Dynamic Amnesty Programs

Sam Kapon[*]

23rd November 2022

Abstract

A regulator faces a stream of agents engaged in crimes with stochastic returns. The regulator designs an amnesty program, committing to a time-path of punishments for criminals who report their crimes. In an optimal program, time-variation in the returns from crime can generate time-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, after which the cycle resets. Agents engaged in high return crime report at the end of each cycle, while agents engaged in low return crime report always.

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 program that offers lenient treatment to self-reporting cartel members, 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 (Figes, 1990). Australia's gun buy-back of 1997 collected more than 650,000 weapons (Leigh and Neill, 2010). The Chieu

---

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

[2]https://www.justice.gov/atr/leniency-program

Hoi program offered amnesty to defectors during the Vietnam war, enticing over 100,000 (Wosepkat, 1971).[3]

An extensive theoretical literature has investigated the use of self-reporting programs in one-shot regulation.[4] Less attention has been paid to the intertemporal 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 (but less frequent) policy has historically been applied to desertion in French militaries.[5] The Brazilian gun buyback program has been run four times since 2013.[6] The U.S. has operated a number of tax-related self-reporting programs, often on a repeated, time-limited, basis.[7] 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*? I study a 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 punishments for agents who self-report before they are detected that applies uniformly to all agents. The range of possible punishment is bounded and agents 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.

A key feature of the environment is that returns from crime change from high to low over time and this can lead to optimal self-reporting terms that change over time. To see why, compare two extreme policies. The first is a *static policy*, offering the same terms for self-reporting at all times. This policy lets the agent benefit both 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 returns 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

---

[3]The exact number is not entirely clear, with possibly inflated statistics.

[4]See Andreoni (1991), Malik (1993) and Kaplow and Shavell (1994) for early contributions.

[5]See Wright (2012) for desertion in the Red Army and Forrest (1989) for desertion in French militaries.

[6]See Macinko et al. (2007).

[7]Luitel and Sobel (2007), OECD (2015)

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. This trade-off is explored in the remainder of the paper.

The basic features of the model are motivated by the following observations. First, returns from crime often accrue slowly over time. For instance, deserters value each moment they avoid military duties, 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, 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 permission, a change in tax payment can spark IRS scrutiny, and in general the cessation of crime may spark increased scrutiny.[8] 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 punishments for any agent who self-reports at a given time, motivating the regulator's problem as a choice of a time path of punishments that applies uniformly to all agents.[9]

The main result (Theorem 1) characterizes an optimal amnesty policy and shows that it takes a *cyclical* form; the punishment upon self-reporting declines over time until it hits the minimum feasible punishment, after which it jumps upward, and this process repeats itself. The declining path induces agents with low returns from crime to immediately report, and ensures they are indifferent between immediately reporting and delaying reporting until the end of the cycle. Agents with high returns from crime report at the end of each cycle, when the self-reporting policy is at its most generous (i.e. the punishment upon self-reporting is at its minimum). The frequency of cycles increases with the risk of detection, the maximum punishment, and the rate of transition from high to low return crime.

To develop intuition for why this is optimal, consider a simplified class of policies—at any time, the regulator can only offer the minimum or the maximum punishment—and consider a simplified problem—the regulator only loses value from the operation of high return crime (i.e. does not care about low return crime). In this case, the regulator's only choice is the frequency with which to offer the minimum punishment to self-reporters; *too*

---

[8]See for instance Baer (2008) on such intertemporal linkages in corporate fraud.

[9]In Section VI, the features of the model are further discussed and interpreted.

frequently and only agents with low returns from crime will report, because agents with high returns know that once they transition to the low return state they will only have to wait a short time before an opportunity to report for the minimum punishment arises, but *not frequently enough* and the mass of criminals with high returns from crime will grow large in the intervening time. As a result, the regulator offers the minimum punishment as frequently as possible without violating the incentives of agents with high returns from crime to report at such times.

In this simplified problem but without the constraint to use only the minimum or maximum punishment, it turns out that it is optimal not to use punishments strictly between the minimum and maximum. The value of using such interior punishments in the simplified problem would be to adjust the delay between times at which high types report. For instance, the regulator could offer rapidly increasing punishments for reporting, followed by a long period in which the maximum punishment is imposed for reporting, which then repeats. This can induce reporting by high types at high frequency (i.e. quickly after they arrive) while punishments initially increase, but this will be followed by a long period of time with no reporting. Such a policy trades long-term losses—a long future period with no reporting—for short-term gains—high frequency reporting for a short period of time. It turns out this trade is never beneficial for the regulator. The basic reason is a *backloading* motive on the part of the regulator. Agents face a risk of detection, which acts like additional time discounting; in contrast, because the regulator is constantly facing newly arriving agents, she does not face any such additional discounting. This effectively makes agents less patient than the regulator so that trading long-term losses for short-term gains is never optimal.[10]

To solve the general problem where the regulator loses value from the operation of both high *and* low return crime, I show that the solution to the simplified problem can be transformed into one in which low types report immediately and high type behavior is unchanged, by setting the punishment equal to the low type's value of mimicking the behavior of high types; this generates the declining punishments along each cycle's path in the optimal policy.

In Section V, 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 forces highlighted in the model could plausibly have contributed to the observed use of dynamic policies. 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

---

[10]The backloading motive is further clarified in Online Appendix G, where I extend the model to allow agents to arrive at a time-inhomogeneous rate, and show how this can reverse the backloading motive.

programs in tax collection.[11]

After reviewing the literature, I introduce the model in Section I and analyze it in Sections II and III. In Section IV, I present comparative statics and discuss how investments in features of the environment can complement dynamic amnesty. In Section V, I present applications. Assumptions and modeling choices are discussed in Section VI. I conclude in Section VII.

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 plausible contributor to 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 driven by other forces, 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 design problem, in which a stream of agents arrives over time with stochastic values for an interaction with a regulator and can choose to irreversibly end their interaction at some cost (i.e. amnesty). The most closely related work is in the determination of an optimal price path for a durable goods monopolist facing a stream of buyers arriving over time 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, the intertemporal price discrimination literature in economics and operations research, and the dynamic mechanism design literature.

The early work of Andreoni (1991), Malik (1993), and Kaplow and Shavell (1994) 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 intertemporal 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 studies the role of 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 with ex-ante different but static types, which I discuss in more detail in the tax amnesty applications section. Bagirgan (2020) shows that repeated tax amnesties can arise without commitment in a model with government reputation formation. Wang, Sun

---

[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 alone. In such cases, amnesty may backfire and lead to a deterioration of this moral obligation.

and de Véricourt (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 hazard and the paper focuses on the interaction between the inspection policy and penalties. When the rate of inspection (like the risk of detection in this paper) cannot be chosen and is constant over time, static self-reporting programs are optimal, unlike in this paper. I focus on the role that dynamic self-reporting programs play absent 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 intertemporal price discrimination by a durable goods monopolist, such as Conlisk, Gerstner and Sobel (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. The underlying intuition for why optimal policies may fluctuate is closely related. A fundamental difference between our papers is the limited punishments and rewards 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 non-trivial but precludes the use of techniques along the lines of Garrett (2016). Solving the model in this paper therefore requires a different approach that explicitly incorporates these limited punishments. I provide a more detailed discussion of this difference in Section VI. Preferences in this paper also differ from those of the durable good monopoly setting, and this leads to different intuition underlying the optimal policy. In particular, the regulator is concerned only with *fast* self-reporting by the agents. In the durable goods monopoly setting, it would be as if the monopolist cared only that buyers purchased quickly, but not about the price.

This paper is also more distantly related to a variety of papers in dynamic contracting and mechanism design. Krasikov, Lamba and Mettral (2020) show that unequal discounting in a principal-agent setting with repeated trade and changing values leads to optimal contracts with distortions that cycle over time.[14] Garrett (2017) studies a setting with repeated sales in which the buyer can arrive at a privately known random time and the seller can commit to a mechanism for trade. The buyer's value can change over time and the paper shows how

---

[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 crimes.

[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]A driving force is the tension that the more patient principal faces between backloading the agent's payoff to provide incentives and front-loading the agent's payoff to take advantage of the differential discounting.

optimal mechanisms punish buyers for late arrivals. Krasikov and Lamba (2020) studies a dynamic contracting setting with persistent private information and financial constraints. Similar to this paper, backloading is limited by financial constraints. Grillo and Ortner (2018) study a related environment and show that limited liability has important implications for the types of distortions from efficiency that arise in the optimal contract.

I The Model

A stream of criminal agents (he) decide whether to continue to operate or self-report. A regulator (she) commits to a policy that is relevant for the agents' decisions. Calendar time is continuous, $t \in \mathbb{R}_+$.

A. The Agents

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 time-invariant continuous-time Markov chain, denoted x_t, 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 constant rate λ. Upon arrival, agents are initialized in state x^h.[16] Index x_t by time since arrival so that x_0 is the initial state of an agent upon arrival. I will refer to flow gain x^h (x^l) as the high (low) state and agents with flow gain x^h (x^l) at some time as high (low) types at that time.

Choice and Detection. An agent arriving at time t_0 chooses a (possibly ∞-valued) stopping time with respect to the filtration generated by $(x_t)_{t \geq 0}$, denoted τ — the calendar time at which the agent stops is $t_0 + \tau$. 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. Upon stopping at calendar time t, the agent pays a penalty p_t, and his flow gains stop accruing. A deterministic path of penalties is called a *penalty policy* and is denoted $\mathbf{p} = (p_t)_{t \geq 0}$. An agent is randomly detected by the regulator after he arrives at a constant rate, ρ, independent of $(x_t)_{t \geq 0}$. If the agent is detected, he pays the maximum penalty $\overline{p}$ and his flow gains stop accruing.

---

[15] As in Garrett (2016), I take the "intuitive" approach to aggregating random variables over agents.

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

Payoffs. Agents discount the future at rate r. To compute an agent's value from a stopping time, first define

$$w(x,t) \equiv \mathbb{E}\left[\overbrace{\int_0^t e^{-(\rho+r)s}x_s ds}^{(a)} - \overbrace{(1 - e^{-(\rho+r)t})\frac{\rho}{\rho+r}\overline{p}}^{(b)}\;\middle|\; x_0 = x\right]$$

where the expectation is taken with respect to the distribution of x_t. This is the value of an agent in state x who delays reporting for a deterministic length of time t, excluding the reporting penalty after delay t. Note that λ does not explicitly appear, but rather controls the evolution of x_t. The term (a) is the accrued flow gain discounted by both time discounting and the risk of detection. The term (b) is the agent's expected loss from paying the penalty if exogenously detected before choosing to stop.

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 then,

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

where the expectation is taken with respect to the distribution of x_t.[17] The term (c) is the agent's expected loss from the penalty he pays when stopping before being detected by the regulator. An agent arriving at t_0 in state x solves the problem,

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

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 suppress the dependence of $W(x,t_0,\tau,\mathbf{p})$ on $\mathbf{p}$ and write $W(x,t_0,\tau)$.

