rho + r)\underline{p}) \leq 0$. Then,

$$\max\{W(x^h, \tau^{*,0}, t_0)), W(x^h, \tau^{*,l}, t_0))\} \leq W(x^h, \tau^{*,0}, t_0))$$

As a consequence, the recommendation policy $(a_t)_{t \geq 0}$ with $a_t(x) = 1$ for all (t, x) is an obedient recommendation policy. The regulator achieves first best with $\mathbf{p}^{\underline{p}}$ and $a_t(x) = 1$ for all (t, x). This implies that $\Theta^* \subseteq \{(\rho, r, \lambda, \overline{p}, \underline{p}, x) | x^h - (\rho\overline{p} - (\rho + r)\underline{p}) > 0\}$.

Suppose now that $x^h > \frac{\rho + r + \lambda}{\rho + r}(\rho\overline{p} - (\rho + r)\underline{p})$. The agent's value for $\tau^{*,\infty}$ is

$$W(x, \tau^{*,\infty}, t_0; \mathbf{p}^{\underline{p}}) = \mathbb{E}\left[\int_0^{\infty} x_t e^{-(\rho+r)} dt - \frac{\rho}{\rho + r}\overline{p}\right]$$

$$= \mathbf{1}_{x=x^h}\frac{x^h}{\rho+r+\lambda} - \frac{\rho\overline{p}}{\rho+r}$$

When $x = x^h$, the assumption that $x^h > (\rho + r + \lambda)(\frac{1}{\rho+r})(\rho\overline{p} - (\rho + r)\underline{p})$ implies that this expression is strictly larger than $-\underline{p}$. Because of this, no recommendation policy with $a_t(x^h) = 1$ for some t can be obedient, and so a static policy is optimal. This implies that $\Theta^* \subseteq \{(\rho, r, \lambda, x^h, 0, \overline{p}, \underline{p})|(\rho + r + \lambda)(\frac{\rho\overline{p}-(\rho+r)\underline{p}}{\rho+r}) \geq x^h\}$.

Combining this with our finding above that $\Theta^* \subseteq \{(\rho, r, \lambda, x^h, 0, \overline{p}, \underline{p})|x^h-(\rho\overline{p}-(\rho+r)\underline{p}) > 0\}$, the first inclusion is shown.

I now show $\left\{(\rho, r, \lambda, x^h, 0\overline{p}, \underline{p})|(\rho+r+\lambda)(\frac{\rho\overline{p}-(\rho+r)\underline{p})}{\rho+r})\overline{p} \geq x^h > \rho\overline{p} - (\rho+r)\underline{p}\right\} \subseteq \Theta^*$. Let θ be an arbitrary element on the left-hand side. Consider the policy $\mathbf{p}^{\underline{p}}$. Using the equations above and the assumption $x^h > \rho\overline{p} - (\rho + r)\underline{p}$,

$$W(x^h, \tau^{*,l}, t_0) - W(x^h, \tau^{*,0}, t_0) = \frac{x - \rho\overline{p} - \lambda\underline{p}}{\rho + r + \lambda} + \underline{p} > 0$$

In this case then, the regulator receives his *worst possible* payoff; no agent ever reports until reaching the low state $x^l = 0$. Thus, all I need to do to conclude the proof is demonstrate a policy which induces a positive mass of high types to report.

To this end, consider a *one-time policy*: $\tilde{p}_t = \mathbf{1}_{t=T}\underline{p} + (1 - \mathbf{1}_{t=T})\overline{p}$ for some $T > 0$, $a_T(x^l) = a_T(x^h) = 1$ and otherwise $a_t(x^l) = a_t(x^h) = 0$. Then, observe that,

$$W(x^h, \tau^{*,\infty}, T) = \frac{x^h}{\rho + r + \lambda} - \frac{\rho\overline{p}}{\rho + r} \leq -\underline{p} = W(x^h, \tau^{*,0}, T)$$

where the inequality follows by assumption that $(\rho + r + \lambda)\frac{\rho\overline{p}-(\rho+r)\underline{p}}{\rho+r} \geq x^h$. Thus, the recommendation $a_t(x) = 1$ if and only $t = T$ is obedient. Since $T > 0$, this policy induces a strictly positive mass of high types to stop by T, generating a strict improvement of the regulator's value over any static policy. So I find,

$$\{(\rho, r, \lambda, x^h, 0, \overline{p}, \underline{p})|(\rho + r + \lambda)(\frac{\rho\overline{p} - (\rho + r)\underline{p}}{\rho + r}) \geq x^h > \rho\overline{p} - (\rho + r)\underline{p}\} \subseteq \Theta^*$$

and the proof is concluded by combining this with the reverse inclusion. □

Proof of Theorem 1: Suppose now that $x^l > 0$ or $\alpha_l > 0$. Suppose first that $x^l > \rho\overline{p} - (\rho + r)\underline{p}$. Then, $W(x^l, \tau^{*,\infty}, t; \mathbf{p}) > 0$ for any $t, \mathbf{p}$ where $\tau^{*,\infty} = \infty$. In that case, the only obedient recommendation is $a_t(x^h) = a_t(x^l) = 0$ for all t, in which case the policy $\mathbf{p}$ has no effect on behavior. Because of this, $\Theta^* \cap \{(\rho, r, \lambda, x^h, x^l, \overline{p}, \underline{p})|x^l > \rho\overline{p} - (\rho + r)\underline{p}\} = \emptyset$.

Suppose now that $x^l \leq \rho\overline{p} - (\rho + r)\underline{p}$. By Lemma 1,

$$V^* = \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}} V(\mathbf{p}, \mathbf{a}) = \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{L}} V(\mathbf{p}, \mathbf{a}).$$

Since the right hand side is independent of α_l, it is without loss of generality to prove the result for the $x^l \geq 0$ but $\alpha_l = 0$. To this end, set $\tilde{x}_h = x^h - x^l$, $\tilde{x}_l = 0$ and $\tilde{\overline{p}} = \overline{p} - \frac{x^l}{\rho}$. An agent's value for a stopping time can then be re-written,

$$
\begin{aligned}
W(x, t_0, \tau; \mathbf{p}) &= \mathbb{E}\left[\int_0^{\tau} e^{-(\rho+r)t} x_t dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho+r}\overline{p} - e^{-(\rho+r)\tau} p_{\tau+t_0}\right] \\
&= \mathbb{E}\left[\int_0^{\tau\wedge\tau_l} e^{-(\rho+r)t}(x^h - x^l)dt + \int_0^{\tau} e^{-(\rho+r)t}(x^l)dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho+r}\overline{p} - e^{-(\rho+r)\tau} p_{\tau+t_0}\right] \\
&= \mathbb{E}\left[\int_0^{\tau} e^{-(\rho+r)t} \tilde{x}_t dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho\tilde{\overline{p}}}{\rho+r} - e^{-(\rho+r)\tau} p_{\tau+t_0}\right]
\end{aligned}
$$

So, an agent's value for a stopping time is the same across parameterizations $(\rho, r, \lambda, \tilde{\overline{p}}, \underline{p}, \tilde{x^h}, \tilde{x^l})$ and $(\rho, r, \lambda, \overline{p}, \underline{p}, x^h, x^l)$. The result then follows from the application of Proposition D.1 to the parameterization $\theta = (\rho, r, \lambda, \tilde{x^h}, \tilde{x^l}, \tilde{\overline{p}}, \underline{p})$, followed by application of the construction in Lemma 1 to ensure that low types report everywhere as in an optimal static policy. $\square$

E. Proofs of Section 3.4

Recall that $P = [\underline{p}, -W^*(x^h, 0, \tau^\infty)]$. Plugging in the definition of $W^*(x^h, 0, \tau^\infty)$, $P = [\underline{p}, \frac{\rho\overline{p} - x^l}{\rho+r} - \frac{x^h}{\rho+r+\lambda}]$. Before proceeding with proofs of Section 3.4, I provide a useful calculation of μ_h^t.

Lemma E.1. *Fix $t \in \mathbb{R}^+$ and suppose that for $\underline{t} < t$, $a_{\underline{t}}(x^h) = 1$ and $a_s(x^h) = 0$ for each $s \in (\underline{t}, t]$. Then, the discounted measure of high type agents is:*

$$e^{-rt}\mu_t^h = e^{-rt}\frac{1 - e^{-(\rho+\lambda)(t-\underline{t})}}{\rho+\lambda}$$

Proof of Lemma 2: This follows directly from arguments in the text. For any policy $(\mathbf{p}, \mathbf{a})$, define a new policy $\tilde{p}_t = \overline{p}$ if $a_t(x^h) = 0$ and $\tilde{p}_t = p_t$ otherwise. Define $\tilde{a}_t(x^h) = \tilde{a}_t(x^l) = a_t(x^h)$. The resulting policy, $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$, is a member of $\mathcal{M}$ and delivers the regulator the same value as $(\mathbf{p}, \mathbf{a})$ when $\alpha_l = 0$. $\square$

Proof of Lemma 3: For any policy $(\mathbf{p}, \mathbf{a})$ and let $\mathcal{T}(\mathbf{a}) \subset \mathbb{R}_+$ such that $a_t(x^h) = 1$ if and only if $t \in \mathcal{T}(\mathbf{a}) \subset \mathbb{R}_+$.

