Dynamic Amnesty Programs

Sam Kapon*

[Click here for the latest version]

19th August 2021

Abstract

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

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

1. Introduction

To stop ongoing crime, a regulator can offer preferable treatment to criminals who self-report. These amnesty, or self-reporting programs, appear in such diverse contexts as illegal gun ownership, collusion, desertion in war, tax evasion, espionage¹, civil conflict, and corruption.

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

¹Such as the amnesties offered to British informants by the Irish Republican Army in the 1980’s.
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 Hoi program offered amnesty to defectors during the Vietnam war, enticing over 100,000 (Wosepakat, 1971).

An extensive theoretical literature has investigated the use of self-reporting programs in one-shot regulation.\(^3\) 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 program has historically been applied to desertion in French militaries.\(^4\) The Brazilian gun buyback program has been run four times since 2013.\(^5\) The U.S. has operated a number of tax-related self-reporting programs, often on a repeated, time-limited, basis.\(^6\) 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 mechanism design problem in which criminal agents arrive at a time-homogeneous rate and their returns from crime are private, idiosyncratic and evolve over time. In particular, criminals can transition from a high return state of crime to a low return state of crime. A regulator commits to a time-path of penalties for agents who self-report before they are detected that applies uniformly to all agents. The range of possible penalties is bounded and agents may be exogenously detected, at which point the regulator applies the maximum penalty 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

\(^2\)https://www.justice.gov/atr/leniency-program
\(^3\)See Kaplow and Shavell (1994), Malik (1993) and Andreoni (1991) for early contributions.
\(^4\)See Wright (2012) for desertion in the Red Army and Forrest et al. (1989) for desertion in French militaries.
\(^5\)See Macinko et al. (2007).
\(^6\)OECD (2015), Luitel and Sobel (2007)
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 opportunity never self-report. The regulator must then balance two forces: (i) enticing contemporaneous agents to self-report by offering a future with less opportunity for self-reporting and (ii) enticing future agents to self-report by not completely shutting down these opportunities, as is done in the one-time policy. This trade-off is explored in the remainder of the paper.

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

The main result (Theorem 2) characterizes an optimal amnesty policy and shows that it takes a cyclical form; the self-reporting penalty declines until it hits the minimum penalty, 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 penalty is at its minimum. The frequency of cycles increases with the risk of detection, the maximum penalty, and the rate of transition from high to low return crime.

---

7See for instance Baer (2008) on such intertemporal linkages in corporate fraud.
8In Section 7, the features of the model are further discussed and interpreted.
To arrive at this policy, I first provide a key lemma (Lemma 1); any regulatory policy can be transformed into another policy in which (i) high return agents exhibit the same reporting behavior as in the original policy, (ii) low return agents self-report immediately at all times and (iii) low return agents are indifferent between reporting immediately and behaving like high return agents. With this lemma, the regulator’s problem can be solved in two steps: (i) optimize over policies in which agents operating high and low return crimes exhibit the same reporting behavior and the regulator experiences no loss from low return crime, then (ii) apply the construction in Lemma 1 to the policy from the first step to recover an optimal policy for the general case in which the regulator’s loss from agents who operate low return crime is positive.

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

The one-shot incentive compatibility constraint can be satisfied in many different ways. The regulator can combine a long delay with a generous amnesty at the next reporting time. The cost to the regulator is then a long period of time during which criminals are allowed to accumulate. Alternatively, the regulator can combine a short delay with an ungenerous amnesty at the next reporting time. The cost to the regulator is then a commitment to offering an ungenerous amnesty at the next decision node, which enters the incentive compatibility condition at that decision node and constrains her available policies. The crucial step of the main result (Theorem 2) is to show that, whatever the penalty the regulator must immediately deliver (the state), the optimal policy is to offer the minimum penalty at the next reporting time, and delay as long as necessary to satisfy the incentive compatibility constraint. Repeatedly applying this optimal policy in the recursive problem generates the path for an optimal policy described above.

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

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

In Section 5, I discuss a number of empirical settings in which dynamic self-reporting policies play a role. In the case of military desertion, I recount qualitative evidence from a case study of the Red Army’s anti-desertion campaign in Karelia, detailed in Wright (2012), among other sources, to argue that 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 programs in tax collection.

After reviewing the literature, I introduce the model in Section 2 and analyze it in Section 3. In Section 4, I present a number of comparative statics and discuss how investments in other features of the environment can act as complements to dynamic amnesty. In Section 5, I present applications of the model. In Section 6, I extend the model to allow the rate at which agents arrive to decay over time. The assumptions of the model and alternative modeling choices are discussed in Section 7. I conclude in Section 8.

\[9\] In some settings, in particular tax collection, the perceived fairness of enforcement may lead to a moral obligation to comply that generates higher compliance than would be implied by enforcement strength and financial considerations alone. In such cases, amnesty may backfire and lead to a deterioration of this moral obligation. I discuss this further in Section 7.
Contribution. This paper makes two contributions. First, it proposes a novel mechanism through which intertemporal variation in amnesty may be optimal. In certain settings, such as the Red Army’s anti-desertion campaign, qualitative evidence is provided that supports this mechanism as a driver of the decision to vary amnesty over time. In other settings, such as tax and gun amnesty, in which the intertemporal variation in amnesty is more naturally understood as a response to public pressure or short-term budget constraints, the model is used to highlight a possible benefit to such intertemporal variation and to propose potential policy improvements.

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

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

The early work of Kaplow and Shavell (1994), Malik (1993), and Andreoni (1991) studied law enforcement and self-reporting behavior in one-shot settings. Much of the subsequent literature is concerned with one-shot self-reporting settings in which the optimal 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 who are ex-ante heterogeneous. Wang et al. (2016) studies how a regulator should design remediation and inspection policies for environmental hazards that arrive randomly over time e.g. leaks. A firm has an option to delay repair of its environmental hazard and the paper focuses on the interaction between the inspection policy and penalties. As the paper shows, when the rate of inspection (which is like the rate of detection in this paper) cannot be chosen but is instead Poisson, optimal self-reporting programs are always static, unlike in this paper. I focus on the role that dynamic self-reporting programs play absent any control of inspection policies but in the presence of dynamic returns from crime. In this sense, the papers are complementary.\textsuperscript{10}

\textsuperscript{10}In the environmental hazard setting, the authors also show that it is without loss of generality to study
This paper is related to work in dynamic mechanism design such as Battaglini (2005) and the work on intertemporal price discrimination by a durable goods monopolist, such as Conlisk et al. (1984), Deb (2014), Garrett (2016) and Araman and Fayad (2020). The most closely related work is Garrett (2016) who studies a durable goods monopolist choosing a price path for dynamically arriving agents with changing values for a product and finds that cyclical pricing is optimal. The underlying intuition for why optimal policies may fluctuate is the same, and I find, as in Garrett (2016), that cyclical pricing is optimal. A fundamental difference between our papers is the limited penalties the regulator in this paper has at her disposal; the regulator cannot punish above a maximum level, and faces a bound on the self-reporting incentive she can offer. This constraint makes the regulator’s problem non-trivial but precludes the use of techniques from Garrett (2016). Solving the model in this paper therefore requires a different approach that explicitly incorporates these limited penalties. Preferences in this paper also differ from those of the 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.

2. The Model

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

2.1 The Agents

I present below the details of the agents’ environment.

Arrival and Flow Gain. Infinitesimal agents arrive at constant flow rate normalized to 1. Each agent is endowed with an individual flow gain process that follows a standard two-state continuous-time Markov chains, denoted $x_t$, independent across agents, with state space $E = \{x^l, x^h\}$ such that $0 \leq x^l < x^h$. For simplicity, state $x^l$ is absorbing and agents mechanisms that induce immediate reporting and repair of the hazard. In the setting of this paper, the analogue of this is true only for low return agents.

11A 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.

12As in Garrett (2016), I take the intuitive approach to aggregate the uncertainty at the individual level. An earlier version of the paper modeled arrival as a stochastic counting process, and the same results can
transition from state $x^h$ to $x^l$ at Poisson rate $\lambda$. Upon arrival, agents are initialized in state $x^h$.\textsuperscript{13} Index $x_t$ by time since arrival so that $x_0$ is the initial state of an agent upon arrival.

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

**Payoffs.** To compute an agent’s value from a stopping time, define

$$w(x, t_0, t) \equiv \mathbb{E} \left[ \int_0^{\tau_{\rho} \wedge \tau} e^{-rs} x_s ds - e^{-rt_{\rho} \wedge \tau} 1_{\tau_{\rho} < t} \bigg| x_0 = x, t_0 = t \right]$$

where the expectation is taken with respect to the distribution of $\tau_{\rho}$ and $x_t$. This is the value of an agent who arrives at time $t_0$ in state $x$ and delays reporting for a deterministic length of time $t$, net of the reporting penalty paid after delay $t$. Note that $\lambda$ does not explicitly appear, but rather controls the evolution of $x_t$. The term $(a)$ is the discounted accrued flow gain until the minimum of (i) the time the agent chooses to stop and (ii) the time the regulator detects the agent. The term $(b)$ is the penalty the agent pays if exogenously detected, $\overline{p}$, before choosing to stop. Notice that, conditional on $x_0$, the expression is independent of $t_0$, since $x_s$ is identically distributed for all agents conditional on $x_0$. So I drop the dependence on $t_0$, and simply refer unambiguously to $w(x, t_0, t)$ as $w(x, t)$.

An agent’s expected payoff from stopping time $\tau$ when arriving at time $t_0$ in state $x_0$ under penalty policy $p = (p_t)_{t \geq 0}$ is then,

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

where the expectation is taken with respect to the joint distribution of $\tau_{\rho}$, $x_t$ and $\tau$. The be obtained there.

\textsuperscript{13}As I detail in Section 7, this assumption can be relaxed to allow for a time-independent arrival distribution across states without affecting the results.
term (c) is the penalty the agent pays when stopping before he is detected by the regulator. The independence of $\tau_\rho$ from $\tau$ and $x_t$ implies that

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

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

The agent solves the problem,

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

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

### 2.2 The Regulator

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

(i) $a^x(t) \equiv a_t(x)$ is measurable for each $x$ and

(ii) $\tau^a \equiv \inf \{t - t_0 | a_t(x_t) = 1\}$ is an optimal stopping time for an agent arriving at $t_0$.

The stopping time defined in (ii) is called the stopping time induced by $a$ for an agent arriving at $t_0$. Observe that the regulator is restricted to deterministic penalty policies (which I discuss in Section 7). As a result, $a_t$ does not affect the agents’ values, but rather serves as a useful accounting device and allows the regulator to break ties in her favor. Let $\overline{M}$ be the set of policies $(p, a)$ that satisfy only (i), and $M$ the set of policies that satisfy both (i) and (ii).

**Payoffs.** The stopping times induced by an obedient recommendation policy $a$ induce a pair of paths, $(\mu^h_t)_{t \geq 0}$ and $(\mu^l_t)_{t \geq 0}$, which are the mass of agents in states $x^h$ and $x^l$ at each time $t$, respectively. The regulator discounts at the same rate as the agent, $r$, and her payoff from a policy $(p, a)$ is

\[14\text{Note that I have placed no restrictions on } p_t, \text{ which can, for instance, be negative and represent a reward as in the case of gun buybacks. Nevertheless, because } p_t \text{ here is pure money burning from the perspective of the regulator, the most natural cases involve } \underline{p} \geq 0.\]
\[
V(p, a) \equiv -\int_{0}^{\infty} e^{-rt} \left( \mu_{h}^{t} + \alpha_{l} \mu_{l}^{t} \right) dt
\]

where \( \alpha_{l} \geq 0 \). The regulator solves the problem,

\[V^{*} \equiv \sup_{(p, a) \in \mathcal{M}} V(p, a)\]

\((\mathcal{P})\)

A policy that achieves \( V^{*} \) is called \textit{optimal}.