B.   The Regulator

Policies. The regulator commits at time 0 to a choice of the penalty policy, $\mathbf{p} = (p_t)_{t\geq 0}$, with $p_t \in [\underline{p},\overline{p}]$, as well as an *obedient recommendation policy* as a function of calendar time and the state, $\mathbf{a}\colon \mathbb{R}_+ \times \{x^l, x^h\} \to \{0,1\}$, such that $\inf_{t\geq t_0}\{t - t_0 | \mathbf{a}(t, x_{t-t_0}) = 1\}$ is an optimal stopping time for an agent arriving at time t_0—this stopping time is called the *stopping*

---

[17]Recall that in the baseline model, all agents arrive in state x^h, but it is still possible to define the value of an agent arriving in state x^l.

time induced by **a** for an agent arriving at t_0.[18],[19] I will use the notation $a_t(x)$ to denote $\mathbf{a}(t, x)$. Let $\mathcal{M}$ be the set of policies $(\mathbf{p}, \mathbf{a})$, consisting of a penalty policy $\mathbf{p}$ and obedient recommendation policies $\mathbf{a}$.

Payoffs. The stopping times induced by an obedient recommendation policy induce a pair of paths, $(\mu_t^h)_{t \geq 0}$ and $(\mu_t^l)_{t \geq 0}$, equal to the mass of agents in states x^h and x^l at each time t, respectively. The regulator's payoff from a policy $(\mathbf{p}, \mathbf{a})$ is

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

where $\alpha_l \geq 0$. Note that time discounting is at the same rate as the agent, r. The regulator solves,

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

A policy that achieves V^* is called *optimal*.[20]

II Preliminary Results

To arrive at an optimal policy I make three preliminary steps: (i) characterize the optimal static policy, (ii) show that under some conditions it is without loss for the regulator's value to study policies in which low types always report, and (iii) characterize when dynamic policies improve over static policies.

Static Policies. A *static policy* is a pair $(\mathbf{p}, \mathbf{a})$ for which the penalty, $\mathbf{p}$, is constant over time. Denote by $\mathbf{p}^v = (p_t^v)_{t \geq 0}$ the penalty policy in which $p_t^v = v$ for all t. A *dynamic policy* is any policy which is not a static policy. An agent's decision problem under a static policy takes a simple form, since the policy does not exhibit any inter-temporal variation.

Proposition 1. *An optimal static policy is $(\underline{p}^{\underline{p}}, \mathbf{a})$ where $\mathbf{a}$ is constant across time.*

---

[18]Observe that the regulator is restricted to deterministic penalty policies—in Online Appendix F, I provide an extension of the main result to a restricted class of Poisson random policies. Note also I have placed no restrictions on the lower bound of p_t, $\underline{p}$, which can, for instance, be negative and represent a reward as in the case of gun buybacks. Nevertheless, because p_t will not directly affect the regulator's value, the most natural cases involve $\underline{p} \geq 0$.

[19]Both $\mathbf{p}$ and $\mathbf{a}$ are measurable with respect to time, t.

[20]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.

In words, an optimal static policy offers the minimium penalty possible and recommends agents in a given state to take the same action at all times. The proof is given in Appendix D. Intuitively, lowering the reporting penalty in a static policy strengthens the incentive to report early, and since the regulator does not directly value the penalty, it is optimal to choose the minimum penalty.

Low Type Screening. I show, under some conditions, that the regulator can transform any policy into one with reporting by low types at all times, without distorting the reporting incentives of high types. Let

$$\tau^h \equiv \inf_{s \geq t}\{s - t | a_s(x^h) = 1\},$$

i.e. the stopping time for an agent arriving at t that follows the recommendation for high types. Let Δ_l be the difference in payoffs for low types between stopping immediately and never, under $\mathbf{p}^{\underline{p}}$, the most generous policy possible. Formally,

$$\Delta_l \equiv W(x^l, t_0, \tau^0, \mathbf{p}^{\underline{p}}) - W(x^l, t_0, \tau^\infty, \mathbf{p}^{\underline{p}})$$

where τ^0 and τ^∞ denote stopping immediately and never, respectively. Finally, let $\mathcal{L} \subset \mathcal{M}$ be the set of policies such that low types are recommended to stop immediately—that is, $a_t(x^l) = 1$ for all t—and an agent arriving in state x^l is indifferent between τ^0 and τ^h i.e. stopping immediately and following the high type recommendation.[21]

The *Low Type Screening Lemma*, formalized and proved in Appendix A as Lemma A.2, states that if Δ_l is non-negative, any policy can be transformed without loss to the regulator's value into a policy in $\mathcal{L}$. The intuition can be illustrated by describing the transformation of the following one-time policy: p_t is set to the minimum, $\underline{p}$, at some $t = T > 0$, and otherwise is set to the maximum, $\overline{p}$, and both high and low types are recommended to report only at T. To transform this policy into a new one in which low types always report, replace p_t at any t smaller than T with the value to the low type of waiting until T to report. Replace p_t at any t larger than T with the value to the low type of never reporting. The positiveness of Δ_l guarantees that this transformation is feasible. The new policy recommends high types to report at T only, as in the original policy, but low types to report everywhere. For a low type, no stopping time strictly improves over her optimal value in the original policy, while stopping immediately delivers her optimal value in the original policy, so stopping immediately is an optimal policy for the low type. The recommendation for high types remains obedient, even though the new policy may have strictly lower penalty (except at T), because (i) the value upon transitioning to the low state is unchanged and (ii) the lowered

---

[21]Formally, $W(x^l, t, \tau^0, \mathbf{p}) = W(x^l, t, \tau^h, \mathbf{p})$.

penalty is never attractive to the high types, since it replicates the low type's value from following the high type's recommendation in the original policy, which is no larger than the high type's value from following her own recommendation in the new policy.

To see the usefulness of the lemma, note that when Δ_l is negative, low types prefer to never report than to report even for the minimum possible penalty; therefore, no agent ever reports, and so a static policy is optimal (along with every other policy). As a result, either a static policy is optimal or the search for an optimal policy can be restricted without loss of value for the regulator to $\mathcal{L}$. The set $\mathcal{L}$ has useful properties: in any policy in $\mathcal{L}$, low types report everywhere and given the times at which high types report and the penalties at those times, penalties elsewhere are pinned down by the condition that low types are indifferent between immediately reporting and mimicking the behavior of a high type.

The Value of Dynamic Policies. Let θ denote an arbitrary parameterization of the model.[22] Recall that an optimal static policy exists by Proposition 1.

Proposition 2. *The set of parameters θ for which there exists a dynamic policy that strictly improves over an optimal static policy, denoted Θ^*, is non-empty and is characterized by:*

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

Observe that Θ^* contains no parameters for which returns are perfectly persistent, i.e. $\lambda = 0$.[23] The result is proved in Appendix B. The idea is straightforward: if $x^h - x^l$ fails the condition, then one of the following is true: (i) x^l is so high that an agent in state x^l would never report under any policy, (ii) x^h is so low that an agent in state x^h would report immediately under the optimal static policy, or (iii) x^l is low enough that an agent in state x^l reports immediately in the optimal static policy, but x^h is sufficiently high that an agent in state x^h would never report at any time under *any* policy. In any of these cases, a static policy is optimal. Otherwise, the transformed one-time policy—described above under *Low Type Screening*—strictly improves over the optimal static policy; in both the transformed one-time and static policies, low types report everywhere, but high types report only under the transformed one-time policy.

---

[22]That is, $\theta = (\rho, r, \lambda, x^h, x^l, \overline{p}, \underline{p}, \alpha_l)$

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

III   An Optimal Policy

To state the main result, define t^* implicitly as the unique strictly positive value that satisfies

(I)
$$w(x^h, t^*) - e^{-(\rho+r)t^*}\underline{p} = -\underline{p}$$

which exists whenever $\theta \in \Theta^*$.[24] This is a one-shot incentive compatibility condition (at equality) for the high type: the high type should be indifferent between (i) immediate reporting for penalty $\underline{p}$ and (ii) waiting t^* and only then reporting for penalty $\underline{p}$.

Definition 1. *A policy $(\boldsymbol{p}, \boldsymbol{a})$ such that $p_{t_0+t} = p_{t_0+t+s}$ for all $t \geq 0$, some $t_0 \geq 0$ and some $s > 0$, such that $p_{t_0} = \underline{p}$ is called a **minimal delayed cyclical policy with period s**. The times $t_0 + \mathbb{N}s$ are called **reset points**.*

Recall that for any policy in $\mathcal{L}$, an agent arriving in state x^l is indifferent between τ^0 and τ^h, and both are optimal.

Theorem 1. *If $\theta \in \Theta^*$, there exist a minimal delayed cyclical policy with period t^* defined by equation (I), denoted by $(\boldsymbol{p}^*, \boldsymbol{a}^*)$, which is (i) optimal, (ii) a member of $\mathcal{L}$, and (iii) has the property*

$$a_t^*(x) = \mathbb{1}_{x=x^l} + \mathbb{1}_{x=x^h}\mathbb{1}_{t \text{ is a reset point}}.$$

If $\theta \notin \Theta^*$, a static policy is optimal.
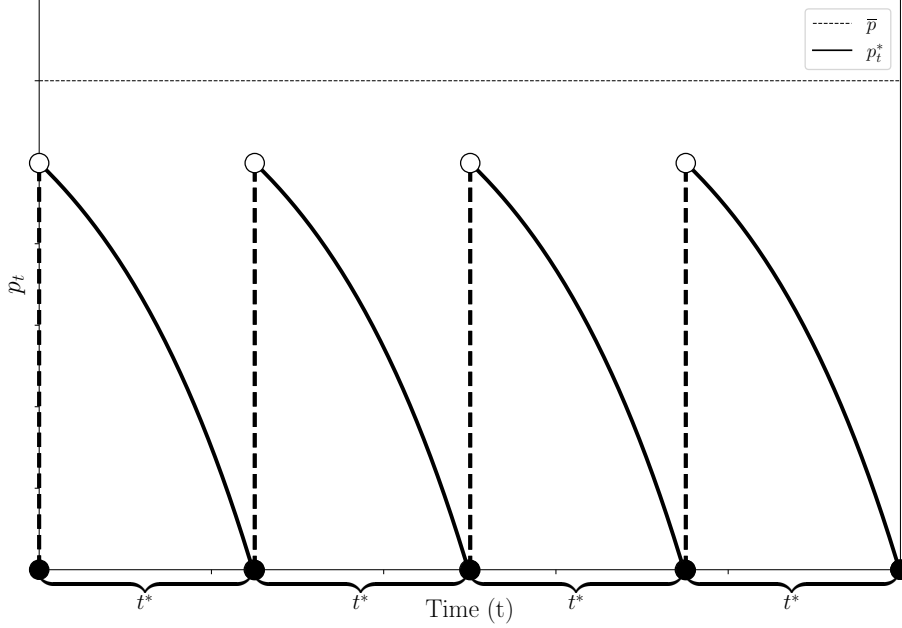
---

[24]See Lemma C.4.

Figure 1: An optimal policy.

The proof of the theorem is given in Appendix C which first gathers a number of preparatory results. The path of $\mathbf{p}^*$ guarantees that a low type is indifferent between immediately reporting anywhere on this path and waiting to report until the next time at which high types report. The delay t_0 is an initial timing choice of the regulator, who is initially unburdened by promise-keeping. The optimal policy beyond t_0 is displayed in Figure 1.