Now, let $\mathcal{M}^0 \subset \mathcal{M}$ be the set of policies $(\mathbf{p}, \mathbf{a})$ such that $a_t(x^h) = 0 \implies p_t = \overline{p}$ and there exists a sequence $\mathbf{t} := (t_i)_{i\in\mathbb{N}}$ such that $a_t(x^h) = 1$ if and only if $t = \sum_{i\leq n} t_i$ for some

$n \in \mathbb{N}$, and $\inf_i t_i > 0$. Let $\mathbf{t}(\mathbf{p}, \mathbf{a})$ be the sequence associated to a policy $(\mathbf{p}, \mathbf{a})$.

First, observe that $V^* = \sup_{(\mathbf{p},\mathbf{a}) \in \mathcal{M}^0} V(\mathbf{p}, \mathbf{a})$. To see, fix any policy $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}$ with $(\mathbf{p}, \mathbf{a}) \notin \mathcal{M}^0$. Generate a new path $\tilde{\mathbf{t}}$ as follows:

- $\tilde{t}_0 = \inf\{t | t \in \mathcal{T}\}$
- $\tilde{t}_{i+1} = \inf\{t \geq \epsilon + \sum_{j \leq i} \tilde{t}_i | t \in \mathcal{T}\} - \sum_{j \leq i} \tilde{t}_i$

Generate associated policy $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$ as follows:

- $\tilde{p}_{\tilde{t}_i} = p_{t_i}$ and otherwise $\tilde{p}_t = \overline{p}$
- $\tilde{a}_t(x^h) = \tilde{a}_t(x^l) = 1$ if $t = \sum_{i \leq n} t_i$ for some n and $\tilde{a}_t(x^h) = \tilde{a}_t(x^l) = 0$ otherwise

Observe that $\tilde{\mathbf{t}} = \mathbf{t}(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$ and that $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}}) \in \mathcal{M}^0$. Furthermore, because the regulator discounts and $\alpha_l = 0$, the loss from $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$ relative to $(\mathbf{p}, \mathbf{a})$ converges to 0 as $\epsilon \to 0$, and so the result follows by taking $\epsilon \to 0$.

To complete the proof, I will show that if $\alpha_l = 0$ and $\mathbf{V}(p)$ satisfies the premise of the lemma with associated policies $(t^*(p), p^*(p))$,

$$\sup_{(\mathbf{p},\mathbf{a}) \in \mathcal{M}^0} V(\mathbf{p}, \mathbf{a}) = \max_{t_0 \geq 0, p_0 \in P} \left\{ -v(t_0) + e^{-rt_0} \mathbf{V}(p_0) \right\}$$

First, I will show that

$$\sup_{(\mathbf{p},\mathbf{a}) \in \mathcal{M}^0} V(\mathbf{p}, \mathbf{a}) \leq \max_{t_0 \geq 0, p_0 \in P} \left\{ -v(t_0) + e^{-rt_0} \mathbf{V}(p_0) \right\}.$$

For any policy $(\mathbf{p}, \mathbf{a})$, let $s_j(\mathbf{p}, \mathbf{a}) = \sum_{i \leq j} \mathbf{t}_i(\mathbf{p}, \mathbf{a})$ and then

$$V(\mathbf{p}, \mathbf{a}) = \sum_{i=0}^{\infty} e^{-rs_{i-1}(\mathbf{p},\mathbf{a})} \int_0^{t_i(\mathbf{p},\mathbf{a})} e^{-rt} \mu^h_{t+s_{i-1}(\mathbf{p},\mathbf{a})} dt$$

where $t_{-1} = 0$. Apply Lemma E.1 to show that for $t \in [s_i(\mathbf{p}, \mathbf{a}), s_{i+1}(\mathbf{p}, \mathbf{a})]$, $\mu^h_t = \frac{1 - e^{-(\rho+\lambda)(t - s_i(\mathbf{p},\mathbf{a}))}}{\rho+\lambda}$.

To complete this step, note that the equivalence between incentive compatibility and one-shot incentive compatibility implies that $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}^0$ if and only if

$$w_h(t) - e^{-(\rho+r)t_{i+1}} p_{s_{i+1}} \leq -p_{s_i}$$

for each $i \in \mathbb{N}$. This then implies that

$$\sup_{(\mathbf{p},\mathbf{a}) \in \mathcal{M}^0} V(\mathbf{p}, \mathbf{a}) \leq \max_{t_0 \geq 0, p_0 \in P} \left\{ -v(t_0) + e^{-rt_0} \mathbf{V}(p_0) \right\}.$$

Next, I argue that

$$\sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^0} V(\mathbf{p},\mathbf{a}) \geq \max_{t_0\geq 0, p_0\in P} \left\{-v(t_0) + e^{-rt_0}\mathbf{V}(p_0)\right\}.$$

Now, define $\mathbf{t}^* = (t^{*,n}(p))_{n\in\mathbb{N}}$ where $t^{*,n}(p) = t^*(p^{*,n-1}(p))$ and $p^{*,n}(p) = p^*(p^{*,n-1}(p))$. Then define $(\mathbf{p}^*, \mathbf{a}^*)$ as follows,

- $p^*_t = p^{*,n}(p)$ if $t = \sum_{i<n} t^*_i$ and $p^*_t = \overline{p}$ otherwise
- $a^*_t(x^h) = a^*_t(x^l) = 1$ if $t \in \{\sum_{i<n} t^*_i\}_{n\in\mathbb{N}}$ and $a^*_t(x^h) = a^*_t(x^l) = 0$ otherwise

Then, $\mathbf{t}(\mathbf{p}^*, \mathbf{a}^*) = \mathbf{t}^*$ and since $\inf_i t^*_i > 0$, $(\mathbf{p},\mathbf{a}) \in \mathcal{M}^0$. This then implies that,

$$\sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^0} V(\mathbf{p},\mathbf{a}) \geq \max_{t_0\geq 0, p_0\in P} \left\{-v(t_0) + e^{-rt_0}\mathbf{V}(p_0)\right\}.$$

Combining this with the previous inequality,

$$\sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^0} V(\mathbf{p},\mathbf{a}) = \max_{t_0\geq 0, p_0\in P} \left\{-v(t_0) + e^{-rt_0}\mathbf{V}(p_0)\right\}.$$

and this is completes the proof. □

Before proceeding to the proof of Proposition 2, I establish a few useful lemmas. Let

$$v^0(t) := \frac{1 - e^{-(\rho+r+\lambda)t}}{(\rho+r)(\rho+r+\lambda)}.$$

Lemma E.2. *If $\alpha_l = 0$, $\theta \in \Theta^*$ then $\inf_{p\in P} t^*(p) > 0$.*

Proof. To see this, first observe that $\lim_{t\to\infty} \left(w_h(t) - e^{-(\rho+r)t}\underline{p}\right) = \frac{x^h}{\rho+r+\lambda} - \frac{\rho\overline{p}-x^l}{\rho+r}$, which is negative since $\theta \in \Theta^*$. Next, observe that

$$\frac{\partial}{\partial t}\left(w_h(t) - e^{-(\rho+r)t}\underline{p}\right) = e^{-(\rho+r)t}\left(x^h e^{-\lambda t} + (\rho+r)\underline{p} - (\rho\overline{p} - x^l))\right)$$

which, because $\theta \in \Theta^*$, is initially strictly positive and changes sign to strictly negative exactly once and remains strictly negative. Finally observe that $w_h(0) - e^{-(\rho+r)0}\underline{p} = \underline{p} \geq -p$ for any $p \in P$. Combining these three leads to $t^*(p) \geq t^*(\underline{p})$ for each p, and $t^*(\underline{p}) > 0$. This leads to the conclusion that $\inf_{p\in P} t^*(p) > 0$. □

Lemma E.3. *Suppose $\theta \in \Theta^*$. Then, $t^*(p)$ is strictly increasing on P.*

Lemma E.4. *Suppose that $\theta \in \Theta^*$. Then*

$$v^0(t) + e^{-rt}\frac{v^0(t^*(\underline{p}))}{1 - e^{-rt^*(\underline{p})}}$$

is decreasing in t for any $t \geq t^*(\underline{p})$. Further,

$$v^0(t^*(p)) + \frac{v^0(t^*(\underline{p}))}{1 - e^{-rt^*(\underline{p})}}$$

is decreasing in p.

Let $\mathbf{V}^*(p)$ be the value function associated to policies $p^{\prime,*}(.) = \underline{p}$ and $t^*(.)$ defined as the unique strictly positive solution to Equation (I).

