Inducing Good Behavior via Reputation*

Ayça Özdoğan† Mehmet Barlo‡

March 8, 2021

Abstract

This paper asks whether or not it is possible to induce agents to good behavior permanently via regulators' reputations and attain perpetual social efficiency. We propose and analyze a repeated incomplete information game with a suitable payoff and monitoring structure between a regulator possessing a behavioral type and an agent. We provide an affirmative answer when a patient regulator faces myopic agents: Reputation empowers the regulator to prevent agents' bad behavior in the long-run with no cost and, hence, attain the social optimum in any Nash equilibrium. These findings are robust to requiring short-lived agents to choose any one of their actions with an arbitrarily small but positive probability. On the other hand, we show that when both parties are long-lived and sufficiently patient, the limiting robust equilibrium cannot be close to perpetual good behavior. The contrast we attain demonstrates the significance of the interaction's longevity and exhibits a novel application of the theory of learning and experimentation in repeated games.

Journal of Economic Literature Classification Numbers: C73

Keywords: Reputation, repeated games, long-lived vs. short-lived agents, regulation.

*We are grateful to Drew Fudenberg, Aldo Rustichini and Larry Samuelson for their invaluable comments and suggestions. We also would like to thank Beth Allen, Sergiu Hart, Martin Hellwig, Johannes Hörner, Christoph Kuzmics, Jan Werner and the participants of LEG 2019 conference for the helpful discussions. All remaining errors are ours.

†TOBB University of Economics and Technology, Department of Economics, Söğütözü Cad. No:43 Söğütözü, Ankara, 06560, Turkey; +90 312 292 4543; aozdogan@etu.edu.tr.

‡**Corresponding Author**: Sabancı University, Orhanlı, Tuzla, 34956, Istanbul, Turkey; +90 216 483 9284; barlo@sabanciuniv.edu.

Table of Contents

1 Introduction **1**

2 Model **7**

3 Dynamic game with short-lived agents **10**

3.1 Nash equilibrium 12

3.2 Markov equilibrium 13

3.2.1 Possible extensions 15

4 Regulator faces a long-lived agent **17**

5 Conclusion **21**

A The proof of Lemma 1 **22**

B The proof of Theorem 1 **22**

C The proof of Theorem 2 **25**

D The proof of Theorem 3 **29**

D.1 The agent's beliefs about her own future behavior 29

D.2 The agent's beliefs about the regulator's future behavior 29

D.3 The regulator's beliefs about the agent's behavior 30

D.4 The agent's beliefs about her own future behavior – revisited 32

D.5 The agent's beliefs about the regulator's future behavior – revisited 32

D.6 The proof of Theorem 3 35

References **36**

1 Introduction

Many unfortunate events involve the misbehavior of agents facing a regulator tasked to audit/investigate if needed. Examples include an investor engaged in fraud by misrepresenting his books to a certified auditor, a construction or mining company neglecting work safety precautions and misreporting its practices, an employee not exerting the promised effort in a business owned by a principal, etc. Such instances are frequently related to the regulator's reputation for being diligent or the lack of it, and regulators' reputation concerns may prevent or lessen the extent and severity of such undesirable outcomes.[1,2]

This paper aims to unravel whether or not regulators' reputation can induce agents to "good" behavior permanently when their repeated interaction is neither observable nor contractable. We analyze a dynamic environment where the regulator (he) is responsible for detecting "bad" behavior via costly auditing yet may not be diligent because of the associated costs. To do this, we propose a repeated incomplete information setup under imperfect public monitoring with a stage game possessing a suitable payoff structure that is played between a regulator (who could be committed to being diligent or is strategic) and an agent (she).

First, we show that when the patient long-lived regulator faces a sequence of myopic agents who play only once and observe the public history of the past play, reputation empowers the regulator to prevent agents' bad behavior with no cost in any Nash equilibrium (NE). In fact, by inducing the agents to behave (well) in the long-run, the patient strategic regulator attains his *maximum* payoff, which coincides with the social optimum. To address

[1]For instance, Bernard Madoff was found guilty of several offenses, including fraud and false statements to the Securities and Exchange Commission (SEC). He began the Ponzi scheme in the early 1990s, yet he was arrested in late 2008 even though the SEC had previously conducted several investigations since 1992. SEC has been criticized for failing to act on Madoff fraud. The SEC inspector confessed: "Despite several examinations and investigations being conducted, a thorough and competent investigation or examination was never performed" (see "SEC criticized for failing to act on Madoff" at *http://business.timesonline.co.uk* by Seib and "Madoff Explains How He Concealed the Fraud" at *www.cbsnews.com*). Yet in another investment fraud charge, against Robert Allen Stanford in 2009, a report of the investigation by the SEC Office of the Inspector General shows that the agency has been following Stanford's companies for much longer and reveals a lack of diligence in the SEC enforcement (see h*ttp://www.sec.gov/news/studies/2010/oig-526.pdf*).

[2]The negligence of regulation may be associated with serious casualties. A mining accident took place in Soma, Turkey, which caused a loss of 301 lives in 2014. In response to a parliamentary question, The General Directorate of Mining Affairs of Turkey (GDMA) said that they could only afford to audit less than one-fourth of all the minefields annually. Meanwhile, many established NGOs (e.g., The Union of Turkish Bar Associations and The Union of Turkish Engineering and Architecture Associations) announced doubts and concerns about GDMA's governance practices in conjunction with this accident. In fact, during the criminal case associated with this accident, it became public information that an auditor of GDMA responsible for that particular mine was also employed by the company owning that mine as a technical supervisor (see Turkish newspaper page at *https://www.hurriyet.com.tr/ekonomi/somada-denetci-skandali-29868799*).

situations with large populations of many long-lived agents who are not able to coordinate on future behavior, rewards, and punishments, we consider the Markovian setting with a myopic (representative) agent.[3] We show that there exists a unique Markov equilibrium (ME) with a value function that is continuous and nondecreasing in the reputation for being diligent. The regulator's value function attains the maximum payoff at the absorbing reputation levels at which the agents exhibit good behavior while the regulator incurs no cost. All these findings are *robust* in the sense that requiring each agent to choose any of her actions with an arbitrarily small but positive probability does not alter these results qualitatively.

On the other hand, a contrasting conclusion emerges when the strategic regulator faces the same long-lived agent. The permanency of good behavior cannot be a robust equilibrium outcome with sufficiently patient players: We prove that, regardless of the initial beliefs, there is no NE in which the agent behaves (well) on average in the long-run on a positive probability set of histories while experimenting with the bad behavior every once in a while.[4]

Our findings display a disparity of robust limiting equilibrium behavior between the short-lived and the long-lived cases. Indeed, social efficiency is approximately sustained as a robust NE payoff when the patient strategic regulator faces myopic agents but not when he encounters a long-lived agent. Therefore, the current paper contributes to the theory of reputation by portraying the significance of the longevity of the interaction among the participants and providing a novel application of the theory of learning and experimentation in repeated games: In our setting, agent's good behavior corresponds to the absorbing case (because then no additional information could emerge and require updating of beliefs), while the strategic regulator would exploit this by refraining from costly auditing (thereby, sustaining efficiency); thus, the problem boils down to discouraging experimentation with the bad behavior in the case of perpetual interaction among the participants.

The repeated game between the regulator and the agent(s) involves unobservable actions on both sides and incomplete information about the regulator's type being *strategic* or *tough*.

[3] There are many such cases where the dismissal of intertemporal coordination among agents is plausible (e.g., a population of taxpayers facing a tax authority). Under some additional restrictions known in the literature, the resulting situation parallels the Markovian case involving a myopic representative agent.

[4] To ensure that the agent chooses each of her actions with a small but positive probability in every period, we discuss a setting where she suffers from one-period amnesia with some small but positive probability at the beginning of every period (in which case she hangs on to her low initial beliefs that the regulator is of commitment type). *Perfection* of Selten (1975) implies our notion of robustness. Sadly it is too powerful: the regulator (the informed player) would be forced to choose each of his actions with some arbitrarily small but positive probability as well. Besides, it creates non-trivial complications. Meanwhile, the *ant colony optimization* (ACO) techniques of computer science pioneered by Dorigo (1992) parallel with our robustness notion.

In the stage game, the agent's actions consist of good (*truthful*) and bad (*untruthful*) behavior while the regulator's of *diligent* and *lazy* actions. The regulator can detect agent's untruthful behavior with some probability determined by the audit quality only if he chooses the costly diligent action. If the regulator detects the agent's untruthfulness, the public signal associated with *detection* occurs. Otherwise, the absence of the public signal indicates that there is *no detection*. The stage game payoffs are so that the agent's best response to the regulator choosing to be diligent is to be truthful. Whereas it is to be untruthful if the agent believes the regulator chooses to be lazy. Meanwhile, the strategic regulator's best response is to be lazy when the agent is truthful and to be diligent if the agent is untruthful. The strategic regulator prefers the agent being truthful, and the agent prefers the regulator being lazy. The tough regulator (Stackelberg type) always chooses the diligent (Stackelberg) action. Hence, the Bayesian Nash equilibrium (BNE) of the stage game is in mixed actions for low values of the regulator's probability to be tough, and otherwise, the agent chooses the truthful action, and the strategic regulator is lazy. In the repeated game, all players observe past signals of detections, the *public history*. We refer to agents' updated beliefs about the regulator's type as the *regulator's reputation*.

Our objective is to analyze whether the strategic regulator can build up a reputation that induces the agent(s) to good behavior permanently. There is no correct model in terms of the longevity of the strategic interaction among the players. Some instances fit situations where the regulator faces different myopic agents each period, while others suit the regulator facing the same agent in every period. To provide an answer and novel insight, we analyze two extremes: (1) a long-lived regulator faces short-lived agents, each observing the public history; (2) a long-lived regulator faces a long-lived agent.

In the *first*, we establish that when the regulator is sufficiently patient, in every NE and for all interior initial common beliefs the agents may have about the regulator's type, in the long-run at almost every history, agents' behavior converges to choosing the truthful action in perpetuity. This finding follows from the result saying that any NE payoff of the patient strategic regulator tends to its maximum level in these cases. Hence, he enjoys a permanent reputation inducing agents' good behavior indefinitely, refrains from costly auditing, and attains perpetual social efficiency. In furtherance, we prove that there is a unique ME with a continuous and nondecreasing value function such that the reputation for being diligent becomes permanent whenever it exceeds a threshold. The reputation above this level implies all the future agents behave while the regulator is lazy permanently, and otherwise, players use mixed actions. As every ME is an NE of the dynamic game, we conclude that the perpetual social optimum is secured in ME as well.

The intuition behind these stems from the short-lived agents only caring about their short-run payoffs and giving myopic best responses to their updated public beliefs. They do not consider the information externality that they could initiate and be helpful to future generations. When an agent is truthful, Bayesian updating does not happen. Thus, the patient strategic regulator finds it optimal to ensure that his reputation eventually reaches a level above which it persists as all subsequent agents would find it optimal to be truthful thereafter. Therefore, good behavior is attained in perpetuity thanks to the patient regulator's reputation and the myopic agents' short-term incentives. Meanwhile, the patient strategic regulator guarantees his maximum payoff, strictly exceeding his Stackelberg returns, in any NE.

Additional complications arise when the regulator (he) faces a patient long-lived agent (she). Both make their choices and update their beliefs according to their private histories. The regulator cannot anticipate the long-lived patient agent's actions since her beliefs are private and she is not giving myopic best responses. In this setting, due to the lack of identifiability conditions, we do not know whether or not under NE, there is a sufficiently high reputation that blocks the avenue leading to aforementioned information externalities. However, even if there were such an NE, we prove it would not be robust. The patient agent would expose the patient strategic regulator's false reputation in the long-run if she were bound to experiment with the untruthful action every period with an arbitrarily small but positive probability. We formalize this notion of robustness via the concept of an α-NE: for any given $\alpha > 0$ but arbitrarily small, an α-NE is an NE in which the agent is restricted to choose each of her actions with at least α probability in every period. Then, we prove the following for all interior initial common beliefs: If $\alpha > 0$ is arbitrarily small and players are sufficiently patient, there is no strictly positive probability set of histories induced by an α-NE such that the agent's limiting equilibrium play converges to choosing the truthful action with a probability of $1 - \alpha$. So, when players are patient, no robust NE induces a strictly positive probability set of events (histories) in which the regulator enjoys the efficient payoff approximately.

The intuition is as follows: Suppose, on the contrary, that there is a set of events with a positive measure on which the agent finds it optimal to be truthful on average after some private history with a very high probability. Thus, in every continuation game following this private history, the agent must be expecting to see diligence with a high probability on average for sufficiently long periods. Thanks to Cripps et al. (2007) and using "conditional identification of the agent" (saying that perpetual diligence identifies the agent's fixed behavior from the frequencies of the public signals), we establish the following: "if the agent's private history implies that she is almost convinced of facing a diligent regulator and behaves accordingly, then this eventually becomes known to the regulator" on a particular set of private histories of the regulator (coinciding with the agent's private beliefs about the regulator's future behavior obtained from the agent's private history as given above).[5] But then, the agent, knowing that her beliefs will eventually become known to the strategic regulator on these particular histories where the regulator is believed to be diligent on average, can infer that the strategic regulator (who can identify the long-run behavior of the agent on those particular private histories of his) would be convinced that the agent believes that the regulator will be diligent thereafter and he would act on it by choosing lazy. However, this may not be enough to convince the agent to switch to bad behavior when the regulator's reputation is high. But, in the long run, the agent draws the irrefutable inference that the regulator is of the strategic type and chooses lazy since she is bound to experiment with the bad behavior every once in a while. Indeed, every time the agent is untruthful in such situations, her private beliefs would be updated accordingly, which the regulator cannot (observe and hence) respond to. Thus, there is a period in which the agent's private beliefs are not compatible with expecting diligence with a high probability on average for long periods; a contradiction.