To arrive at this policy, I use the *Low Type Screening Lemma* to solve the regulator's problem in two steps: step (1), optimize in the special case $\alpha_l = 0$ and, step (2), apply the lemma to transform the policy from step (1) into a policy in which low types always report.

Observe that in step (1), p_t can be set to $\bar{p}$ whenever $a_t(x^h) = 0$ i.e. high types are not recommended to report—this strengthens reporting incentives for high types elsewhere and, while it may deter reporting by low types, this does not cost the regulator since $\alpha_l = 0$. As a result, in the special case of $\alpha_l = 0$, the policy that agrees with the policy in Theorem 1 except that between reset points, $p_t = \bar{p}$, is also optimal. The intuition for why this is optimal can be broken into two components: (i) conditional on offering $p_t = \underline{p}$ whenever $a_t(x^h) = 1$, why is the policy optimal? (ii) why is it optimal to offer $p_t = \underline{p}$ whenever $a_t(x^h) = 1$? I describe intuition for each of these.

For (i), since $p_t = \bar{p}$ when high types are not recommended to report and $p_t = \underline{p}$ otherwise, the only choice left for the regulator is how frequently to offer $\underline{p}$, with the constraint that high types find it optimal to report at such times. To satisfy the constraint, the offers of $p_t = \underline{p}$ must be sufficiently separated in time, otherwise high types prefer to delay reporting

13

until they become low types—they would only need to incur the risk of detection while in the low state for a short period of time before an opportunity for $p_t = \underline{p}$ arrives. But, conditional on satisfying the constraint, the regulator prefers the delay between offers of $p_t = \underline{p}$ to be as short as possible—the shorter the delay, the quicker criminals stop committing crime after arriving. As a result, the solution to the first step, constrained to policies that take value $p_t \in \{\overline{p}, \underline{p}\}$, is to offer the minimum punishment as frequently as possible without violating the incentive constraints of high types—this frequency is t^* as defined by eq. (I).

Point (ii) is more subtle, and it is useful to briefly describe the approach to the proof. The proof of the theorem proceeds by first writing the regulator's problem recursively, with *decision nodes* as the times at which high types report, *state* as the penalty that must be offered immediately, *choices* as (a) the delay until the next decision node and (b) the penalty at that time (which becomes the state at the next decision node) and finally a *constraint* that agents must prefer to report immediately rather than wait until the next decision node and report then. There are many different choices the regulator can make, given the state, that would satisfy the constraint at equality. The constraint forces the regulator to make the following trade-off.[25] Increasing the penalty at the next node introduces slack into the constraint, so the regulator can then decrease delay until the constraint becomes tight, which is valuable for the regulator who prefers crime stops as quickly as possible; but, at the next decision node, the state (i.e., penalty the regulator must incentivize agents to accept) is now higher, and this shrinks the set of choices available because the constraint at that node becomes harder to satisfy (it is harder to incentivize agents to report for a higher penalty). To satisfy the constraint at the next decision node, the regulator may again opt to offer a short delay to the decision node after that, but this requires an even higher subsequent penalty, further constraining her available policies. A sequence of short delays must eventually end with a long delay—short delays require increasing penalties and since penalties are bounded above, this cannot continue indefinitely and eventually the only way to satisfy the constraint is to use a long period with the maximum punishment for reporting, which induces no reporting. This highlights the trade-off the regulator faces: using short delays creates short-term gains but comes with long-term losses—eventually, the only way to satisfy the constraint is to commit to a long period of time with no reporting, which is costly both for agents and the regulator.

With the recursive structure in mind, a strong form of point (ii) is: for *any* state, it is optimal to choose the next penalty to be $\underline{p}$, and the shortest possible delay that satisfies the constraint. A *backloading* motive for the regulator is the driving force behind this result. In particular, at each decision node, the agent discounts payoffs at the next decision node

---

[25]For the purposes of intuition, some subtleties are ignored.

not only by time discounting, as does the regulator, but also by the risk of detection, which acts like additional time discounting. The regulator, on the other hand, always faces a new group of agents at the next decision node, and so incurs no such additional discounting. This effectively makes agents less patient than the regulator. Intuitively then, using short delays to and high penalties at the next decision node—which creates short-term gains but long-term losses—is never optimal, and the regulator instead prefers to incentivize reporting with long delays to and low penalties at the next decision node. Combining this with the result that delay longer than the shortest necessary to satisfy the constraint at equality is never optimal, it is optimal in any state to choose the next penalty to be $\underline{p}$, and the shortest possible delay that satisfies the constraint at equality.[26]

So for step (1) ($\alpha_l = 0$), combining points (i) and (ii) shows that a policy identical to that in Theorem 1, except that between reset points $p_t = \overline{p}$, is optimal. For the general case of $\alpha_l \geq 0$, applying the transformation in the *Low Type Screening Lemma* delivers the form of the optimal policy in Theorem 1 (step (2)), with penalties decreasing along each cycle.

IV   Comparative Statics

It is immediate from Theorem 1 that on the interior of Θ^*, 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 return crime (λ). When the regulator can increase ρ or $\overline{p}$, the regulator can also increase the frequency with which high types self-report in the optimal policy. This 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 jointly with amnesty, the comparative statics nevertheless provide useful insight. In some settings, the regulator may only be able to imperfectly affect the risk of detection; in the case of cartels or desertion during war, much of detection comes from third-parties reporting to the regulator. While the regulator can affect the incentives of third-parties to report, this is less tightly controlled than in the case of, for instance, environmental inspections.[27,28] Alternatively, one can view the results as the second step of a simplified two-step policy-making process: choose a uniform level of detection and punishment, then an amnesty policy.

---

[26]The backloading motive is further clarified in Online Appendix G, where I extend the model to allow agents to arrive at a time-inhomogeneous rate.

[27]As in Wang, Sun and de Véricourt (2016).

[28]For instance, a large share of detection of price-setting cartels 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).

V   Applications

In this section, I discuss an application of the model to military desertion and detail a case of amnesties during the Russian Civil War. Afterwards, I discuss the model's implications for tax amnesties and gun buybacks and amnesties.

A.   Desertion

From 1919 to 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 "the first and principal obstacle to the organization of the army of the Kingdom" (Grab, 1995, p. 37).

Desertion amnesties are often offered during the course of war to entice return and have been applied extensively across history.[29] 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, p. 206), "...the most successful means of combating desertion [in the Red Army] were the amnesty weeks." 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, p. 461). 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, p. 48). 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, 1989, p. 213).

The difficulties that a deserter faces can change over time, and this provides the basis for the dynamic returns modeled in this paper. Forrest (1989) provides, among other things, an account of the uncertain conditions of desertion in France in the early 19th century. Under such circumstance, Forrest (1989) (p. 102) remarks, "[I]t is hardly surprising that considerable numbers of deserters changed their minds."

A.1   The Red Army and the Anti-Desertion Commission

In this section, I argue that the model's basic forces can rationalize the use of intermittent amnesty for deserters from the Red Army during the Russian Civil War.

---

[29]In contrast to desertion amnesties offered after a war, as a method of reconciliation and forgiveness.

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 (p. 145), 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.[30] 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 deserters to reenter the military with no repercussions.

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, p. 474), "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, apparently in order to encourage deserters' return. The following is an extract of a newspaper publication described in Wright (2012, p. 154): 'Deserters, townspeople! Today is the last day to appear before the [anti-desertion] commission. Hurry; present yourself today as tomorrow will be too late!'.

Undoubtedly, understanding the overall amnesty-granting decision requires understanding the relationship between the military, its personnel and the population. As discussed in Wright (2012), the Red Army's overall decision to apply amnesty can be seen both as a way of recovering manpower and as a way of striking a balance between repression and restraint in a bid to win the support of the peasantry (which naturally included family members of deserters). Nevertheless, this paper develops a formal model with a force, echoed in qualitative evidence from the period, that drives towards the particular form that amnesty took, as a response to the uncertainty and variation in the life of a deserter. It is instructive to consider other explanations for intermittent amnesties.

Public Variation.   Aside from the idiosyncratic variation in a deserter's plight as described above, some of the most important time-varying factors were the advances of the White

---

[30]For instance, soldiers returned home to sow their fields.

army and the harvest season. Deserters 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 periodic amnesty weeks during the June-October 1919 period in the Karelia region, so more fine-grained events or a more fine-grained theory, such as the one provided in this paper, is necessary to explain the structure of these amnesties.

Discouraging Desertion. A natural guess is that amnesties should be sufficiently infrequent in order to discourage desertion. To think formally through this possibility, consider a version of the model of this paper with $\lambda = 0$ (perfect persistence) but when agents arrive to the model, they make a decision whether to desert (and cannot delay desertion). In this case, *the optimal policy offers no amnesty, fully punishing any self-reporter*. The reason is that under such a policy, an agent who arrives to the model faces a "one-time amnesty": by not deserting, they receive no punishment, but if they desert they never have a chance at amnesty. Since the value to deserting is constant, an agent who chooses to desert under this policy can never be induced to report.

Although the model is simple, it highlights the reason that an explanation based purely on discouraging desertion is insufficient. When a deserter deserts, they have revealed that unless their or the military's situation changes, they would not accept an amnesty since they could have avoided punishment altogether by not deserting in the first place. For discouraging desertion to be an important determinant of the decision to offer intermittent amnesty, it must be coupled with another force such as the one in this paper.[31]

B. Gun Amnesties and Buybacks

Gun amnesties and buybacks can potentially be improved by taking into account the dynamic considerations of illegal gun owners. 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.[32] Buy-back programs go one step further, offering to pay for each firearm surrendered. During the Argentine buyback of 2007-08, the government collected more than 100,000 weapons (Lenis, Ronconi and Schargrodsky, 2010). During the Brazilian

---

[31]An alternative possibility is that public pressure to forgive deserters grows with the mass of deserters and amnesty is then held whenever the mass of deserters is large enough and hence the benefit of amnesty (releasing public pressure) outweighs its cost (encouraging desertion). It is not clear though, given the short time period during which amnesties were held, that public pressure could respond so quickly to make this a reasonable explanation.

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

buy-backs of 2003, 2009 and 2011, the government collected more than 1 million weapons.[33] 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, Ronconi and Schargrodsky (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 operated temporary amnesty programs in 2007, 2013 and 2018. Mexico operates a permanent buyback program through its Secretariat of Defense (SEDENA). Tasmania has operated a permanent amnesty program for years and all of Australia began to do so in 2021. Denmark operates regular, periodic amnesties, which have been run in 2009, 2013 and 2017. The Africa Union (AU) Assembly declared in Decision 645 (XXIX) that each September from 2017 to 2020 would be "Africa Amnesty Month", in which weapons could be turned in without any risk of punishment. 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 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 policy is optimal.[34]

C.   Voluntary Disclosure and Tax Amnesty Programs

In this section, I argue that the forces of the model are present in voluntary disclosure and tax amnesty programs, and that intermittent amnesties, while more likely driven by other considerations—lack of commitment, short-term revenue considerations, political cycles—have benefits when compared to more permanent disclosure programs which have sometimes been implemented.

Tax amnesty is ubiquitous.[35] The use of tax amnesty is controversial, despite its preval-

[33]https://www.gunpolicy.org/firearms/region/brazil

[34]It must be noted that the 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.