Proof of Proposition 2: To prove the result, I show that $\mathbf{V}^*(p)$ and $(t^*(p), p^{\prime,*}(p))$ satisfy the premise of Lemma 3, which immediately implies the result. That is, I must show that Equation (R^*) is satisfied for $\mathbf{V}^*(.)$ and $\inf_{p \in P} t^*(p) > 0$. This latter requirement is proved in Lemma E.2 and so I need only show the former.

First, observe that

$$\begin{aligned} v(t) &= \frac{1 - e^{-rt}}{r(\rho + \lambda)} - \frac{1 - e^{-(\rho + r + \lambda)t}}{(\rho + \lambda)(\rho + r + \lambda)} \\ &= \frac{1 - e^{-rt}}{r(\rho + \lambda)} - v^0(t) \end{aligned}$$

Plugging in the definition of $(t^*(p), p^{\prime,*}(p))$ to $\mathbf{V}^*(p)$, I find that after rearranging

$$-v(t) + \mathbf{V}^*(p') = \frac{1}{r(\rho + \lambda)} + \left(v^0(t) + e^{-rt} v^0(t^*(p')) + e^{-r(t + t^*(p'))} \frac{v^0(t^*(0))}{1 - e^{-rt^*(0)}} \right)$$

Define $\mathbf{V}^{*,0}(p) = v^0(t^*(p)) + e^{-rt^*(p')} \frac{v^0(t^*(0))}{1 - e^{-rt^*(0)}}$. Then, rather than show Equation (R^*), it is sufficient to show that

$$0 = \begin{cases} \sup_{t, p'} \quad v^0(t) + e^{-rt} \mathbf{V}^{*,0}(p') - \mathbf{V}^{*,0}(p) \\ \text{subject to} \\ w_h(t) - e^{-(\rho + r)t} p' \leq -p \\ p' \in P \end{cases}$$

First, I argue that any choice of $t > t^*(p)$ can be strictly improved by setting $t = t^*(p)$. To see this, note that, given $t \geq t^*(p)$, the optimal choice of p' is $-\underline{p}$, since $\mathbf{V}^{*,0}(p')$ is decreasing in p' by the second part of Lemma E.4. So, then by Lemma E.4, I know that $v^0(t) + e^{-rt} \mathbf{V}^{*,0}(\underline{p})$ is decreasing in t, so among choices $t \geq t^*(\underline{p})$, it is optimal to set $t = t^*(\underline{p})$.

I now argue the more difficult case when $t \leq t^*(\underline{p})$. Let $f := \rho + \lambda + r$ and $g := \rho + r$. To prove the result, it is sufficient to show, after plugging in the definition of $\mathbf{V}^{*,0}(p)$ and canceling, that

$$
\begin{aligned}
&(1-e^{-ft^*(p)}) + e^{-rt^*(p)}(1-e^{-ft^*(\underline{p})}) + e^{-r(t^*(p)+t^*(\underline{p}))}\left(\frac{1-e^{-ft^*(\underline{p})}}{1-e^{-rt^*(\underline{p})}}\right) \\
\geq &\left[1-e^{-ft} + e^{-rt}(1-e^{-ft^*(p')}) + e^{-r(t+t^*(p'))}\frac{1-e^{-ft^*(\underline{p})}}{1-e^{-rt^*(\underline{p})}}\right] \qquad \text{(P*)}
\end{aligned}
$$

for any t, p, p' such that inequality

$$
x^h \frac{1-e^{-ft}}{f} - \frac{\rho\overline{p} - x^l}{\rho + r}(1-e^{-gt}) - p' e^{-gt} \leq -p
$$

holds.

The next step manipulates this latter inequality into a form that looks like inequality (P*). Use the definition of $t^*(p')$ to plug in for p' on the left hand side, let $c_A := \frac{\rho\overline{p}-x^l-(\rho+r)\underline{p}}{g} f$ and rearrange to find,

$$
x^h(1-e^{-ft}) + x^h e^{-gt}(1-e^{-ft^*(p')}) - c_A(1-e^{-g(t+t^*(p'))}) \leq -fp
$$

By definition, this holds with equality for $(t, p') = (t^*(p), \underline{p})$, so this inequality is equivalent to

$$
\begin{aligned}
& x^h(1-e^{-ft}) + x^h e^{-gt}(1-e^{-ft^*(p')}) - c_A(1-e^{-g(t+t^*(p'))}) \\
\leq \quad & x^h(1-e^{-ft^*(p)}) + x^h e^{-gt^*(p)}(1-e^{-fT^*(\underline{p})}) - c_A(1-e^{-g(t+t^*(\underline{p}))})
\end{aligned}
$$

The definition of $t^*(\underline{p})$ implies that $c_A = x^h \frac{1-e^{-ft^*(\underline{p})}}{1-e^{-gt^*(\underline{p})}}$. Plugging this in on both sides, rearranging and canceling x^h,

$$
\begin{aligned}
&1-e^{-ft} + e^{-gt}(1-e^{-ft^*(\ell')}) + e^{-g(t+t^*(\ell'))}\frac{1-e^{-ft^*(\underline{p})}}{1-e^{-gt^*(\underline{p})}} \\
\leq &1-e^{-ft^*(\ell)} + e^{-gt^*(\ell)}(1-e^{-ft^*(\underline{p})}) + e^{-g(t^*(\ell)+t^*(\underline{p}))}\frac{1-e^{-ft^*(\underline{p})}}{1-e^{-gt^*(\underline{p})}} \qquad \text{(A*)}
\end{aligned}
$$

Thus, to prove the result, it is sufficient to show that for any t, p', p s.t (A^*) holds, (P^*) holds as well. For any t, p', let

$$
z := e^{-ft^*(\underline{p})}, \quad z_p := e^{-ft^*(p)}, \quad u := e^{-ft}, \quad y := e^{-ft^*(p')}
$$

By assumption, $z \geq z_p$ and $y \leq z$ and by the assumption that $t \leq t^*(p)$, $u \geq z_p$. Since $t^*(p)$ is strictly increasing in t, we have $y = z$ if and only if $u = z_p$ or $u = 1$. In this case, the deviation is in fact the conjectured optimal choice, and so the inequalities (A^*) and (P^*) hold at equality. Thus, I need only show the result assuming that $1 > u > z_p$ and $y < z$.

Plug these definitions into (A^*) and multiply both sides by $1 - z^{\frac{g}{f}}$ to get

$$
\begin{aligned}
&(1-u)(1-z^{\frac{g}{f}}) + u^{\frac{g}{f}}(1-y)(1-z^{\frac{g}{f}}) + (uy)^{\frac{g}{f}}(1-z) \\
\leq &(1-z_p)(1-z^{\frac{g}{f}}) + z_p^{\frac{g}{f}}(1-z)(1-z^{\frac{g}{f}}) + (zz_p)^{\frac{g}{f}}(1-z)
\end{aligned}
$$

Similarly, plug in to (P^*) and multiply both sides by $1 - z^{\frac{r}{f}}$ to get

$$
\begin{aligned}
&(1-u)(1-z^{\frac{r}{f}}) + u^{\frac{r}{f}}(1-y)(1-z^{\frac{r}{f}}) + (uy)^{\frac{r}{f}}(1-z) \\
\leq &(1-z_p)(1-z^{\frac{r}{f}}) + z_p^{\frac{r}{f}}(1-z)(1-z^{\frac{r}{f}}) + (zz_p)^{\frac{r}{f}}(1-z)
\end{aligned}
$$

Note than in neither case does this multiplication change the direction of the inequality, since $\frac{r}{f}, \frac{g}{f}, z \in (0,1)$. Each of these is a special case of the inequality, for arbitrary a,

$$
\begin{aligned}
&(1-u)(1-z^a) + u^a(1-y)(1-z^a) + (uy)^a(1-z) \\
\leq &(1-z_p)(1-z^a) + z_p^a(1-z)(1-z^a) + (zz_p)^a(1-z)
\end{aligned}
$$

After rearranging and canceling terms, we arrive at:

$$
\begin{aligned}
&0 \leq (u-z_p) - u^a(1-y) + (uz)^a(1-y) - (uy)^a(1-z) \\
&\qquad + z^a(1-u) + (z_p z^a - z_p^a z) + (z_p^a - z^a) \\
\implies &0 \leq (u-z_p) + (uz)^a(1-y) + z_p^a(1-z) - (uy)^a(1-z) - u^a(1-y) - z^a(u-z_p)
\end{aligned}
$$

Denote the RHS of the inequality by $h(a; u, y, z, z_p)$. The crucial step is the following claim:

$$
\begin{aligned}
&\text{if } 0 \leq h(a; u, y, z, z_p) \textit{ at some } a \in (0,1), \\
&\text{then } 0 < h(a; u, y, z, z_p) \text{for all } 0 < a' < a.
\end{aligned}
\tag{C}
$$

The agent's version of $h(.)$ sets $a = \frac{g}{f}$, while the regulator's version sets $a = \frac{r}{f} < \frac{g}{f}$. Thus, if the agent's version holds – i.e. IC holds – then the regulator's version holds – i.e. value to the deviating choice t, p' is lower than the conjectured optimal choice.