Reexamining our results with myopic agents using α-NE, $\alpha > 0$ but arbitrarily small, documents that there are no significant qualitative changes to equilibrium behavior and payoffs. This is because the BNE of the stage game does not change significantly. Agents observe only the public history, which the regulator also sees. So, the beliefs are public, and the regulator can predict short-lived agents' choices. Hence, if the regulator's reputation strictly surpasses the threshold obtained from the stage game, then the corresponding agent chooses the truthful action with probability $1 - \alpha$ and the strategic regulator is lazy—the identifiability of Cripps et al. (2004) fails. Thus, if the current reputation is high, the probability that tomorrow's reputation is high is high. The rest of the argument follows from continuity.

[5]The conditional identification of the agent enables us to use the techniques of Cripps et al. (2007) on a particular set of regulator's private histories and bypass the complications due to private beliefs. When proving disappearing private reputations, Cripps et al. (2007, pp.289) shows that "when the uninformed player's private history induces her to act as if she is convinced of some characteristic about the informed player, the informed player must eventually be convinced that such a private history did indeed occur."

Early literature on reputation focuses on settings where a long-lived player faces a sequence of myopic players observing past play. These studies provide the Stackelberg payoff as the lower bound on the patient long-lived player's average limiting payoff given that there is a commitment type always choosing the Stackelberg action.[6] Cripps et al. (2004), on the other hand, shows that a long-lived informed player, both against myopic and long-lived uninformed opponents, can maintain a permanent reputation for playing a commitment action in a game with imperfect public monitoring only if that action appears in an NE of the complete information stage game.[7] Cripps et al. (2007) extends their disappearance of noncredible reputations result by allowing for private beliefs.

Our findings concerning the asymptotic equilibrium behavior and the permanency of reputation with myopic uninformed players diverge from those of Cripps et al. (2004) as our setting violates both their full-support and full-rank conditions.[8]

Another important work related to our analysis with short-lived agents involves bad reputations. Building on Ely and Välimäki (2003)'s motorist-mechanic example and bad reputation result, Ely et al. (2008) characterizes a class of games with the following details: The short-run uninformed players decide whether or not to participate in a game with the long-run player (he), while each of his actions inducing the short-run players to participate "has a chance of being interpreted as a signal that the long-run player is bad." Thus, the equilibrium payoffs of the patient long-run player are close to his utility from the short-run players' exit decision. Our result with myopic agents parallels that of Ely et al. (2008) in terms of equilibrium payoffs when their participation games are such that the exit action provides the long-run player his maximum payoff: Both studies establish persistent reputations. Their public signals satisfy our conditional identification of the long-lived informed player, and the myopic

[6]See Fudenberg and Levine (1989) (perfect monitoring), Fudenberg and Levine (1992) (imperfect public monitoring), and Gossner (2011) (imperfect private monitoring). Moreover, such results arise also with two long-lived players: Schmidt (1993a) (conflicting interests with asymmetric discount factors); Celentani et al. (1996) and Aoyagi (1996) (imperfect monitoring and asymmetric discount factors); Cripps et al. (2005) (strictly conflicting interests with equal discount factors); Atakan and Ekmekci (2012) and Atakan and Ekmekci (2015) (locally nonconflicting or strictly conflicting interests with equal discount factors); Chan (2000) (equal discounting and commitment being dominant).

[7]Benabou and Laroque (1992) also provides a model of repeated strategic communication with a long-lived insider trader who has noisy private information about the value of an asset and aims to manipulate asset prices. They focus on the stationary ME and show that insider traders reveal their true type asymptotically in any ME. Moreover, Özdoğan (2014) extends the disappearing reputations result to games with two long-lived players with incomplete information on both sides.

[8]In particular, detection happens and is informative about the regulator's behavior only when the agent is untruthful (conditional identification of the regulator) and a bad signal following the agent's untruthfulness is probable only when the regulator is diligent (conditional identification of the agent).

agents do not find it optimal to experiment and unravel the type of the long-lived player.

On the other hand, the two signaling structures differ in significant ways. In their setup, there are exit signals that occur with probability one if the myopic players choose an exit action, which cannot be observed if the short-run players decide to participate and are not affected by the action of the long-lived player. However, in our model, the no detection signal that occurs with probability one if the short-lived agents choose to be truthful ("exit") can also be generated when the agent chooses to be untruthful ("participate"), the probability of which then depends on the regulator's action. This structure gives rise to "the conditional identification of the agent" that is the key condition in analyzing the two long-lived player case, which is left as an open question in Ely et al. (2008).[9, 10]

The organization is as follows: Section 2 presents the model. The descriptions of the repeated games and the results with the short-lived and long-lived agent cases are provided in Section 3 and 4, respectively. Section 5 concludes. The proofs are presented in the Appendix.

2 Model

We model the agent and regulator's strategic interaction through a simultaneous-move *stage game*. The agent (*she*) can be either *truthful* or *untruthful* in her interaction with the regulator (he). Thus, her action set is $A = \{T, U\}$ where $a \in A$. The mixed action of the agent is given by $\sigma_A \in \Delta(A)$ where $\Delta(A)$ is the probability simplex on A; with abuse of notation, we denote the probability that she chooses T also by σ_A. The regulator can detect deviations from the truthful behavior via costly auditing. He chooses to be *diligent* or *lazy* in auditing the agent. His choice generates different detection probabilities of the agent's untruthfulness, provided that she is indeed untruthful. The regulator's action set is $R = \{D, L\}$ while his mixed action is $\sigma_R \in \Delta(R)$. As before, σ_R also denotes the probability of him choosing D.[11]

[9] Ely and Välimäki (2003) constructs a sequential equilibrium that shows their bad reputation result may not hold with two long-lived players in the motorist-mechanic example.

[10] Another strand of related reputation literature involves recent studies featuring continuous-time models that analyze monitoring in employment contracts (e.g., Halac and Prat (2016)) and certification of quality in product-quality choice settings (e.g., Marinovic et al. (2018) and Dilmé (2019) following Board and Meyer-ter Vehn (2013)). While only Halac and Prat (2016) and Marinovic et al. (2018) endogenize the costly learning, the former focuses on dynamics, and the latter analyzes costly voluntary certification as a means to build a reputation in Markov Perfect Equilibrium (MPE). Indeed, that study sustains permanency of reputation in MPE with a stage game based on Board and Meyer-ter Vehn (2013) when "the industry manages to coordinate on a good certification standard."

[11] Our stage game parallels the one in Özdoğan (2016) while the following version is in line with those in some papers on monitoring in employment contracts, e.g., Halac and Prat (2016): There is a business owned by a principal (he) who has to employ an agent (she) to operate. The principal cannot observe the agent's

The set of public signals is $I_d = \{0, 1\}$ where 1 stands for detection and 0 for no detection. The audit quality is given by the following probability distribution on I_d conditional on $A \times R$, which is denoted by ρ where $\rho(i_d \mid a, r)$ is the probability of i_d given $(a, r) \in A \times R$:

$$\rho(1 \mid U, D) = 1 - \rho(0 \mid U, D) = \beta \qquad \rho(1 \mid T, D) = 1 - \rho(0 \mid T, D) = 0$$
$$\rho(1 \mid U, L) = 1 - \rho(0 \mid U, L) = 0 \qquad \rho(1 \mid T, L) = 1 - \rho(0 \mid T, L) = 0$$

where $\beta \in (0, 1)$ is the probability of detecting an agent who has chosen U if the regulator chooses D. Notice that no detection must occur whenever the agent has chosen T.

A player's action is not observable to the other. Yet, the public signals, informative about agents' choices, become commonly observable at the end of the corresponding period. Public signals are statistically informative about a player's behavior conditional on the other one choosing a particular action: the regulator can infer the fixed action chosen by the agent from the signals' frequencies only when he has been diligent; the agent can identify the regulator's fixed action from the frequency of the detections only when she has been untruthful. These are summarized in Remarks 1 and 2 also establishing that in our model, the full support assumption, typically presumed in many studies in the literature, does not hold.

Remark 1. *The* conditional identification of the agent's actions *holds as* $|A|$ *columns in the matrix* $[\rho(i_d \mid a, D)]_{a=U,T;\ i_d=0,1}$ *are linearly independent. And* $\rho(0|U, L) = \rho(0|T, L) = 1$.

Remark 2. *The* conditional identification of the regulator's actions *holds as* $|R|$ *columns in the matrix* $[\rho(i_d \mid U, r)]_{r=D,L;\ i_d=0,1}$ *are linearly independent. And* $\rho(0|T, D) = \rho(0|T, L) = 1$.

We normalize the agent's payoff to zero when she chooses T. If she chooses U, she pays a fine of l if detected and otherwise receives a gain of g. So, $u_A(T, D) = u_A(T, L) = 0$, $u_A(U, L) = g$, and $u_A(U, D) = \ell = g - \beta(g + l)$. The following ensures that her unique best response to D is T:

Assumption 1. *The parameter values satisfy* $\frac{g}{g+l} < \beta$.

The regulator's payoff is also normalized to zero if he chooses L and the agent T. This is the maximum payoff the regulator can attain. Given that the agent chooses U, the regulator's gain is d if U is detected, and otherwise, his expected loss is f. The regulator incurs a cost

performance. His options are to monitor the agent intensively (I) or not (N). If the agent chooses high effort (H) the outcome has to be good, g, regardless of whether or not the principal monitors intensively. If she chooses low effort (L), there is a probability that the bad outcome, b, occurs, which can be detected *only when* the principal monitors the agent intensively. Otherwise, he observes g even though the agent has chosen L.

of c if he chooses D. Thus, regulator's expected payoffs are: $u_R(T, L) = 0$, $u_R(T, D) = -c$, $u_R(U, D) = -e = \beta d - (1 - \beta) f - c$, and $u_R(U, L) = -f$.[12]

The resulting ex-ante (expected) stage game payoffs are presented in Table 1.

	D	L
T	$0, -c$	$0, 0$
U	$-\ell, -e$	$g, -f$

Table 1: Ex ante stage game payoffs under complete information

We employ the following restriction on the regulator's payoffs.

Assumption 2. *The parameter values satisfy* $\frac{c}{d+f} < \beta < \frac{f}{d+f}$.

The first inequality implies $u_R(U, D) > u_R(U, L)$ and the second $u_R(T, D) > u_R(U, D)$. Thus, the regulator's expected payoffs are ordered as follows: $0 = u_R(T, L) > u_R(T, D) > u_R(U, D) > u_R(U, L) = -f$. Under this construction, no matter what the regulator chooses, he prefers the agent to be truthful as the implied expected loss in case of untruthfulness, f, is higher than the cost of being diligent, c. Thus, the regulator would like to convince the agent to be diligent to induce truthfulness, which is the regulator-preferred action. However, the regulator wants to be lazy if he thinks that the agent is truthful, while he has an incentive to be diligent if he believes that the agent is going to be untruthful.

Additionally, we assume that $g < f$ so that the regulator's payoff maximizing action profile, (T, L), also maximizes total welfare.

Consequently, the unique NE is in mixed actions:

$$\sigma_A^* = 1 - \frac{c}{\beta(d + f)} \quad \text{and} \quad \sigma_R^* = \frac{g}{\beta(g + l)}. \tag{1}$$

Next, we discuss some properties of the ex-ante stage game payoff structure. First, the *minmax payoffs* (both in pure and mixed actions) are as follows: *0* for the agent with (T, D) being the pure action profile that minmaxes the agent; $-e$ for the regulator with (U, D) being the pure action profile that minmaxes the regulator. Second, the regulator's pure Stackelberg action is D and D mixed-action minmaxes the agent. Thus, following Schmidt (1993b), the stage game described in the current paper has conflicting interests. The regulator's preferred

[12] Our payoff specifications differ from some those used in the literature, in which players' ex post payoffs depend on their own actions and the public signals, and ex ante payoffs equal the expectation of ex post payoffs taken over opponents' actions. This type of specification would imply $u_R(U, L) = u_R(T, L)$ as the regulator chooses the same action and receives the same signal of no detection with probability one. However, then, the forgone societal loss due to the agent being untruthful would not be captured.

opponent action is T, which is also the unique best response to the Stackelberg action D, whereas the agent's preferred opponent action is L.

To introduce reputation as in Harsanyi (1967-68), Kreps and Wilson (1982) and Milgrom and Roberts (1982), we consider two types of regulators: *tough* or *strategic*. The tough regulator is committed to being diligent (the pure Stackelberg action of the strategic type), whereas the strategic regulator's preferences are as above. The regulator knows his true type while the belief of the agent that the regulator is tough (i.e., the *reputation of the regulator*) is given by $\gamma \in (0,1)$. The agent's equilibrium behavior depends on her belief about the regulator's type. Let $\pi(\gamma, \sigma_R)$ be the expected probability of detection, i.e., $\pi \equiv \pi(\gamma, \sigma_R) = \gamma\beta + (1-\gamma)\sigma_R\beta$. Then, the agent's problem is

$$\max_{\sigma_A \in [0,1]} (1-\sigma_A)\,[(1-\pi)g - \pi l] \tag{2}$$

There is a cutoff value of detection, $\pi^* = \frac{g}{g+l}$, determining the optimal behavior of the agent: her best response equals $\{U\}$ if $\pi(\gamma, \sigma_R) < \pi^*$ and $\{T\}$ if $\pi(\gamma, \sigma_R) > \pi^*$. The Bayesian Nash equilibrium (BNE) of the incomplete information stage game is presented in Lemma 1.

Lemma 1. *The following action profile* (σ_A, σ_R) *constitutes an BNE,*

(*i*) $\sigma_A = 1$ *and* $\sigma_R = 0$ *if* $\gamma \geq \gamma^*$,

(*ii*) $\sigma_A = 1 - \frac{c}{\beta(d+f)}$ *and* $\sigma_R = \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)} = \frac{\pi^*-\gamma\beta}{(1-\gamma)\beta}$ *if* $\gamma < \gamma^*$,

where the cutoff value of the belief is $\gamma^* = \frac{g}{\beta(g+l)} \in (0,1)$.