3. Model Analysis

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

(i) **Static Policies.** I define static policies as policies in which the terms of self-reporting are constant over time and demonstrate their implications for optimal stopping.

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

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

(iv) **Recursive Representation.** I formulate and analyze a recursive representation of the problem.

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

3.1 Static Policies

A static policy is one which is constant over time.

**Definition 1.** A static policy is a pair \((p, a)\) such that \(p_{t} = v\) for all \(t\) for some \(v \in [p, \bar{p}]\).

A dynamic policy is any policy which is not a static policy. Denote by \(p^{v}\) the penalty policy in which \(p_{t} = v\) for all \(t\).

\[15\] 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.
An agent’s decision problem under a static policy takes a particularly simple form. Since the policy does not exhibit any inter-temporal variation, the agent’s problem in each state of $x_t$ is exactly the same at any time. It follows that the only stopping policies relevant for computing an agent’s value are (i) never stop in either state, (ii) stop immediately in both the high and low states, and (iii) wait until the low state and then stop immediately. The following proposition follows readily from this observation.

**Proposition 1.** An optimal static policy is $(p^2, a)$ where $a_t(x) = a_{t+s}(x)$ for each $x \in \{x^h, x^l\}$ and $t, s \in \mathbb{R}_+$.

The proof is given in Appendix E. To see the intuition, suppose that an agent considers reporting at time $t$ or $t+s$, under $p^v$. Lowering $v$ has two effects from the agent’s perspective: reporting at $t$ becomes more profitable and so does reporting at $t+s$. However, the value to reporting at $t$ increases by more than the value to reporting at $t+s$, because the increase at $t+s$ is scaled down by discounting and the risk of detection, $\rho + r$. Decreasing $v$ as much as possible, i.e. to $p$, is therefore optimal. Note that it is critical for the result that $v$, the penalty for reporting, does not enter the regulator’s objective function explicitly.

### 3.2 Low Type Screening

In this section, I provide a lemma that allows me to focus on a sub-class of policies in the search for an optimal policy. I show that the regulator can induce reporting by low return agents at all times without distorting the reporting incentives of high return agents. Let

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

i.e. a stopping time for an agent arriving at $t$ that follows the recommendation for high return agents.

**Definition 2.** Let $L$ be the set of policies $(p, a) \in M$ such that for all $t$, (i) $a_t(x^l) = 1$ and (ii) $W(x^l, t, 0, p) = W(x^l, t, \tau^h, p)$.

The first requirement is that an agent in the low state, $x^l$, chooses to report immediately. The second requirement is that an agent in the low state is indifferent between immediately reporting and instead following the high return agent’s strategy as prescribed by $a$.

Let $\tau^\infty$ denote the policy that never reports to the regulator and,

$$\Delta_t \equiv (-p) - W(x^l, 0, \tau^\infty, p^2).$$

---

16I abuse notation slightly and denote by 0 a stopping time that places probability 1 on immediate stopping upon arrival.
The value $\Delta_l$ is the difference in payoffs for an agent in the low state between immediate reporting for penalty $p$ and never reporting i.e. $\tau^\infty$. Notice that this value is measured at time 0, but is the same for any arrival time $t_0$.

The following lemma states that, as long as $\Delta_l$ is positive, any policy can be transformed without loss to the regulator’s value into a policy in $\mathcal{L}$. When $\Delta_l$ is positive, an agent in state $x^l$ prefers self-reporting and paying the minimum penalty ($\tilde{p}$) to never reporting. As I show in Theorem 1, if $\Delta_l$ is negative, a static policy is optimal. 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 $\mathcal{L}$.

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

To see an example, consider the policy in the left panel of Figure 1. In this policy, the high and low return agents report on the black regions, while the low return agents may report elsewhere. There may also be times at which nobody reports. To clarify the transformation in the lemma, consider performing the following two-step procedure on the policy $(p, a)$, in Figure 1. Initialize $(\tilde{p}, \tilde{a}) = (p, a)$. First, replace $\tilde{p}_t = p$ and $\tilde{a}_t = 0$ on the gray region. This only strengthens the incentives for agents to report on the black regions where $\tilde{a}_t(x^h) = 1$, since this step (weakly) worsens any alternative to immediate reporting on the black region. For the second step, replace $\tilde{a}_t(x^l) = 1$ and wherever $a_t(x^h) = 0$ replace $\tilde{p}_t = -W(x^l, t, \tau^h, p)$ i.e. the negative of the low return agent’s value from waiting until the next black point to report. The result of the transformation is depicted in the right panel.
of Figure 1. The key idea here is that by offering the low return agent a penalty equal to their value from behaving like a high return agent, high return agents are not incentivized to delay at any previous date. Indeed, the value under $\tilde{p}$ from reporting on the gray region can be recreated under $p$ by waiting until the next black point to report. Since the penalty at black points is no different under $\tilde{p}$, high return agents still find it optimal to report on the black region, while low return agents find it optimal to report everywhere. As a result, the new recommendation policy, $\tilde{a}_t(x^h) = 1$ on the black region and $\tilde{a}_t(x^l) = 1$ everywhere, is an obedient recommendation policy. The resulting policy $(\tilde{p}, \tilde{a})$ is a member of $L$ that induces the same reporting behavior by high return agents as $(p, a)$.

### 3.3 The Value of Dynamic Policies

In this section, I characterize the set of parameters for which dynamic policies improve over static policies. First, the lemma below describes a set of parameters for which static policies are optimal.

**Lemma 2.** If $(\rho + r + \lambda)\Delta_l < x^h - x^l$ or $x^h - x^l \leq (\rho + r)\Delta_l$, then a static policy is optimal.

The proof is in Appendix C. The intuition for the result is as follows. Observe that since the low return state is absorbing, low return agents face the same decision at any point in time under a static policy. Suppose first that $x^h - x^l \leq (\rho + r)\Delta_l$. In this case $\Delta_l \geq 0$, so low return agents prefer immediately reporting and paying $p$ to never reporting, and so the static policy $p^L$ induces reporting by low return agents at all times. Under the static policy $p^L$, the gain of operating crime in the high state is not enough to offset the loss from possible detection, given immediate reporting in the low state. As a result, $p^L$ also induces reporting by high return agents at all times. The static policy $p^L$ therefore achieves the regulator’s highest possible value and so dynamic policies are unnecessary.

Suppose now that $(\rho + r + \lambda)\Delta_l < x^h - x^l$. Consider a one-time policy at $T$: $p_t = p1_{t=T} + \overline{p}1_{t\neq T}$. Under the parameter restriction, an agent in the high state at time $T$ would not report at $T$. Since the high return agent does not report for the most generous penalty when it is followed by the least generous continuation, no policy can ever induce reporting by high return agents. In this case, the regulator’s policy can only hope to induce low return agents to report. If $\Delta_l \geq 0$, then as in the previous case low return agents are induced to report at all times by $p^L$, and so a dynamic policy cannot improve the regulator’s value. If instead $\Delta_l < 0$, a low return agent prefers never reporting to immediately reporting for the minimum penalty possible, and so no reporting policy can ever induce reporting by low return agents. In this case, all policies deliver the regulator’s worst possible value, and so dynamic policies cannot improve over static policies.
Let \( \theta \equiv (\rho, r, \lambda, x^h, x^l, \overline{p}, p) \) denote an arbitrary parameterization. I will say that *dynamic policies strictly improve over static policies* under \( \theta \) if

\[
\theta \in \Theta^* \equiv \left\{ \theta \ \bigg| \ V^* - \sup_{\{(v,a):(p^v,a)\in\mathcal{M}\}} V(p^v, a) > 0 \right\}.
\]

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

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

Observe that \( \Theta^* \) is empty when returns are perfectly persistent, i.e. \( \lambda = 0 \).\(^{17}\) The result is proved in Appendix C. I provide here some intuition. First, the result that

\[
\Theta^* \subseteq \left\{ \theta \ \bigg| \ (\rho + r + \lambda)\Delta_l \geq x^h - x^l > (\rho + r)\Delta_l \right\}
\]

is a direct consequence of Lemma 2.

For the reverse inclusion, suppose that \( x^h - x^l \in (\Delta_l(\rho + r), \Delta_l(\rho + r + \lambda)) \). For any \( T > 0 \), recall the one-time policy at \( T \), defined by \( p_t = \overline{p}1_{t=T} + \underline{p}1_{t \neq T} \). When \( x^h - x^l \leq (\rho + r + \lambda)\Delta_l \), an agent prefers to report and pay \( p \) than to never report, and so the agent would immediately report for penalty \( p_t = \underline{p} \) if the only other available choice was to never report. By offering the one-time policy at \( T \), the regulator effectively forces the agent to make this decision at time \( T \), since an agent never reports if he must pay \( p \).

As long as \( T > 0 \), the regulator is therefore able to induce reporting by a strictly positive mass of high return agents. On the other hand, because \( x^h - x^l > (\rho + r)\Delta_l \), no static policy induces reporting by high return agents. Applying Lemma 1 to the one-time policy at \( T \) generates a policy in which low return agents report everywhere and high return agents report somewhere, which is a strict improvement over any policy in which no high return agents report, including any static policy.

### 3.4 Recursive Representation

A corollary of Theorem 1 is that if a dynamic policy is necessary to achieve the regulator’s optimal value, that is \( \theta \in \Theta^* \), then \( 0 < \Delta_l \). This is the condition required for the application

---

\(^{17}\)This is not a consequence of the assumption that agents arrive in state \( x^h \) which, as I discuss in Section 7, can be generalized at no cost to the results and just minor cost to the notation.
of Lemma 1, which states that any policy can be transformed into one in which low return agents report everywhere without altering the reporting behavior of high return agents. As a result, the optimal policy can be solved in two steps:

(i) Solve the regulator’s problem when $\alpha_l = 0$, in which case the regulator does not lose any value from low return crime,

(ii) Apply Lemma 1 and the construction therein to the resulting policy to find an optimal policy for $\alpha_l \geq 0$.

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

**Lemma 3.** When $\alpha_l = 0$, then

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

The result is proved in Appendix D. Since $a_t(x^h) = a_t(x^l)$ for each $t$, I drop the dependence of $a_t$ on $x$ while I study the case $\alpha_l = 0$. To state the problem recursively, let

$$\tau^{prev}(t,a) \equiv \sup\{s \mid s \leq t \text{ and } a_s = 1\}$$

for any recommendation policy $a$. Given $t$, this is the most recent time that agents reported, according to $a$. It is straightforward to show that the measure of high types at $t$, when agents choose the stopping time induced by $a$, is $\mu^h_t = \frac{1-e^{-(\rho+\lambda)(t-\tau^{prev}(t,a))}}{\rho+\lambda}$. Then, the regulator’s
value for a policy \((p, a)\) is

\[
V(p, a) = -\int_0^\infty e^{-rt} \frac{1 - e^{-(\rho + \lambda)(t - t_{\text{prev}}(t, a))}}{\rho + \lambda} dt.
\]

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

\[t_{\text{next}}(t, a) \equiv \inf\{s | s > t \text{ and } a_s = 1\} \]

This is the delay until the next time, \(s\), such that \(a_s = 1\). Then, a one-shot deviation available to an agent is, rather than report at \(t\), report instead at \(t_{\text{next}}(t, a)\). A consequence of obedience of \(a\) is that one-shot deviations are not profitable. That is, for any \(t\) such that \(a_t = 1\) (and assuming \(a_{t_{\text{next}}(t, a)} = 1\)), it must be that

\[-p_t \geq w(x, t_{\text{next}}(t, a) - t) - e^{-(\rho + r)(t_{\text{next}}(t, a) - t)} p_{t_{\text{next}}(t, a)}.\]

The regulator’s problem can be formulated recursively in way similar to Spear and Srivastava (1987), with (i) decision nodes as the times at which agents report, (ii) the state as the penalty that must be offered at the decision node, \(p_t\), (iii) choices as the delay until the next such time (the analogue of \(t_{\text{next}}(t, a) - t\)) and the penalty at that time (the analogue of \(p_{t_{\text{next}}(t, a)}\)) and (iv) the constraint as the one-shot incentive compatibility constraint. The regulator’s value, \(V^*\), can then be computed by optimizing over the first calendar time at which agents report and the penalty at that time, followed by the solution to the recursive formulation of the regulators problem in which that penalty is the state.