To prove this claim, there are then three cases to consider.

Case 1 - $y > 0$ and $uy < z_p$: In this case, instead of showing Property (C) for h, I will show it for $\tilde{h}(a) := \frac{h}{(uy)^a}$, which can be written,

$$
\tilde{h}(a) = \frac{(u-z_p)}{(uy)^a} + (\frac{z}{y})^a(1-y) + (\frac{z_p}{uy})^a(1-z) - (1-z) - (\frac{1}{y})^a(1-y) - (\frac{z}{uy})^a(u-z_p)
$$

from which Property (C) for h can be recovered immediately. Note that $\tilde{h}$ is smooth in a for any feasible choices of u, y, z, z_p with $y > 0$. Going forward, when it is clear, I will suppress the dependence of h on all inputs but a.

First, given our baseline ordering of u, z, z_p, y and under the additional assumption that $uy < z_p$, $\tilde{h} \to \infty$ as $a \to \infty$. Second, since $uy < z_p$ and $y < z$, $\tilde{h} \to -(1-z)$ as $a \to -\infty$. Third, $\tilde{h}(1) = \tilde{h}(0) = 0$. Given these observations, if I can show that $\frac{\partial \tilde{H}^2}{\partial a^2}$ intersects zero at most twice, the result will follow. To see this, first observe that this implies there can be no interval such that $h = 0$ on the interval. Given this, observe that to violate the property, there must be points $a_0 < 0 < a_1 < a_2 < 1 < a_3$ such that $\tilde{h}(a_0) < 0$, $\tilde{h}(a_1) \leq 0$, $\tilde{h}(a_2) \geq 0$, $\tilde{h}(a_3) > 0$, and $\tilde{h}(0) = \tilde{h}(1) = 0$. Since there is no interval on which $\tilde{h} = 0$, connecting these points in an infinitely differentiable way (as required by the definition of H) requires the existence of points $b_0 < b_1 < b_2 < b_3$ such that b_0 and b_2 are local maxima with strictly negative second derivative while b_1 and b_3 are local minima with strictly positive second derivative. This implies that $\frac{\partial \tilde{H}^2}{\partial a^2}$ crosses zero at least three times.

So, as long as I show that $\frac{\partial^2 h(a)}{\partial a^2}$ has at most two zeros, the proof will be complete.

Twice differentiate $\tilde{h}$,

$$\begin{aligned}\frac{\partial \tilde{h}^2}{\partial a^2} =& \frac{(u-z_p)}{(uy)^a} ln(\frac{1}{uy})^2 + (\frac{z}{y})^a (1-y) ln(\frac{z}{y})^2 \\ &+ (\frac{z_p}{uy})^a (1-z) ln(\frac{z_p}{uy})^2 - (\frac{1}{y})^a (1-y) ln(\frac{1}{y})^2 - (\frac{z}{uy})^a (u-z_p) ln(\frac{z}{uy})^2\end{aligned}$$

We want to show that this object has at most 2 zeros. The zeros of this function are the same as the zeros of the function $G := y^a \frac{\partial \tilde{H}^2}{\partial a^2}$. Computing, we get,

$$\begin{aligned}G =& \frac{(u-z_p)}{(u)^a} ln(\frac{1}{uy})^2 + (z)^a (1-y) ln(\frac{z}{y})^2 + (\frac{z_p}{u})^a (1-z) ln(\frac{z_p}{uy})^2 \\ &- (1-y) ln(\frac{1}{y})^2 - (\frac{z}{u})^a (u-z_p) ln(\frac{z}{uy})^2\end{aligned}$$

Then, to show that G has at most two zeros, it would be enough to show that $\frac{\partial G}{\partial a}$ has at most one zero. Differentiating:

$$\begin{aligned}\frac{\partial G}{\partial a} =& \frac{(u-z_p)}{(u)^a} ln(\frac{1}{uy})^2 ln(\frac{1}{u}) + (z)^a (1-y) ln(\frac{z}{y})^2 ln(z) \\ &+ (\frac{z_p}{u})^a (1-z) ln(\frac{z_p}{uy})^2 ln(\frac{z_p}{u}) - ln(\frac{z}{u})(\frac{z}{u})^a (u-z_p) ln(\frac{z}{uy})^2\end{aligned}$$

I apply one more transformation: $\frac{\partial G}{\partial a}$ has the same number of zeros as $J := \frac{u^a \frac{\partial G}{\partial a}}{z^a}$. Thus, if $\frac{\partial J}{\partial a}$ is either always negative or always positive, then J has at most one zero and the result will be proved.

Differentiating yields

$$\begin{aligned}\frac{\partial J}{\partial a} =&\frac{(u-z_p)}{(z)^a}ln(\frac{1}{uy})^2ln(\frac{1}{u})ln(\frac{1}{z})+(u)^a(1-y)ln(\frac{z}{y})^2ln(z)ln(u)\\&+(\frac{z_p}{z})^a(1-z)ln(\frac{z_p}{uy})^2ln(\frac{z_p}{u})ln(\frac{z_p}{z})\end{aligned}$$

Recalling that $1 > u > z_p > 0$ and $1 \geq z \geq z_p > 0$ implies that all the terms in the RHS of the above equation are positive and the first is strictly positive. Thus, J is a strictly increasing function and has at most one zero. That implies that the same is true of $\frac{\partial G}{\partial a}$. This therefore implies that G has at most two zeros and hence the same is true of $\frac{\partial \tilde{h}^2}{\partial a^2}$.

Case 2 - $uy \geq z_p$: In this case, I show that for any $a \in (0,1)$, $h(a) < 0$ (and so the claim is proved).[36]

Suppose first that $z_p = uy$. Then, we have:

$$\begin{aligned}h =&(u-z_p)+(uz)^a(1-y)+z_p^a(1-z)-(uy)^a(1-z)-u^a(1-y)-z^a(u-z_p)\\=&(1-y)(u-u^a)(1-z^a)\end{aligned}$$

and since $u < u^a$ for $a \in (0,1)$, this is strictly negative. Next, I show that $\frac{\partial h}{\partial z_p} \geq 0$ on $z_p < uy$. This implies that on $z_p \leq uy$, H is maximized at $z_p = uy$, which we've already seen is negative and so the proof will be concluded.

Let $G(z) := \frac{\partial h}{\partial z_p} = z^a + \frac{a}{{z_p}^{1-a}}(1-z) - 1$ and differentiate to get $\frac{\partial G}{\partial z} = \frac{a}{z^{1-a}} - \frac{a}{z_p^{1-a}}$. Since $G(1) \geq 0$ and $\frac{\partial G}{\partial z}(z) \leq 0$ for all $z \geq z_p$, we find that $G(z) \geq 0$ for all $z \geq z_p$.

This implies by definition that $\frac{\partial h}{\partial z_p}(x) \geq 0$ for all $z \in [z_p, 1]$ and since $z \in [z_p, 1]$ by assumption, this concludes the proof for this case.

Case 3: $y = 0$: Observe now that $h(0) = 1-z$, $h(1) = 0$ and $h(\infty) = u - z_p > 0$. Then, to violate the claim, $\frac{\partial h(a)}{\partial a}$ must have at least three zeros on $a \geq 0$. So, I prove here that $\frac{\partial h(a)}{\partial a}$ has at most two zeros on $a \geq 0$ when $y = 0$.

$$\begin{aligned}\frac{\partial}{\partial a}h(a) &= ln(uz)(uz)^a + ln(z_p)z_p^a(1-z) - ln(u)u^a - ln(z)z^a(u-z_p)\\&= z_p^a\left(ln(uz)(\frac{uz}{z_p})^a + ln(z_p)(1-z) - ln(u)(\frac{u}{z_p})^a - ln(z)(\frac{z}{z_p})^a(u-z_p)\right)\end{aligned}$$

[36]That is, there can never be a pair (u,y) s.t. $uy \geq z_p$ and (u,y) satisfies IC. Intuitively, it would be as if the regulator said, in between 0 and $t^*(p)$, you will have two opportunities for reduced penalties and the second opportunity will have a penalty of $\underline{p}$. By the definition of $t^*(p)$, such a policy cannot be incentive compatible.