[35]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 more than 20 have implemented three or more (Mikesell and Ross, 2012).

ence (Le Borgne and Baer, 2008). While programs that reduce the original tax amount often fail to deliver benefits exceeding costs, 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.

While tax amnesties are often offered to raise revenue in the short-term, they are also used to increase compliance in the medium and long-term, the view taken in the model of this paper.[36] A basic question relevant for examining tax amnesties and voluntarily disclosure programs is, why do people apply? Although direct evidence regarding motivation is not widely available, one source of evidence on this question comes from Ritsema, Thomas and Ferrier (2003), who implemented a survey of participants in the 2003 Arkansas tax amnesty program. Income, ease of evasion and inability to pay were three important determinants in the decision to evade taxes. In any setting in which these factors vary substantially over time, the model can speak to the design of voluntary disclosure programs.

As detailed in OECD (2015), there are many examples of both permanent and temporary but repeated 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 taxpayers who are ex-ante heterogeneous in their value for for avoiding sanctions (there are two types: compliance-prone and evasion-prone) — types are *static*, regulatory preferences are to maximize revenue, and the intuition provided for periodic amnesty is similar to the use of periodic discounts by a monopolist seeking to occasionally entice low value buyers to purchase the product. The setting studied in this paper is different — types are *dynamic*, the regulator cares about stopping crime as quickly as possible, and the dynamics in amnesty are used to ensure criminals do not *delay* reporting, which they do only because their types are dynamic. In the model of this paper, when types are static, optimal amnesty is static. Although the model in the present paper abstracts from important features of tax evasion and collection (importantly, by assuming 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. In this context, when the value to evasion is persistent but 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.

---

[36]See Le Borgne and Baer (2008), Luitel and Tosun (2014), and OECD (2015).

VI   Discussion and Extensions

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

Deterrence.   The model ignores the decision to become a criminal—criminals arrive criminals. Consider instead a setting identical to the one presented in Section I, except that agents make a once-and-for-all decision whether to begin committing crime at the moment they arrive to the model. The optimal policy in this case is to either never offer amnesty or offer static amnesty that induces reporting by low types only. The reason is that never offering amnesty effectively gives an agent a "once-and-for-all" opportunity to avoid punishment at the moment of arrival by not engaging in crime. If such a policy deters high types, then the regulator achieves her first-best. If instead high types begin committing crime then there is no policy that can induce reporting by high types. As a result, the regulator focuses her policy on inducing low types to report, which can be accomplished with a static policy.

To jointly study deterrence and amnesty then, it is necessary to enrich the model. Suppose agents arrive in a third, very high state, at which point they choose whether to become criminals. If this third state is sufficiently high, and transition from this state to the lower states is irreversible, then the results of Section II remain unchanged. A more satisfactory model with deterrence would allow agents to arrive in multiple states and decide whether to become criminals; agents arriving in the highest states cannot be deterred, while agents arriving in the intermediate or low states can be. In such a model, an optimal policy will trade off *deterrence* and *ex-post detection via self-reporting*. When the main motivation is deterrence, self-reporting policies will be counter-productive. When the main motivation is ex-post detection, self-reporting policies like the ones studied in this paper will be useful.

Arrival Time.   Amnesty policies are valuable when the enforcement environment is too weak to fully deter crime. In light of this, arrival to the model can be interpreted as the time at which returns from crime have fallen far enough, relative to their initial level when the agent began committing crime, to consider reporting. In some cases, like gun amnesty and buybacks, the regulator has no information on the time at which crime began, and so cannot directly make use of arrival time.[37]   But in some settings, the time criminals begin committing crime is perfectly observed. Though in many such settings amnesties are not conditioned on the time at which crime began, it is instructive to consider a variation

---

[37]Though, as previously mentioned, in a more general mechanism design approach, the principal may be able to elicit arrival time information.

of the model:[38] the regulator chooses a time path for amnesty conditional on the time of initial commission of the crime, crime is initiated in a third, very high state, and the agent cannot be deterred by any policy. This last condition is a reflection of the weak enforcement environment. At some Poisson rate the agent transitions from the very high state to the high state and never returns. The regulator does not observe the transition time from the very high state to the high state. This version of the model is equivalent to the model described in Online Appendix G, in which the arrival distribution of agents is generalized, and so results from there can be applied to show that when transition from the very high to the high state is not too fast, a cyclical policy is optimal.[39]

Initial Distribution of Values. Agents arrive in state x^h. Results are unchanged if I allow for a time-independent distribution of arriving values across x^h and x^l, since an optimal policy induces low types to report either always or never.

Value of Penalties. The regulator's objective function attaches a value of zero to penalties i.e. p_t for self-reporters and $\bar{p}$ for those detected. In some cases, like tax amnesty, this is an important omission; to incorporate revenue considerations, generalize the regulator's objective function to be a weighted combination of the loss from crime (x_t) and the profit from penalties. Weights are context specific e.g. in the case of tax amnesty, the weight is positive but should account for the cost of collecting penalties and proving guilt, which is administratively expensive (Franzoni, 1996). Although the proof of Theorem 1 does not generalize to such settings, the main force at work remains intact.

In the case of desertion, one of the main methods of punishment is prison time. Militaries have found ways of preserving manpower while still imposing punishment, such as random punishment (Becker (1968), Chen et al. (2020)), 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 imposing punishment, since by imprisoning a deserter, the military sacrifices manpower. The fundamental force of the paper is therefore strengthened in this context. This extension complicates the analysis and how it affects the results is left as an open question.

A related point concerns the inability of the regulator in the model to break the lower bound on $\underline{p}$ by using funds to pay criminals to stop. For instance, the regulator may be able

---

[38]In the Russian Civil War, I am not aware of evidence to suggest the Red Army's desertion amnesties restricted the application of amnesty by desertion date. In other desertion settings, there have been such restrictions, but they appear 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).

[39]Online Appendix G also characterizes an optimal policy (which takes a different form) when the transition from the very high to the high state is fast.

to set any penalty $p \in [\underline{p}, \overline{p}]$ at no cost, but can set $p < \underline{p}$ by making a transfer to a criminal who self-reports and stops—this transfer then enters somehow into the regulator's payoff. The effect of this extension will depend on the weight the regulator places on the loss from crime—for small enough weight, using payments will not be profitable. In some settings, such as desertion, there are rarely instances in which the military pays deserters to return.[40] The regulator may not have funds and be unable to borrow. In other settings, there are payments, but the relevant constraints on payment may be forces outside the model—for instance, payments in gun buybacks should not be so large as to entice people to buy guns only to immediately turn them in during a buyback.[41] Nevertheless, this extension would be valuable, and there will exist a large enough weight on the loss from crime that cyclical policies that use transfers at the end of each cycle to shorten the length of cycles will improve over policies without transfers. An analysis of optimal amnesty in this case is left to future work.

Quitting. One of the real-life features motivating the model is that certain aspects of crime are irreversible without approval from the authority, like desertion. Nevertheless, it is interesting to think about a case in which an agent is given an option to "quit" without self-reporting, which is important in cases like gun amnesties and buybacks. When quitting is not free (e.g. 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, and the basic forces remain.

Increasing Penalties. In the model, maximum penalties do not increase with the severity of the crime. In many cases, there are upper bounds on the severity of penalties. For instance, in the case of desertion, the military may be unable to credibly commit to frequent executions, since the legitimacy of its campaign relies on public support.[42] In other cases, like illegal gun ownership, the regulator has no information on when the crime began.

Absorbing Low State. The model does not allow for the possibility that agents in the low state transition to the high state. This assumption is made for tractability. The *Low Type Screening Lemma* generalizes to this case, so a policy similar to the policy in Theorem 1 is

---

[40]This may be because it would encourage desertion or harm morale, forces excluded from the model.

[41]See Plotkin (1996) on the 1991 St. Louis gun buyback.

[42]See, for instance, the support of the peasantry during the Russian Civil War discussed in Wright (2012).

approximately optimal when transitions to the high state are infrequent, with the maximum loss possible as compared to the optimal policy shrinking with the size of the transition rate.

Relationship to Garrett (2016). I detail here the main technical difference to the most closely related paper, Garrett (2016). The paper studies a monopolist choosing a price path for randomly arriving buyers who can have either *high value* for the product (analogue of *low returns* to crime) or *low value* for the product (analogue of *high returns* to crime), and can transition between these states. To make the parallel clear, let an *allocation* be the times at which the good is purchased by low value buyers in the monopoly setting, or the times at which high types report in the regulatory setting of this paper. Key to the solution method in the monopoly setting is that the monopolist's problem is *separable*: at any point in time at which low value buyers are allocated the good, the optimal choice of future allocations is independent of the history of allocations. In contrast, in the regulatory setting, the regulator's problem is *not* separable; at any point in time at which high types report, the optimal continuation policy *depends on history*—this dependence is summarized in the penalty offered for reporting at that time. The resulting dynamic program requires tracking this penalty as a state variable, and solving it requires a detailed verification of the proposed optimal policy.

To see why one problem is separable and the other is not, note that in the monopolist's problem, any allocation can be implemented by some price path. The monopolist's problem is then formulated as an unconstrained choice of allocation, given an optimal price path that implements it.[43] So, the history of allocations does not impose any feasibility constraints on future allocations. In contrast, in the regulatory setting, the optimal allocation would always be $\mathbb{R}_+$ if not for the bound on penalties. The bound on penalties must therefore be explicitly incorporated as a constraint in the regulator's problem, through which the history of allocations constrains future allocations. The solution to the regulator's problem reflects the interaction of the *backloading motive* with the *bound on penalties*.

Note that Garrett (2016) does not require either state to be absorbing, in contrast to this paper. This is a result of the solution method, as well as the difference in preferences.[44]

---

[43]A minor point is that negative prices are required to implement some allocations. In the monopolist's problem however, non-negativity constraints on prices can be ignored, so that the problem behaves as if any allocation is feasible. Indeed, this problem always results in non-negative optimal prices.

[44]More precisely, I show that when the regulator's problem is recursively formulated, there is a linear relationship between (i) the difference in the regulator's value between the proposed optimal policy and a deviation from the proposed optimal policy between any two decision nodes, and (ii) the difference in an agent's gains from crime between the proposed optimal policy and the deviation from the proposed optimal policy between the same two decision nodes. This linearity facilitates the verification of the proposed optimal policy, but no longer holds once agents can transition from low to high return crime.

Intermediate Amnesty. In some settings, one may observe that amnesty is offered either not at all, or as a full reduction of penalties—that is, intermediate levels of amnesty are not explicitly offered. It would be natural that a regulator treats criminals who reports their crime to the regulator in between amnesties more leniently than those who are detected, but less leniently than those who report during amnesties.[45]

In other settings, there may be a permanent program that reduces penalties below those imposed when the criminal is detected, interspersed with more generous short-term programs (see OECD (2015) for U.S. offshore compliance, or the permanent provision in the Swedish Weapons Act combined with periodic, more generous amnesties (Hofverberg, 2018)).