The next lemma presents a recursive problem, and states that if a solution exists and is associated with a policy that ensures that the time between decision nodes is uniformly bounded below by some \(\epsilon > 0\), then the solution to the recursive problem solves the regulator’s problem. This uniform bound ensures that the times at which agents are recommended to report, according to the solution to the recursive problem, contains no limit points. Let \(w^h(t) \equiv w(x^h, t)\) and define

\[
v(t) \equiv \int_0^t e^{-rt} \frac{1 - e^{-(\rho + \lambda)t}}{\rho + \lambda} dt,
\]

which will be the per-period loss of the regulator in the recursive representation of her problem. Let \(P \equiv [p, -W(x, 0, \tau^\infty)]\). This is the domain on which the regulator’s value function will be defined; the agent can always guarantee the payoff \(W(x, 0, \tau^\infty)\) by never reporting, and so \(-W(x, 0, \tau^\infty)\) is the maximum penalty that an agent would ever be willing to pay.
Lemma 4. Suppose that $\alpha_l = 0$, $\theta \in \Theta^*$, and $V : \mathcal{P} \to \mathbb{R}$ satisfies

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

and there is some policy, $(t(p), p'(p))$ that achieves the value $V(p)$ and has $\inf_{p \in \mathcal{P}} t(p) > 0$. Then,

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

The proof is given in Appendix D. To solve for a fixed point in (3), I conjecture an optimal policy and associated value function $V(p)$, and then verify the conjecture. In particular, I conjecture that an optimal policy is $p'(p) = p$ and $t(p)$ is defined as the unique strictly positive solution to the equation

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

which is guaranteed to exist for each $p \in \mathcal{P}$ whenever $\theta \in \Theta^*$.\(^{18}\) This $t(p)$ guarantees that a one-shot incentive compatibility constraint holds at equality: a high return agent is indifferent between immediately reporting for penalty $p$ and reporting for penalty $p$ after a deterministic amount of time $t(p)$.

I provide now a sketch of the proof of the verification. Recall that $\alpha_l = 0$. For the purposes of this sketch, I assume $x^l = p_0 = 0$ and I will describe the computations necessary for verification when the state is $p = 0$, but the general case is similar.

Let $V^*(p)$ be the value associated to the policy $(t(p), p'(p))$. Plugging in the definition of the conjectured optimal policy into (3), verification requires showing

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

It is straightforward to show that setting $t > t(0)$ cannot deliver an improvement over the

\(^{18}\)This is established in the appendix, in Lemma D.2.
conjecture, so I focus on choices \( t \in (0, t(0)). \)

Integrate \( v(t) \) to find that,

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

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

Integrate \( w^h(t) \) and plug in the definition of \( v^0(t) \) to get,

\[
w^h(t) = x^h(\rho + \lambda)v^0(t) - \frac{\rho p}{\rho + r}(1 - e^{-(\rho + r)t})
\]

Plugging this into (4) at \( p = 0 \) leads to

\[
\rho p x^h(\rho + \lambda)v^0(t(0)) = v^0(t(0)) - e^{-rt(0)}.
\]

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

Equation (9) can be rewritten

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

where

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

\(^{19}\)If \( t = 0 \) or \( t = t(0) \), the objective is 0.

\(^{20}\)In Section 6, I show the left-hand side of the constraint is in fact the regulator’s objective under the alternative assumption that agents arrive to the model at time-inhomogeneous rate \( e^{-\rho t} \).

18
To complete the verification, I show that \( h(a, t, t') \) satisfies a single-crossing property. In particular, I first define a set of conditions on \( t' \) that must hold for any \( t \in (0, t(0)) \), and show that whenever \( t' \) satisfies these conditions, then

\[
 h(a, t, t') \leq 0 \quad \text{for} \quad a \in (0, \rho + r + \lambda) \implies h(\tilde{a}, t, t') < 0 \quad \text{for any} \quad \tilde{a} \in (0, a)
\]

Plugging in \( a = \rho + r \) and \( \tilde{a} = r \) to this condition completes the verification, since \( h(r, t, t(0)) = 0 \). The following proposition summarizes the result.

**Proposition 2.** Suppose that \( \alpha_l = 0 \) and \( \theta \in \Theta^* \). Further suppose that \( p'(p) = \underline{p}, \; t(.) \) is the unique strictly positive solution to equation (4), and \( V^*(p) \) is the associated value function. Then

\[
 V^* = \max_{t_0 \geq 0, p_0 \in P} \{-v(t_0) + e^{-rt_0} V^*(p_0)\}.
\]

The proof is given in Appendix D and proceeds along the lines described above to show (5) and then applies Lemma 4 after observing that \( \inf_{p \in P} t(p) \geq t(\underline{p}) > 0 \) when \( \theta \in \Theta^* \).

### 3.5 An Optimal Policy

In the result that follows, I describe an optimal policy for the general case when \( \alpha_l \geq 0 \), which is the outcome of combining Proposition 2 with Lemma 1.

**Theorem 2.** If \( \theta \in \Theta^* \), an optimal policy is \((p^*, a^*)\) defined as,

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

where \( t(p) \) is the unique strictly positive solution to (4). If \( \theta \notin \Theta^* \), a static policy is optimal.
The proof is given in Appendix D. For $\theta \in \Theta^*$ and any $t \notin \{t_0 + nt(p)\}_{n \in \mathbb{N}}$, the path of $p_t$ guarantees that an agent in state $x^j$ is indifferent between immediately reporting anywhere on this path and waiting until the next $t \in \{t_0 + nt(p)\}_{n \in \mathbb{N}}$ to report. The $t_0$ in the theorem is an initial timing choice of the regulator, who is initially unburdened by incentives of prior agents. The optimal policy beyond $t_0$ is displayed in Figure 2.

In the recursive formulation of the problem in the previous section, I argued that the agents’ and regulator’s values from a deviation from the proposed optimal policy are closely connected. In particular, the agent’s value is a distorted version of the regulator’s value, applying an extra discount factor of $\rho$ across the times at which high return agents report. In fact, it can be shown that, for any policy that satisfies incentive compatibility constraints at equality, the agent’s value is a linear function of the regulator’s value under the alternative assumption that agents arrive to the model at time-inhomogeneous rate $e^{-\gamma t}$. Since incentive compatibility constraints would be unchanged under this alternative assumption, this suggests that if the regulator faced a population arriving at time-inhomogeneous rate $e^{-\gamma t}$ for arbitrary $\gamma \geq 0$, a new result could be established: when $\gamma \leq \rho$, the optimal policy in Theorem 2 remains optimal, but if $\gamma > \rho$, this will no longer be true and a new optimal policy will arise. This indeed turns out to be true: in Section 6, I show that when $\gamma \leq \rho$, the optimal policy remains optimal, but when $\gamma > \rho$ a different policy is optimal and I provide a characterization.

Before studying this extension, I discuss comparative statics of the baseline model in Section 4 and then applications of the model in Section 5.
4. Comparative Statics

It is immediate from Theorem 2 that on the interior of $\Theta^*$, the frequency of cycles increases in the risk of detection ($\rho$), the maximum penalty ($\bar{p}$) and the rate of transition from high to low return crime ($\lambda$). When the regulator can invest to increase $\rho$ or $\bar{p}$ ex-ante, the regulator can also increase the frequency with which high return agents 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 dynamically and jointly with amnesty, the comparative static nevertheless provides useful insight. In some settings, the regulator may only be able to imperfectly affect the rate of detection; for instance, in the case of cartels or desertion during war, much of detection comes from third-parties reporting to the regulator. While the regulator can potentially affect the incentives of third-parties to report, this is a much less tightly controlled process than in the case of, for instance, environmental inspections where the regulator directly controls the rate of inspection. Second, in some of the examples I discuss in the paper, policy-makers first implemented a blanket strengthening of enforcement and only then implemented an amnesty policy. One way to approximate this decision making process is by adding a pre-play stage to Section 2 in which the regulator chooses, at some cost, $\rho$ and $\bar{p}$. The comparative statics provide insight as to what would happen in such a model.

5. 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.

5.1 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 as “the first and principal obstacle to the organization of the army of the Kingdom” (Grab, 1995).

\footnote{As in Wang et al. (2016).}
Desertion amnesties are often offered during the course of war in an effort to entice return and have been applied extensively across history.\textsuperscript{22} The Red Army created its anti-desertion commission in 1918 – it increased punishments, strengthened enforcement (for instance, dispatching armed groups to search for deserters) and implemented periodic amnesties to entice deserters back to their units (Wright, 2012). As noted in Figes (1990), “...the most successful means of combating desertion [in the Red Army] were the amnesty weeks.” During and surrounding the Argentine War of Independence, the military engaged in “alternating carrot and stick”, offering amnesties to deserters in December 1813, September 1815, and September 1821 (Slatta, 1980). In Napoleonic Italy, “the government’s repressive policy was mitigated by frequent amnesties designed to entice deserters and draft dodgers back to the army” (Grab, 1995). French Militaries in the 18th and 19th centuries offered periodic amnesties “interspersed with periods of severe repression, in an attempt to lure waverers back to their units” (Forrest et al., 1989).

The difficulties that a deserter faces can change over time, and this provides the basis for the dynamic returns modeled in this paper. Forrest et al. (1989) provides, among other things, an account of desertion in France in the early 19th century. Deserters were often hungry unless they were lucky enough to receive help from local people. Snow could block passages through the mountains and the cold could be deadly. Deserters were “forced into the surrounding countryside [of their village], searching out caves and hiding-places that would offer protection until the forces of law and order had passed through.” Under such circumstance, Forrest et al. (1989) remarks, “[I]t is hardly surprising that considerable numbers of deserters changed their minds.”

### 5.1.1 The Red Army and the Anti-Desertion Commission

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.

Wright (2012) offers an account of the anti-desertion effort in the Red Army during the years 1918-1920, along with a detailed case study of the anti-desertion experience in Karelia. As noted in the case study, historians have deemed material shortages – ‘uniforms, linen, tea, tobacco, and soap’ – a primary reason for mass desertion during the Russian Civil War. Other important factors were the intensity of fighting, proximity to the White Army, and seasonality.\textsuperscript{23} 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

\textsuperscript{22}In contrast to desertion amnesties offered after a war, as a method of reconciliation and forgiveness.
\textsuperscript{23}For instance, the harvest season led soldiers to return home to sow their fields.
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), “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): ‘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, 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. 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 possible explanations for the intermittent nature of amnesties.