This lemma establishes that there is no equilibrium in which the regulator chooses to be diligent with probability one. If the belief that the regulator is tough is above a threshold, then the agent is truthful with probability one; anticipating this, the regulator chooses to be lazy with probability one. Otherwise, players go for the mixed actions specified in the lemma. Moreover, the equilibrium actions are monotone in the prior belief.

3 Dynamic game with short-lived agents

The game is infinitely repeated where the periods are $t = 0, 1, \ldots$. The regulator is the long-lived player with a discount factor $\delta \in (0,1)$, and the agents are short-lived (myopic) players. The agent of a period t, agent t, plays only in that period and cares only about her own payoff. In each period, the players simultaneously choose actions from their action sets.

The reputation affects behavior only when the short-lived agents have information about past detections. Hence, we suppose that in every t, agent t observes the public history of signals h^t (while h^0 stands for the unique null history) which consists of whether or not each of the preceding agents have been detected, i.e., $h^t = (i_{d0}, i_{d1}, ..., i_{dt-1}) \in H^t$. We let h_R^t be the private history of the regulator which is composed of h^t and his past actions up to time t and hence $h_R^t = ((r_0, i_{d0}), (r_1, i_{d1}), ..., (r_{t-1}, i_{dt-1})) \in H_R^t \equiv (R \times I_d)^t$. The filtration on $(R \times I_d)^\infty$ induced by the regulator's private histories are given by $\{\mathcal{H}_{Rt}\}_{t=0}^\infty$, while $\{\mathcal{H}_t\}_{t=0}^\infty$ is the filtration on $(I_d)^\infty$. We let $K = \{tough, strategic\}$ be the type space for the regulator. The regulator's type is determined once and for all before the beginning of the game, and the common prior belief about the regulator being tough is $\gamma_0 \in (0, 1)$.

Then, the strategy of the regulator, σ_R, is a sequence of maps $\sigma_{Rt} : H_R^t \times K \to \Delta(R)$. We let $\sigma_R \equiv (\hat{\sigma}_R, \tilde{\sigma}_R)$ where $\hat{\sigma}_R$ is the strategy of the tough type, who always plays diligent (action D) with probability one regardless of his private history, and $\tilde{\sigma}_R$ is the strategy of the strategic type. Agent t's strategy, σ_{At}, is a function $\sigma_{At} : H^t \to \Delta(A)$, while $||\sigma_{At} - \sigma'_{At}|| = \sup_{h^t \in H^t} |\sigma_{At}(h^t) - \sigma'_{At}(h^t)|$. The prior belief γ_0, $\sigma_R \equiv (\hat{\sigma}_R, \tilde{\sigma}_R)$, and $\sigma_A \equiv (\sigma_{At})_{t=0,1,...}$ induce a probability measure Q on $\Omega \equiv K \times (R \times A \times I_d)^\infty$, illustrating how the game evolves for an uninformed outsider. The profiles $\hat{\sigma} \equiv (\sigma_A, \hat{\sigma}_R)$ and $\tilde{\sigma} \equiv (\sigma_A, \tilde{\sigma}_R)$ induce probability measures $\hat{Q}$ and $\tilde{Q}$ on Ω, describing the evolution of the game when the regulator is following the strategy of the tough type, $\hat{\sigma}_R$, and strategic type, $\tilde{\sigma}_R$, respectively. The expectation taken with respect to Q is E and the expectations associated with $\hat{Q}$ and $\tilde{Q}$ are $\hat{E}$ and $\tilde{E}$, respectively. $\hat{E}[\,\cdot \mid \mathcal{H}_t]$ and $\tilde{E}[\,\cdot \mid \mathcal{H}_t]$ identifies agent t's expectation based on public history up to time t when the regulator uses the strategy $\hat{\sigma}_R$ and $\tilde{\sigma}_R$, respectively.

The posterior belief of agent t at the beginning of period t is $\gamma_t(h^t)$ with $\gamma_0(h^0) = \gamma_0$. When the meaning is clear, we shorten $\gamma_t(h^t)$ to γ_t. If agent t choses U, then Bayesian updating is needed at the end of period t (see Remark 2). Otherwise, $\gamma_{t+1} = \gamma_t$. Then, given agent t's choice U, the reputation after the signal $i_d \in \{0, 1\}$ is calculated as follows:

$$\gamma_{t+1} = \begin{cases} \gamma_{t+1}^+ = \frac{\gamma_t \beta}{\pi(\gamma_t, \tilde{\sigma}_{Rt})} = \frac{\gamma_t \beta}{\gamma_t \beta + (1-\gamma_t)\tilde{\sigma}_{Rt}\beta} & \text{if} \quad i_d = 1, \\ \gamma_{t+1}^- = \frac{\gamma_t(1-\beta)}{1-\pi(\gamma_t, \tilde{\sigma}_{Rt})} = \frac{\gamma_t(1-\beta)}{\gamma_t(1-\beta)+(1-\gamma_t)[\tilde{\sigma}_{Rt}(1-\beta)+(1-\tilde{\sigma}_{Rt})]} & \text{if} \quad i_d = 0. \end{cases} \tag{3}$$

where $\pi(\gamma_t, \tilde{\sigma}_{Rt})$ is agent t's assessment of the probability of detection at t given $\tilde{\sigma}_{Rt} \in [0, 1]$:

$$\pi(\gamma_t, \tilde{\sigma}_{Rt}) \equiv \gamma_t \beta + (1 - \gamma_t) \tilde{\sigma}_{Rt} \beta. \tag{4}$$

Bayesian updating implies $\{\gamma_t\}_t$ is a martingale: $E[\gamma_s(h^s) \mid \mathcal{H}_t] = \gamma_t(h^t)$ for all h^s following h^t. Then, $E[\,\cdot \mid \mathcal{H}_t] = \gamma_t \hat{E}[\,\cdot \mid \mathcal{H}_t] + (1 - \gamma_t)\tilde{E}[\,\cdot \mid \mathcal{H}_t]$.

Given a strategy profile σ, the prior belief γ_0, and a public history h^t that has positive probability under σ, we can find the conditional probability of the long-lived strategic player's action that depends on the public history. Thereby, we restrict attention to public strategies.

A *Nash equilibrium* is a strategy profile $\sigma = (\sigma_A, \hat{\sigma}_R, \tilde{\sigma}_R)$ such that

(i) for all t and all positive probability public histories (PPPH) h^t, $\sigma_{At}(h^t)$ is a best response of agent t against $(\hat{\sigma}_R, \tilde{\sigma}_R)$; i.e., for all t and all PPPH h^t, $E[u_A(\sigma_{At}(h^t), \sigma_{Rt}(h^t)) \mid \mathcal{H}_t] \geq E[u_A(\sigma'_{At}(h^t), \sigma_{Rt}(h^t)) \mid \mathcal{H}_t]$ for all $\sigma'_{At}(h^t) \in \Delta(A)$, and

(ii) $\tilde{E}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_R(\sigma_{At}(h^t), \tilde{\sigma}_{Rt}(h^t))] \geq \tilde{E}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_R(\sigma_{At}(h^t), \tilde{\sigma}'_{Rt}(h^t))]$, for all $\tilde{\sigma}'_R$; i.e., $\tilde{\sigma}_R$ is a best response of the strategic regulator against σ_A.

As each agent is short-lived, her decision depends only on the updated reputation of the regulator and the strategic regulator's expected behavior at that period. Indeed, if $\gamma_t \geq \gamma^*$ (which is as given in Lemma 1), agent t chooses T and the strategic regulator L delivering each a payoff of zero and sustaining efficiency. If $\gamma_t < \gamma^*$, then agent t chooses U with some probability only if the strategic regulator is diligent with no more probability than $\frac{\pi^* - \gamma_{t-1}\beta}{(1-\gamma_{t-1})\beta}$.

3.1 Nash equilibrium

Below, we show that if the regulator is sufficiently patient, in every strictly positive probability set of histories induced by an NE, agents' limiting behavior converges to choosing the truthful action in perpetuity given any interior initial common beliefs the agents may have about the regulator's type. As a result, the strategic regulator shies away from being diligent in the long-run, and hence social efficiency is obtained. We let the *commitment strategy*, $\hat{\sigma}_A$, be defined by $\hat{\sigma}_{At}(h^t) = 1$ for all h^t.

Theorem 1. *Suppose that Assumptions 1 and 2 hold. Then, for all* $\gamma_0 \in (0,1)$*, there is* $\delta^* \in (0,1)$ *such that for all* $\delta > \delta^*$ *and for all* $\mathcal{A} \subset \Omega$ *with* $Q(\mathcal{A}) > 0$ *induced by an NE* (σ^*_A, σ^*_R)*,* $\lim_{t\to\infty} \|\hat{\sigma}_{At} - \sigma^*_{At}\| = 0$*, for all* $\omega \in \mathcal{A}$*.*

At the heart of the proof of Theorem 1 lies Özdoğan (2016, Theorem 1) which establishes that when agents are short-lived, reputation helps the patient strategic regulator to achieve the maximum attainable payoff for any prior belief agents may have about regulator's types. Specifically, for any prior belief $\gamma_0 > 0$, the minimum payoff of the strategic regulator across all NE converges to zero, his maximum utility, as δ approaches one. This outcome provides the strategic regulator *strictly more* than his Stackelberg payoff of $-c$.[13]

[13] Applying the payoff bound of Gossner (2011) in the current context, we see that the lower bound of the regulator's NE payoffs approach his Stackelberg utility, $-c$, as he gets more patient (since the unique 0-entropy confirming best response of the agent to the Stackelberg action D is T).

The techniques used in the proof of Theorem 1 parallel those in Ely and Välimäki (2003). The short-lived agents only care about their own payoffs and give myopic best responses to their updated beliefs about the regulator's type. So, the agent plays truthfully if and only if her belief about the regulator being diligent is above a threshold. If so, there is no learning, and the regulator attains his maximum payoff as there is no need to engage in costly auditing (Lemma 2). In histories where the agent is not yet convinced of the regulator being diligent, she has an incentive to be untruthful. This incentivizes the strategic regulator to be diligent with some probability in order to induce an increase in consequent agents' beliefs.[14] After probable subsequent detections, the reputation would eventually reach a level above which all the consequent agents find it optimal to be truthful. Hence, the regulator receives a payoff lower than his maximum for a finite number of periods, while, in the long-run, the sufficiently patient strategic regulator captures all the surplus, thereby sustaining social efficiency.

3.2 Markov equilibrium

Now, we consider short-lived agents who are restricted to use Markov strategies. This situation also corresponds to cases in which agent t is one of a continuum of long-lived agents, coordination among agents is not possible, and all agents observe the same public history.[15]

We characterize the ME with the reputation of the regulator being the Markov state variable and strategies, $\sigma_A(\gamma)$ and $\sigma_R(\gamma)$, are functions of only the current reputation level (and neither the public history nor the time index). We let $\tilde{V}(\gamma)$ denote the expected life-time payoff to the strategic regulator from $(\sigma_A, \tilde{\sigma}_R) \equiv (\sigma_{At}, \tilde{\sigma}_{Rt})_t$ where $(\sigma_{At}, \tilde{\sigma}_{Rt}) = (\sigma_A(\gamma), \tilde{\sigma}_R(\gamma))$ for all t. Then, this equilibrium is defined via the following value function $\tilde{V}(\gamma)$:

$$\tilde{V}(\gamma) = \begin{array}{l} (1-\delta)\{\tilde{\sigma}_R[(1-\sigma_A)(\beta d - (1-\beta)f) - c] - (1-\tilde{\sigma}_R)(1-\sigma_A)f\} + \delta\sigma_A\tilde{V}(\gamma) \\ \quad + \delta(1-\sigma_A)\tilde{\sigma}_R\beta\tilde{V}\left(\frac{\gamma\beta}{\pi(\gamma,\tilde{\sigma}_R)}\right) + \delta(1-\sigma_A)(1-\tilde{\sigma}_R\beta)\tilde{V}\left(\frac{\gamma(1-\beta)}{1-\pi(\gamma,\tilde{\sigma}_R)}\right). \end{array} \tag{5}$$

[14] Lemma 3 displays that every NE continuation path starting from h^t with $\gamma_t < \gamma^*$ includes the play of diligence with some probability, and hence involves positive probability of detection.

[15] Coordination among agents and the regulator for future punishments/rewards may be hard to sustain if agents do not receive the same signal and or individual signals are not public (see, Mailath and Samuelson (2006, Remark 18.1.3) and Mailath and Samuelson (2015) for an equilibrium with coordinated punishments using idiosyncratic and public signals in the context of Mailath and Samuelson (2001)). In such cases, it is innocuous to assume that agents receive independently drawn private signals. This eliminates the coordination among agents and the regulator. However, the idiosyncrasy of signals causes technical complications (e.g., Al-Najjar (1995)) and diverts attention away from reputation. To abstract away from these complications and to capture agents' myopic incentives due to lack of coordination in a large population environment, we consider a continuum of agents who cannot coordinate among each other but receive the same public signal.