VII Conclusion

In this paper, I studied the problem of a regulator who designs amnesty programs to induce self-reporting of crime. I show (Proposition 2) 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 1 establishes that a cyclical policy is optimal and describes its form. After an initial period, the minimum possible penalty for reporting ($\underline{p}$) 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 agents with a low return from crime report at all times. A *backloading* motive on the part of the regulator drives the optimality of this form of amnesty. Agents discount more than the regulator across reporting times of high types; both agents and the regulator time discount, but agents also discount by the risk of detection. The regulator therefore finds it optimal to incentivize reporting by high types at a given time by committing the next amnesty to be the minimum penalty possible with the minimum delay necessary to satisfy incentive constraints.

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 the risk of detection. Second, new insights might result from incorporating political economy constraints. Third, the time-homogeneous risk of detection is unrealistic in some cases—a larger stock of past crimes changes the risk of detection and criminals become better at evading detection over time—and relaxing this could be a fruitful direction. Finally, it would be interesting to further examine the deterrence margin. For instance, when the regulator cannot condition her policy on the time at which

---

[45]Evidence for this, at least in the case of desertion, is scant, possibly due to the informal nature of the arrangement. In the case of desertion, there may be a period of time immediately after desertion during which, if the deserter turns themselves in, they will be treated leniently relative to those who remain deserters for longer. How this interacts with formal desertion amnesties is unclear.

crime begins, randomization becomes clearly 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

Andreoni, James. 1991. "The Desirability of a Permanent Tax Amnesty." *Journal of Public Economics*, 45(2): 143–159.

Araman, Victor F, and Bassam Fayad. 2020. "Intertemporal price discrimination with time-varying valuations." *Operations Research.*

Baer, Miriam H. 2008. "Linkage and the deterrence of corporate fraud." *Va. L. Rev.*, 94: 1295.

Bagirgan, Onursal. 2020. "Repeated Tax Amnesties." *Working Paper.*

Becker, Gary S. 1968. "Crime and punishment: An economic approach." In *The economic dimensions of crime.* 13–68. Springer.

Chen, Daniel L, et al. 2020. "The Deterrent Effect of the Death Penalty? Evidence from British Commutations During World War I." Toulouse School of Economics (TSE).

Conlisk, John, Eitan Gerstner, and Joel Sobel. 1984. "Cyclic Pricing by a Durable Goods Monopolist." *The Quarterly Journal of Economics*, 99(3): 489–505.

Connor, John M. 2007. *Global price fixing.* Springer.

Deb, Rahul. 2014. "Intertemporal Price Discrimination with Stochastic Values." *Unpublished manuscript, University of Toronto.*

Dreyfus, Pablo, Luis Eduardo Guedes, Ben Lessing, Antônio Rangel Bandeira, Marcelo de Sousa Nascimento, and Patricia Silveira Rivero. 2008. "Small Arms in Rio de Janeiro." *The Guns, the Buypack, and the Victims.*

Figes, Orlando. 1990. "The Red Army and Mass Mobilization during the Russian Civil War 1918-1920." *Past & Present*, , (129): 168–211.

Forrest, Alan I. 1989. *Conscripts and Deserters: The Army and French Society During the Revolution and Empire.* Oxford University Press on Demand.

Franzoni, Luigi A. 1996. "Punishment and grace: On the economics of permanent amnesties." *Public Finance/Finances Publiques*, 3: 353–368.

Garrett, Daniel F. 2016. "Intertemporal Price Discrimination: Dynamic Arrivals and Changing Values." *American Economic Review*, 106(11): 3275–99.

Garrett, Daniel F. 2017. "Dynamic mechanism design: dynamic arrivals and changing values." *Games and Economic Behavior*, 104: 595–612.

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

Grab, Alexander. 1995. "State power, brigandage and rural resistance in Napoleonic Italy." *European History Quarterly*, 25(1): 39–70.

Grillo, Edoardo, and Juan Ortner. 2018. "Dynamic contracting with limited liability constraints." *Collegio Carlo Alberto and Boston University*.

Harrington, Joseph E. 2005. "Detecting cartels." working paper.

Hofverberg, Elin. 2018. "Sweden: Three-Month Weapons Amnesty Implemented." *https://www.loc.gov/item/global-legal-monitor/2018-02-07/sweden-three-month-weapons-amnesty-implemented/*.

Kaplow, Louis, and Steven Shavell. 1994. "Optimal Law Enforcement with Self-Reporting of Behavior." *Journal of Political Economy*, 102(3): 583–606.

Krasikov, Ilia, and Rohit Lamba. 2020. "A Theory of Dynamic Contracting with Financial Constraints." *Available at SSRN 2909334*.

Krasikov, Ilia, Rohit Lamba, and Thomas Mettral. 2020. "Implications of unequal discounting in dynamic contracting." *Available at SSRN 3142932*.

Le Borgne, Eric, and Katherine Baer. 2008. *Tax Amnesties: Theory, Trends, and Some Alternatives*. International Monetary Fund.

Leigh, Andrew, and Christine Neill. 2010. "Do Gun Buybacks Save Lives? Evidence from Panel Data." *American Law and Economics Review*, 12(2): 509–557.

Lenis, David, Lucas Ronconi, and Ernesto Schargrodsky. 2010. "The Effect of the Argentine Gun Buy-back Program on Crime and Violence."

Luitel, Hari Sharan, and Mehmet Serkan Tosun. 2014. "A reexamination of state fiscal health and amnesty enactment." *International Tax and Public Finance*, 21(5): 874–893.

Luitel, Hari Sharan, and Russell S Sobel. 2007. "The Revenue Impact of Repeated Tax Amnesties." *Public budgeting & finance*, 27(3): 19–38.

Macinko, James, et al. 2007. "Reducing Firearm Injury: Lessons from Brazil." *LDI issue Brief*, 12(7): 1–4.

Malik, Arun S. 1993. "Self-reporting and the Design of Policies for Regulating Stochastic Pollution." *Journal of Environmental Economics and Management*, 24(3): 241–257.

Marchese, Carla, and Alberto Cassone. 2000. "Tax Amnesty as Price-Discriminating Behavior by a Monopolistic Government." *European Journal of Law and Economics*, 9(1): 21–32.

Mikesell, John L, and Justin M Ross. 2012. "Fast money? The contribution of state tax amnesties to public revenue systems." *National Tax Journal*, 65(3): 529–562.

Mullin, Wallace P. 2001. "Will Gun Buyback Programs Increase the Quantity of Guns?" *International Review of Law and Economics*, 21(1): 87–102.

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

Plotkin, Martha R. 1996. Police Executive Research Forum Washington, DC.

Rendle, Matthew. 2014. "Mercy Amid Terror? The Role of Amnesties During Russia's Civil War." *Slavonic & East European Review*, 92(3): 449–478.

Ritsema, Christina M, Deborah W Thomas, and Gary D Ferrier. 2003. "Economic and Behavioral Determinants of Tax Compliance: Evidence from the 1997 Arkansas Tax Penalty Amnesty Program." Citeseer.

Shiryaev, Albert N. 2007. *Optimal Stopping Rules.* Vol. 8, Springer Science & Business Media.

Slatta, Richard W. 1980. "Rural Criminality and Social Conflict in Nineteenth-Century Buenos Aires Province." *Hispanic American Historical Review*, 60(3): 450–472.

Wang, Shouqiang, Peng Sun, and Francis de Véricourt. 2016. "Inducing Environmental Disclosures: A Dynamic Mechanism Design Approach." *Operations Research*, 64(2): 371–389.

Willmer, MAP. 1971. "On the Estimation of the Number of Persons Entering and Leaving Crime Annually." *Journal of Research in Crime and Delinquency*, 8(2): 175–181.

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

Wright, Alistair S. 2012. "Stemming the Flow: The Red Army Anti-Desertion Campaign in Soviet Karelia (1919)." *Revolutionary Russia*, 25(2): 141–162.

Appendix

A Low Type Screening Lemma

Lemma A.1. *For any parameterization* $(\rho, r, \lambda, x^h, x^l, \overline{p}, \underline{p}, \alpha_l)$,

(A.1)
$$\Delta_l = \frac{\rho \overline{p} - x^l}{\rho + r} - \underline{p}.$$

Lemma A.2 (Low Type Screening Lemma). *Suppose* $0 \leq \Delta_l$. *Then, for any policy* $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}$, *there is another policy* $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}}) \in \mathcal{L}$ *s.t.* $a_t(x^h) = \tilde{a}_t(x^h)$.

Proof. Fix a policy $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}$. Observe first that the alternative policy $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$ defined by (i) $\tilde{a}_t(x^h) \equiv a_t(x^h)$, (ii) $\tilde{a}_t(x^l) \equiv \tilde{a}_t(x^h)$ and, (iii) $\tilde{p}_t \equiv (\mathbb{1}_{a_t(x^h)=1}) p_t + (1 - \mathbb{1}_{a_t(x^h)=1})\overline{p}$ is an element of $\mathcal{M}$ such that $a_t(x^h) = \tilde{a}_t(x^h)$. So, to prove the result, it is sufficient to show that for any policy $(\mathbf{p}, \mathbf{a})$ with $a_t(x^h) = a_t(x^l)$ and $p_t = \overline{p}$ if $a_t(x^h) = 0$, there is a policy $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}}) \in \mathcal{L}$ such that $a_t(x^h) = \tilde{a}_t(x^h)$. To this end, fix such a policy $(\mathbf{p}, \mathbf{a})$ and consider the alternative policy, $(\tilde{\mathbf{p}}, \tilde{\mathbf{a}})$, defined by (i) $\tilde{a}_t(x^l) \equiv 1$, (ii) $\tilde{a}_t(x^h) \equiv a_t(x^h)$ and (iii) $\tilde{p}_t \equiv -W^*(x^l, t, \mathbf{p})$.

The definition of W^* and the assumption that $0 \leq \Delta_l = \frac{\rho \overline{p} - (\rho+r)\underline{p} - x_l}{\rho + r}$ implies that $\tilde{p}_t \in [\underline{p}, \overline{p}]$, where I have plugged in equation (A.1) for Δ_l.

First, I argue that $\tilde{\mathbf{p}}$ is measurable. Let $\tau^h(t) \equiv \inf_{s \geq t} \{s - t | \tilde{a}_s(x^h) = 1\}$. Observe that $\tau^h(t)$ is measurable in t since $\tilde{\mathbf{a}}$ is, and that

$$\tilde{p}_t = -W^*(x^l, t, \mathbf{p}) = \frac{x^l - \rho \overline{p}}{\rho + r}(1 - e^{-(\rho+r)\tau^h(t)}) - e^{-(\rho+r)\tau^h(t)} p_{t + \tau^h(t)}$$

which is measurable in t since $\tau^h(t)$ is.