Public Randomness. Aside from the general idiosyncratic variation in a deserter’s plight as described earlier, some of the most important time-varying factors were the advances of the White army and the harvest season. In particular, deserters from the Red Army often returned at the end of their harvests, which is responsible for the success of some amnesties (Wright, 2012). A theory based on a public end to the harvest season would be able to account for annual amnesties, but even at a relatively small regional level, amnesties were more frequent — Wright (2012) describes a number of amnesty weeks separated by periods with no amnesty during the June-October 1919 period in the Karelia region. As a result, 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 $\lambda = 0$ (no intertemporal fluctuations in value to desertion) but with a decision to desert: that is, when agents arrive to the model, they make a once-and-for-all decision whether to desert. 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 essentially faces the most generous incentive not to desert, 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. As a result, the policy that never offers amnesty cannot be improved.

Although the model is stylized, it highlights the reason that an explanation based purely on discouraging desertion is insufficient. When a deserter deserts, they calculate that desertion is preferred to remaining in the army. At that point they have revealed that, unless their or the military’s situation changes, they would not accept an amnesty since they could have avoided penalties altogether by not deserting in the first place.

Organizational Instability and Fluctuating Preferences. A final possibility is that the decision to offer amnesties was not decided by a long-run, forward-looking and unified anti-desertion commission, but rather by a commission with fluctuating preferences or priorities, or one in which those in the leadership were replaced frequently. While certainly plausible, the relatively short time-horizon over which the amnesties were offered makes this story less compelling than it is in, for instance, the case of tax amnesties held every few years, which are subject to a natural political cycle.

5.2 Gun Amnesties and Buybacks

In this section, I argue that 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.\footnote{The exact content of ‘no-questions’ asked varies from program to program.} Buy-back programs go one step further, offering to pay for each firearm surrendered. During the Argentinian buyback of 2007-08, the government collected more than 100,000 weapons (Lenis et al., 2010). During the Brazilian buy-backs of 2003, 2009 and 2011, the government collected more than 1 million weapons. When operated on a small scale, the evidence,
especially in the U.S., points to the lack of any effect of gun buybacks on gun violence (Plotkin, 1996). On a large scale, however, these programs can potentially be effective (Lenis et al. (2010), Macinko et al. (2007)), especially when coupled with changes to the enforcement environment.

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

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

Whether the optimal policy takes a dynamic form and represents a significant improvement over the optimal static policy depends on parameters of the environment. While some parameters like the detection rate and penalties can be estimated from available data, the speed at which people transition from high to low value gun ownership cannot be. Many gun amnesty and buyback programs are accompanied with anonymous surveys of participants (McGuire et al., 2011). One way to estimate these parameters is to add two questions which

25 However, it must be noted that this change in value cannot come from a malfunction in the gun itself. As shown in Mullin (2001), such a change in value will lead gun owners to turn their gun in during a buy-back only to turn around and use the money to buy a new gun.
are often left out of these surveys. The first is “how long have you owned your gun?” Answers to this question provide information on $\lambda$, the rate of transition from high to low value gun ownership. This can then inform the frequency and form of the optimal policy. The second is “if you owned it during the last buyback, why didn’t you turn it in then?” Answers to this question can provide direct evidence on the motives of the participants. For instance, some may have not known about the program, or may have been otherwise misinformed about its conditions. Failing to account for such delay would overstate the value of a dynamic policy.

5.3 Voluntary Disclosure and Tax Amnesty Programs

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 most likely driven by political considerations, have benefits when compared to more permanent disclosure programs which have sometimes been implemented.

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

Le Borgne and Baer (2008) discusses the two main motivations for implementing a tax amnesty or voluntary disclosure program: “The two primary reasons for introducing tax amnesties are (i) to raise revenue in the short-term, and/or (ii) to increase compliance (e.g., by encouraging taxpayers to declare and pay previously undeclared tax, file tax returns, or register to pay taxes, so as to increase revenue and horizontal equity in the medium term).” It it often argued, especially recently, that tax amnesties have been implemented with an eye to (i). The model I present in this paper shows that despite this focus, the intermittent nature of these programs, as compared to a permanent policy, is also valuable from the perspective of (ii), increasing long-term compliance, when the value to tax evasion changes.

26 See Luitel and Tosun (2014).
over time.

A basic question relevant for examining tax amnesties and voluntarily disclosure programs is, why do people apply? In the model presented, the change in profits from tax evasion leads evaders to self-report. Although direct evidence regarding motivation is not widely available, one source of evidence on this question comes from Ritsema et al. (2003), who implemented a survey of participants in the 2003 Arkansas tax amnesty program. The authors find that income, ease of evasion and inability to pay were three important determinants in the decision to evade taxes. 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 ex-ante honest heterogeneous taxpayers. Although the model in the present paper abstracts from important features of tax evasion and collection (importantly, that the regulator does not care about collected penalties), it offers a new take on the relative value of permanent versus repeated, temporary programs, focusing on how such programs reinstate those who have already decided to evade i.e. (ii) in the taxonomy of Le Borgne and Baer (2008). In this context, when the value to evasion is persistent and changes over time, it is sub-optimal to offer a static program and a cyclical program can provide stronger incentives for agents to self-report.

Andreoni (1991), in a one-shot setting, argues that a permanent partially forgiving voluntary disclosure program can be valuable as it provides agents with insurance for negative income shocks. From a design perspective, the model presented in this paper suggests that such a program may be further improved when agents’ values for tax evasion are imperfectly persistent, by offering only occasional opportunities for voluntary disclosure.

6. Generalizing the Arrival Distribution

In this section, I assume that the regulator faces a stream of agents arriving at time-inhomogeneous rate $e^{-\gamma t}$ for some $\gamma \in [0, \infty)$. The model studied in Section 3 corresponds to $\gamma = 0$, while $\gamma > 0$ corresponds to a setting in which the distribution of arrival is weighted towards time 0. I show that when $\gamma < \rho$, the main theorem of Section 3 still holds; an optimal policy consists of amnesty cycles that take the form described in Theorem 2. When

\[27\text{One of the purposes of that paper is to rationalize some permanent programs observed in the United States and Canada.}\]
instead $\gamma > \rho$, a new optimal policy can be described as follows: after an initialization period as in Theorem 2, the regulator offers an interval with an increasing self-reporting penalty, and after this interval offers a fixed penalty forever.

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

**Assumption 1.** $\underline{p} = x' = 0$

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

$$V_{\gamma}^* \equiv \sup_{(p, a) \in \mathcal{M}} V_{\gamma}(p, a).$$

The steps for proving Theorem 2 apply with little adjustment to $V_{\gamma}^*$, as long as $\gamma \leq \rho$.

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

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

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

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

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

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

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

$$t^I \equiv \ln \left( \frac{x^h - (p+\sigma)x^h}{\rho+x^h+\lambda} \right) \frac{1}{\rho+r}$$

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

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

- $p_t^* = e^{-(\rho+r)(t_0-t)}p + (1 - e^{-(\rho+r)(t_0-t)})\left(\frac{\underline{p}}{\rho+r}\right)$ for $t < t_0$
\[ p_t^* = (e^{(\rho + r)(t - t_0)} - 1)^{\frac{\rho}{\rho + r}} \] if \( t_0 \leq t \leq t_0 + t^I \) and
\[ p_t^* = \frac{\rho \bar{p}}{\rho + r} \] for \( t \geq t_0 + t^I \).

\[ a_t^*(x^h) = 1 \ if \ and \ only \ if \ t_0 \leq t \leq t_0 + t^I \]
\[ a_t^*(x^l) = 1 \ for \ all \ t \]

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

![Figure 3: An Example of the Optimal Policy in Proposition 4](image)

7. 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 many 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 2, except that agents make a once-and-for-all decision whether to begin committing crime at the moment they arrive to the model. In this case, it is straightforward to show that the optimal policy...
is to either never offer an amnesty or offer a static amnesty that induces reporting by low return criminals only. The reason is that never offering amnesty effectively gives an agent a “once-and-for-all” opportunity to avoid penalties at the moment of arrival by not engaging in crime. If such a policy deters high return crime, then the regulator achieves her first-best. If instead agents in the high return state choose to begin committing crime then there is no policy that can induce reporting by high return agents. As a result, the regulator focuses her policy on inducing low return agents to report, which can be accomplished with a static policy.

To jointly study deterrence and the amnesty granting decision then, it is necessary to enrich the model. A simple modification of the model can restore the value of dynamic amnesty policies even in the presence of deterrence. In particular, introduce a third, very high state in which the agent arrives and chooses whether to become a criminal. If this third state is sufficiently high, and transition from this state to the lower states can never be reversed, then the results of Section 3 remain unchanged. A more satisfactory model with deterrence would have multiple states in which an agent can arrive to the model and decide whether to become a criminal; agents arriving in the highest states cannot be deterred from crime, while agents arriving in the intermediate and low states can be. As agents transition from the high states down into the intermediate and low states, they engage with self-reporting policies. In this model, the 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. This can happen when the distribution of arriving values is concentrated below the level deterred by shutting down self-reporting programs. When the main motivation is ex-post detection, then self-reporting programs like the ones studied in Section 3 will be useful. This can happen when the distribution of arriving values is concentrated above the level deterred by shutting down self-reporting programs.

I investigate the deterrence issue further in Appendix A.2, and show that the insights developed can be partially extended to a model that allows the regulator to express a deterrence motive.

**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. Two versions of this interpretation, suitable for different environments, are offered below.

1. First, in some settings the time criminals begin committing crime is perfectly observed. In this case, the model applies to situations in which potential criminals can initiate
crime in a very high return state \((x_{hh})\), and arrival to the model is the time that returns reach an intermediate level \((x_h)\). If returns are private, arrival time to the model is then naturally viewed as private information of the criminal.\(^{28}\)

Consider, for instance, the case of military desertion. A military will know, within a few hours, that a soldier has deserted. Therefore, the military can condition on desertion time when designing the amnesty program. For instance, the military could say that anyone who returns within two days will not be labeled a deserter and will be punished only mildly. Such programs could entice deserters who quickly decided they made a mistake, but not those who made calculated decisions and chose to remain deserters for longer periods.\(^{29}\) In the Russian Civil War, there is no 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).

To formalize this interpretation of the model, (i) suppose the regulator chooses a time path for amnesty conditional on the time of initial commission of a crime and (ii) crime is initiated in state \(x_{hh} > x_h\), where \(x_{hh}\) is so large that the agent cannot be deterred by any policy. This last condition is a reflection of the weak enforcement environment. At some Poisson rate, \(\gamma\), the agent transitions from \(x_{hh}\) to \(x_h\) and never returns to \(x_{hh}\). While the regulator is allowed to condition her amnesty path on the time at which the agent begins crime, she does not observe the time of transition from \(x_{hh}\) to \(x_h\). This version of the model is formally equivalent to the model described in Section 6, and so the results from that setting can be applied.

2. In other cases, it may be more natural to suppose that the regulator observes neither the private return nor the actual time at which the crime begins. This is the natural interpretation in, for instance, the gun amnesty and buyback case. In this case, we can again introduce a very high state \(x_{hh}\) in which crime is initiated and which cannot be deterred. The agent transition from \(x_{hh}\) to \(x_h\) at some rate \(\gamma\). Since the regulator does not observe timing of the initial commission of crime, the analysis is identical to that in Section 3.

Each of these modeling approaches accord with the view that enforcement is too weak to deter crime ex ante and so must focus on detection ex post through self-reporting. Nevertheless,

\(^{28}\)As previously mentioned, more complex mechanisms could elicit such private information.