Let $h^1(a) := ln(uz)(\frac{uz}{z_p})^a + ln(z_p)(1-z) - ln(u)(\frac{u}{z_p})^a - ln(z)(\frac{z}{z_p})^a(u-z_p)$. To conclude, I show that $h^1(a)$ has at most two zeros. Differentiate to get

$$\begin{aligned}\frac{\partial}{\partial a}h^1(a) &= ln(\frac{uz}{z_p})(\frac{uz}{z_p})^a - ln(u)ln(\frac{u}{z_p})(\frac{u}{z_p})^a - ln(z)ln(\frac{z}{z_p})(\frac{z}{z_p})^a(u-z_p)\\ &= \left(\frac{uz}{z_p}\right)^a\left(ln(\frac{uz}{z_p}) - ln(u)ln(\frac{u}{z_p})(\frac{1}{z})^a - ln(z)ln(\frac{z}{z_p})(\frac{1}{u})^a(u-z_p)\right).\end{aligned}$$

Since $z_p < 1$ and $u > z_p$, the term on the right-hand side is increasing in a for $a > 0$. This implies that $\frac{\partial}{\partial a}h^1(a)$ has at most one zero, and so $h^1(a) = \frac{\partial h(a)}{\partial a}$ has at most two zeros on $a \geq 0$. This concludes the proof. □

F. Proof of Theorem 2

Proof of Theorem 2: I first prove the result in case $\alpha_l = 0$ and then prove the result for $\alpha_l \geq 0$.

$\boldsymbol{\alpha_l = 0}$**.** Suppose $\alpha_l = 0$ and $\theta \in \Theta^*$. The result follows by applying Proposition 2, observing that $\mathbf{V}^*(p)$ is maximized at $p = \underline{p}$, so that

$$V^* = \max_{t_0}\{v(t_0) + e^{-rt_0}\mathbf{V}^*(\underline{p})\}$$

and repeatedly substituting in for $\mathbf{V}^*(\underline{p})$.

$\boldsymbol{\alpha_l \geq 0}$. Now, I move on to the case in which $\alpha_l \geq 0$. When $\theta \notin \Theta^*$, an application of Proposition 1 leads to the result.

Suppose instead that $\theta \in \Theta^*$. To prove this result, I transform the regulator's problem into one with $\alpha_l = 0$, to which I will then apply the result in the case $\alpha_l = 0$.

Fix some parameters of the model, $\theta \in \Theta^*$. Rather than studying problem $\mathcal{P}$, consider the problem

$$V_h^* := \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}} \int_0^t \mu_t^h \qquad (\mathcal{P}^h)$$

This problem differs from ($\mathcal{P}$) only in that losses from the low type agent do not enter the objective function.

Notice that the problem $\mathcal{P}^h$ is exactly the problem studied when $\alpha_l = 0$. Therefore, an optimal policy in problem $\mathcal{P}^h$ is $(\mathbf{p}^*, \mathbf{a}^*)$ defined by:

- $p^*_{t_0+nt^*(\underline{p})} = 0$ for some t_0 with $t^*(\underline{p})$ defined by Equation (I)

- $p_t^* = \overline{p}$ otherwise

- $a_t^*(x) = 1$ if and only if $t \in \{t_0, nt^*(\underline{p})\}_{n\in\mathbb{N}}$

Since $\theta \in \Theta^*$, Theorem 1 implies that $0 \leq (\rho + r)\Delta_l = \rho\overline{p} - x^l - (\rho + r)\underline{p}$. So, we can apply Lemma 1 to transform $(\mathbf{p}^*, \mathbf{a}^*)$ into $(\tilde{\mathbf{p}}^*, \tilde{\mathbf{a}}^*) \in \mathcal{L}$ which has the properties:

- $\tilde{a}_t^*(x^l) = 1$ for all $t \geq 0$

- $\tilde{a}_t^*(x^h) = a_t(x^h)$

- $\tilde{p}_t^* = e^{-(\rho+r)(t_0-t)}\underline{p} + (1 - e^{-(\rho+r)(t_0-t)})\frac{(\rho\overline{p}-x^l)}{\rho+r}$ for $t < t_0$

- $\tilde{p}_t^* = e^{-(\rho+r)(t^*(\underline{p})\lceil\frac{t}{t^*(\underline{p})}\rceil - t)}\underline{p} + (1 - e^{-(\rho+r)(T^*\lceil\frac{t}{t^*(\underline{p})}\rceil - t})\frac{(\rho\overline{p}-x^l)}{\rho+r}$ for $t > t_0$ and $t \notin \{t_0 + nt^*(\underline{p})\}_{n\in\mathbb{N}}$

where the last two lines translate the last requirement of an element in set $\mathcal{L}$ to the policy $(\mathbf{p}^*, \mathbf{a}^*)$. $(\mathbf{p}^*, \mathbf{a}^*)$ is an optimal policy for problem $\mathcal{P}^h$ and the policy in the statement of the theorem. Lemma 1 further implies that

$$V^* = \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{L}} V(\mathbf{p}, \mathbf{a}) = V_h^*$$

which concludes the proof. $\square$

ONLINE APPENDIX

F.1 Proofs of Lemmas in Appendix F

Proof of Lemma E.1: Observe that for $s \in [T^h(t), t]$, μ_s^h faces an outflow rate of $(\rho+\lambda)\mu_s^h$ and an inflow rate of 1, where ρ is the risk of detection and λ the risk of transition to the low state. That is

$$\frac{\partial \mu_s^h}{\partial s} = 1 - (\rho + \lambda)\mu_s^h.$$

Solving this simple differential equation with the initial condition $\mu_0^h = 0$ leads to the result. $\square$

Proof of Lemma E.3: First, observe that $w_h(0) = 0$. Differentiate $w_h(t) - e^{-(\rho+r)t}\underline{p}$ with respect to t to get:

$$\frac{\partial \left(w_h(t) - e^{-(\rho+r)t}\underline{p}\right)}{\partial t} = e^{-(\rho+r)t}(x^h e^{-\lambda t} - (\rho\overline{p} - x^l - (\rho + r)\underline{p}))$$

Since $\theta \in \Theta^*$, $x^h > \rho\overline{p} - x^l - (\rho + r)\underline{p}$, and so $w_h(t) - e^{-(\rho+r)t}\underline{p}$ is first strictly increasing and then decreasing, converging to $\frac{x^h}{\rho+r+\lambda} - \frac{\rho\overline{p}-x^l}{\rho+r}$. Since the expression is 0 at $t = 0$, this implies that there is a unique strictly positive solution to $w_h(t) - e^{-(\rho+r)t}\underline{p} = -p$ for any $p \in [\underline{p}, \frac{\rho\overline{p}-x^l}{\rho+r} - \frac{x^h}{\rho+r+\lambda}]$ and it is increasing in p. $\square$

Proof of Lemma E.4: Let $f = \rho + r + \lambda$. Plugging definitions, we get:

$$v^0(t) + e^{-rt}\frac{v^0(t^*(\underline{p}))}{1 - e^{-rt^*(\underline{p})}} = \frac{1}{\rho + \lambda}\left(\frac{1 - e^{-ft}}{f} + \frac{1 - e^{-ft^*(\underline{p})}}{f(1 - e^{-rt^*(\underline{p})})}\right)$$

Differentiate the term in parentheses on the RHS with respect to t to get:

$$\begin{aligned}
\frac{\partial}{\partial t}\frac{1 - e^{-ft}}{f} + \frac{1 - e^{-ft^*(\underline{p})}}{f(1 - e^{-rt^*\underline{p}})} &= e^{-ft} - \frac{r}{f}e^{-rt}\frac{1 - e^{-ft^*(\underline{p})}}{1 - e^{-rt^*(\underline{p})}} \\
&\leq e^{-ft} - \frac{r}{f}e^{-rt}\frac{1 - e^{-ft(\underline{p})}}{1 - e^{-rt(\underline{p})}} \\
&= \frac{1 - e^{-rt}}{1 - e^{-ft}}\left[\frac{fe^{-ft}}{1 - e^{-ft}} - \frac{re^{-rt}}{1 - e^{-rt}}\right] \\
&= \frac{1 - e^{-rt}}{1 - e^{-ft}}\left[\phi(f) - \phi(r)\right]
\end{aligned}$$

where $\phi(a) := \frac{ae^{-at}}{1-e^{-at}}$ and the third line follows for any $t \geq t^*(\underline{p})$ because $\frac{1-e^{-ft}}{1-e^{-rt}}$ is decreasing in t. To see that the last term is negative, observe that it is equal to 0 at $t = 0$ and

$$\frac{\partial}{\partial a}\phi(a) = \frac{e^{at}\left(1 - (at + e^{-at})\right)}{(e^{at} - 1)^2}$$
$$\leq 0$$

for any $at \geq 0$ since $z + e^{-z} > 1$ for any $z \geq 0$. This concludes the first part of the lemma.

To see the second part of the lemma, observe that, by Lemma E.3, $t^*(p)$ is increasing in p. Applying the first part of this lemma to then leads to the result. □

F.2 Proof of Theorem A.1

I prove here that simple Poisson randomized policies do not improve over the deterministic optimal policy.