Definition 1. *A Markov equilibrium consists of* $\sigma^* \equiv (\sigma_A^*(\gamma), \hat{\sigma}_R^*(\gamma), \tilde{\sigma}_R^*(\gamma))$ *and the corresponding beliefs such that for all* $\gamma \in [0,1]$*:*

1. *Given the expected probability of detection* $\pi(\gamma, \tilde{\sigma}_R^*)$ *induced by* $\tilde{\sigma}_R^*(\gamma)$*,* $\sigma_A^*(\gamma)$ *maximizes the agent's problem given in (2); and*
2. *Given* $\sigma_A^*(\gamma)$*,* $\tilde{\sigma}_R^*(\gamma)$ *maximizes the associated value function* $\tilde{V}(\gamma)$ *given in (5), while* $\hat{\sigma}_R^*(\gamma') = 1$ *for all* γ'*; and*
3. *Posterior beliefs are determined via Bayes' rule whenever possible (i.e., when* $\sigma_A^* < 1$*) according to (3).*

The following is our result for the Markov case:

Theorem 2. *Suppose Assumptions 1 and 2 hold. Then, there is a unique ME,* σ^**, possessing a continuous and nondecreasing value function* $\tilde{V}$ *such that* $\gamma \leq \gamma^*$ *implies*

$$\sigma_A^*(\gamma) = 1 - \frac{(1-\delta)c}{\beta\Big((1-\delta)(f+d) + \delta[\tilde{V}(\gamma^+) - \tilde{V}(\gamma^-)]\Big)} \quad \text{and} \quad \tilde{\sigma}_R^*(\gamma) = \frac{\pi^* - \gamma\beta}{(1-\gamma)\beta}$$

and $\gamma \geq \gamma^*$ *implies* $\sigma_A^*(\gamma) = 1$ *and* $\tilde{\sigma}_R^*(\gamma) = 0$*, where* $\gamma^* = \frac{g}{\beta(g+l)}$ *and* $\pi^* = \frac{g}{g+l}$*. Moreover, if* $\gamma \geq \gamma^*$*, then* $\tilde{V}(\gamma)$ *attains its maximum level of* 0*.*

The ME induces an NE of the dynamic game.[16] When γ crosses the threshold level γ^*, the absorbing state is attained, and both equilibria specify the same pure action profile thereafter. But, if $\gamma < \gamma^*$, the ME specifies a totally mixed action profile so that every public history, apart from some in the absorbing states, is reached with a strictly positive probability. So, the ME defined on the resulting histories constitutes an NE in the dynamic game. Thus, by Theorem 1, in the long run, agents' ME strategies converge to the commitment strategy, $\hat{\sigma}_A$, at every induced set of histories with strictly positive probability when the strategic regulator is sufficiently patient. Consequently, the efficient payoff is sustained in ME.[17]

Theorem 2 establishes that in transient states, i.e., when $\gamma < \gamma^*$, the regulator and each short-lived agent play totally mixed actions that result in probable consecutive detections. Corollary 1 identifies an upper bound on the number of consecutive detections that can be

[16] In the complete-information case, i.e., when $\gamma = 0$, the only ME consists of the repetition of the stage-game equilibrium as given in (1).

[17] $\{\gamma_t\}_t$ is a bounded martingale which can be verified in the Markovian context by Theorem 2 and Lemma 7. Therefore, the *evolution of beliefs* is such that, no matter what has happened in the past (and regardless of whether or not the agent is short-lived), the expectation of future beliefs about the regulator being the tough type conditional on the current information must equal today's value. By employing Lemma 7, we also observe that γ_{t+1} cannot equal γ_t, as it must either be $\gamma^+(\gamma_t)$ or $\gamma^-(\gamma_t)$ with some probabilities specified by Theorem 2 and equation (3), provided that the agent has chosen U in t.

observed at each reputation level, provided that the agent chooses U whenever she is indifferent between her actions. This bound is also the minimum number of periods that the regulator has to invest to build up absorbing reputation at that state.[18]

Corollary 1. *Suppose that Assumptions 1 and 2 hold and consider the ME given in Theorem 2. Let h^τ be a PPPH, h^t be a PPPH following h_τ and involving k consecutive detections starting date τ, $t \geq \tau + k$, and $\gamma_\tau < \gamma^*$. Then, k can be at most the smallest integer that exceeds $k^*_{\gamma_\tau}$ where $k^*_{\gamma_\tau} = \frac{\log(\gamma^*)-\log(\gamma_\tau)}{\log(\beta)-\log(\pi^*)}$.*

To see why note that the posterior probability when the regulator chooses $\sigma^*_R(\gamma_\tau) = \frac{\pi^*-\gamma_\tau\beta}{(1-\gamma_\tau)\beta}$ is derived from (3) and equals to $\gamma_{\tau+1} = \frac{\gamma_\tau\beta}{\pi^*} > \gamma_\tau$ upon observing a detection. After k consecutive detections starting at τ, we obtain $\gamma_{\tau+k} = \gamma_\tau(\frac{\beta}{\pi^*})^k$. Since detection is possible only when agent chooses U, which requires the posterior beliefs to be less than γ^*, we obtain: $\gamma_{\tau+k} \leq \gamma^*$ implies $\gamma_\tau(\frac{\beta}{\pi^*})^k \leq \gamma^*$ and hence our conclusion.

A special case emerges when $\beta = 1 - \pi^*$: an observation of detection followed by no detection or vice versa does not change the posterior belief. Hence, the posterior probability depends only on the number of different public signals in history and not on their order. Thus, the continuation of any history that involves at least $k^*_{\gamma_0}$ more detections results in a persistent reputation exceeding γ^*.[19]

3.2.1 Possible extensions

Given the seminal result of Cripps et al. (2004) establishing the impermanency of reputation effects (obtained when the long-lived player's action is imperfectly observed but all the signals are statistically informative about the long-lived informed player's behavior), in the literature, the survival of the reputation effects is mainly generated by two means: (1) unobserved replacements of the long-lived player with a new copy, and this introduces persistent

[18] Suppose that the parameters are given as $\gamma_0 = 1/2$, $\beta = 3/4$ and $\pi^* = g/(g + l) = 2/3$. The threshold reputation level at these values becomes $\gamma^* = 8/9$. The ME specifies $\sigma^*_R(\gamma) = (8 - 9\gamma)/(9 - 9\gamma)$ for $\gamma \leq \gamma^*$. Under Markov strategy σ^*_R, $\gamma^+(\gamma) = 9/8\gamma$ after a detection and $\gamma^-(\gamma) = 3/4\gamma$ after no detection. The smallest k, at which the reputation exceeds $\gamma^* = 8/9 \cong 0.89$ is 5.

[19] It would be interesting to compute the expected time until the agent stops being untruthful, i.e., the expected hitting time until the Markov chain starting from γ_0 reaches the absorbing state γ^*. Suppose that $\beta = 1 - \pi^*$, which implies $\pi^* < \frac{1}{2}$ as $\beta > \pi^*$. Then, we get a Markov chain with infinitely countable states where $\gamma_{k^*_{\gamma_0}}$, that is the reputation level after $k^*_{\gamma_0}$ many detections, is the only absorbing state and all other states are transient. One can construct an example in which, starting from γ_0, it is sufficient to observe only one detection to reach the absorbing state. But, when this is the case, the ME requires that the regulator choose to be diligent with a small probability. The expected hitting (absorption) time becomes unboundedly large as the transition probability puts higher weight on the lower levels of reputation.

changes in the type of long-lived player (e.g., Benabou and Laroque (1992), Gale and Rosenthal (1994), Holmström (1999), Mailath and Samuelson (2001), Phelan (2006), Wiseman (2008) and Ekmekci, Gossner, and Wilson (2012)); (2) limited observability of histories, i.e., the bounded memory of short-lived uninformed players (e.g., Liu (2011), Ekmekci (2011) and Liu and Skrzypacz (2014)). Below, we discuss how our permanency of reputation result changes if we extend our Markov model to these directions.

First, we consider a setting with unobserved replacements and changing types: Suppose that in each period, the regulator survives to the next period with probability λ and otherwise is replaced with a new regulator who could be a behavioral type with probability $\hat{\gamma}$. To simplify exposition, we let $\gamma_0 = (1-\lambda)\hat{\gamma}$. Then, when the agent chooses U, the posterior belief that the regulator is of behavioral type in period t conditional the signal $i_d \in \{0,1\}$ is

$$\gamma_{t+1} = \begin{cases} \gamma_{t+1}^{+} = \lambda \frac{\gamma_t \beta}{\gamma_t \beta + (1-\gamma_t)\tilde{\sigma}_{Rt}\beta} + (1-\lambda)\hat{\gamma} & \text{if} \quad i_d = 1 \\ \gamma_{t+1}^{-} = \lambda \frac{\gamma_t (1-\beta)}{\gamma_t (1-\beta) + (1-\gamma_t)[\tilde{\sigma}_{Rt}(1-\beta) + (1-\tilde{\sigma}_{Rt})]} + (1-\lambda)\hat{\gamma} & \text{if} \quad i_d = 0, \end{cases} \tag{6}$$

while $\gamma_{t+1} = \lambda\gamma_t + (1-\lambda)\hat{\gamma}$ (***) when the agent is truthful.

Then, we get the following observation saying that frequent replacements prevent the planner from investing in building and attaining an absorbing reputation. Therefore, social efficiency cannot be obtained in the Markov case with frequent replacements.

Proposition 1. *There is a unique ME with the replacement of the regulator that possesses a continuous and nondecreasing value function V^{rep}. Moreover, if the survival rate of the regulator is low so that $\lambda < \gamma^* - \gamma_0$, then the posterior beliefs are always below γ^*, there is no absorbing state, and the agent is never truthful with probability one. And for any $\gamma \in (0, \gamma^*)$,*

$$\sigma_A^*(\gamma) = 1 - \frac{(1-\delta\lambda)c}{\beta\{(1-\delta\lambda)(f+d) + \delta\lambda[V^{rep}(\gamma^+) - V^{rep}(\gamma^-)]\}} \quad \textit{and} \quad \tilde{\sigma}_R^*(\gamma) = \frac{\pi^* - \gamma\beta}{(1-\gamma)\beta}.$$

The proof of Proposition 1 is omitted as it parallels that of Theorem 2.[20] The arguments in that proof also show that with a sufficiently patient regulator and λ sufficiently close to one (the case of infrequent replacements), continuity properties enable us to see that investing into absorbing reputation (and hence social efficiency) reemerges in ME.[21]

[20] In this case, the resulting dynamic programming problem is very similar to the one in Appendix C. The posterior probabilities stated in (6) and (***) should be substituted into (11) and (12); δ must be replaced by $\lambda\delta$. Also, $\lambda < \gamma^* - \gamma_0$ implies that the posterior beliefs cannot exceed the threshold γ^*. Moreover, if $\lambda = 0$, we get the repetition of the stage game Bayesian Nash equilibrium.

[21] Our observations parallel Ekmekci et al. (2012) showing that the long-lived player's replacement can generate permanent reputation effects if the replacements are arbitrarily infrequent and the long-lived player is arbitrarily patient. They provide lower bounds on equilibrium payoffs in every continuation game, which co-

Second, we consider a setting in which the agents have access only to the recent piece of the history (rather than the entire history) of play:[22] Suppose that each agent t is born with the same prior belief γ_0 and observes only the last k entries of the public history. Hence agents' behavior depends on the k-tail of the public history. To eliminate the possibility of deriving complicated inferences with bounded memory (see Barlo et al. (2016, Section 5)), we suppose that agents' behavior does not depend on calendar time.

When k is strictly less than $k^*_{\gamma_0}$ (see Corollary 1), there is no hope of the regulator to attain absorbing reputation by inducing agents' posterior beliefs to exceed γ^*. Thus, the possibility of absorbing reputation, desired by the regulator, disappears. To counteract, he would want to announce the relevant part of the public history.[23] On the other hand, when k is sufficiently high, we conjecture that in our Markov model with bounded but long memory, one could establish that reputation effects would prevail.[24]

4 Regulator faces a long-lived agent

Now, we assume that the agent is long-lived and uses the same discount factor $\delta \in (0, 1)$. Each long-lived player observes the realization of the public signals and his or her own previous actions. Then, $h^t_A = ((a_0, i_{d0}), (a_1, i_{d1}), ..., (a_{t-1}, i_{dt-1})) \in H^t_A \equiv (A \times I_d)^t$ identifies the long-lived agent's private histories up to period t. The set of full histories up to t is $H^t_f \equiv (A \times R \times I_d)^t$ while the filtration on $(A \times R \times I_d)^\infty$ induced by private and public histories are denoted by $\{\mathcal{H}_{it}\}_{t=0}^\infty$ for $i = \{A, R\}$ and $\{\mathcal{H}_t\}_{t=0}^\infty$, respectively. The long-lived agent's strategy, σ_A, is a sequence of maps $\sigma_{At} : H^t_A \to \Delta(A)$.

incides with those of Fudenberg and Levine (1989), Fudenberg and Levine (1992) and Gossner (2011), as the discount rate goes to one at faster than the replacement rate goes to zero. This payoff bound corresponds to the regulator's Stackelberg payoff, which is $-c$ in our setting. Indeed, in our model, the patient regulator could do better when λ, his replacement rate, is sufficiently close to one.

[22] It may be that the short-lived players do not observe any of the previous outcomes without exerting time, effort, or cost. Liu (2011) constructs a class of equilibria that exhibits reputation cycles in a perfect-monitoring product-choice game incorporating costly discovery of past actions.

[23] For instance, if $\beta = 1 - \pi^*$, the regulator would like to announce any part of the history that has involved at least $k^*_{\gamma_0}$ more detections than no detections to each agent.

[24] With bounded but long memory, the analysis becomes more complicated. Liu and Skrzypacz (2014) analyzes a variation of perfect-monitoring product-choice games with limited but long records of the history. Their equilibria feature recurrent reputation bubbles sustained by limited memory. Ekmekci (2011), on the other hand, examines a version of the product-choice game with imperfect public monitoring where the public signals are observed by a rating agency announcing one of the finite numbers of ratings to the short-lived players. Ekmekci (2011) shows that there exists a finite rating system that induces a perfect Bayesian equilibrium, in which the sufficiently patient long-lived player's payoff is close to the Stackelberg levels after every history that implies permanent reputation effects.

A *Nash equilibrium* is $\sigma = (\sigma_A, \sigma_R)$ with $\sigma_R = (\hat{\sigma}_R, \tilde{\sigma}_R)$ satisfying both of the following:

(*i*) $E[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_A(\sigma_A(h_A^t), \sigma_R(h_R^t))] \geq E[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_A(\bar{\sigma}_{At}(h_A^t), \sigma_{Rt}(h_R^t))]$, for all $\bar{\sigma}_A$,

(*ii*) $\tilde{E}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_R(\sigma_A(h_A^t), \tilde{\sigma}_R(h_R^t))] \geq \tilde{E}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_R(\sigma_{At}(h_A^t), \tilde{\sigma}'_{Rt}(h_R^t))]$, for all $\tilde{\sigma}'_R$.