\(^{29}\)For those, enforcement is too weak and offering an immediate amnesty is like the military asking them not to desert, which it already does.
deterrence is a first-order concern in most settings, and so I discuss it in greater depth in Appendix A.2.

**Exogenous Detection Penalty.** The penalty the agent pays when exogenously detected is $p$. In this setting, if the regulator could freely choose to set the penalty when detected at any point in $[p, \bar{p}]$, she would always set the maximum possible, $\bar{p}$. As a result, it is without loss of generality to suppose penalty upon detection is $\bar{p}$.

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

**Initial Distribution of Values.** Arriving agents are initialized in state $x^h$. Results are unchanged if I allow for a time-independent distribution of arriving values across the high and low states. As I have shown, an optimal policy induces low value agents to report either always or never. Allowing for a time-independent distribution of arriving values then just scales the regulator’s problem by a constant factor, leaving the optimal policy unchanged.

**Uncontrollable Rate of Detection.** In the model, the regulator does not control the rate of detection. In Wang et al. (2016), a self-reporting problem is studied in the context of environmental regulation in which the regulator controls the rate of detection. In that setting, in the absence of dynamic values, the paper shows that manipulating the risk of detection can amplify the value to dynamically adjusted self-reporting penalties. The model presented in this paper then fulfills a role complementary to Wang et al. (2016) by showing how self-reporting programs should behave in the absence of inspection control but in the presence of dynamic values to crime. It addresses the design of self-reporting programs in settings in which the rate of inspection is indeed more difficult to control, unlike the environmental audit setting considered in Wang et al. (2016). For instance, in the case of price-fixing cartels, much of detection comes from buyer complaints which the anti-trust authorities do not directly control (Harrington, 2005). In the case of the anti-desertion campaigns in the Red Army, enforcement was locally delegated but deserters could be caught anywhere or discovered by people other than those tasked with explicit enforcement (Wright, 2012). In general, at least some detection typically comes from third-party reporting which the regulator cannot directly control.
Absorbing Low State. The model does not allow for the possibility that agents in the low state transition back to the high state. This assumption is made for tractability. Lemma 1 generalizes to this case and so a policy similar to the policy in Theorem 2 is approximately optimal when transitions back to the high state are infrequent, with the loss relative to optimality shrinking with the size of the transition.

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

In the case of tax amnesty, where one of the main motivations is short-term revenue, this issue is salient. One way to incorporate revenue considerations is to generalize the regulator’s objective function to be a weighted combination of the loss from tax evasion ($x_t$) and the profit from penalties. When the weight on collected penalties is equal to the weight on the loss from crime, the game between regulator and agents is zero-sum: in this case the regulator minimizes the agents’ value, which is achieved by setting $p_t = \bar{p}$ for all $t$. When instead the regulator places a lower weight on the profit from penalties, there is scope for self-reporting to benefit the regulator. When penalties are financial, a lower weight represents the cost of collecting penalties and proving guilt, which is administratively expensive (Franzoni et al., 1996). When the penalty represents prison time, a non-positive weight on penalties is apt. Although the proof of Theorem 2 does not generalize to this setting, the main force at work remains intact.

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

In other cases, it is more natural to ignore the profits from collected penalties. In cartels, for instance, the penalties may be viewed as pure transfers, while anti-competitive behavior represents a dead weight loss (Motta and Polo, 2003). Valuing penalties from illegal gun ownership may be accommodated in a way similar to tax evasion, with different weights on

---

$^{30}$See Le Borgne and Baer (2008) for a discussion of this issue.
penalties and behavior. But, gun ownership generates externalities through misuse, theft and unregulated sale (Cook and Ludwig, 2006); these are not internalized by the gun owner in weak enforcement environments by definition. In this case, the weight the regulator places on collected penalties may be small. Gun ownership penalties also involve prison time, which should not be treated positively in the regulator’s objective function.

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

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

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

**Uniqueness.** The policy described in Theorem 2 when \( \theta \in \Theta^* \) and \( \alpha_l > 0 \) is the unique optimal policy among all policies \((p, a)\) for which \( \{t|a_t(x^h) = 1\} \) has at least one isolated point not equal to \( p \). This follows from the proof of Theorem 2. The policy described in
Proposition 4, in the case $\theta \in \Theta$ and $\alpha_l > 0$, is the unique optimal policy among all policies $(p, a)$ for which any limit point of $\{t|a_t(x^h) = 1\}$ is a member of an interval.

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

**Deterrence Without Dynamics.** As mentioned in Section 5 in the discussion on desertion, a deterrence motive alone is not sufficient to generate dynamics in optimal amnesty. Indeed, even by enriching the model to allow agents to have more than two possible values from committing crime, a static amnesty policy is optimal. The fundamental reason is that without variation in the value from committing crime, the downsides of static policies disappear: agents do not value the option of reporting in the future.

**Increasing Penalties.** A consideration missing from the model is that, in cases when the severity of crime can be freely learned by the regulator, penalties increase with the severity of the crime. In many cases, there are upper bounds on the severity of crime that a regulator can impose on criminals. 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.\(^{31}\) Long prison sentences may be subject to similar constraints, and recent scholarship (albeit in a very different context) suggests that there are diminishing returns in deterrence to longer prison sentences. In such cases, maximum punishments may increase up to a point, after which they are forced to flatten, either due to political constraints or psychological factors. In other cases, such as illegal gun ownership, the regulator has no information on the length of ownership and so punishment cannot be conditioned on this dimension of severity.

Nevertheless, introducing a maximum punishment that increases in the severity of crime provides further impetus for time-variation in amnesty policy. In particular, a deterministic, version of the logic takes hold: with a static policy, agents would time their reporting to exactly the moment that the marginal benefit of continuing is outweighed by the marginal cost, while dynamic policies would force agents to report earlier.

---

\(^{31}\)See, for instance, the support of the peasantry during the Russian Civil War discussed in Wright (2012).
8. Conclusion

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

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

There are a number of avenues for future work. First, it would be useful to study a version of the problem in which the regulator can control, at some cost, the rate of detection. Second, new insights might result from incorporating political economy constraints into the model. Third, it would be interesting to further examine the deterrence margin beyond what has been discussed in Section 7. For instance, when the regulator cannot condition her policy on the time at which crime begins, it becomes clear that randomization can be useful — by randomizing the timing of amnesty, agents cannot take advantage by initiating crime at times close to attractive amnesties. In that case, how should the regulator randomize amnesty?

References


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


Appendix

A. Extensions

In this section I discuss a number of extensions to the baseline model.
A.1 Poisson Randomization

Until now, I have restricted the regulator to deterministic policies. Although I do not characterize the optimal policy for general random policies, I expand the model to allow for a limited class of random policies and show that the deterministic optimal policy remains optimal. Extend $V$ linearly to random policies.

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

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

The set of $\gamma$-Poisson policies for any $\gamma > 0$ is denoted $\Gamma$.

These policies feature inter-arrival times of minimum penalties that are exponentially distributed with mean $\frac{1}{\gamma}$. I restrict to the setting in which $\bar{p} = x^d = \alpha_1 = 0$ and argue that the policy in Theorem 2 remains optimal when allowing the regulator to choose from $\Gamma$. Let $\mathcal{M}^\Gamma \equiv \mathcal{M} \cup \Gamma$.

$$V^\Gamma \equiv \sup_{(p,a) \in \mathcal{M}^\Gamma} V(p, a)$$

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

I prove the result below for the case of $\alpha_t = x^d = \bar{p} = 0$, but it extends readily to the general case.

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

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

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

The proof is given in Online Appendix G. Random mechanisms in $\Gamma$ which are incentive compatible for the agent require putting substantial probability on short inter-arrival times of amnesty, relative to the deterministic optimal policy in Theorem 2. For the same reason as in Section 3, this is relatively more costly for the regulator than the agent.

A.2 Deterrence

In this section, I allows agents to decide whether to begin committing crime.
Model. To formally allow for a decision to enter crime, I study exactly the model of Section 2 but introduce a third state for values, \( x^{hh} > x^h \). This state \( x^{hh} \) is so high that agents can be neither induced to self-report nor deterred from entering using any policy. Upon arrival to the model, agents make a once-and-for-all decision whether to begin committing crime and agents can arrive in states \( x^h \) and \( x^{hh} \). To this end, let \( \mu^0 \) be a time independent arrival distribution across states \( \{x^{hh}, x^h\} \). If \( \mu^0 \) places probability 1 on \( x^h \), or equivalently if \( x^{hh} = x^h \), a static policy is optimal. If \( p_t = \overline{p} \), then giving agents the option to not begin committing crime is like offering a once-and-for-all amnesty. If this does not deter \( x^h \) types from entering, then no self-reporting policy can induce self-reporting by \( x^h \) types after they have entered. As a result, either the static policy \( p_t = \overline{p} \) for all \( t \) deters \( x^h \) types from entering, or it does not and the static policy \( p_t = p \) for all \( t \) is optimal.

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

Assumption 2. \( \overline{p} = x^l = 0 \).

Agents can transition only from \( x^{hh} \) to \( x^h \) or from \( x^h \) to \( x^l \). Transitions from \( x^{hh} \) to \( x^h \) occur at rate \( \lambda_{hh} \) and, as before, transitions from \( x^h \) to \( x^l \) occur at rate \( \lambda \).

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

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

Assumption (3) guarantees that agents arriving in state \( x^{hh} \) always begin committing crime, even under the least forgiving policy that sets \( p_t = \overline{p} \) for all \( t \). Assumption (4) guarantees that an agent in state \( x^h \) can be induced to report, but not by a static policy.

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

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

Since \( x^h \in (\rho \overline{p}, (\rho + r + \lambda) \frac{\rho \overline{p}}{\rho+r}) \), there exists a \( v \in [\rho \overline{p}, \overline{p}] \) such that agents arriving in state

\(^{32}\)As in the baseline model of Section 2, allowing for arrival in state \( x^l \) does not change results.
$x^h$ choose not to enter if $p_t \geq v$. Let $v$ be the smallest such $v$. Then, as long as $p_t \geq v$, agents in state $x^h$ will never choose to begin committing crime.

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

![Figure 4: A optimal policy when $p_t \geq v$.](image)

The policy depicted in Figure 4 can be thought of as a version of the optimal policy in Theorem 2 but shifted upward to guarantee deterrence of intermediate value crime.

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

It is straightforward to find numerical examples that can achieve the same level of deterrence as in the cyclical policy above — agents in states $x^h$ never begin committing crime — but improve the regulator’s value. For instance, consider the Poisson random policies studied in Appendix A.1. Any such policy that satisfies the agent’s incentive compatibility

42
condition also deters agents in state $x^h$ from entering, because the arrival of an amnesty is i.i.d. over time. The description of an optimal policy in this setting with fully general random policies is left open.

B. Proof of Lemma 1

Proof of Lemma 1: Fix a policy $(p, a) \in \mathcal{M}$. Observe first that the alternative policy $(\tilde{p}, \tilde{a})$ defined by

$$
\tilde{a}_t(x^h) \equiv a_t(x^h) \\
\tilde{a}_t(x^l) \equiv a_t(x^h) \\
\tilde{p}_t \equiv (1_{a_t(x^h)=1}) p_t + (1 - 1_{a_t(x^h)=1}) \overline{p}
$$

is also 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 $(p, a)$ with $a_t(x^h) = a_t(x^l)$ and $p_t = \overline{p}$ if $a_t(x^h) = 0$, there is a policy $(\tilde{p}, \tilde{a}) \in \mathcal{L}$ such that $a_t(x^h) = \tilde{a}_t(x^h)$. To this end, fix such a policy $(p, a)$ and consider the alternative policy, $(\tilde{p}, \tilde{a})$, defined by

$$
\tilde{a}_t(x^l) \equiv 1, \\
\tilde{a}_t(x^h) \equiv a_t(x^h), \\
\tilde{p}_t \equiv -W^*(x^l, t, p).
$$

The definition of $W^*$ and the assumption that $0 \leq \Delta_t = \rho \overline{p} - (\rho + r) \underline{p} - x_t$ imply that $\tilde{p}_t \in [\underline{p}, \overline{p}]$.