Second, I argue that the recommendation policy $\tilde{\mathbf{a}}$ is obedient.[46] Fix an arbitrary stopping

---

[46] Observe that $\tau^{\tilde{\mathbf{a}}} = \inf_{t \geq t_0} \{t - t_0 | \tilde{a}_t(x_{t-t_0}) = 1\}$ is indeed a stopping time since $\tilde{\mathbf{a}}$ is measurable.

time τ for an agent arriving at time t. Let $\sigma \equiv \tau \mathbb{1}_{a_\tau(x^h)=0} + \infty(1 - \mathbb{1}_{a_\tau(x^h)=0})$. Then,

$$
\begin{aligned}
W(x, t, \tau, \tilde{\mathbf{p}}) = \ & \mathbb{E}\left[w(x, \tau) + e^{-(\rho+r)\tau} W^*(x^l, \tau, \mathbf{p}) \right] \\
= \ & \mathbb{E}\left[w(x, \tau \wedge \sigma) - e^{-(\rho+r)(\tau \wedge \sigma)} \left(\mathbb{1}_{\tau<\sigma} p_\tau - \mathbb{1}_{\sigma \le \tau} W^*(x^l, \sigma, \mathbf{p}) \right) \right] \\
\le \ & \mathbb{E}\left[w(x, \tau \wedge \sigma) - e^{-(\rho+r)(\tau \wedge \sigma)} \left(\mathbb{1}_{\tau<\sigma} p_\tau - \mathbb{1}_{\sigma \le \tau} W^*(x_\sigma, \sigma, \mathbf{p}) \right) \right] \\
\le \ & W^*(x, t, \mathbf{p})
\end{aligned}
$$

where the third line follows from $x^l \le x_\sigma$. Conversely, if an agent arriving at time t uses stopping time $\tau^h = \inf_{s \ge t}\{s - t | a_s(x^h) = 1\}$, the agent guarantees himself $W^*(x, t, \mathbf{p})$.[47] So I conclude that $W^*(x, t, \mathbf{p}) = W^*(x, t, \tilde{\mathbf{p}})$ for $x \in \{x^h, x^l\}$. As a result, $\tilde{\mathbf{a}}$ is an obedient recommendation policy with $\tilde{a}_t(x^l) = 1 \ \forall t$ and $W^*(x^l, t, \tilde{\mathbf{p}}) = W(x^l, t, \tau^0, \tilde{\mathbf{p}}) = W(x^l, t, \tau^h, \tilde{\mathbf{p}})$.
□

B Proof of Proposition 2 (and 2 preliminary results)

Lemma B.1. *If $(\rho + r + \lambda)\Delta_l < x^h - x^l$ or $x^h - x^l \le (\rho + r)\Delta_l$, a static policy is optimal.*

Proof. By Proposition 1, we can constrain to static policies with obedient recommendation policies $\mathbf{a}$ that give the agent the maximum of (i) $\tau^0 \equiv 0$, (ii) $\tau^\infty \equiv \infty$ and (iii) $\tau^l \equiv \inf_{t \ge t_0}\{t - t_0 | x_{t-t_0} = x^l\}$ (for an agent arriving at t_0). The values under penalty policy $\mathbf{p}^{\underline{p}}$ of τ^0, τ^l and τ^∞ for an agent arriving at t_0 in state x are

$$
W^l(x) \equiv W(x, t_0, \tau^l, \mathbf{p}^{\underline{p}}) = \mathbb{E}_{x_0 = x}\left[w(x, t_0) - e^{-(\rho+r)\tau^l}\underline{p} \right] = \mathbb{1}_{x=x^h}\left(\frac{x^h - \rho\bar{p} - \lambda\underline{p}}{\rho + r + \lambda} \right) + \mathbb{1}_{x=x^l}(-\underline{p})
$$

$$
W^0(x) \equiv W(x, t_0, \tau^0, \mathbf{p}^{\underline{p}}) = -\underline{p}
$$

$$
W^\infty(x) \equiv W(x, t_0, \tau^\infty, \mathbf{p}^{\underline{p}}) = \mathbb{1}_{x=x^h}\left(\frac{x^h - x^l}{\rho + r + \lambda} \right) - \frac{\rho\bar{p} - x^l}{\rho + r}
$$

Suppose first that $x^h - x^l \le (\rho+r)\Delta_l$. Then, $\max\{W^\infty(x), W^l(x)\} \le W^0(x)$ for $x \in \{x^h, x^l\}$. As a result, the regulator achieves her highest feasible value, $V^* = 0$, with the static policy $(\mathbf{p}^{\underline{p}}, \mathbf{a})$ such that $a_t(x) = 1$ for all $(t, x) \in \mathbb{R}_+ \times \{x^h, x^l\}$.

---

[47]This follows from the fact that the original policy $\mathbf{a}$ is obedient and I argued that it is without loss to restrict to policies $\mathbf{a}$ with the property that $a_t(x^h) = a_t(x^l)$.

Suppose now that $x^h - x^l > (\rho + r + \lambda)\Delta_l$. The agent's value for τ^∞ when $x = x^h$ is

$$W^\infty(x^h) = \frac{x^h - x^l}{\rho + r + \lambda} - \frac{\rho\bar{p} - x^l}{\rho + r} = \frac{x^h - x^l}{\rho + r + \lambda} - (\underline{p} + \Delta_l) > -\underline{p} \geq W^0(x^h)$$

for any $\mathbf{p}$, where the second equality follows by plugging in equation (A.1) and the first inequality follows by our assumption $x^h - x^l > (\rho + r + \lambda)\Delta_l$. Since the value of τ^∞ is larger than the value of τ^0 for penalty $\underline{p}$, any obedient recommendation policy must have $a_t(x^h) = 0$ for all t. In this case, I claim that the static penalty policy $\mathbf{p}^{\underline{p}}$ is optimal. If $\Delta^l \geq 0$, then $\max\{W^\infty(x^l), W^l(x^l)\} \leq W^0(x^l)$ and $a_t(x) = \mathbb{1}_{x=x_l}$ for all t is obedient for $\mathbf{p}^{\underline{p}}$. If instead $\Delta^l < 0$, then $W^\infty(x^l) > -\underline{p}$. Since $W^\infty(x^l) = W(x^l, \tau^\infty, t_0, \mathbf{p}^{\underline{p}})$ is independent of the penalty policy and $\underline{p}$ is the minimum penalty, $a_t(x) = 0$ is the only obedient recommendation for all penalty policies, so any penalty policy is optimal. $\qquad\square$

Before proving Proposition 2, I prove an intermediate result when $x^l = \alpha_l = 0$. Define[48]

$$\widetilde{\Theta} \equiv \left\{ (\rho, r, \lambda, x^h, 0, \bar{p}, \underline{p}, 0) \,\middle|\, \frac{\rho + r + \lambda}{\rho + r}\left(\rho\bar{p} - (\rho + r)\underline{p}\right) \geq x^h > \rho\bar{p} - (\rho + r)\underline{p} \right\}.$$

Proposition B.1. $\Theta^* \cap \{\theta | x^l = \alpha_l = 0\} = \widetilde{\Theta}$.

Proof. First, note that Lemma B.1 implies (using equation (A.1)) that $\Theta^* \cap \{\theta | x^l = \alpha_l = 0\} \subseteq \widetilde{\Theta}$. I show the reverse inclusion, $\widetilde{\Theta} \subseteq \Theta^* \cap \{\theta | x^l = \alpha_l = 0\}$. Fix an arbitrary $\theta \in \widetilde{\Theta}$. By Proposition 1 it is sufficient to demonstrate a policy that strictly improves over $\mathbf{p}^{\underline{p}}$. Observe,

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

where the inequality follows by the assumption that $\theta \in \widetilde{\Theta}$. In this case then, the regulator receives her *worst possible* payoff; no agent ever reports until reaching the low state $x^l = 0$. Thus, to conclude the proof it is sufficient to demonstrate a policy which induces a positive mass of high types to report. Consider a *one-time policy* $(\mathbf{p}, \mathbf{a})$: (i) $p_t = (\mathbb{1}_{t=T})\underline{p} + (1 - \mathbb{1}_{t=T})\bar{p}$ for some $T > 0$, (ii) $a_t(x) = \mathbb{1}_{t=T}$ for each $x \in \{x^h, x^l\}$. Then, observe that,

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

where the inequality follows by assumption that $\theta \in \widetilde{\Theta}$. Thus, the recommendation $a_t(x) = \mathbb{1}_{t=T}$ 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. Therefore, $\widetilde{\Theta} \subseteq \Theta^* \cap \{\theta | x^l = \alpha_l = 0\}$ and combining with the reverse inclusion completes the proof. $\qquad\square$

---

[48]Recall, Θ^* is the set of parameters for which dynamic policies improve over an optimal static policy.

Proof of Proposition 2: Suppose now that $x^l > 0$ or $\alpha_l > 0$. If $\Delta_l < 0$, then Lemma B.1 implies that $\Theta^* \cap \{\theta | \Delta_l < 0\} = \emptyset$. Suppose instead that $\Delta_l \geq 0$. By Lemma A.2,

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

Since the right hand side is independent of α_l, it is wlog to prove the result for $\alpha_l = 0$. To this end, for any $\theta = (\rho, r, \lambda, \overline{p}, \underline{p}, x^h, x^l, 0)$, let $\widetilde{\theta}(\theta) \equiv (\rho, r, \lambda, \widetilde{\overline{p}}, \underline{p}, \widetilde{x}^h, \widetilde{x}^l, 0)$ where $\widetilde{x}^h = x^h - x^l$, $\widetilde{x}^l = 0$ and $\widetilde{\overline{p}} = \overline{p} - \frac{x^l}{\rho}$ (note that $\widetilde{\overline{p}} \geq \underline{p}$ since $\Delta_l \geq 0$). An agent's value for stopping time τ can then be re-written,

$$W(x, t_0, \tau, \mathbf{p}) = \mathbb{E}\left[\int_0^\tau e^{-(\rho+r)t} x_t dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho+r} \overline{p} - e^{-(\rho+r)\tau} p_{\tau+t_0} \right]$$

$$= \mathbb{E}\left[\int_0^{\tau \wedge \tau_l} e^{-(\rho+r)t}(x^h - x^l)dt + \int_0^\tau e^{-(\rho+r)t}(x^l)dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho+r} \overline{p} - e^{-(\rho+r)\tau} p_{\tau+t_0} \right]$$

$$= \mathbb{E}\left[\int_0^\tau e^{-(\rho+r)t} \widetilde{x}_t dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho \widetilde{\overline{p}}}{\rho+r} - e^{-(\rho+r)\tau} p_{\tau+t_0} \right]$$

where $\widetilde{x}_t = (x^h - x^l) \mathbb{1}_{x_t = x^h}$. So, an agent's value for a stopping time under any policy $\mathbf{p}$ is the same for θ and $\widetilde{\theta}(\theta)$. As a result, $\theta \in \Theta^* \iff \widetilde{\theta}(\theta) \in \Theta^*$. By Proposition B.1,

$$\widetilde{\theta}(\theta) \in \Theta^* \iff \widetilde{x}^h \in \left(\rho\widetilde{\overline{p}} - (\rho+r)\underline{p}, \frac{\rho+r+\lambda}{\rho+r}(\rho\widetilde{\overline{p}} - (\rho+r)\underline{p}) \right]$$

$$\iff x^h - x^l \in \left(\rho\overline{p} - (\rho+r)\underline{p} - x^l, \frac{\rho+r+\lambda}{\rho+r}(\rho\overline{p} - (\rho+r)\underline{p} - x^l) \right]$$

$$\iff x^h - x^l \in \left((\rho+r)\Delta^l, (\rho+r+\lambda)\Delta^l \right]$$

where the last line follows by equation (A.1), and the result follows. $\square$

C   Proof of Theorem 1 (and preliminary results)

Let $\tau^{\mathbf{a}} \equiv \inf_{t \geq t_0} \{t - t_0 | a_t(x_{t-t_0}) = 1\}$ i.e., the stopping time induced by $\mathbf{a}$. Let $\overline{\mathcal{M}}$ denote the set of policies $(\mathbf{p}, \mathbf{a}) \in \mathcal{M}$ but without the requirement that the stopping time induced by $\mathbf{a}$ is optimal for each agent (so $\mathcal{M} \subset \overline{\mathcal{M}}$). Let

(C.1)
$$t^{next}(t, \mathbf{a}) \equiv \inf_{s > t} \{s | a_s(x^h) = 1\}$$

(C.2)
$$w^h(t) \equiv w(x^h, t)$$

(C.3)
$$v(t) \equiv \int_0^t e^{-rs} \frac{1 - e^{-(\rho+\lambda)s}}{\rho + \lambda} ds$$

(C.4)
$$v^0(t) \equiv \frac{1 - e^{-(\rho+\lambda+r)t}}{(\rho + r + \lambda)(\rho + \lambda)}$$

(C.5)
$$\mathcal{P} \equiv [\underline{p}, -W(x_h, 0, \tau^\infty)]$$

where $W(x_h, 0, \tau^\infty) \equiv W(x_h, 0, \tau^\infty, \mathbf{p})$ for any $\mathbf{p}$. Plugging in the definition of $W(x^h, 0, \tau^\infty)$,

$$\mathcal{P} = \left[\underline{p}, \frac{\rho\bar{p} - x^l}{\rho + r} - \frac{x^h}{\rho + r + \lambda} \right].$$

The following lemmas are proved in Online Appendix E.