Proof of Theorem A.1: Let $f := \rho + r + \lambda$. Recall that $v(t) = \frac{1-e^{-ft}}{f}$ and $w_h(t) = x^h \frac{1-e^{-ft}}{f} - \frac{\rho\overline{p}}{\rho+r}(1-e^{-(\rho+r)t})$. Let $\mathbb{E}_\gamma[.]$ denote the expectation with respect to the exponential distribution with rate parameter γ. In a γ-Poisson policy, the recommendation $a_t(x^h) = \mathbf{1}_{p_t=0}$ is incentive compatible if and only if

$$\mathbb{E}_\gamma\left[w_h(t)\right] \leq 0 \tag{$\mathcal{A}^\gamma$}$$

Recalling the linear relationship between $\mathbf{V}$ and v from Section 3.4, the theorem will follow if I can show that any γ-Poisson policy satisfying Equation ($\mathcal{A}^\gamma$) is such that

$$\max_{t_0 \geq 0} v(t_0) + e^{-rt_0}\frac{-\mathbb{E}_\gamma\left[v(t)\right]}{\mathbb{E}_\gamma\left[1 - e^{-rt}\right]} < \max_{t_0 \geq 0} v(t_0) + e^{-rt_0}\frac{-v(t^*)}{1 - e^{-rt^*}}$$

where t^* is defined at the unique strictly positive solution to Equation (I) at $p = \underline{p}$ (when $\theta \in \Theta^*$), the right-hand side is (a linear function of) the regulator's value from the optimal deterministic policy, and the left-hand side is the regulator's value from a γ-Poisson policy. Since the choice of t_0 has the same domain in both problems, it is sufficient to show that

$$\frac{\mathbb{E}_\gamma\left[v(t)\right]}{\mathbb{E}_\gamma\left[1 - e^{-rt}\right]} < \frac{v(t^*)}{1 - e^{-rt^*}} \tag{$\mathcal{P}^\gamma$}$$

First, I will manipulate Equation ($\mathcal{A}^\gamma$). From the definition of t^*, we have $\frac{x^h v(t^*)}{1-e^{-(\rho+r)t^*}} = \frac{\rho\overline{p}}{\rho+r}$. Recalling that $w(t) = x^h v(t) - \frac{\rho\overline{p}}{\rho+r}(1 - e^{-(\rho+r)t})$, we have

$$\mathbb{E}_\gamma\left[x^h v(t) - \frac{\rho\overline{p}}{\rho + r}(1 - e^{-(\rho+r)t})\right] \leq 0$$

$$\iff \mathbb{E}_\gamma\left[x^h v(t) + e^{-(\rho+r)t}\frac{x^h v(t^*)}{1-e^{-(\rho+r)t^*}}\right] \le \frac{x^h v(t^*)}{1-e^{-rt^*}} \qquad (\tilde{\mathcal{A}}^\gamma)$$

Plugging in for $v(t)$ and rearranging Equation ($\mathcal{P}^\gamma$), we find

$$\frac{\mathbb{E}_\gamma[v(t)]}{\mathbb{E}_\gamma[1-e^{-rt}]} < \frac{v(t^*)}{1-e^{-rt^*}}$$

$$\iff \mathbb{E}_\gamma\left[1-e^{-ft} + e^{-rt}\frac{1-e^{-ft^*}}{1-e^{-rt^*}}\right] \le \frac{1-e^{-ft^*}}{1-e^{-rt^*}} \qquad (\tilde{\mathcal{P}}^\gamma)$$

After canceling constants, observe that $\tilde{\mathcal{A}}^\gamma$ and $\tilde{\mathcal{P}}^\gamma$ are special cases of the inequality:

$$\mathbb{E}_\gamma\left[(1-e^{-ft})(1-(e^{-ft^*})^a) + (e^{-ft})^a(1-e^{-ft^*})\right] - (1-e^{-ft^*}) \le 0 \qquad (C^\gamma(a))$$

where $a = \frac{r}{\rho+r+\lambda}$ for the regulator and $a = \frac{\rho+r}{\rho+r+\lambda}$ for the agent and I've multiplied both sides by $(1-(e^{-(\rho+r+\lambda)t^*})^a)$. Denote the left-hand side by $h(a;z,\gamma)$. As in Theorem 2, the crucial step is the following:

$$\text{if } h(a;z,\gamma) \le 0 \textit{ at some } \overline{a} \in (0,1), \text{ then } h(a;z,\gamma) < 0 \text{ for each } 0 < a \le \overline{a}. \qquad (C^\gamma)$$

With this the proof will be concluded. Let $z = e^{-ft^*}$. Integrating with respect to t, the inequality becomes

$$(1-z^a)(1-\frac{\gamma}{\gamma+f}) + \frac{\gamma}{\gamma+fa}(1-z) - (1-z) \le 0$$

Rather than show Property (C^γ) for h, I will show it for $\tilde{h} = h(\gamma + (\rho+r+\lambda)a)$, from which the property for h can be recovered (since for $a \in [0,1]$, $sgn(h) = sgn(\tilde{h})$). Computing $\tilde{h}$, we see that:

$$\tilde{h} = (1-z^a)(\gamma+fa) - \gamma\frac{\gamma+fa}{\gamma+f}(1-z^a) + \gamma(1-z) - (1-z)(\gamma+fa)$$

I claim that if $\frac{\partial^2\tilde{h}}{\partial a^2}$ has at most one 0, then Property C^γ will be verified and the proof will be complete. To see this, observe that as $a \downarrow -\infty$, $\tilde{h} \uparrow \infty$. Observe also that $\tilde{h}(0) = \tilde{h}(1) = 0$. To violate the property, there must exist points $a_1 < 0 < a_2 < a_3 < 1$ such that $\tilde{H}(a_1), \tilde{H}(a_2) > 0$, $\tilde{h}(a_3) < 0$, while $\tilde{h}(0) = \tilde{h}(1) = 0$. This requires $\frac{\partial^2\tilde{h}}{\partial\tilde{h}^2}$ to pass through 0 *at least twice*. So, I proceed to show that $\frac{\partial^2\tilde{h}}{\partial a^2}$ has at most one 0.

Twice differentiating $\tilde{h}$ leads to:

$$\frac{\partial^2\tilde{h}}{\partial a^2} = -\left(\frac{f^2}{\gamma+f}\right) z^a ln(z)\left[2 + ln(z)a\right]$$

Since $z > 0$, this has exactly one zero at $a = -\frac{2}{ln(z)}$. So, I conclude that $\frac{\partial^2 \tilde{h}}{\partial a^2}$ crosses 0 at most once so that Property C^γ holds, and the conclusion follows. $\square$

F.3 Proofs for Section 6

Proof of Proposition 3: Suppose $\gamma < \rho$.

Lemma 2 proceeds in exactly the same way. The statement of Lemma 3 must now be altered so that rather than applying a discount of $e^{-(\gamma+r)t}$ the regulator applies a discount of $e^{-\gamma t}$, but is otherwise identical. To see this, fix some $t \geq 0$ and policy $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}^0$. Recall that, by the definition of $\mathcal{M}^0$, there exists a sequence $\mathbf{t}$ such that $a_t(x^h) = 1$ if and only if $t = s_i := \sum_{j \leq i} t_i$ for some j. Then, under the new arrival distribution, integrating and rearranging yields,

$$V(\mathbf{p}, \mathbf{a}) = -\sum_{i=0}^{\infty} \int_0^{\mathbf{t}_i(\mathbf{p},\mathbf{a})} e^{-(\gamma+r)t} \left(\int_t^{\mathbf{t}_i(\mathbf{p},\mathbf{a})} e^{-(\rho+r+\lambda)(s-t)} ds \right) dt$$
$$= -\frac{1}{(\gamma+r)(\rho+\lambda-\gamma)} + \frac{1}{\rho+\lambda-\gamma} \sum_{i=0}^{\infty} e^{-(r+\gamma)s_{i-1}(\mathbf{p},\mathbf{a})} \frac{1 - e^{-(\rho+r+\lambda)t_i(\mathbf{p},\mathbf{a})}}{\rho+r+\lambda}$$

So a version of Lemma 3 holds using the recursive equation

$$\mathbf{V}(p) = \begin{cases} \sup\limits_{t,p'} \quad v(t) + e^{-(r+\gamma)t}\mathbf{V}(p') \\ \text{subject to} \\ w_h(t) - e^{-(\rho+r)t}p' \leq -p \\ p' \in [\underline{p}, \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}] \end{cases}$$

As long as $\gamma < \rho$, Proposition 2 can be derived with exactly the same steps, and the result of Theorem 2 follows. $\square$

Proof of Proposition 4: Let V^{cont} be the regulator's value associated to the policy described in the proposition. To prove the result, expand the set $\mathcal{M}^0$ to include the policies $(\mathbf{p}, \mathbf{a})$ such that there exists $\epsilon > 0$ such that:

- If t is not a limit-point of the set $\{t|a_t(x^h) = 1\}$, then $a_{[t-\epsilon,t+\epsilon]}(x^h) = 0$.
- If t is a limit-point of $\{t|a_t(x^h) = 1\}$, then $t \in I \subset \{t|a_t(x^h) = 1\}$ where I is an interval.
- If t' is a right-endpoint of an interval $[t, t']$ such that $a_{[t,t']}(x^h) = 1$, then $a_{[t',t'+\epsilon]}(x^h) = 0$.