I first argue that $\tilde{p}$ is measurable. Let

$$
t^n(t) \equiv \inf_{s \geq t} \left\{ s \mid \tilde{a}_s(x^h) = 1 \right\}
$$

which is measurable since $\tilde{a}$ is. Then, observe that

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

which is measurable since $t^n(t)$ is.

I now argue that the recommendation policy $\tilde{a}$ is obedient. Fix an arbitrary stopping policy $\tau$ for an agent arriving at time $t$. Let

$$
\sigma \equiv \tau 1_{a_\tau(x^h)=0} + \infty(1 - 1_{a_\tau(x^h)=0}).
$$

The random variable $\sigma$ takes the value $\tau$ if an agent stops when $a$ recommends against
stopping in state $x^h$ and infinity otherwise. Then,

$$W(x, t, \tau, \bar{p}) = \mathbb{E} \left[ w(x, \tau) + e^{-r(\tau \wedge \tau_0)} (1_{\tau < \tau_0} W^*(x^l, \tau, \bar{p})) \right]$$

$$= \mathbb{E} \left[ w(x, \tau) - e^{-r(\tau \wedge \tau_0)} (1_{\tau < \tau_0 \wedge \sigma} p_{\tau} - 1_{\sigma \leq \tau_0 \wedge \sigma} W^*(x^l, \sigma, \bar{p})) \right]$$

$$\leq \mathbb{E} \left[ w(x, \tau) - e^{-r(\tau \wedge \tau_0 \wedge \sigma)} (1_{\tau < \tau_0 \wedge \sigma} p_{\tau} - 1_{\sigma \leq \tau_0 \wedge \sigma} W^*(x^l, \sigma, \bar{p})) \right]$$

$$\leq W^*(x, t, p)$$

where the third line is a result of the fact that $x^l \leq x_\sigma$. Conversely, if an agent arriving at time $t_0$ uses the strategy $\tau \equiv \inf_{t \geq t_0} \{t - t_0 | a_{t_0+t}(x^h) = 1\}$, the agent guarantees himself $W^*(x, t, p)$, and so I conclude that $W^*(x, t, p) = W^*(x, t, \bar{p})$ for $x \in \{x^h, x^l\}$. As a result, $\bar{a}$ is an obedient recommendation policy such that $\bar{a}_t(x^l) = 1$ for all $t$ and $W(x^l, t, 0, \bar{p}) = W(x^l, t, t^h, \bar{p})$. □

C. Proof of Lemma 2 and Theorem 1

Proof of Lemma 2: Plugging in the definition of $\Delta_t$, we find that

$$\Delta_t = \frac{\rho \bar{p} - x^l}{p + r} - p.$$ \hspace{1cm} (C.1)

This expression will be useful in what follows.

By Proposition 1, it is without loss of generality to suppose that, for static penalty policies $p^v$, the regulator’s recommendation is constant i.e. $a_t(x) = a_s(x)$ for each $t, s \geq 0$ and $x \in \{x^h, x^l\}$. A stopping time induced by obedient recommendation policy $a$ must replicate the value an agent places on one of the three stopping times: (i) $\tau^0 \equiv 0$, (ii) $\tau^\infty \equiv \infty$ and (iii) $\tau^l \equiv \inf \{t | x_t = x^l\}$. Observe that, for the case of $x_t$ considered here with $x^l = 0$, the regulator is indifferent between policies $\tau^\infty$ and $\tau^l$.

Fix an arbitrary set of parameters $\theta$. The values under penalty policy $p^2$ of the three stopping times defined above for an agent arriving at $t_0$ in state $x$ are

$$W(x, \tau^l, t_0, p^2) = \mathbb{E}_{x_0 = x} \left[ w(x, t_0) - e^{-(\rho + r)\tau^l} p \right]$$

$$= 1_{x = x^h} \left( \frac{x^h - \rho \bar{p} - \lambda p}{\rho + r + \lambda} \right) + 1_{x = x^l} (-p)$$

$$W(x, \tau^0, t_0, p^2) = -p$$

$$W(x, \tau^\infty, t_0, p^2) = 1_{x = x^h} \left( \frac{x^h - x^l}{\rho + r + \lambda} \right) - \frac{\rho \bar{p} - x^l}{\rho + r}$$
Suppose first that \(x^h - x^l \leq (\rho + r)\Delta l\). Then,

\[
\max \left\{ W(x, \tau^\infty, t_0, p^2), W(x, \tau^l, t_0, p^2) \right\} \leq W(x, \tau^0, t_0, p^2)
\]

for \(x \in \{x^h, x^l\}\). As a result, the recommendation policy \(a\) with \(a_t(x) = 1\) for all \((t, x)\) is an obedient recommendation policy. The regulator achieves her highest possible value, \(V^* = 0\), with \(p^2\) and \(a_t(x) = 1\) for all \((t, x)\), and so a static policy is optimal.

Suppose now that \(x^h - x^l > (\rho + r + \lambda)\Delta l\). The agent’s value for \(\tau^\infty\) when \(x = x^h\) is

\[
W(x^h, \tau^\infty, t_0, p^2) = \frac{x^h - x^l}{\rho + r + \lambda} - \frac{\rho p - x^l}{\rho + r} = \frac{x^h - x^l}{\rho + r + \lambda} - (p + \Delta l) \geq -p
\]

where the second line follows by plugging in (C.1) and the third line follows by our assumption \(x^h - x^l > (\rho + r + \lambda)\Delta l\). Since the value of \(\tau^\infty\) is larger than the value of \(\tau^0\) for penalty \(p\), any obedient recommendation policy must have \(a_t(x^h) = 0\) for all \(t\). In this case, I claim that the static policy \(p^2\) is optimal. If \(\Delta l \geq 0\), then

\[
\max \{W(x^l, \tau^\infty, t_0, p^2), W(x^l, \tau^l, t_0, p^2)\} \leq W(x^l, \tau^0, t_0, p^2).
\]

If instead \(\Delta l < 0\), then \(W(x^l, \tau^\infty, t_0, p^2) > -p\). Since \(W(x^l, \tau^\infty, t_0, p^2)\) is independent of the penalty policy and \(p\) is the minimum possible penalty, no penalty policy can ever induce reporting by low return agents. As a result, all policies deliver the same value to the regulator and so any static policy is optimal. \(\square\)

Before proving Theorem 1, I prove an intermediate result when \(x^l = \alpha_l = 0\). Define

\[
\tilde{\Theta} \equiv \left\{ (\rho, r, \lambda, x^h, 0, \overline{p}, p) \left| \frac{\rho + r + \lambda}{\rho + r} (\rho \overline{p} - (\rho + r)p) \geq x^h > \rho \overline{p} - (\rho + r)p \right. \right\}.
\]

**Proposition C.1.** Suppose \(\alpha_l = 0\). Then,

\[
\Theta^* \cap \left\{ \theta \left| x^l = 0 \right. \right\} = \tilde{\Theta}.
\]

**Proof.** First, note that Lemma 2 implies that

\[
\Theta^* \subseteq \tilde{\Theta}.
\]

I now show the reverse inclusion,

\[
\tilde{\Theta} \subseteq \Theta^*.
\]
Fix an arbitrary $\theta \in \bar{\Theta}$. By Proposition 1 it is sufficient to demonstrate a policy that strictly improves over $p^\theta$. To that end observe that,

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

where the inequality follows by the assumption that $\theta \in \bar{\Theta}$. In this case then, the regulator receives his 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 $(p, a)$: (i) $p_t = 1_{t=T} \bar{p} + (1 - 1_{t=T})\bar{p}$ for some $T > 0$, (ii) $a_t(x) = 1_{t=T}$ for each $x \in \{x^h, x^l\}$. Then, observe that,

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

where the inequality follows by assumption that $\theta \in \bar{\Theta}$. Thus, the recommendation $a_t(x) = 1$ if and only if $t = T$ is obedient. Since $T > 0$, this policy induces a strictly positive mass of high types to stop by $T$, generating a strict improvement of the regulator’s value over any static policy. So I find,

$$\bar{\Theta} \subseteq \Theta^*$$

and the proof is concluded by combining this with the reverse inclusion. \hfill \square

**Proof of Theorem 1:** Suppose now that $x^l > 0$ or $\alpha^l > 0$. Suppose first that $\Delta_l < 0$. Then, $W(x^l, \tau^\infty, t, p) > -p$ for any $t, p$. In that case, the only obedient recommendation is $a_t(x^h) = a_t(x^l) = 0$ for all $t$, in which case the policy $p$ has no effect on behavior. Because of this,

$$\Theta^* \cap \left\{ \theta \left| \Delta_l < 0 \right. \right\} = \emptyset.$$

Suppose now that $\Delta_l \geq 0$. By Lemma 1,

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

Since the right hand side is independent of $\alpha_l$, it is without loss of generality to prove the result for $\alpha_l = 0$. To this end, for any $\theta = (\rho, r, \lambda, \bar{p}, p, x^h, x^l)$, let $\bar{\theta}(\theta) \equiv (\rho, r, \lambda, \bar{p}, p, \bar{x}^h, \bar{x}^l)$ where $\bar{x}^h = x^h - x^l$, $\bar{x}^l = 0$ and $\bar{p} = \bar{p} - \frac{x^l}{p}$. An agent’s value for stopping time $\tau$ can then
be re-written,
\[
W(x, t_0, \tau, p) = \mathbb{E} \left[ \int_0^\tau e^{-(\rho+r)t}x_tdt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho + r} p - e^{-(\rho+r)\tau} p_{r+t_0} \right]
\]
\[
= \mathbb{E} \left[ \int_0^{\tau \wedge \tau} e^{-(\rho+r)t}(x^h - x^l)dt + \int_0^\tau e^{-(\rho+r)t}(x^l)dt - (1 - e^{-(\rho+r)\tau}) \frac{\rho}{\rho + r} p - e^{-(\rho+r)\tau} p_{r+t_0} \right]
\]
\[
= \mathbb{E} \left[ \int_0^\tau e^{-(\rho+r)t}\tilde{x}_tdt - (1 - e^{-(\rho+r)\tau}) \frac{\rho p}{\rho + r} - e^{-(\rho+r)\tau} p_{r+t_0} \right]
\]
where \(\tilde{x}_t = (x^h - x^l)1_{x_t = x^h} + x^l1_{x_t = x^l}\). So, an agent’s value for a stopping time under any policy \(p\) is the same across parameterizations \(\theta\) and \(\tilde{\theta}(\theta)\). As a result,
\[
\theta \in \Theta^* \iff \tilde{\theta}(\theta) \in \Theta^*.
\]
By Proposition C.1,
\[
\tilde{\theta}(\theta) \in \Theta^* \iff \tilde{x}^h \in \left(\rho \tilde{p} - (\rho + r)p, \frac{\rho + r + \lambda}{\rho + r} (\rho \tilde{p} - (\rho + r)p)\right)
\]
\[
\iff x^h - x^l \in \left(\rho \tilde{p} - (\rho + r)p - x^l, \frac{\rho + r + \lambda}{\rho + r} (\rho \tilde{p} - (\rho + r)p - x^l)\right)
\]
\[
\iff x^h - x^l \in \left(\rho + r)\Delta^l, (\rho + r + \lambda)\Delta^l\right)
\]
and the result follows. \(\square\)

### D. Proofs of Section 3.4 and Theorem 2

Recall that \(\mathcal{P} = [p, -W(x^h, 0, \tau^\infty)]\). Plugging in the definition of \(W(x^h, 0, \tau^\infty)\),
\[
\mathcal{P} = \left[ \frac{\rho \tilde{p} - x^l}{\rho + r} - \frac{x^h}{\rho + r + \lambda} \right].
\]
Before proceeding with proofs of Section 3.4, I provide a useful calculation of \(\mu^h_t\), the measure of high return agents at time \(t\), which is proved in Appendix F.