The analysis of NE with a long-lived agent demands some identification conditions our current setting lacks. To recover these identification conditions intuitively, in what follows, we concentrate on NE, in which the agent has to choose each of her actions with a small but positive probability. This, in turn, delivers a robustness notion that we refer to as α-NE with $\alpha > 0$ and arbitrarily small: An *α-Nash equilibrium* is an NE in which the agent is restricted to choose any one of her actions with at least α probability.[25]

Ours is a direct approach. Instead, we could adopt the following formulation involving *one-period amnesia*:[26] Suppose that the initial belief, $\gamma_0 \in (0, \gamma^*)$ where γ^* is as in Lemma 1. In every period t, the agent may experience one-period amnesia with a probability of $\vartheta > 0$ arbitrarily small. While it is common knowledge that she will recover at the end of the period, whether or not she suffers from one-period amnesia in a given period is her private information. If there is no amnesia, it is business as usual: The agent observes her private history h_A^t (hence, γ_t) and chooses accordingly. However, in case of amnesia when choosing her period t action, she observes neither her private nor the public history, h_A^t, and hence cannot infer γ_t. Thus, from her perspective, it is indistinguishable from the start of the game apart from the calendar time t. To avoid serious complications (see Barlo et al. (2016)), we consider strategies that do not use the calendar time in these cases: her action, hence, cannot depend on t. In this contingency, we require her to behave according to the ME of Theorem 2 hanging on to her initial belief $\gamma_0 < \gamma^*$. This provides a *consistent* formulation because players' behavior depends only on the level of reputation (and no other aspect related to the past play) in that equilibrium. Hence, her choice would be U with a probability of $(1-\sigma_A^*(\gamma_0))$ as $\gamma_0 < \gamma^*$ where $\sigma_A^*(\gamma_0)$ is as described in Theorem 2. At the end of t, she recovers from amnesia, observes h_A^t along with her period t choice a_t, whether or not there has been a detection in t, performs the Bayesian updating if there was detection in t, records these as

[25] The existence of an α-NE with $\alpha > 0$ and sufficiently small follows from the compactness of the action space of the stage game $A_\alpha \times R \equiv [\alpha, 1-\alpha] \times [0, 1]$ and standard continuity properties.

[26] One may think of the following detailed scenario: The agent uses reading glasses to keep a notebook that contains her records. In the morning (the beginning) of the period t, there is an ϑ chance that she cannot find her glasses. If they are not misplaced, she uses them to check her notebook, observe her private history, and choose her action by noontime accordingly. But if her glasses cannot be found, she cannot check her notebook by noon and hence has to choose an action without knowing the past and caring about the calendar time. The glasses do not get lost. At the end of the day, she finds them and uses them to record today's observations, also performing the Bayesian updating if needed.

h_A^{t+1} and hence identifies γ_{t+1}, and gets ready for tomorrow. If $\vartheta > 0$ is arbitrarily small, the incentives of the strategic regulator and the agent do not get affected. Thus, an NE of this formulation with $\vartheta > 0$ is an α-NE with $\alpha = \vartheta(1 - \sigma_A^*(\gamma_0)) > 0$.

When $\alpha > 0$ is arbitrarily small, and players are sufficiently patient, the following holds for all interior initial beliefs of the agent: there is no strictly positive probability set of events (histories) induced by an α-NE with the agent's limiting equilibrium behavior converging to playing T with probability $1 - \alpha$. Then, robust NE cannot induce strictly positive probability sets of events in which the regulator attains the efficient payoff approximately. In this context, the agent's commitment strategy, $\hat{\sigma}_A$, is $\hat{\sigma}_{At}(h_A^t) = 1$ for all h_A^t.

Theorem 3. *Suppose Assumptions 1 and 2 and let $\alpha > 0$ and arbitrarily small. Then for all $\gamma_0 \in (0, 1)$, there is $\delta_\alpha \in (0, 1)$ such that for all $\delta > \delta_\alpha$ there is no $\mathcal{A} \subset \Omega$ with $\tilde{Q}(\mathcal{A}) > 0$ induced by an α-NE $(\tilde{\sigma}_A, \hat{\sigma}_R, \tilde{\sigma}_R)$ with $\lim_{t\to\infty} \|\hat{\sigma}_{At} - \tilde{E}[\tilde{\sigma}_{At} \mid \mathcal{H}_{At}]\| = \alpha$, for all $\omega \in \mathcal{A}$.*

We adapt the identification technique of Cripps et al. (2007) to our setting as follows: Thanks to Remark 1, the regulator can identify private histories of the agent in which she plays the truthful action T with a constant probability in the long run if the regulator were to concentrate on histories in which he is diligent (Lemma 11). On the other hand, Remark 2 empowers the agent to identify private histories of the regulator who plays diligently with a constant probability when the agent restricts attention to histories in which she plays U and such histories are sustained in an α-NE (Lemma 10).

The intuition behind Theorem 3 is as follows: Suppose on the contrary that there is a set of events with strictly positive measure, $\mathcal{A}$, on which the agent finds it optimal to play the truthful action T with a probability close to $1 - \alpha$ in all continuation histories after some period $\bar{t}$. The agent finds it optimal to play T with a high probability indefinitely implies that she expects to see the diligent action D with a sufficiently high probability on average for a long enough period after every $s \geq \bar{t}$ observing her private history. The key step in our proof is Lemma 11 which says that "if the agent's private history ensured that she is almost convinced that she faces a diligent regulator and behaves according to that belief, then this eventually becomes inferred by the regulator" on a particular private history where the regulator is choosing D. Therefore, the strategic regulator would find it optimal to deviate and play the lazy action L on those histories. At first, the agent may act as the regulator wishes if his reputation is at a high level. However, in every period, there is $\alpha > 0$ chance that the agent *tests* the regulator's reputation. Every time this happens, the reputation level of the strategic regulator gets updated. Indeed, thanks to Remark 2 (saying that the fixed action of the regulator can be inferred by the agent when she chooses U), there is a period when the agent (restricting attention to her private histories with her choosing U) deduces that her opponent is not choosing D but L. Hence, we get a contradiction on such a set of events, $\mathcal{A}$, with a positive probability measure.

This notion of robustness does not imply major qualitative changes to our results with myopic agents in terms of equilibrium behavior and payoffs. If short-lived players are restricted to choose any one of their actions with a probability $\alpha > 0$ and sufficiently small, the only change to the BNE of the stage game described in Lemma 1 involves revising (i) of this lemma by $\sigma_A = 1 - \alpha$. As a result, when $\alpha > 0$ and sufficiently small, the regulator's best response does not change and calls for D with some positive probability for histories with $\gamma_t \leq \gamma^*$ and L (with probability 1) otherwise. Hence, our findings presented in Section 3.1 continue to hold with some small modifications to their statements and proofs: Due to short-lived agents conditioning only on the public histories, arbitrarily infrequent and mandatory experimentation does not suffice to dismiss agent's conditional identification property. Particularly, thanks to the continuity properties of players' utilities, compactness of feasible payoffs, and the short-lived agents conditioning only on the public histories, letting $\tilde{V}_\alpha(\gamma_0, \delta)$ be the minimum α-NE payoff of the strategic regulator for given γ_0 and δ when facing short-lived agents while Assumptions 1 and 2 hold, we conclude that for any $\alpha > 0$ and arbitrarily small, and any prior belief $\gamma_0 > 0$, $\lim_{\delta \to 1} \tilde{V}_\alpha(\gamma_0, \delta) = -\alpha f$.[27] Thus, the restatement of Theorem 1 with α-NE becomes: Suppose that Assumptions 1 and 2 hold. Then, for all $\gamma_0 \in (0, 1)$, there is $\delta^* \in (0, 1)$ such that for all $\delta > \delta^*$ and for all $\mathcal{A} \subset \Omega$ with $Q(\mathcal{A}) > 0$ induced by an α-NE (σ_A^*, σ_R^*) with $\alpha > 0$ and arbitrarily small, $\lim_{t\to\infty} \|\hat{\sigma}_{At} - \sigma_{At}^*\| = \alpha$, for all $\omega \in \mathcal{A}$.[28] We remark that even if $\alpha > 0$ is arbitrarily small, the use of α-NE results in the dismissal of the absorbing reputation with myopic agents. Still, social efficiency is approximately sustained.

[27] We note that agent t, $t \in \mathbb{N}$, only observes public history h^t and naturally we consider h^t being a PPPH. So, if $\gamma_t > \gamma^*$, Lemmas 1 and 2 hold due to continuity and hence at PPPH h^t, $\sigma_{At} = 1 - \alpha$ implies $\sigma_{Rt} = 0$. Thus, in all PPPH h^τ preceding h^t with $\gamma_\tau > \gamma^*$, agent t is aware that he would not have observed a detection as the regulator would have been choosing L in those periods no matter what the realized choice of agent τ in $\{U, T\}$ has been. As a result, the identifiability of Cripps et al. (2004) does not hold. Then, agent t infers that, as his predecessors have been restricted to choose U with $\alpha > 0$ but arbitrarily small probability, γ_τ has been updated (and gradually decreased) to $\gamma_{\tau+1}$ accounting for the probability that agent τ had to choose U with α probability. So, as $\alpha > 0$ is arbitrary small, $\gamma_{t+1} > \gamma^*$ with a high probability.

[28] A similar conclusion also holds in the Markov case: the modification implied in Theorem 2 involves changing its statement so that $\sigma_A^*(\gamma) = 1 - \alpha$ for any $\gamma \geq \gamma^*$ while the values of γ^*, π^*, and $\tilde{\sigma}_R^*(\gamma)$ do not change and the upperbound of $\tilde{V}(\gamma)$ needs a slight alteration.

Combining these observations with Theorem 3 delivers a gap between the cases with short and long-lived agents in terms of limiting equilibrium behavior of agents who are required to experiment with the bad behavior every once in a while. This also implies a gap in terms of limiting equilibrium payoffs of the strategic regulator. Therefore, we conclude that with mandatory but infrequent experimentation, social efficiency is approximately sustained as a limiting equilibrium payoff with short-lived agents but not with a long-lived agent.

5 Conclusion

This paper analyzes the long-run equilibrium behavior of uninformed players (agents) in a repeated regulatory environment with incomplete information and imperfect public monitoring. It asks whether or not agents can be induced to good behavior permanently by the regulator's (informed player's) reputation. We provide a positive answer when a patient long-lived regulator faces a sequence of short-lived agents for any one of their interior initial common beliefs: Using his reputation, the regulator prevents agents' bad behavior in the long-run with no cost in every set of histories that is induced by an NE and has a strictly positive probability measure. As a result, reputation secures perpetual social efficiency. These conclusions are robust to requiring short-lived agents to choose any one of their actions with a small but positive probability. On the other hand, when both parties are long-lived and sufficiently patient, for all interior initial beliefs of the agent, no robust equilibrium induces a strictly positive set of histories in which the agent's limiting behavior is close to perpetual truthful play. That is why robust NE fails to induce strictly positive probability sets of events in which the regulator obtains the efficient payoff approximately. This contrast demonstrates the significance of the strategic interaction's longevity and provides a novel insight into the importance of learning and experimentation in repeated games.

Appendix

A The proof of Lemma 1

As $u_A(T,\sigma_R) = 0$, $u_A(U,\sigma_R) = \gamma[(1-\beta)g-\beta l]+(1-\gamma)\sigma_R[(1-\beta)g-\beta l]+(1-\gamma)(1-\sigma_R)g$, $u_R(\sigma_A, D) = (1-\sigma_A)[\beta d-(1-\beta)f]-c$, and $u_R(\sigma_A, L) = -(1-\sigma_A)f$, agent's and strategic regulator's best responses are as follows:

$$BR_A(\sigma_R) = \begin{cases} 1 & \text{if} \quad \sigma_R > \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)} \\ [0,1] & \text{if} \quad \sigma_R = \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)} \\ 0 & \text{if} \quad \sigma_R < \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)} \end{cases}, \quad BR_R(\sigma_A) = \begin{cases} 1 & \text{if} \quad \sigma_A < 1-\frac{c}{\beta(d+f)} \\ [0,1] & \text{if} \quad \sigma_A = 1-\frac{c}{\beta(d+f)} \\ 0 & \text{if} \quad \sigma_A > 1-\frac{c}{\beta(d+f)}. \end{cases}$$

From this, we deduce the cutoff prior beliefs. The mixed action of the regulator that makes the agent indifferent between T and U, $\sigma_R = \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)}$, is greater than 0 if $\gamma < \gamma^* = \frac{g}{\beta(g+l)}$ and equals 0 if $\gamma = \gamma^*$. If $\gamma > \gamma^*$, then $BR_A(\sigma_R) = 1$ for all σ_R. Agent's mixed action making the regulator indifferent, $\sigma_A = 1 - \frac{c}{\beta(d+f)} > 0$ if $\beta > \frac{c}{f+d}$. Thus, we conclude the following:

Case 1. $\gamma > \gamma^*$: In this case, $BR_A(\sigma_R) = 1$ for any σ_R. The unique fixed point of the best response correspondences is $\sigma_A = 1$ and $\sigma_R = 0$.

Case 2. $\gamma = \gamma^*$: The action that makes the agent indifferent is $\sigma_R = 0$. For $\sigma_R > 0$, $BR_A(\sigma_R) = 1$. But, $\sigma_R > 0$ cannot be a best response against $\sigma_A = 1$. Thus, the BNE are $\sigma_A \in [1-\frac{c}{\beta(d+f)}, 1]$ and $\sigma_R(D) = 0$. As we assume that the agent is truthful for sure when she is indifferent $\sigma_A = 1$ and $\sigma_R = 0$.

Case 3. $\gamma < \gamma^*$: The unique intersection of the best response correspondences in this case is when $\sigma_A = 1 - \frac{c}{\beta(d+f)}$ and $\sigma_R = \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)}$.