The first requirement says that reporting times must be separated by an ϵ if they are not in an interval. The second requirements say that if t is a limit-point of the set of high type reporting times, then it is an element of an interval. The third says that if t is an endpoint of such an interval, there is a uniform lower bound (across time) for how long the policy must wait until another reporting time.

Denote by $\mathcal{M}^{cont}$ the set of policies that satisfy these requirements. Recall from Lemma 2, that $V^* := \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^0} V(\mathbf{p},\mathbf{a})$, where M^0 was the set of policies that satisfied the first requirement above and had no limit points in reporting time. Since, $M^0 \subset M^{cont}$, then $V^* := \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^{cont}} V(\mathbf{p},\mathbf{a})$. To prove the result, it is then sufficient to show that

$$V^{cont} = \sup_{(\mathbf{p},\mathbf{a})\in\mathcal{M}^{cont}} V(\mathbf{p},\mathbf{a}).$$

Recursive Representation. I first write the regulator's problem recursively in almost the same way as in the main text. The only difference now is that the regulator can also choose an interval of time during which to set $a_t(x^h) = 1$. The regulator's problem is written recursively:

- A *decision node* of the regulator is any reporting time of the high type *that is not an interior point of an interval* of reporting times
- The *choice* of the regulator is now either
  - *next reporting time* and *penalty at next reporting time* OR
  - *length of an interval* on which to continuously induce reporting by high types, as well as the penalty offered at the end of this interval

  Let $d = 0$ indicate the regulator is choosing the former and $d = 1$ the latter.
- The *state* of the regulator is the reporting penalty that he must offer to the agent immediately
- The *constraint* of the regulator is
  - if $d = 0$, the one-shot incentive compatibility condition we have seen
  - if $d = 1$,
    - penalty at the end of the interval > penalty at the beginning i.e. the state
    - a maximum length of the interval, to be defined below.

In case $d = 1$, the penalty must be larger to continuously induce the high type agent to report. There is a maximum length of the interval, as a function of the initial and final penalty of the interval, because the penalty p_t must increase at a minimum speed to ensure reporting by the high type agent.

Let $t^I(p',p)$ denote the the maximum length of the interval when the interval starts with $p_t = p$ and ends with $p_{t+t^I(p,p')} = p'$.

Let $P(p) = [\underline{p}, \min\{p, \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}\}]$. The recursive version of the regulator's problem can now be stated as follows:

$$\mathbf{V}_\gamma(p) = \begin{cases} \sup\limits_{t,p',d} \quad \mathbf{1}_{d=0}\left(\int\limits_0^t e^{-(\gamma+r)s}\left(\int\limits_s^t e^{-(\rho+r+\lambda)q}dq\right)ds\right) + e^{-rt}\mathbf{V}_\gamma(p') \\ \text{subject to} \\ \text{if } d=0, w_h(t) - e^{-(\rho+r)t}p' \leq -p \\ \text{if } d=1, 0 \leq t \leq t^I(p',p) \\ p' \in P(\overline{p}) \end{cases} \tag{4}$$

The requirement that $0 \leq t$ when $d=1$ guarantees that $p \leq p'$. Similarly to Lemma 3, if $V^\gamma(p)$ solves this equation and the policy function $t(p)$ is such that $\inf\limits_{p\in P(\overline{p})} t(p) > 0$, then

$$V_\gamma^* = \max_{t_0,p_0} \int\limits_0^{t_0} e^{-(\gamma+r)s}\left(\int\limits_s^{t_0} e^{-(\rho+r+\lambda)q}dq\right)ds + e^{-rt_0}V_\gamma(p_0).$$

Notice that if $d=1$, it is always optimal to set t as large as possible. Given this, and after integrating the objective, we find:

$$\mathbf{V}_\gamma(p) = \begin{cases} \sup\limits_{t,p',d} \quad \mathbf{1}_{d=0}\left(\frac{-(1-e^{-(\gamma+r)t})}{(\gamma+r)(\rho+\lambda-\gamma)}\right) + \frac{1-e^{-(\rho+\lambda+r)t}}{(\rho+\lambda-\gamma)(\rho+\lambda+r)} + e^{-rt}\mathbf{V}_\gamma(p') \\ \text{subject to} \\ \text{if } d=0, w_h(t) - e^{-(\rho+r)t}p' \leq -p \\ \text{if } d=1, 0 \leq t = t^I(p',p) \\ p' \in P(\overline{p}) \end{cases} \tag{$R^{\gamma,*}$}$$

Computing t^I In this step I derive the maximum length of an interval that the regulator can continuously induce reporting by high types, given a penalty p that the regulator must deliver at the beginning of the interval, i.e. the state, as well as the penalty chosen for the end of the interval, p'.

To compute the maximum length of this interval, it is sufficient to compute the path of penalties with starting point p and ending point p' such that the agent in the high state is exactly indifferent between reporting at any point on the interval. Fix some time t and suppose that the regulator wants to ensure that the agent in the high state is indifferent over all reporting times on $[t, t+s]$ for some $s > 0$.

Let τ^l be the time to transition from the high to the low state. For an agent that arrives to the model at time t_0, let τ_q be the deterministic stopping time that stops with probability 1 at the minimum of t_0+q and τ^l. Then, if the high type agent is indifferent over all stopping times on $[t, t+s]$, then it is sufficient to find a path for $p_{[t,t+s]}$ such that

$$\frac{\partial W(x^h, t, \tau_q)}{\partial q} = 0$$

for all $q \in [0, s]$. We can compute,

$$\begin{aligned}
0 &= \frac{\partial W(x^h, t, \tau_q)}{\partial q} \\
&= \frac{\partial}{\partial q}\left(\int_0^q \lambda e^{-\lambda t}\left[\frac{1-e^{-gt}}{g}(x^h - \rho\overline{p}) - e^{-gq}p_t\right]dt + e^{-\lambda t}\left(\frac{1-e^{-gq}}{g}(x^h - \rho\overline{p}) - e^{-gq}p_q\right)\right) \\
&= \frac{\partial}{\partial q}\left(\frac{x^h - \rho\overline{p}}{f}(1-e^{-fq}) - \lambda\int_0^q e^{-ft}p_t - e^{-fq}p_q\right) \\
&= (x^h - \rho\overline{p})e^{-fq} - \lambda e^{-fq}p_q + fe^{-fq}p_q - e^{-fq}\frac{\partial p_q}{\partial q} \\
&= (x^h - \rho\overline{p})e^{-fq} + ge^{-fq}p_q - e^{-fq}\frac{\partial p_q}{\partial q}
\end{aligned}$$

Solving this equation with an initial condition $p_t = p$ implies that $p_{t+q} = \frac{x^h - \rho\overline{p}}{g}(e^{gq} - 1) + e^{gq}p$.[37] So, after rearranging, we find that

$$t^I(p', p) = \frac{1}{g}ln\left(\frac{p' + \frac{x^h - \rho\overline{p}}{g}}{p + \frac{x^h - \rho\overline{p}}{g}}\right).$$

Guess solution to Equation ($R^{\gamma,*}$). I guess that an optimal policy in Equation ($R^{\gamma,*}$) is

$$\begin{cases}\text{if } p < \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}, & d^*(p) = 1, \quad p'^{,*} = \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}, \quad t^*(p) = t^I(p'^{,*}(p), p) \\ \text{if } p = \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}, & d^*(p) = 0, \quad p'^{,*}(p) = \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}, \quad t^*(p) = \infty\end{cases} \tag{5}$$

Let $V^*_\gamma(p)$ be the value function associated to this policy. The second part of the guess can be immediately verified, since $t^*(p) = \infty$ and $d^*(p) = 0$ is the only feasible policy available

[37] It can be verified that p_q must be differentiable. One can with only the knowledge that $\lambda\int_0^q e^{-ft}p_q dt + e^{-fq}p_q$ is differentiable, which follows immediately from differentiability of $W(x^h, t, \tau_q)$ and $\frac{x^h-\rho\overline{p}}{f}(1-e^{-fq})$, and arrive at the same conclusion.

when $p = \frac{\rho\overline{p}}{\rho+r} - \frac{x^h}{\rho+r+\lambda}$, and the choice of $p^{\prime,*}(p)$ in this case does not enter the regulator's value.

So, I need to verify the first part of the guess.