Lemma C.1. *Fix $t \in \mathbb{R}_+$ and $\underline{t} < t$ with $a_{\underline{t}}(x^h) = 1$. If $a_s(x^h) = 0$ for all $s \in (\underline{t}, t]$ then,*

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

Lemma C.2. *When $\alpha_l = 0$,*

(C.6)
$$V^* = \begin{cases} \sup_{(\mathbf{p},\mathbf{a}) \in \overline{\mathcal{M}}} V(\mathbf{p}, \mathbf{a}) \\ \text{subject to} \\ W(x^h, t, \tau^a, \mathbf{p}) \geq W^*(x^h, t, \mathbf{p}) \text{ for each } t \\ p_t = \bar{p} \text{ if } a_t(x^h) = 0 \\ a_t(x^h) = a_t(x^l) \text{ for each } t \end{cases}$$

Since $a_t(x^h) = a_t(x^l)$ for each t, I drop the dependence of a_t on x for the case $\alpha_l = 0$.

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

(C.7)
$$\mathbf{V}(p) = \begin{cases} \sup_{\substack{t \geq 0, \\ p' \in \mathcal{P}}} -v(t) + e^{-rt} \mathbf{V}(p') \\ \text{subject to} \\ w^h(t) - e^{-(\rho+r)t} p' \leq -p \end{cases}$$

and there is a policy, $(t_V(\cdot), p'_V(\cdot))$, that achieves value $\mathbf{V}(p)$ and has $\inf_{p \in \mathcal{P}} t_V(p) > 0$. Then,

$$V^* = \max_{t_0 \geq 0} \left\{ -v(t_0) + e^{-rt_0} \mathbf{V}(\underline{p}) \right\}.$$

Define $t(p)$ implicitly by,

(C.8)
$$w^h \circ t(p) - e^{-(\rho+r)t(p)} \underline{p} = -p.$$

33

Lemma C.4. *If $\theta \in \Theta^*$, then equation (C.8) has a unique strictly positive solution, $t(p)$, which is strictly increasing for $p \in \mathcal{P}$, and $\inf\limits_{p \in \mathcal{P}} t(p) > 0$. Further, for any $p \in \mathcal{P}$ and $t \geq t(p)$,*

$$w^h(t) - e^{-(\rho+r)t}\underline{p} \leq -p.$$

Lemma C.5. *Suppose that $\theta \in \Theta^*$. For any $\underline{t} > 0$,*

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

is decreasing in t for any $t \geq \underline{t}$, where $v^0(t)$ is defined by equation (C.4). Further,

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

is decreasing in p.

Proposition C.1. *Suppose that $\alpha_l = 0$ and $\theta \in \Theta^*$. Let $p' : \mathcal{P} \to \mathcal{P}$ be defined by $p'(p) = \underline{p}$ and $t : \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ be defined, for each $p \in \mathcal{P}$, as the unique strictly positive solution to (C.8), and suppose $\mathbf{V}^*(p)$ is the value function associated with $(t(\cdot), p'(\cdot))$. Then,*

$$V^* = \max_{t_0 \geq 0}\left\{-v(t_0) + e^{-rt_0}\mathbf{V}^*(\underline{p})\right\}.$$

Proof of Proposition C.1: To prove the result, I show that $\mathbf{V}^*(p)$ and $(t(\cdot), p'(\cdot))$ satisfy the premise of Lemma C.3, i.e. (C.7) is satisfied for $\mathbf{V}^*(.)$ and $\inf\limits_{p \in \mathcal{P}} t(p) > 0$. The latter follows from Lemma C.4, so I need only show the former.

Observe that

$$\text{(C.9)} \qquad v(t) = \frac{1 - e^{-rt}}{r(\rho + \lambda)} - \frac{1 - e^{-(\rho+r+\lambda)t}}{(\rho + \lambda)(\rho + r + \lambda)} = \frac{1 - e^{-rt}}{r(\rho + \lambda)} - v^0(t).$$

Let $\mathbf{V}^{*,0}(p) \equiv v^0(t(p)) + e^{-rt(p)}\frac{v^0(t(\underline{p}))}{1-e^{-rt(\underline{p})}}$, and observe further that,

$$\text{(C.10)} \qquad \mathbf{V}^*(p) = -v(t(p)) - e^{-rt(p)}\frac{v(t(\underline{p}))}{1 - e^{-rt(\underline{p})}} = \frac{-1}{r(\rho + \lambda)} + \mathbf{V}^{*,0}(p)$$

where the first equality follows by definition of $\mathbf{V}^*(p)$ and the second by plugging in (C.9) and the definition of $\mathbf{V}^{*,0}(p)$. Plugging (C.9) and (C.10) into (C.7), it is sufficient to show

$$\text{(C.11)} \qquad 0 = \begin{cases} \sup\limits_{\substack{t \geq 0, \\ p' \in \mathcal{P}}} \quad v^0(t) + e^{-rt}\mathbf{V}^{*,0}(p') - \mathbf{V}^{*,0}(p) \\[2mm] \text{subject to} \\[1mm] w^h(t) - e^{-(\rho+r)t}p' \leq -p \end{cases}$$

for any $p \in \mathcal{P}$. To prove this, it is sufficient to argue that the right-hand side of (C.11) is attained at $p' = \underline{p}$ and $t = t(p)$, since in this case the objective is 0. To compute the

34

value of the right-hand side of (C.11), it is sufficient to constrain to $t \leq t(p)$. To see this, observe that for any $t \geq t(p)$, an optimal choice of p' is $\underline{p}$ (feasible by Lemma C.4), since $\mathbf{V}^{*,0}(p')$ is decreasing in p' by the second part of Lemma C.5. By the first part of Lemma C.5, $v^0(t) + e^{-rt}\mathbf{V}^{*,0}(\underline{p})$ is decreasing in t, so among choices $t \geq t(\underline{p})$, it is optimal to set $t = t(\underline{p})$.

Next, I argue the objective on the right-hand side of (C.11) is bounded above by 0 in three simple cases. First, if $t = t(p)$, Lemma C.5 implies the objective in (C.11) is maximized over p' at $p' = p$ (where it is 0). Second, if $t = 0$, the incentive constraint in (C.11) requires that $p' \leq p$, in which case the objective on the right-hand side of (C.11) is bounded above by 0, again by Lemma C.5. Third, if $p' = p$, the incentive constraint in (C.11) requires that $t = t(p)$, and the objective equals 0.[49] Hence, any feasible choice (t, p') such that $t = t(p)$, $p' = p$, or $t = 0$, leads to a value of the objective in (C.11) no larger than 0.

I proceed now to remaining cases, assuming that $t < t(p)$, $p' > p$ and $t > 0$. Let $f \equiv \rho + \lambda + r$ and $g \equiv \rho + r$. Also, let $\phi(b) \equiv (1 - e^{-fb})$. To prove the result, it is sufficient to show, after plugging in the definition of $\mathbf{V}^{*,0}(p)$ and rearranging, that

(C.12)
$$\phi(t(p)) + e^{-rt(p)}\phi(t(\underline{p})) + \frac{e^{-r(t(p)+t(\underline{p}))}}{1 - e^{-rt(\underline{p})}}\phi(t(\underline{p})) \geq \phi(t) + e^{-rt}\phi(t(p')) + \frac{e^{-r(t+t(p'))}}{1 - e^{-rt(\underline{p})}}\phi(t(\underline{p}))$$

for any t, p, p' such that

(C.13)
$$(x^h - x^l)\frac{1 - e^{-ft}}{f} - \frac{\rho\bar{p} - x^l}{\rho + r}(1 - e^{-gt}) - p'e^{-gt} \leq -p.$$

Substitute for p' on the left-hand side of (C.13) using (C.8) to get,

(C.14)
$$\phi(t) + e^{-gt}\phi(t(p')) - c(1 - e^{-g(t+t(p'))}) \leq -\left(\frac{f}{x^h - x^l}\right)(p - \underline{p})$$

where $c \equiv \frac{f(\rho\bar{p} - x^l - (\rho+r)\underline{p})}{(x^h - x^l)g}$. By definition, (C.14) holds with equality for $(t, p') = (t(\underline{p}), \underline{p})$, so (C.14) can be written

$$\phi(t) + e^{-gt}\phi(t(p')) - c(1 - e^{-g(t+t(p'))}) \leq \phi(t(p)) + e^{-gt(p)}\phi(t(\underline{p})) - c(1 - e^{-g(t(p)+t(\underline{p}))})$$

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

(C.15)
$$\phi(t) + e^{-gt}\phi(t(p')) + \frac{e^{-g(t+t(p'))}}{1 - e^{-gt(\underline{p})}}\phi(t(\underline{p})) \leq \phi(t(p)) + e^{-gt(p)}\phi(t(\underline{p})) + \frac{e^{-g(t(p)+t(\underline{p}))}}{1 - e^{-gt(\underline{p})}}\phi(t(\underline{p}))$$

So to prove the result, it is sufficient to show that for any t, p', p such that (C.15) holds, (C.12)

---

[49]The only exception is if $p = \underline{p}$, in which case $t = 0$ is feasible. In that case, the objective is 0 as well.

holds as well. Observe that (C.15) and (C.12) are each a special case of the inequality,

(C.16)
$$\phi(t) + e^{-aft}\phi(t(p')) + \frac{e^{-af(t+t(p'))}}{1 - e^{-aft(\underline{p})}}\phi(t(\underline{p})) \leq \phi(t(p)) + e^{-aft(p)}\phi(t(\underline{p})) + \frac{e^{-af(t(p)+t(\underline{p}))}}{1 - e^{-aft(\underline{p})}}\phi(t(\underline{p}))$$

at $a = \frac{g}{f}$ and $a = \frac{r}{f}$, respectively.