B The proof of Theorem 1

The proof employs a result that appeared in a conference proceeding Özdoğan (2016, Theorem 1) the restatement and proof of which are presented below for completeness purposes. Let $\tilde{V}(\gamma_0, \delta)$ be the strategic regulator's minimum NE payoff given a prior belief $\gamma_0 \in (0,1)$.

Theorem 4 (Theorem 1 of Özdoğan (2016))**.** *Suppose Assumptions 1 and 2 hold. Then, for any prior belief* $\gamma_0 \in (0,1)$, $\lim_{\delta\to 1} \tilde{V}(\gamma_0, \delta) = 0$.

Proof of Theorem 4. Fix an arbitrary NE in public strategies, $\sigma = (\sigma_A, \hat{\sigma}_R, \sigma_R)$ with σ_R denoting the strategic regulator's strategy; each PPPH and posterior belief that are considered are going to be with respect to this NE. For each PPPH h^t, we let $v(h^t)$ denote the *expected continuation value* to the strategic regulator starting from h^t. If T has a positive probability under $\sigma_{At}(h^t)$ and D has a positive probability under $\sigma_{Rt}(h^t)$, then $v(h^t; T, D) \equiv (1-\delta)u_R(T, D) + \delta \sum_{i_d} \rho(i_d \mid T, D) v(h^t, i_d)$. The definition of $v(h_t; \sigma_{At}(h^t), \sigma_{Rt}(h^t))$ is done in the natural way.

Our proof uses the following results: The first tells that if the agent is truthful in this NE at a PPPH, then the regulator must be lazy on that history.

Lemma 2. *If h^t is a PPPH with $\sigma_{At}(h^t) = 1$, then $\sigma_{Rt}(h^t) = 0$.*

Proof. If the agent chooses $\sigma_{At}(h^t) = 1$ at h^t, then the regulator choosing D or L generates the same distribution of public signals and hence the same continuation payoffs $v(h^t, i_d = 0)$. As $u_R(T, L) = 0 > u_R(T, D) = -c$, $\sigma_{Rt}(h^t) = 0$ due to the one-shot deviation principle. ■

The next lemma establishes that every NE continuation path starting from a PPPH h^t must include the play of D with some positive probability if $\gamma_t(h^t) < \gamma^*$.

Lemma 3. *If h^t is a PPPH with $\gamma_t(h^t) < \gamma^*$, then h^t has a positive probability continuation history h^τ such that $\sigma_{R\tau}(h^\tau) > 0$.*

Proof. Let h^t be a PPPH and suppose, for a contradiction, for every PPPH h^τ following h^t, $\sigma_{R\tau}(h^\tau) = 0$. Therefore, there are no detections in any continuation PPPH following h^t since $\gamma_t(h^t) < \gamma^*$. Then, by Lemma 1, all the myopic agents in such continuation histories choose U. Thus, the regulator's expected continuation payoff at such histories equals $-f$, which is strictly less than the minmax payoff $-e$, delivering the desired contradiction. ■

The following result identifies the agents' smallest posterior belief that the regulator is though after a detection is observed.

Lemma 4. *If detection occurs at a PPPH h^t, then $0 < \sigma_{Rt}(h^t) \leq \frac{\pi^* - \beta\gamma_t(h^t)}{\beta(1-\gamma_t(h^t))}$.*

Proof. A detection at a PPPH h^t implies the myopic agent being untruthful at h^t. For that to happen in an NE, by Lemma 1, it must be that $\gamma_t(h^t) < \gamma^*$ and $\sigma_{Rt}(h^t) \leq \frac{\pi^* - \beta\gamma_t(h^t)}{\beta(1-\gamma_t(h^t))}$. Moreover, $\sigma_{Rt}(h^t) > 0$, or else there cannot be any detections. ■

We note that if $\gamma_0 \geq \gamma^*$, then the NE must such that $\sigma_{At}(h^t) = 1$ and $\sigma_{Rt}(h^t) = 0$ for all positive probability h^t. Moreover, if $\gamma_0 < \gamma^*$, then at a PPPH h^t with $\gamma_t(h^t) \geq \gamma^*$, due to Lemmas 1 and 2, the NE has to be such that all the consequent myopic agents choose $\sigma_{A\tau}(h^\tau) = 1$ and the regulator $\sigma_{R\tau}(h^\tau) = 0$ at any continuation history h^τ following h^t. Thus, in all these cases, $\lim_{\delta \to 1} \tilde{V}(\gamma_0, \delta) = 0$.

When $\gamma_0 < \gamma^*$ and h^t being a PPPH with $\gamma_t(h^t) < \gamma^*$ and the myopic agent t choosing $\sigma_{At}(h^t) = 1$ implies (thanks to Lemma 2) the NE is such that $\sigma_{Rt}(h^t) = 0$ at h^t. Thus, consider a PPPH h^t with $\gamma_t \equiv \gamma_t(h^t) < \gamma^*$ and $\sigma_{At}(h^t) < 1$. This time, the reputation is updated according to (3). Note that $\sigma_{At}(h^t) < 1$ implies the expected probability of detection $\pi(\gamma_t, \sigma_{Rt}(h^t)) \leq \pi^*$, and this requires $\sigma_{Rt}(h^t) \leq \frac{\pi^* - \gamma_t\beta}{(1-\gamma_t)\beta}$ (recall that $\pi^* \equiv \frac{g}{g+l}$). Then, by Lemma 4, we define the smallest posterior probability of the regulator being tough conditional on the identification of a detection by $\Gamma(\gamma_t) \equiv \frac{\gamma_t\beta}{\pi^*}$. By Assumption 1, $\Gamma(\gamma) > \gamma$ for all $\gamma \in (0, \gamma^*)$, i.e., Γ is strictly increasing and continuous. As in Ely and Välimäki (2003), let $\{p_n\}_n$ be a decreasing sequence of beliefs by $p_1 \equiv \gamma^*$ and $p_n \equiv \Gamma^{-1}(p_{n-1})$ for $n > 1$, and note that $p_n \searrow 0$. Then, there exists a sequence of lower bounds on payoffs of the strategic regulator $\{\tilde{V}_n(\delta)\}_n$ with $\lim_{\delta\to 1} \tilde{V}_n(\delta) = 0$ and $\tilde{V}(\gamma', \delta) \geq \tilde{V}_n(\delta)$ for all $\gamma' > p_n$ for all $n \in \mathbb{N}$. This is due to the following: Note that the assertion in the previous sentence holds for $n = 1$, as $\gamma > p_1 = \gamma^*$. Thus, we assume this relation holds for n and want to show that it holds for $n + 1$. By hypothesis, we have $\sigma_{At}(h^t) < 1$ and this implies that $\sigma_{Rt}(h^t) \leq \frac{\pi^* - \gamma_t\beta}{(1-\gamma_t)\beta} < 1$ (due to Assumption 1). As $\gamma_t < \gamma^*$, Lemma 3 implies that there exists a PPPH h^τ following h^t such that $\sigma_{R\tau}(h^\tau) > 0$ and hence $\sigma_{A\tau}(h^\tau) < 1$ thanks to Lemma 2. Without loss of generality, assume that h^τ is the "first" of such continuation histories and $\gamma_\tau(h^\tau) \equiv \gamma_\tau > p_{n+1}$ for some $n \in \mathbb{N}$. $\sigma_{R\tau}(h^\tau) \in (0, 1)$ implies that the strategic regulator must be indifferent (in the long-run) between D or L at h^τ. So, $(1 - \delta)(-f) + \delta Z_L(\gamma_\tau) = (1 - \delta)[\sigma_{A\tau}(h^\tau)(-f - c) + (1 - \sigma_{A\tau}(h^\tau))(-e)] + \delta Z_D(\gamma_\tau)$ implies

$$\tilde{V}_{n+1}(\delta) \equiv (1 - \delta)(-f) + \delta Z_L(\gamma_\tau) \geq (1 - \delta)(-f - c) + \delta Z_D(\gamma_\tau) \tag{7}$$

where $Z_D(\gamma_\tau)$ and $Z_L(\gamma_\tau)$ are the lower bounds on the regulator's expected continuation payoffs from D and L at h^τ, respectively. When the regulator chooses L, the posterior decreases to $\gamma_\tau^- = \frac{\gamma_\tau(1-\beta)}{\gamma_\tau(1-\beta)+(1-\gamma_\tau)}$ with probability 1. Thus, $Z_L(\gamma_\tau) = \tilde{V}(\gamma_\tau^-, \delta)$. When, the strategic regulator chooses D at h^τ, detection occurs with probability β and the posterior probability that regulator is though conditional on detection at h^τ (i.e., the threshold for agent τ's decision) is no less than $\Gamma(\gamma_\tau)$, which is at least p_n as $\gamma_\tau > p_{n+1}$. Hence,

$$\beta\tilde{V}_n(\delta) + (1 - \beta)\tilde{V}(\gamma_\tau^-, \delta) \leq Z_D(\gamma_\tau). \tag{8}$$

Combining (7) and (8) implies $(1 - \delta)(-f - c) + \delta\beta\tilde{V}_n(\delta) + \delta(1 - \beta)\tilde{V}(\gamma_\tau^-, \delta) \leq (1 - \delta)(-f) + \delta\tilde{V}(\gamma_\tau^-, \delta)$; so, $(1 - \delta)(-c) + \delta\beta\tilde{V}_n(\delta) \leq \delta\beta\tilde{V}(\gamma_\tau^-, \delta)$; so, $\tilde{V}_n(\delta) \leq \tilde{V}(\gamma_\tau^-, \delta)$ since $c, \beta, \delta > 0$. As $\tilde{V}_{n+1}(\delta) = (1 - \delta)(-f) + \delta Z_L(\gamma_\tau) = (1 - \delta)(-f) + \delta\tilde{V}(\gamma_\tau^-, \delta)$ and $\tilde{V}_n(\delta) \leq \tilde{V}(\gamma_\tau^-, \delta)$,

$$(1 - \delta)(-f) + \delta\tilde{V}_n(\delta) \leq \tilde{V}_{n+1}(\delta). \tag{9}$$

By hypothesis, $\lim_{\delta\to 1}\tilde{V}_n(\delta) = 0$; ergo, the limit of the left-hand side of (9) is zero as δ tends to one. Thus, $\lim_{\delta\to 1}\tilde{V}_{n+1}(\delta) = 0$ as the regulator's payoffs are bounded above by 0. This implies that the PPPH h^τ is such that $\lim_{\delta\to 1}\tilde{V}(\gamma_\tau,\delta) = 0$ with $\gamma_\tau \leq \gamma_t < \gamma^*$.

Having established the existence of $\{\tilde{V}_n(\delta)\}_n$ with $\lim_{\delta\to 1}\tilde{V}_n(\delta) = 0$ and $\tilde{V}(\gamma',\delta) \geq \tilde{V}_n(\delta)$ for all $\gamma' > p_n$ for all $n \in \mathbb{N}$, we observe that for any $\gamma_0 \in (0,\gamma^*)$, there exists $N \in \mathbb{N}$ such that $\tilde{V}(\gamma_0,\delta) \geq \tilde{V}_n(\delta)$ and $\lim_{\delta\to 1}\tilde{V}_n(\delta) = 0$ for all $n \geq N$. Thus, as the regulator's payoffs are bounded above by 0, $\lim_{\delta\to 1}\tilde{V}(\gamma_0,\delta) = 0$. This completes the proof of Theorem 4. ■

Proof of Theorem 1. First, to establish the existence of the limit, we note that agent t uses the information contained in public history h^t and hence her posterior beliefs at h^t are given by $\mathcal{H}_t$ - measurable random variable $\gamma_t \equiv Q(tough \mid \mathcal{H}_t) : \Omega \to [0,1]$. γ_t is a bounded martingale with respect to $\{\mathcal{H}_t\}_t$ and measure Q, and $E[\gamma_t \mid \mathcal{H}_t] = \gamma_t$. Therefore, in any equilibrium, γ_t converges Q-almost surely to a random variable γ_∞. Agents are myopic and $E[u_A(\cdot) \mid \mathcal{H}_t]$ is continuous, so by Lemma 1, their best responses are in the form of threshold strategies such that σ^*_{At} equals 1 if $\gamma_t > \gamma^*$ and some mixed action if $\gamma_t \leq \gamma^*$. And hence, for any NE $(\sigma^*_A, \hat{\sigma}^*_R, \tilde{\sigma}^*_R)$, $\{\sigma^*_{At}\}_t$ Q-converges in $[0,1]$.

As we already have established the existence of the limit and $\hat{\sigma}_A$ is a constant sequence, we suppose, for a contradiction, that there is $\delta^* \in (0,1)$ such that for all $\delta > \delta^*$ we let $(\sigma^*_A, \hat{\sigma}^*_R, \tilde{\sigma}^*_R)$ be an NE inducing $\mathcal{A}^* \subset \Omega$ with $\tilde{Q}(\mathcal{A}^*) > 0$ and $\lim_{t\to\infty}\sigma^*_{At} \neq \hat{\sigma}_{At}$. Therefore, there exists τ such that (without a loss of generality) for all $s \geq \tau$, $\sigma^*_{As} < 1$, for a.e. $\omega \in \mathcal{A}^*$. But, this is at odds with Theorem 4. This concludes the proof of Theorem 1. ■

C The proof of Theorem 2

Let C_+ denote the space of continuous and nondecreasing value functions endowed with the sup norm. An equilibrium value function $\tilde{V} : [0,1] \to [-f,0]$ should satisfy the Bellman equation given by (5). So, it should be a fixed point of the operator T which maps any continuation value function W into $\bar{W}$ as follows: $\bar{W}(\gamma) = TW(\gamma)$ and

$$\bar{W}(\gamma) = \begin{array}{l} (1-\delta)\{\tilde{\sigma}_R[(1-\sigma_A)(\beta d - (1-\beta)f) - c] - (1-\tilde{\sigma}_R)(1-\sigma_A)f\} + \delta\sigma_A W(\gamma) \\ + \delta(1-\sigma_A)\tilde{\sigma}_R\beta W\left(\frac{\gamma\beta}{\pi(\gamma,\tilde{\sigma}_R)}\right) + \delta(1-\sigma_A)(1-\tilde{\sigma}_R\beta)W\left(\frac{\gamma(1-\beta)}{1-\pi(\gamma,\tilde{\sigma}_R)}\right) \end{array} \tag{10}$$

where $\tilde{\sigma}_R(\gamma)$ maximizes the right-hand side of the equation (10) given $\sigma_A(\gamma)$ while $\sigma_A(\gamma)$ maximizes (2) given the expected probability of detection $\pi(\gamma,\tilde{\sigma}_R)$ implied by $\tilde{\sigma}_R(\gamma)$.