Verification. To verify the first part of the guess, it is sufficient to consider only deviations with $d = 0$. Any deviation with $d = 1$ but $p' < p^{\prime,*}(p)$, delivers the regulator exactly the same value as the guess, because it generates exactly the same policy, since the regulator returns to the guess after the one-shot deviation.[38]

Let $f = \rho + r + \lambda$ and $g = \rho + r$. The regulator's value under the guessed policy is

$$V^*_\gamma(p) = -e^{-(\gamma+r)(t^I(p,p^{\prime,*}(p)))} \int_0^\infty \left(e^{-(\gamma+r)t} \int_t^\infty e^{-(\rho+\lambda+r)s} ds \right) dt$$

$$V^*_\gamma(p) = -e^{-(\gamma+r)(t^I(p,p^{\prime,*}(p))} \frac{1}{f(\gamma+r)}$$

Then, to verify the guess, I must show that:

$$\frac{-e^{-(\gamma+r)(t^I(p,p^{\prime,*}(p))}}{f(\gamma+r)} = \begin{cases} \sup\limits_{t,p'} & \frac{-(1-e^{-(\gamma+r)t})}{(\gamma+r)(\rho+\lambda-\gamma)} + \frac{1-e^{-(\rho+\lambda+r)t}}{(\rho+\lambda-\gamma)f} - e^{-(r+\gamma)(t+t^I(p',p^{\prime,*}(p'))}\frac{1}{f(\gamma+r)} \\ \text{subject to} & \\ w_h(t) - e^{-(\rho+r)t}p' \leq -p & \\ p' \in P(\overline{p}) & \end{cases}$$

Since $t^I(p', p^{\prime,*}(p')$ is decreasing in p', the optimal choice of p' given t is the maximum p' that satisfies incentive constraints. This implies that, given t, we can recover p',

$$p' = (w_h(t) + p)e^{(\rho+r)t}$$

Now, let $s(t,p) = t^I\left((w_h(t) + p)e^{(\rho+r)t}, p^{\prime,*}((w_h(t) + p)e^{(\rho+r)t})\right)$. Plugging this into our equation, I must show

$$\frac{-e^{-(\gamma+r)s(t,p)}}{f(\gamma+r)} = \begin{cases} \sup\limits_{t,p'} & \frac{-(1-e^{-(\gamma+r)t})}{(\gamma+r)(\rho+\lambda-\gamma)} + \frac{1-e^{-(\rho+\lambda+r)t}}{(\rho+\lambda-\gamma)f} - \frac{1}{f(\gamma+r)}e^{-(r+\gamma)(t+s(t,p))} \\ \text{subject to} & \\ p' \in P(\overline{p}) & \end{cases} \tag{$R_0^{\gamma,*}$}$$

Denote by $v^\gamma(t)$ the objective on the right-hand side. Since the left-hand side is the objective

[38] To see this, observe that for any $p' \in [p, p'']$, $t^I(p,p'') = t^I(p,p') + t^I(p',p'')$.

of the right-hand side evaluated at $t = 0$, the verification will be complete if we can show that the derivative of the objective on the right-hand side is negative everywhere.

Plugging in the definition of $p^{\prime,*}$,

$$s(t,p) = \frac{1}{g} ln \left(\frac{x^h(\frac{1}{g} - \frac{1}{f})}{e^{gt}\left(\frac{1-e^{-ft}}{f}x^h - \frac{\rho\overline{p}}{g}\left(1 - e^{-gt}\right) + p\right) + \frac{x^h - \rho\overline{p}}{g}} \right)$$

$$= \frac{1}{g} ln \left(\frac{x^h(\frac{1}{g} - \frac{1}{f})}{e^{gt}\left(\frac{1-e^{-ft}}{f}x^h - \frac{\rho\overline{p}}{g} + p\right) + \frac{x^h}{g}} \right)$$

Then,

$$v^\gamma(t) = \frac{-(1 - e^{-(\gamma+r)t})}{(\gamma + r)(\rho + \lambda - \gamma)} + \frac{1 - e^{-ft}}{(\rho + \lambda - \gamma)f} - \frac{1}{f(\gamma + r)} e^{-(r+\gamma)(t+s(t,p))}$$

$$= \frac{-(1 - e^{-(\gamma+r)t})}{(\gamma + r)(\rho + \lambda - \gamma)} + \frac{1 - e^{-ft}}{(\rho + \lambda - \gamma)f} - \frac{e^{-(r+\gamma)t}}{f(\gamma + r)} \left(\frac{e^{gt}\left(\frac{1-e^{-ft}}{f}x^h - \frac{\rho\overline{p}}{g} + p\right) + \frac{x^h}{g}}{x^h(\frac{1}{g} - \frac{1}{f})} \right)^{\frac{\gamma+r}{\rho+r}}$$

$$= \frac{-(1 - e^{-(\gamma+r)t})}{(\gamma + r)(\rho + \lambda - \gamma)} + \frac{1 - e^{-ft}}{(\rho + \lambda - \gamma)f} - \frac{1}{f(\gamma + r)} \left(\frac{\frac{1-e^{-ft}}{f}x^h - \frac{\rho\overline{p}}{g} + p + e^{-gt}\frac{x^h}{g}}{x^h(\frac{1}{g} - \frac{1}{f})} \right)^{\frac{\gamma+r}{\rho+r}}$$

The last term in parentheses in the second line is always positive for feasible p', i.e. $p' \in P(\overline{p})$, and is smaller than 1. In the last line then, the last term in parentheses is always positive for feasible p' and is smaller than $e^{-(\rho+r)t}$. Differentiating $v^\gamma(t)$, we find:

$$\frac{\partial v^\gamma(t)}{\partial t} = \frac{1}{(\rho + \lambda - \gamma)} \left(e^{-ft} - e^{-(\gamma+r)t}\right) - \frac{e^{-ft} - e^{-gt}}{\lambda} \left(\frac{p + \frac{x^h e^{-gt}}{g} + \frac{1-e^{-ft}}{f}x^h - \frac{\rho\overline{p}}{g}}{\frac{x^h\gamma}{fg}} \right)^{\frac{\gamma-\rho}{\rho+r}}$$

$$\leq \frac{1}{\rho + \lambda - \gamma} \left(e^{-ft} - e^{-(\gamma+r)t}\right) - \frac{1}{\lambda} \left(e^{-ft} - e^{-gt}\right) e^{-(\gamma-\rho)t}$$

where the inequality follows since, as noted above, the last term in parentheses in the first line is positive and strictly smaller than e^{-gt}. Rearranging, we find

$$\frac{\partial v^\gamma(t)}{\partial t} \leq \frac{1}{\rho + \lambda - \gamma} \left(e^{-ft} - e^{-(\gamma+r)t}\right) - \frac{1}{\lambda} \left(e^{-ft} - e^{-gt}\right) e^{-(\gamma-\rho)t}$$

$$= e^{-(\gamma+r)t} \left(\frac{1}{\lambda} - \frac{1}{\rho + \lambda - \gamma} \right) + e^{-ft} \left(\frac{1}{\rho + \lambda - \gamma} - \frac{e^{-(\gamma-\rho)t}}{\lambda} \right)$$

$$= e^{-(\gamma+r)t} \left(\frac{1}{\lambda} - \frac{1}{\rho + \lambda - \gamma} \right) + e^{-ft} \left(\frac{1}{\rho + \lambda - \gamma} - \frac{1}{\lambda} \right)$$

$$= \frac{e^{-ft}}{\lambda}\left[\overbrace{\frac{\lambda(1-e^{-(\gamma-\rho)t})+(\rho-\lambda)(e^{(\rho+\lambda-\gamma)t}-e^{(\rho-\gamma)t})}{\rho+\lambda-\gamma}}^{c(t):=}\right]$$

The term $c(t)$ is 0 at $t=0$ and we can differentiate to find,

$$\begin{aligned}\frac{\partial c(t)}{\partial t} &= \frac{\lambda(\gamma-\rho)e^{-(\gamma-\rho)t}+(\rho-\gamma)^2e^{-(\gamma-\rho)t}(e^{\lambda t}-1)+\lambda(\rho-\gamma)e^{(\lambda+\rho-\gamma)t}}{\rho+\lambda-\gamma}\\ &= \frac{(\gamma-\rho)e^{-(\gamma-\rho)t}\left[\lambda+(\gamma-\rho)[e^{\lambda t}-1]-e^{\lambda t}\lambda\right]}{\rho+\lambda-\gamma}\\ &= (\gamma-\rho)e^{-(\gamma-\rho)t}(1-e^{\lambda t})\\ &\leq 0\end{aligned}$$

where the last line is a result of the fact that $\gamma > \rho$ and $1-e^{\lambda t} \leq 0$. This implies then that

$$\frac{\partial v^{\gamma}(t)}{\partial t} \leq 0$$

which subsequently implies that $V^*_\gamma(p)$ solves Equation ($R_0^{\gamma,*}$). This completes the verification and so $V^*_\gamma(p)$ solves Equation ($R^{\gamma,*}$).

Conclusion. Given our solution to the recursive representation $V^*_\gamma(p)$, and observing that $V^*_\gamma(p)$ is maximized at $p=\underline{p}$, we know that

$$V^*_\gamma = \max_{t_0}\int_0^{t_0} e^{-(\gamma+r)s}\left(\int_s^{t_0} e^{-(\rho+r+\lambda)q}dq\right)ds + e^{-rt_0}V^*_\gamma(\underline{p}).$$

Applying the policy functions associated with V^*_γ that were verified above then generates the optimal path described in the theorem. $\square$