For any t, p', let

(C.17)
$$z \equiv e^{-ft(\underline{p})}, \quad z_p \equiv e^{-ft(p)}, \quad u \equiv e^{-ft}, \quad y \equiv e^{-ft(p')}$$

By Lemma C.4, $z \geq \max\{z_p, y\}$. If $z = 0$, $\mathcal{P}$ is a singleton and there is nothing to prove, so I proceed under the assumption that $z > 0$. I assumed earlier that $t < t(p)$, $p' > \underline{p}$ and $t > 0$, which implies that $1 > u > z_p$ and $y < z$.

Plug the definitions in (C.17) into (C.16) and multiply both sides by $1 - z^a$ to get,[50]

(C.18)
$$(1 - u)(1 - z^a) + u^a(1 - y)(1 - z^a) + (uy)^a(1 - z)$$
$$\leq (1 - z_p)(1 - z^a) + z_p^a(1 - z)(1 - z^a) + (zz_p)^a(1 - z).$$

After rearranging and canceling terms,

$$0 \leq (u - z_p) - u^a(1 - y) + (uz)^a(1 - y) - (uy)^a(1 - z)$$
$$+ z^a(1 - u) + (z_p z^a - z_p^a z) + (z_p^a - z^a)$$
$$\iff 0 \leq \underbrace{(u - z_p) + (uz)^a(1 - y) + z_p^a(1 - z) - (uy)^a(1 - z) - u^a(1 - y) - z^a(u - z_p)}_{h(a;u,y,z,z_p)\equiv}$$

The crucial step is the following claim:

(C*)   if $0 \leq h(a; u, y, z, z_p)$ for $a \in (0, 1)$, then $0 < h(a; u, y, z, z_p)$ for all $0 < a' < a$.

A direct implication is that (C.15) implies (C.12), and so (C.11) follows. To prove (C*), there are three cases to consider.

Case 1: $y > 0$ **and** $uy < z_p$: In this case, instead of showing (C*) for h, I will show it for $\tilde{h}(a) \equiv \frac{h(a)}{(uy)^a}$, 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 (C*) for h can be recovered. Note that $\tilde{h}$ is smooth for any feasible choices of u, y, z, z_p with $y > 0$. When it is clear, I suppress the dependence of h on all inputs but a.

---

[50]Note that this multiplication does not change the direction of the inequality since $z \in (0, 1)$.

To prove (C*), it is sufficient to show that $\frac{\partial \tilde{h}^2}{\partial a^2}$ intersects zero at most twice. To see this, three observations prove key: (i) $1 > u > z_p$ implies $\tilde{h}(a) \to \infty$ as $a \to \infty$, (ii) $uy < z_p$ and $y < z$ imply $\tilde{h}(a) \to -(1-z)$ as $a \to -\infty$ and (iii) $\tilde{h}(1) = \tilde{h}(0) = 0$.

Violating (C*) requires the existence of $a_0 < 0 < a_1 < a_2 < 1 < a_3$ such that $\tilde{h}(a_0) < 0$ (observation (ii)), $\tilde{h}(a_1) \leq 0$, $\tilde{h}(a_2) \geq 0$, and $\tilde{h}(a_3) > 0$ (observation (i)). As a result, ensuring that observation (iii) is satisfied (while maintaining the smoothness of h) requires the existence of points $b_0 < b_1 < b_2 < b_3$ such that at $\tilde{h}''(b_0), \tilde{h}''(b_2) < 0$ while $\tilde{h}''(b_1), \tilde{h}''(b_3) > 0$, the existence of which implies that $\frac{\partial \tilde{h}^2}{\partial a^2}$ intersects zero at least three times.

To show that $\frac{\partial \tilde{h}^2}{\partial a^2}$ intersects zero at most twice, write

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

the zeros of which are the same as the zeros of the function $G(a) \equiv y^a \frac{\partial \tilde{h}^2}{\partial a^2}$. Plugging in,

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

To show G has at most two zeros, I show that $\frac{\partial G(a)}{\partial a}$ has at most one zero. Differentiating,

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

Finally, $\frac{\partial G}{\partial a}$ has the same number of zeros as $J \equiv \frac{u^a \frac{\partial G}{\partial a}}{z^a}$. Differentiating J,

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

Recalling that $1 > u > z_p > 0$ and $1 \geq z \geq z_p > 0$, all the terms on the right-hand side are positive and the first is strictly positive. As a result, J is a strictly increasing function with at most one zero. The same is then true of $\frac{\partial G(a)}{\partial a}$. So, G has at most two zeros, as does $\frac{\partial \tilde{h}^2}{\partial a^2}$.

Case 2: $uy \geq z_p$: I show that for $a \in (0,1)$, $h(a) < 0$.[51] Suppose first that $z_p = uy$. Then,

$$h(a; u, y, z, z_p) = (u - z_p) + (uz)^a(1 - y) + z_p^a(1 - z) - (uy)^a(1 - z) - u^a(1 - y) - z^a(u - z_p)$$
$$= (1 - y)(u - u^a)(1 - z^a) < 0$$

where the inequality follows from $u < u^a$ for any $a \in (0,1)$. Next, I show that $\frac{\partial h}{\partial z_p} \geq 0$ if $z_p < uy$. This will imply that for $z_p \leq uy$, h is maximized at $z_p = uy$, where it is negative, and so the proof will be concluded.

Let $G(z) \equiv \frac{\partial h}{\partial z_p} = z^a + \frac{a}{z_p^{1-a}}(1 - z) - 1$ 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 $z \in [z_p, 1]$, then $G(z) \geq 0$ for all $z \in [z_p, 1]$. As a result, $\frac{\partial h}{\partial z_p} \geq 0$ for all $z \in [z_p, 1]$ and since $z \in [z_p, 1]$, this concludes the proof.

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

$$\frac{\partial h}{\partial a} = ln(uz)(uz)^a + ln(z_p)z_p^a(1 - z) - ln(u)u^a - ln(z)z^a(u - z_p)$$
$$= z_p^a \underbrace{\left(ln(uz)(\frac{uz}{z_p})^a + ln(z_p)(1 - z) - ln(u)(\frac{u}{z_p})^a - ln(z)(\frac{z}{z_p})^a(u - z_p) \right)}_{h^1(a)}$$

To conclude, I show that $h^1(a)$ has at most two zeros. Differentiate to get

$$\frac{\partial h^1}{\partial a}(a) = ln(uz)ln(\frac{uz}{z_p})(\frac{uz}{z_p})^a - ln(u)ln(\frac{u}{z_p})(\frac{u}{z_p})^a - ln(z)ln(\frac{z}{z_p})(\frac{z}{z_p})^a(u - z_p)$$
$$= \left(\frac{uz}{z_p}\right)^a \underbrace{\left(ln(uz)ln(\frac{uz}{z_p}) - ln(u)ln(\frac{u}{z_p})(\frac{1}{z})^a - ln(z)ln(\frac{z}{z_p})(\frac{1}{u})^a(u - z_p) \right)}_{A(a)}.$$

Since $z_p \leq z$ and $1 > u > z_p$, $A(a)$ is increasing in a for $a \geq 0$. As a result, $\frac{\partial h^1(a)}{\partial a}$ has at most one zero for $a \geq 0$, and so $\frac{\partial h(a)}{\partial a} = h^1(a)$ has at most two zeros on $a \geq 0$. $\square$

Proof of Theorem 1: For $\theta \in \Theta^*$, I first prove the result in case $\alpha_l = 0$.

$\alpha_l = 0$. Suppose $\alpha_l = 0$ and $\theta \in \Theta^*$. The result follows by applying Proposition C.1, so that $V^* = \max_{t_0}\{v(t_0) + e^{-rt_0}V^*(\underline{p})\}$, and by repeatedly substituting in for $V^*(\underline{p})$.

---

[51] 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 $t(p)$, there will be 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.

$\underline{\alpha_l \geq 0}$. When $\theta \notin \Theta^*$, Proposition 2 leads to the result. Suppose instead that $\theta \in \Theta^*$.

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

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

Problem $(\mathcal{P}^h)$ is simply $(\mathcal{P})$ when $\alpha_l = 0$. Applying the result for $\alpha_l = 0$, an optimal policy in $(\mathcal{P}^h)$ is $(\mathbf{p}^*, \mathbf{a}^*)$ defined by: (i) $p_{t_0+nt(\underline{p})}^* = 0$ for some t_0 with $t(\underline{p})$ defined by (C.8), (ii) $p_t^* = \overline{p}$ otherwise, and (iii) $a_t^*(x) = 1$ if and only if $t \in \{t_0, t_0 + nt(\underline{p}))\}_{n \in \mathbb{N}}$.

Since $\theta \in \Theta^*$, Proposition 2 implies that $0 \leq (\rho + r)\Delta_l = \rho\overline{p} - x^l - (\rho + r)\underline{p}$. So, I can apply Lemma A.2 to transform $(\mathbf{p}^*, \mathbf{a}^*)$ into $(\tilde{\mathbf{p}}^*, \tilde{\mathbf{a}}^*) \in \mathcal{L}$ which has the properties:

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

- $p_{nt(\underline{p})+t_0}^* = \underline{p}$ for $n \in \mathbb{N}$ for some $t_0 \geq 0$

- $p_t^* = e^{-(\rho+r)t^{next}(t,\mathbf{a}^*)}\underline{p} + (1 - e^{-(\rho+r)t^{next}(t,\mathbf{a}^*)})\frac{(\rho\overline{p}-x^l)}{\rho+r}$ for all $t \notin \{t_0, t_0 + nt(\underline{p})\}_{n \in \mathbb{N}}$

where the last two lines translate the indifference requirement of an element in $\mathcal{L}$ to $(\mathbf{p}^*, \mathbf{a}^*)$. Lemma A.2 implies that $V^* = \sup_{(\mathbf{p},\mathbf{a}) \in \mathcal{L}} V(\mathbf{p},\mathbf{a}) = V_h^*$, so $(\mathbf{p}^*, \mathbf{a}^*)$ is an optimal policy. $\qquad\square$

D   Proof of Proposition 1

Proof of Proposition 1:   Because $\mathbf{p}^v$ is constant and hence continuous, Theorem 3 in Ch. 3 in Shiryaev (2007) can be applied to show that there exists some $D \subset \{x^h, x^l\}$ such that (i) $\tau_v^* \equiv \inf_{t \geq t_0}\{t - t_0 | x_{t-t_0} \in D\}$ and (ii) if τ is any other optimal stopping time for the agent, then $\mathbb{P}(\tau_v^* \leq \tau) = 1$.[52] As a result, it is 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 proposition, 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 agent's value can be computed by considering only three possibilities (i) $\tau_v^* = \tau^0 \equiv 0$, (ii) $\tau_v^* = \tau^\infty \equiv \infty$, or (iii) $\tau_v^* = \tau^l \equiv \inf_{t \geq t_0}\{t - t_0 | x_{t-t_0} = x^l\}$. The agent's value for τ^∞ is independent of v. The agent's values for τ^0 and τ^l are

$$\mathbb{E}\left[W(x,t,\tau^0,\mathbf{p}^v)\right] = -v$$

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

---

[52]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.

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 of an optimal stopping time from τ^∞ to one of the other two, or from τ^l to τ^0. The conclusion follows. $\quad\square$