Our proof is built upon results that parallel Benabou and Laroque (1992). We start by taking a continuous and nondecreasing value function $W \in C_+$ as given, call the associated problem involving this continuation payoff a *short-term* game, and show that there exists a *unique* profile $\sigma_A(\gamma; W)$ and $\sigma_R(\gamma; W)$ such that the following holds: $\sigma_R(\gamma; W)$ maximizes the right-hand side of (10) given $\sigma_A(\gamma; W)$, and $\sigma_A(\gamma; W)$ maximizes (2) given the expected probability of detection $\pi(\gamma, \sigma_R)$ induced by $\sigma_R(\gamma; W)$. We call this unique equilibrium for any given $W \in C_+$ as the *temporary equilibrium* of the short-term game. Then, we consider the operator that maps the continuation valuation W into current valuation resulting from the outcomes of the optimization of the short-term game $T(W) : \gamma \in [0, 1] \to \bar{W}(\gamma, W)$ and show that the resulting $\bar{W}(\gamma)$ is also a continuous and nondecreasing function, i.e., $\bar{W} \in C_+$. We also show $T : C_+ \to C_+$ is a contraction and hence has a unique fixed point in C_+, $\tilde{V}$.

Let $W \in C_+$ be given. The value function in the short-term game when the regulator choose D and L are:

$$\begin{aligned} W_D(\gamma) &= (1-\delta)\{(1-\sigma_A)[\beta d - (1-\beta)f] - c\} + \delta\sigma_A W(\gamma) \quad (11) \\ &\qquad + \delta(1-\sigma_A)\beta W(\gamma^+) + \delta(1-\sigma_A)(1-\beta)W(\gamma^-), \\ W_L(\gamma) &= -(1-\delta)(1-\sigma_A)f + \delta\sigma_A W(\gamma) + \delta(1-\sigma_A)W(\gamma^-). \quad (12) \end{aligned}$$

The regulator is indifferent when $W_D(\gamma) - W_L(\gamma) = 0$, so $(1-\sigma_A)\beta\{(1-\delta)(f+d) + \delta[W(\gamma^+) - W(\gamma^-)]\} = (1-\delta)c$. If $W_D(\gamma) > W_L(\gamma)$, the regulator chooses D, so $\pi(\gamma) = \beta$. If $W_D(\gamma) < W_L(\gamma)$, the regulator chooses L, thus $\pi(\gamma) = \gamma\beta$. These prove the following:

Lemma 5. *Given $W \in C_+$, an equilibrium of the short-term game induces a detection probability for the regulator as a function of his reputation, $\pi : [0, 1] \to [0, \beta]$ (his strategy then can be deduced from (4)), and an implied strategy for the agent $\sigma_A(\pi)$ that maximizes the agent's problem (2) at π, with an associated value function $\bar{W} : [0, 1] \to [-f, 0]$ such that for any $\gamma \in [0, 1]$, $\bar{W}(\gamma) = \max\{W_D(\gamma), W_L(\gamma)\}$ where $W_D(\gamma), W_L(\gamma)$ are as in (11) and (12) and*

$$\begin{aligned} W_D(\gamma) &> W_L(\gamma) \quad \textit{implies that} \quad \pi(\gamma) = \beta, \\ W_D(\gamma) &= W_L(\gamma) \quad \textit{implies that} \quad \gamma\beta \le \pi(\gamma) \le \beta, \\ W_D(\gamma) &< W_L(\gamma) \quad \textit{implies that} \quad \pi(\gamma) = \gamma\beta. \end{aligned}$$

Combining these, we see $F(\sigma_A(\pi); \gamma, W) \equiv W_D(\sigma_A(\pi); \gamma, W) - W_L(\sigma_A(\pi); \gamma, W)$ equals

$$(1-\sigma_A(\pi))\beta\left\{(1-\delta)(f+d) + \delta\left[W\left(\frac{\gamma\beta}{\pi}\right) - W\left(\frac{\gamma(1-\beta)}{1-\pi}\right)\right]\right\} - (1-\delta)c. \quad (13)$$

Lemma 6. *For any $\gamma \in [\gamma^*, 1]$ and $W \in C_+$, $\pi(\sigma_R, \gamma) \ge \pi^*$ for any value of σ_R. Then, at the unique temporary equilibrium of the short-term game $(\bar{\sigma}_A(\pi), \bar{\sigma}_R(\gamma))$, $\bar{\sigma}_A(\pi) = 1$ solves the* *agent's problem in (2), $F(\bar{\sigma}_A(\pi);\gamma,W) = -(1-\delta)c < 0$ and $\bar{W}(\gamma) = \max\{W_D(\gamma), W_L(\gamma)\} = W_L(\gamma)$ with $\bar{\sigma}_R(\gamma) = 0$, and $\gamma^+(\gamma) = \gamma^-(\gamma) = \gamma$ and $\bar{W}(\gamma) = 0$.*

The proof follows from using $\gamma^* = \frac{g}{\beta(g+l)}$ and $\pi^* = \frac{g}{g+l}$ in Lemma 5.

Lemma 7. *For any $\gamma \in [0,\gamma^*)$ and $W \in C_+$, $F(\sigma_A(\pi);\gamma,W)$ is nonincreasing in $\pi \in [0,\beta]$ and strictly decreasing in $\sigma_A(\pi)$. Moreover, there exists a unique temporary equilibrium of the short-term game $(\bar{\sigma}_A(\pi),\bar{\sigma}_R(\gamma))$, given by mixed actions $\bar{\sigma}_A(\pi^*) : (\gamma,W) \to (0,1)$ that is continuous in (γ,W) and $\bar{\sigma}_R(\gamma) = \frac{\pi^*-\gamma\beta}{(1-\gamma)\beta}$ that induces the perceived detection probability $\pi = \pi^*$, which together satisfy $F(\bar{\sigma}_A(\pi^*);\gamma,W) = 0$ and solve the agent's problem in (2). The associated posterior probabilities are $\gamma^+(\gamma,\bar{\sigma}_R) = \frac{\gamma\beta}{\pi^*}$ and $\gamma^-(\gamma,\bar{\sigma}_R) = \frac{\gamma(1-\beta)}{1-\pi^*}$.*

Proof. Take any $\gamma \in [0,\gamma^*)$ and $W \in C_+$. For $\pi = 0$, the strategy that solves the agent's problem in (2) dictates that $\sigma_A(\pi) = 0$ and thus $F(\sigma_A(\pi);\gamma,W) > 0$ by Assumption 2 and by W being nondecreasing. And, for $\pi = \beta$, $\sigma_A(\pi) = 1$ solves the agent's problem and the corresponding $F(\sigma_A(\pi);\gamma,W) = -(1-\delta)c < 0$. As $F > 0$ for $\sigma_A = 0$ and $F < 0$ for $\sigma_A = 1$ and F is continuous and strictly decreasing in σ_A, there exists unique $\bar{\sigma}_A(\pi;\gamma,W) \in (0,1)$ that ensures $F(\bar{\sigma}_A(\pi);\gamma,W) = 0$. As for all (γ,W), $\bar{\sigma}_A(\pi;\gamma,W)$ is unique, we let $\bar{\sigma}_A(\pi) = \bar{\sigma}_A(\pi;\gamma,W)$. Also, since F is continuous in (γ,W), $\bar{\sigma}_A(\pi;\gamma,W)$ is also continuous in (γ,W).

Next, we argue that $\bar{\sigma}_A(\pi^*) \in (0,1)$ and $\bar{\sigma}_R(\gamma)$ constitute a unique temporary equilibrium. Suppose for a contradiction that $\bar{\sigma}_A(\pi) = 0$ for some (γ,W). Then, $W_D(\bar{\sigma}_A(\pi);\gamma,W) > W_L(\bar{\sigma}_A(\pi);\gamma,W)$ and $\bar{\sigma}_R(\gamma) = 1$. But, this implies that the perceived probability of detection is β and thus $\bar{\sigma}_A(\pi) = 0$ does not solve the agent's problem in (2). Suppose on the contrary that $\bar{\sigma}_A(\pi) = 1$ for some (γ,W). Then $W_L(\bar{\sigma}_A(\pi);\gamma,W) > W_D(\bar{\sigma}_A(\pi);\gamma,W)$ and $\bar{\sigma}_R(\gamma) = 0$, which suggests that $\pi = \gamma\beta$, and in turn would imply $\bar{\sigma}_A(\pi;\gamma,W) = 0$ as $\gamma < \gamma^*$. Hence, we conclude that $\bar{\sigma}_A(\pi^*;\gamma,W) \in (0,1)$ and $\bar{\sigma}_R(\gamma) = \frac{\pi^*-\gamma\beta}{(1-\gamma)\beta}$ is the unique temporary equilibrium, since for $\bar{\sigma}_A(\pi)$ to be a totally mixed strategy, the detection probability must be set to π^* by the regulator employing the strategy $\bar{\sigma}_R(\gamma) = \frac{\pi^*-\gamma\beta}{(1-\gamma)\beta}$. Lastly, the associated posterior beliefs are calculated from (3) by using Bayes' rule. ■

Hereafter, we refer to the temporary equilibrium of the agent for a given W by $\sigma_A^*(\gamma,W) \equiv \bar{\sigma}_A(\pi^*;\gamma,W)$ and that of the regulator by $\sigma_R^*(\gamma) \equiv \bar{\sigma}_R(\gamma)$ (as in Lemmas 6 and 7); define $\bar{W}(\gamma,W) \equiv W_D(\sigma_A^*(\gamma,W);\gamma,W) = W_L(\sigma_A^*(\gamma,W);\gamma,W)$ as the value function evaluated at the temporary equilibrium of the short-term game for $\gamma < \gamma^*$. We want to show that $\bar{W}$ is continuous and nondecreasing in γ, which depend on the behavior of $\sigma_A^*(\gamma,W)$ as well as

W.[29] As the value of W changes (in a nondecreasing way as γ increases), we should consider both of the arguments of $\bar{W}$ to investigate whether or not it is nondecreasing in γ.

Lemma 8. *$\bar{W}(\gamma, W)$ is continuous and nondecreasing in γ for any $W \in C_+$.*

Proof. As at $\sigma_A^*(\gamma, W)$, $F(\sigma_A^*(\gamma, W); \gamma, W) = 0$, both (i) $\bar{W}(\gamma, W) = W_D(\sigma_A^*(\gamma, W); \gamma, W)$ and (ii) $\bar{W}(\gamma, W) = W_L(\sigma_A^*(\gamma, W); \gamma, W)$ hold. Note that, by (12), (ii) implies,

$$\bar{W}(\gamma, W) = -(1-\delta)(1-\sigma_A^*(\gamma, W))f + \delta W(\gamma^-(\gamma)) + \delta\sigma_A^*(\gamma, W)\left[W(\gamma) - W(\gamma^-(\gamma))\right]. \quad (14)$$

Multiplying (ii) by $1-\beta > 0$ and then subtracting this from (i) result in the following expression for $\bar{W}(\gamma, W)$:

$$\bar{W}(\gamma, W) = (1-\delta)(1-\sigma_A^*(\gamma, W))d - \frac{(1-\delta)c}{\beta} + \delta W(\gamma^+(\gamma)) - \delta\sigma_A^*(\gamma, W)\left[W(\gamma^+(\gamma)) - W(\gamma)\right]. \quad (15)$$

For any $(\gamma_1, W_1) \leq (\gamma_2, W_2)$; if $\sigma_A^*(\gamma_2, W_2) \geq \sigma_A^*(\gamma_1, W_1)$, expression (14) implies that $\bar{W}$ is nondecreasing in (γ, W). This is because, $\sigma_A^*(\gamma, W) \in (0, 1)$ is assumed to be nondecreasing and W is nondecreasing in γ. As $[W(\gamma) - W(\gamma^-)] \geq 0$ for any W_1, W_2 and γ_1, γ_2, the right-hand side of (14) is nondecreasing. Hence, $\bar{W}(\gamma_2, W_2) \geq \bar{W}(\gamma_1, W_1)$ when $(\gamma_2, W_2) \geq (\gamma_1, W_1)$. If, on the other hand, we suppose that $\sigma_A^*(\gamma_2, W_2) < \sigma_A^*(\gamma_1, W_1)$, then expression (15) will imply that $\bar{W}$ is nondecreasing. To see that, suppose $\sigma_A^*(\gamma, W)$ is nonincreasing in γ and $[W(\gamma^+) - W(\gamma)] \geq 0$ for any W_1, W_2 and γ_1, γ_2 as before, the right-hand side of (15) is nondecreasing which implies $\bar{W}(\gamma_2, W_2) \geq \bar{W}(\gamma_1, W_1)$ when $(\gamma_2, W_2) \geq (\gamma_1, W_1)$.

The continuity of $\bar{W}$ in (γ, W) is a direct implication of the continuity of $\sigma_A^*(\gamma, W)$, which is established by Lemma 7, and $W \in C_+$. ■

Now, we can redefine the operator that maps the next period's continuation value to today's with the equilibrium outcome of the short-term game as $T(W) : [0, 1] \times W \to \bar{W}(\gamma, W)$. The last step is to argue that T is a contraction that maps continuous and nondecreasing functions on $[0, 1]$ into itself. Lemma 8 shows that the operator T defined above maps C_+ into C_+ and hence monotonicity is satisfied. Moreover, $T(W + k) = T(W) + \delta k$ for any constant k. Then, by Blackwell's theorem, T is a contraction mapping on a complete metric space (C_+ with the sup norm); and hence, it has a unique fixed point $\tilde{V}$. Finally, the equilibrium strategies and the value function $\tilde{V}$ of Theorem 2 follow from Lemmas 6 and 7.