**Lemma D.1.** Fix \(t \in \mathbb{R}_+\) and \(\bar{t} < t\) with \(a_{\bar{t}}(x^h) = 1\). If \(a_s(x^h) = 0\) for all \(s \in (\bar{t}, t]\) then,
\[
\mu^h_t = \frac{1 - e^{-(\rho+\lambda)(t-\bar{t})}}{\rho + \lambda}.
\]
Proof of Lemma 3: For any policy \((p, a) \in \mathcal{M}\), define \(\tilde{p}_t \equiv (1_{a_t(x^h) = 0}) \overline{p} + (1 - 1_{a_t(x^h) = 0}) p_t\) and \(\tilde{a}_t(x) \equiv a_t(x^h)\) for \(x \in \{x^h, x^l\}\). The resulting policy, \((\tilde{p}, \tilde{a})\), satisfies the constraints on the right-hand side of Equation (1) and delivers the regulator the same value as \((p, a)\) when \(\alpha_t = 0\). Conversely, any policy that satisfies the constraints on the right-hand side of Equation (1) is an element of \(\mathcal{M}\), and the result follows.

Proof of Lemma 4: For any policy \((p, a) \in \mathcal{M}\), let \(\mathcal{T}(a) \equiv \{t | a_t(x^h) = 1\}\). Let \(\mathcal{M}^0 \subset \mathcal{M}\) be the set of policies \((p, a)\) such that

(i) \((1 - a_t(x^h)) \overline{p} = (1 - a_t(x^h)) p_t\) and

(ii) \(\inf_{(t, s) \in (\mathcal{T}(a))^2} |t - s| > 0\).

Let \(t(a) \equiv (t_i(a))_{i \in \mathbb{N}}\) be the increasing sequence such that \(\bigcup_{i \in \mathbb{N}} t_i(a) = \mathcal{T}(a)\). I first show that

\[
V^* = \sup_{(p, a) \in \mathcal{M}^0} V(p, a). \tag{D.1}
\]

To see this, fix any policy \((p, a) \in \mathcal{M} \cap (\mathcal{M}^0)^C\). Choose recursively a sequence,

- \(\bar{t}_0 \in \left[ \inf \mathcal{T}(a), \varepsilon + \inf \mathcal{T}(a) \right] \cap \mathcal{T}(a)\)
- \(\bar{t}_{i+1} \in \left[ \inf (\mathcal{T}(a) \cap [\bar{t}_i + \varepsilon, \infty)) , \varepsilon + \inf (\mathcal{T}(a) \cap [\bar{t}_i + \varepsilon, \infty)) \right] \cap \mathcal{T}(a)\).

and generate policy \((\tilde{p}, \tilde{a})\) as follows:

- \(\tilde{p}_{\bar{t}_i} \equiv \overline{p} + (p_t - \overline{p})1_{t \in \bar{t}_i}, \forall i \in \mathbb{N}\)
- \(\tilde{a}_t(x) = 1_{t \in \bar{t}_i}, \forall x \in \{x^h, x^l\}\).

Observe that \((\tilde{p}, \tilde{a}) \in \mathcal{M}^0\). Since the regulator discounts at rate \(r > 0\) and \(\alpha_t = 0\),

\[|V(p, a) - V(\tilde{p}, \tilde{a})| \rightarrow 0\]

and Equation (D.1) follows.

To complete the proof, I will show that if \(\alpha_t = 0\) and \(V(p)\) satisfies the premise of the lemma with associated policies \((t(p), p'(p))\) then,

\[
\sup_{(p, a) \in \mathcal{M}^0} V(p, a) = \max_{t_0 \geq 0, p_0 \in \mathcal{P}} \left\{-v(t_0) + e^{-r t_0} V(p_0)\right\}. \tag{D.2}
\]

Applying Equation (D.1) then leads to the result.

To prove equation (D.2) holds, I first show that

\[
\sup_{(p, a) \in \mathcal{M}^0} V(p, a) \leq \max_{t_0 \geq 0, p_0 \in \mathcal{P}} \left\{-v(t_0) + e^{-r t_0} V(p_0)\right\}. \tag{D.3}
\]
For any policy \((p, a)\),

\[ V(p, a) = \sum_{i=0}^{\infty} e^{-rt_{i-1}(a)} t_i(a) - t_{i-1}(a) \int_{0}^{t_i(a) - t_{i-1}(a)} e^{-rt} \mu_t \, dt \]

\[ = \sum_{i=0}^{\infty} e^{-rt_{i-1}(a)} t_i(a) - t_{i-1}(a) \int_{0}^{t_i(a) - t_{i-1}(a)} e^{-rt} \frac{1 - e^{-(\rho+\lambda)t}}{\rho + \lambda} \, dt \]

\[ = \sum_{i=0}^{\infty} e^{-rt_{i-1}(a)} (t_i(a) - t_{i-1}(a)) \]

where \(t_{-1}(a) = 0\), the second line follows from Lemma D.1 and the third line follows from equation (2). Inequality (D.3) follows from the observation that \((p, a) \in \mathcal{M}^0\) if and only if

\[ w^h(t) - e^{-(\rho+r)(t_{i+1}(a) - t_i(a))} p_{i+1}(a) \leq -p_t(a) \]

for each \(i \in \mathbb{N}\).

Next, I argue that

\[ \sup_{(p, a) \in \mathcal{M}^0} V(p, a) \geq \max_{t_0 \geq 0, p_0 \in \mathcal{P}} \left\{ -v(t_0) + e^{-rt_0} V(p_0) \right\}. \]  \hspace{1cm} (D.4)

For any \(p_0 \in \mathcal{P}\), recursively define \(p^i(p_0) \equiv p_i(p^{i-1}(p_0))\) and \(p^0(p_0) = p_0\). For any \(p_0 \in \mathcal{P}\) and \(t_0 \geq 0\), recursively define \(t^i(p_0) = t^{i-1}(p_0) + t(p^{i-1}(p_0))\) and \(t^0(p_0) = t_0\). Then, for any choice \(p_0\) and \(t_0\) on the right-hand side of (D.4), define \((p, a)\) as follows,

\[ p_t = p + \sum_{i \in \mathbb{N}} 1_{t = t_i(p_0)} (p^i(p_0) - p) \] \hspace{1cm} (D.5)

\[ a_t(x) = \sum_{i \in \mathbb{N}} 1_{t = t_i(p_0)} \] \hspace{1cm} (D.6)

Then since \( \inf_{i \in \mathbb{N}} (t^i(p_0) - t^{i-1}(p_0)) > 0\), \((p, a) \in \mathcal{M}^0\). Further, repeatedly substituting yields \(V(p, a) = -v(t_0) + e^{-rt_0} V(p_0)\). As a result, inequality (D.4) is satisfied, and combining with (D.3) completes the proof. \(\square\)

Before proceeding to the proof of Proposition 2, I establish a few useful lemmas, which are proved in Online Appendix F.

**Lemma D.2.** If \(\theta \in \Theta^*\), \(t(p)\) defined by equation (4) exists, is unique and is strictly increasing on \(\mathcal{P}\). Further, \(\inf_{p \in \mathcal{P}} t(p) > 0\).
Lemma D.3. Suppose that \( \theta \in \Theta^* \). Then

\[
v(t) = e^{-rt} \frac{v^0(t)}{1 - e^{-rt(p)}}
\]

is decreasing in \( t \) for any \( t \geq t(p) \), where \( v^0(t) \) is defined by equation (6) and \( t(p) \) is defined by equation (4). Further,

\[
v^0(t(p)) = e^{-rt(p)} \frac{v^0(t(p))}{1 - e^{-rt(p)}}
\]

is decreasing in \( p \).

Let \( V^*(p) \) be the value function associated to policies \( p'(\cdot) = p \) and \( t(\cdot) \) defined by equation (4).

Proof of Proposition 2: To prove the result, I show that \( V^*(p) \) and \( (t(p), p'(p)) \) satisfy the premise of Lemma 4, i.e. equation (3) is satisfied for \( V^*(\cdot) \) and \( \inf_{p \in \mathcal{P}} t(p) > 0 \). The latter follows from Lemma D.2, so I need only show the former.

To this end, observe that

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

Plug in the definition of \((t(p), p'(p))\) to \( V^*(p) \), and rearrange to find

\[
-v(t) + V^*(p'(p)) = \frac{1}{r(\rho + \lambda)} + (v^0(t) + e^{-rt}V^*,0(p'(p)))
\]

where \( V^*,0(p) \equiv v^0(t(p)) + e^{-rt(p)} \frac{v^0(t(p))}{1 - e^{-rt(p)}} \). Plugging in (D.7) to equation (3), it is sufficient to show

\[
0 = \begin{cases} 
\sup_{t,p'} v^0(t) + e^{-rt}V^*0(p') - V^*,0(p) \\
\text{subject to} \\
w^h(t) - e^{-(\rho+r)t}p' \leq -p \\
p' \in \mathcal{P}
\end{cases}
\]

for any \( p \in \mathcal{P} \). To prove this, it is sufficient to argue that the right-hand side of (D.8) is attained at \( p' = p \) and \( t = t(p) \). To compute the value of the right-hand side of (D.8), it is sufficient to constraint to \( t \leq t(p) \). To see this, observe that for any \( t \geq t(p) \), an optimal choice of \( p' \) is \(-p\), since \( V^*,0(p') \) is decreasing in \( p' \) by the second part of Lemma D.3. By the first part of Lemma D.3, \( v^0(t) + e^{-rt}V^*,0(p) \) is decreasing in \( t \), so among choices \( t \geq t(p) \),
it is optimal to set \( t = t(p) \).

Suppose now that \( t \leq t(p) \). 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 \( V^{s,6}(p) \) and rearranging, that

\[
\phi(t(p)) + e^{-rt(p)}\phi(t(p)) + \frac{e^{-r(t(p)+t(p))}}{1 - e^{-rt(p)}}\phi(t(p)) \geq \phi(t) + e^{-rt}\phi(t(p)) + \frac{e^{-r(t+t(p'))}}{1 - e^{-rt(p)}}\phi(t(p))
\]

(D.9)

for any \( t, p, p' \) such that

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

(D.10)

Substitute for \( p' \) on the left hand side of (D.10) using the definition of \( t(p') \),

\[
\phi(t) + e^{-gt}\phi(t(p')) - c(1 - e^{-g(t+t(p'))}) \leq -\left( \frac{f}{x^h - x^l} \right) p
\]

(D.11)

where \( c \equiv \frac{f(\rho x - x^l - (\rho + r)p)}{(x^h - x^l)g} \). By definition, (D.11) holds with equality for \( (t, p') = (t(p), p) \), so (D.11) can be written

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

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

\[
\phi(t) + e^{-gt}\phi(t(p')) + \frac{e^{-g(t+t(p'))}}{1 - e^{-gt(p)}}\phi(t(p)) \leq \phi(t(p)) + e^{-gt(p)}\phi(t(p)) + \frac{e^{-g(t(p)+t(p))}}{1 - e^{-gt(p)}}\phi(t(p))
\]

(D.12)

So to prove the result, it is sufficient to show that for any \( t, p', p \) such that (D.12) holds, (D.9) holds as well. Observe that (D.12) and (D.9) are each special cases of the inequality,

\[
\phi(t) + e^{-at}\phi(t(p')) + \frac{e^{-a(t+t(p'))}}{1 - e^{-at(p)}}\phi(t(p)) \leq \phi(t(p)) + e^{-at(p)}\phi(t(p)) + \frac{e^{-a(t(p)+t(p))}}{1 - e^{-at(p)}}\phi(t(p))
\]

(D.13)

at \( a = \frac{g}{2} \) and \( a = \frac{f}{2} \), respectively.

For any \( t, p' \), let

\[
z \equiv e^{-ft(p)}, \quad z_p \equiv e^{-ft(p)}, \quad u \equiv e^{-ft}, \quad y \equiv e^{-ft(p')}
\]

(D.14)
By definition \( z \geq \max\{z_p, y\} \) and by assumption \( t \leq t(p) \). As a result, \( u \geq z_p \). Since \( t(p) \) is strictly increasing in \( p, y = z \) if and only if \( u = z_p \) or \( u = 1 \). In such cases, (D.12) and (D.9) hold at equality, and so I proceed under the assumption that \( 1 > u > z_p \) and \( y < z \). If \( z = 0 \), then \( \mathcal{P} \) is a singleton and there is nothing to prove, so I also proceed under the assumption that \( 1 > u > z_p \) and \( y < z \).

Plug the definitions in (D.14) into (D.13) and multiply both sides by \( 1 - z^a \) to get,

\[
(1 - u)(1 - z^a) + u^a(1 - y)(1 - z^a) + (uy)^a(1 - z) \leq (1 - z_p)(1 - z^a) + z_p^a(1 - z)(1 - z^a) + (z z_p)^a(1 - z). \tag{D.15}
\]

After rearranging and canceling terms,

\[
0 \leq (u - z_p) - u^a(1 - y) + (uz)^a(1 - y) - (uy)^a(1 - z) + z^a(1 - u) + (z_p z^a - z_p^a z) + (z_p^a - z^a)
\]

\[
\Rightarrow 0 \leq (u - z_p) + (uz)^a(1 - y) + z^a(1 - z) - (uy)^a(1 - z) - u^a(1 - y) - z^a(u - z_p)
\]

Denote the right-hand side of the inequality by \( h(a; u, y, z, z_p) \). The crucial step is the following claim:

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 \( 0 < a' < a \). \tag{C*}

A direct implication of this is that (D.12) implies (D.9), and so (D.8) 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,

\[
\tilde{h}(a) = \frac{(u - z_p)}{(uy)^a} + \left(\frac{z}{y}\right)^a(1 - y) + \left(\frac{z}{uy}\right)^a(1 - z) - (1 - z) - \left(\frac{1}{y}\right)^a(1 - y) - \left(\frac{z}{uy}\right)^a(u - z_p)
\]

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

To prove (C*), it is sufficient to show that \( \frac{\partial^2 \tilde{h}}{\partial a^2} \) 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 \( a \to -\infty \)

\[33\]Note that this multiplication does not change the direction of the inequality for since \( z \in (0, 1) \).
(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 \( \tilde{h} \)) requires the existence of points \( b_0 < b_1 < b_2 < b_3 \) such that \( b_0 \) and \( b_2 \) have strictly negative second derivative while \( b_1 \) and \( b_3 \) has strictly positive second derivative, the existence of which implies that \( \frac{\partial^2 \tilde{h}}{\partial a^2} \) intersects zero at least three times.

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

\[
\frac{\partial^2 \tilde{h}}{\partial a^2} = \frac{(u - z_p)(1 - y)\ln(z/y)}{(uy)^a} + \left(\frac{z_p}{uy}\right)^a(1 - z)\ln(z/y)^2 - \left(\frac{z_p}{u}\right)^a(1 - y)\ln(z/y)^2 - \left(\frac{z_p}{uy}\right)^a(u - z_p)\ln(z/y)^2
\]

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

\[
G(a) = \frac{(u - z_p)(1 - y)\ln(z/y)}{(uy)^a} + \left(\frac{z_p}{u}\right)^a(1 - z)\ln(z/y)^2 - \left(\frac{z_p}{u}\right)^a(u - z_p)\ln(z/y)^2
\]

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

\[
\frac{\partial G}{\partial a} = \frac{(u - z_p)(1 - y)\ln(z/y)^2\ln(z)}{(uy)^a} + \left(\frac{z_p}{u}\right)^a(1 - z)\ln(z/y)^2\ln(z)
+ \left(\frac{z_p}{u}\right)^a(u - z_p)\ln(z/y)^2
\]

Finally, \( \frac{\partial G}{\partial a} \) has the same number of zeros as \( J \equiv \frac{\partial^2 \tilde{h}}{\partial a^2} \). Differentiating \( J \),

\[
\frac{\partial J}{\partial a} = \frac{(u - z_p)(1 - y)\ln(z/y)^2\ln(z)}{(uy)^a} + \left(\frac{z_p}{u}\right)^a(1 - z)\ln(z/y)^2\ln(z)\ln(u)
+ \left(\frac{z_p}{u}\right)^a(z_p/y)\ln(z/y)^2\ln(z/y)\ln(z)
\]

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 \( \frac{\partial G(a)}{\partial a} \). As a result, \( G \) has at most two zeros, as does \( \frac{\partial^2 \tilde{h}}{\partial a^2} \).
Case 2 - $uy \geq z_p$: In this case, I show that for any $a \in (0,1)$, $h(a) < 0$, so that the result is vacuously true.\footnote{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 $p$. By the definition of $t(p)$, such a policy cannot be incentive compatible.} 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 last line 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 $G(z) \equiv \frac{\partial h}{\partial z_p} = z^a + \frac{a}{1-a}(1-z) - 1$ and differentiate to get $\frac{\partial G}{\partial z} = \frac{a}{1-a}z^{1-a} - \frac{a}{z^a}$. 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^a), $\frac{\partial h(a)}{\partial a}$ 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 $a \geq 0$.

\[
\frac{\partial h}{\partial a} = ln(uz)(uz)^a + ln(z_p)z_p^a(1 - z) - ln(u)u^a - ln(z)z^a(u - z_p) = z_p^a z_p(z_p)^a + ln(z_p)(1 - z) - ln(u)u^a - ln(z)z^a(u - z_p)
\]

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

\[
\frac{\partial h^1}{\partial a} = ln(\frac{uz}{z_p})(\frac{uz}{z_p})^a - ln(u)ln(\frac{u}{z_p})\left(\frac{u}{z_p}\right)^a - ln(z)ln(\frac{z}{z_p})\left(\frac{z}{z_p}\right)^a(u - z_p) = \left(\frac{uz}{z_p}\right)^a \left(\frac{ln(u)}{z_p} - ln(u)ln\left(\frac{u}{z_p}\right)\left(\frac{1}{z}\right)^a - ln(z)ln\left(\frac{z}{z_p}\right)\left(\frac{1}{u}\right)^a(u - z_p)\right).\]

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

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

\footnotetext[34]{}
\( \alpha_l = 0 \). Suppose \( \alpha_l = 0 \) and \( \theta \in \Theta^* \). The result follows by applying Proposition 2, observing that \( V^*(p) \) is maximized at \( p = \overline{p} \), so that
\[
V^* = \max_{t_0} \{ v(t_0) + e^{-r_0} V^*(p) \}
\]
and repeatedly substituting in for \( V^*(p) \).

\( \alpha_l \geq 0 \). Now, I move on to the case in which \( \alpha_l \geq 0 \). When \( \theta \notin \Theta^* \), an application of Proposition 1 leads to the result. Suppose instead that \( \theta \in \Theta^* \). To prove the result in this case, I transform the regulator’s problem into one with \( \alpha_l = 0 \), to which I will then apply the result in the case \( \alpha_l = 0 \).

Fix some parameters of the model, \( \theta \in \Theta^* \). Rather than studying problem \((P)\), consider
\[
V^*_h \equiv \sup_{(p,a) \in M} \int_0^t \mu^h_t \quad (P^h)
\]
This problem differs from \((P)\) only in that losses from the low type agent do not enter the objective function. Problem \((P^h)\) is simply the regulator’s problem when \( \alpha_l = 0 \). Therefore, an optimal policy in problem \((P^h)\) is \((\bar{p}^*, a^*)\) defined by:
- \( p^*_t = 0 \) for some \( t_0 \) with \( t(p) \) defined by Equation (4)
- \( p^*_t = \bar{p} \) otherwise
- \( a^*_t(x) = 1 \) if and only if \( t \in \{t_0, t_0 + nt(p)\} \) \( n \in \mathbb{N} \)

Since \( \theta \in \Theta^* \), Theorem 1 implies that \( 0 \leq (\rho + r)\Delta_l = \rho \bar{p} - x^l - (\rho + r)p \). So, we can apply Lemma 1 to transform \((\bar{p}^*, a^*)\) into \((\bar{p}^*, \bar{a}^*)\) \( \in L \) which has the properties:
- \( \bar{a}^*_t(x^l) = 1 \) for all \( t \geq 0 \)
- \( \bar{a}^*_t(x^h) = a_t(x^h) \)
- \( p^*_{nt(p)+t_0} = \bar{p} \) for \( n \in \mathbb{N} \) for some \( t_0 \geq 0 \)
- \( p^*_t = e^{-(\rho+r)t_nxt(t,a^*)} \bar{p} + (1 - e^{-(\rho+r)t_nxt(t,a^*)}) \frac{(\rho \bar{p} - x^l)}{\rho + r} \) for all \( t \notin \{t_0, t_0 + nt(p)\} \) \( n \in \mathbb{N} \)

where the last two lines translate the last requirement of an element in set \( L \) to the policy \((\bar{p}^*, a^*)\). Lemma 1 implies that
\[
V^* = \sup_{(p,a) \in L} V(p,a) = V^*_h
\]
and so \((\bar{p}^*, a^*)\) is an optimal policy. \( \square \)
E. Proof of Proposition 1

Proof of Proposition 1: Because $p^v$ is constant and hence continuous, Theorem 3 in Shiryaev (2007) can be applied to show that there exists some $D \subset \{x^h, x^l\}$ such that (i) $\tau^*_v \equiv \inf\{t \geq t_0| x_t \in D\}$ and (ii) if $\tau$ is any other optimal stopping time for the agent, then $\mathbb{P}(\tau^*_v \leq \tau) = 1$. As a result, it is without loss of generality for the regulator to restrict to recommendation policies $a$ 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 $\tau^*_v$.

To prove the lemma, it is thus sufficient to argue that $\tau^*_p \leq \tau^*_v$ for all $v \geq p$. Given the characterization of $\tau^*_v$ 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) $\tau^*_v = \tau^l \equiv \inf\{t| x_t = x^l\}$. The agent’s value for $\tau^\infty$ is independent of $v$. The agent’s values for $\tau^0$ and $\tau^l$ are

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

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

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

---

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