[29] Recall that $\bar{W}$ is constant at zero for $\gamma \geq \gamma^*$ for any $W \in C_+$, thus it is continuous and nondecreasing at these values.

D The proof of Theorem 3

D.1 The agent's beliefs about her own future behavior

The following shows that if there is a strictly positive probability set of states on which the agent plays T with a very high probability in the long-run in an α-NE, then, given any one of her private histories, there must be a strictly positive probability subset of these states such that there is a period after which the agent plays T on average almost all the time in any continuation game.

Lemma 9. *Suppose that there exists $\mathcal{A} \subset \Omega$ with $\tilde{Q}(\mathcal{A}) > 0$ such that for all $\omega \in \mathcal{A}$, $\lim_{t\to\infty} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At} \mid \mathcal{H}_{At}]\| = \alpha$ given an α-NE $(\tilde{\sigma}_A, \tilde{\sigma}_R)$, for some arbitrarily small $\alpha > 0$. Then, there exists $\mathcal{F} \subset \mathcal{A}$ with $\tilde{Q}(\mathcal{F}) > 0$ such that, for any $\xi > \alpha$, there exists $\bar{t}_\alpha$ for which*

$$\tilde{E}[\sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\| \mid \mathcal{H}_{At}] < \xi, \quad \forall t \geq \bar{t}_\alpha \tag{16}$$

for all $\omega \in \mathcal{F}$; and for some $\psi > \alpha$,

$$\tilde{Q}(\alpha \leq \sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\| < \psi \mid \mathcal{H}_{At}) \to 1, \quad \forall t \geq \bar{t}_\alpha \tag{17}$$

where the convergence is uniform on $\mathcal{F}$.

Proof. By hypothesis, $\|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At} \mid \mathcal{H}_{At}]\|$ converges $\tilde{Q}$-almost surely to α on $\mathcal{A}$. So, the random variables $Z_t \equiv \sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\|$ also converge $\tilde{Q}$-almost surely to α on $\mathcal{A}$. Thus, on $\mathcal{A}$, $\tilde{E}[Z_t \mid \mathcal{H}_{At}] \to \alpha$, $\tilde{Q}$-almost surely (by Lemma 4.24 Hart (1985)). Egorov's Theorem (Chung (1974)) then suggests that there exists $\mathcal{F} \subset \mathcal{A}$ with $\tilde{Q}(\mathcal{F}) > 0$ on which the convergence of $\tilde{E}[Z_t \mid \mathcal{H}_{At}]$ is uniform. This implies that, for any $\xi > \alpha$, there exists $\bar{t}_\alpha$ such that on $\mathcal{F}$, $\alpha \leq \tilde{E}[Z_t \mid \mathcal{H}_{At}] \equiv \tilde{E}[\sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\| \mid \mathcal{H}_{At}] < \xi$ for all $t \geq \bar{t}$.

Finally, the last expression in Lemma 9 follows from Chebyshev-Markov inequality. Fix $\psi > 0$ so that $\xi = \epsilon\psi$ for some $\epsilon > 0$. Since Z_t has a finite mean and $Z_t \geq \alpha$, $\tilde{Q}(Z_t \geq \psi \mid \mathcal{H}_{At}) \leq \frac{\tilde{E}[Z_t|\mathcal{H}_{At}]}{\psi} < \frac{\xi}{\psi}$. As $\psi > 0$ and $\xi = \epsilon\psi$, we obtain $\tilde{Q}(Z_t \geq \psi) < \epsilon$, which indicates that $\tilde{Q}(\alpha \leq Z_t < \psi \mid \mathcal{H}_{At}) > 1 - \epsilon$ for all $t > \bar{t}_\alpha$ on $\mathcal{F}$. This implies (17). ■

D.2 The agent's beliefs about the regulator's future behavior

The next lemma states that if there is a set of states with a positive measure on which the agent plays T with a very high probability on average in every continuation game after some period, then she must be convinced to see D in every continuation game from then on with a high probability. This follows from the uniqueness of the agent's best response against the regulator's repeated strategy of playing D.

Lemma 10. *Suppose that there exists* $\mathcal{F} \subset \Omega$ *with* $\tilde{Q}(\mathcal{F}) > 0$ *such that, for any* $\xi > 0$ *and* $\psi > 0$*, there exists* $\bar{t}_\alpha$ *for which (16) and (17) stated in Lemma 9 hold for all* $\omega \in \mathcal{F}$ *for a given* α*-NE* $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ *with* $\alpha > 0$ *arbitrarily small. Then, for some* $\zeta_\alpha > 0$*,*

$$\tilde{Q}(\sup_{t'>t} \|\hat{\sigma}_R - E[\tilde{\sigma}_{Rt'} \mid \mathcal{H}_{At'}]\| < \zeta_\alpha \mid \mathcal{H}_{At}) > 1 - \zeta_\alpha, \quad \forall t \geq \bar{t}_\alpha, \forall \omega \in \mathcal{F}. \tag{18}$$

Proof. Let $Z_t \equiv \sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\|$ and (16) and (17) hold by the hypothesis. For any $\epsilon > 0$ (so that $\xi = \epsilon\psi$ as given in Lemma 9), $\tilde{Q}(Z_t < \psi \mid \mathcal{H}_{At}) > 1 - \epsilon$ for all $t > \bar{t}_\alpha$ on $\mathcal{F}$. This means that the agent chooses the strategy that is the unique best response to the commitment strategy of the regulator with a very high probability not only in the current periods but in every continuation game after $t > \bar{t}_\alpha$ on $\mathcal{F}$ on the given α-NE.

Fix some $s > 0$ and a private history h_{At}, where $h_{A\bar{t}_\alpha}$ is the initial segment of h_{At}, in $\mathcal{F}$. Since the agent is discounting (one can identify $\delta_\alpha \in (0, 1)$ such that for all discount factors strictly exceeding δ_α), there exists $s' > s$ and $\zeta > 0$ such that for all $t' = t, ..., t + s'$ and $h_{At'}$ for which $\|\hat{\sigma}_R - E[\tilde{\sigma}_{Rt'} \mid h_{At'}]\| < \zeta$, the continuation strategy $\tilde{\sigma}_A$ of the agent (after the initial history $h_{A\bar{t}}$) agrees with $\hat{\sigma}_A \equiv BR_A(\hat{\sigma}_R)$ for the next s periods. In other words, if the agent expects to see D with a very high probability for s' number of periods after some private history, then he would be playing T with a very high probability for s periods. Since by hypothesis, $\tilde{\sigma}_A$ agrees with $\hat{\sigma}_A$ for every $t' > t$ and for all $t \geq \bar{t}_\alpha$, $\|\hat{\sigma}_R - E[\tilde{\sigma}_{Rt'} \mid h_{At'}]\| < \zeta$ must hold for all $t' > t$ and for all $t \geq \bar{t}_\alpha$. Then, we obtain (18). ■

D.3 The regulator's beliefs about the agent's behavior

Next, we show that the regulator eventually becomes convinced that the agent plays a best response to T in the continuation game on a class of private histories that involve successive plays of D. To become convinced about the agent's behavior and beliefs, he does not need to know her private history when he has been diligent.

To prove this result, we follow the footsteps of Cripps et al. (2007, Lemma 3); let the σ-algebra of the regulator who has played D up to (and not including) period s be $\mathcal{H}^D_{Rs}$. Then, regulator's information set at time s (given this particular filtration of private histories) if he were to know the private history of the agent at time t can be described by $\varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At})$, the smallest σ-algebra containing the σ-algebras $\mathcal{H}^D_{Rs}$ and $\mathcal{H}_{At}$.

Lemma 11. *For any given* α*-NE* $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ *with* $\alpha > 0$ *arbitrarily small and for any* $t > 0$ *and* $\tau \geq 0$, $\lim_{s\to\infty} \left\|\tilde{E}[\tilde{\sigma}_{A,s+\tau} \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At})] - \tilde{E}[\tilde{\sigma}_{A,s+\tau} \mid \mathcal{H}^D_{Rs}]\right\| = 0$, $\tilde{Q} - a.s.$

Proof. We prove for $\tau = 0$. The case of $\tau \geq 1$ can be proven by induction and making the appropriate modifications in the proof of Lemma 3 of Cripps et al. (2007) provided in Appendix A.3. Suppose that $K \subset A^t$ is a set of t-period agent action profiles $(a_0, a_1, ..., a_{t-1})$, which also denotes the corresponding event. By Bayes' rule, we can derive the conditional probability of the event K given that the regulator has played diligent D, i.e., after the private history $\hat{h}_{R,s+1} \equiv (\hat{h}_{Rs}, D, i_d)$ with $i_d \in I_d$, as follows:

$$\begin{aligned}\tilde{Q}[K \mid \hat{h}_{R,s+1}] &= \tilde{Q}[K \mid \hat{h}_{Rs}, D, i_d] = \frac{\tilde{Q}[K \mid \hat{h}_{Rs}]\tilde{Q}[D, i_d \mid K, \hat{h}_{Rs}]}{\tilde{Q}[D, i_d \mid \hat{h}_{Rs}]} \\ &= \frac{\tilde{Q}[K \mid \hat{h}_{Rs}] \sum_{a\in A} \rho(i_d \mid a, D)\tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid K, \hat{h}_{Rs}]}{\sum_{a\in A} \rho(i_d \mid a, D)\tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid \hat{h}_{Rs}]}.\end{aligned}$$

Subtracting $\tilde{Q}[K \mid \hat{h}_{Rs}]$ from both sides we get $\tilde{Q}[K \mid \hat{h}_{R,s+1}] - \tilde{Q}[K \mid \hat{h}_{Rs}]$ equals

$$\frac{\tilde{Q}[K \mid \hat{h}_{Rs}] \sum_{a\in A} \rho(i_d \mid a, D)\Big(\tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid \hat{h}_{Rs}]\Big)}{\sum_{a\in A} \rho(i_d \mid a, D)\tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid \hat{h}_{Rs}]}.$$

Note that the term $\sum_{a\in A} \rho(i_d \mid a, D)\tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid \hat{h}_{Rs}]$ (equals $\beta\tilde{E}[\tilde{\sigma}_A^U(h_{As}) \mid \hat{h}_{Rs}]$ when $i_d = 0$ and $\tilde{E}[\tilde{\sigma}_A^U(h_{As}) \mid \hat{h}_{Rs}](1-\beta) + \tilde{E}[\tilde{\sigma}_A^T(h_{As}) \mid \hat{h}_{Rs}] * 1$ when $i_d = 1$) is strictly positive and less than or equal to one by assumption, for all $i_d \in I_d$. Hence, $\left|\tilde{Q}[K \mid \hat{h}_{R,s+1}] - \tilde{Q}[K \mid \hat{h}_{Rs}]\right| \geq \tilde{Q}[K \mid \hat{h}_{Rs}]\left| \sum_{a\in A} \rho(i_d \mid a, D)(\tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{E}[\tilde{\sigma}_A^a(h_{As}) \mid \hat{h}_{Rs}])\right|$.

As each of the random variables $\{\tilde{Q}[K \mid \mathcal{H}_{Rs}^D]\}_s$ is a martingale with respect to $(\{\mathcal{H}_{Rs}^D\}_s, \tilde{Q})$, it converges to a non-negative limit as $s \to \infty$. Thus, the LHS of the above inequality goes to zero $\tilde{Q}$-a.s. Let

$$\Pi_{A,I_d} = \begin{bmatrix} 0 & \beta \\ 1 & (1-\beta) \end{bmatrix}, \text{ and}$$

$$\tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid \hat{h}_{Rs}] = \begin{pmatrix} \tilde{E}[\tilde{\sigma}_A^T(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{E}[\tilde{\sigma}_A^T(h_{As}) \mid \hat{h}_{Rs}] \\ \tilde{E}[\tilde{\sigma}_A^U(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{E}[\tilde{\sigma}_A^U(h_{As}) \mid \hat{h}_{Rs}] \end{pmatrix}.$$

to rewrite the RHS as

$$\tilde{Q}[K \mid \mathcal{H}_{Rs}^D] \left\|\Pi_{A,I_d} \cdot \Big(\tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid \hat{h}_{Rs}]\Big)\right\|.$$

As there exists a strictly positive constant x such that

$$\left\|\Pi_{A,I_d} \cdot \Big(\tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid \hat{h}_{Rs}]\Big)\right\| \geq x\left\|\Big(\tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid \hat{h}_{Rs}]\Big)\right\|$$

we get $\lim_{s\to\infty} \tilde{Q}[K \mid \mathcal{H}_{Rs}^D]\left\|\tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid \hat{h}_{Rs}]\right\| = 0$, $\tilde{Q}$-a.s on K, where $\varphi(\mathcal{H}_{Rs}^D, K)$ is the smallest σ-algebra containing the σ-algebra $\mathcal{H}_{Rs}^D$ and event K. Since, $\tilde{Q}[K \mid \mathcal{H}_{R\infty}^D](\omega) \geq \tilde{Q}[K \mid \mathcal{H}_{Rs}^D](\omega) > 0$ for all $s > t$ for $\tilde{Q}$-almost all $\omega \in K$,

$$\lim_{s\to\infty} \left\|\tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \varphi(\mathcal{H}_{Rs}^D, K)] - \tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \mathcal{H}_{Rs}^D]\right\| = 0, \ \tilde{Q}\text{-a.s on } K.$$

As this holds for all $K \in \mathcal{H}_{At}$, $\lim_{s\to\infty} \left\| \tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At})] - \tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \mathcal{H}^D_{Rs}] \right\| = 0$, $\tilde{Q}$ − a.s. ■

D.4 The agent's beliefs about her own future behavior – revisited

The following result establishes that for any given α-NE with $\alpha > 0$, there exists a period $\hat{t}_\alpha$ such that in any history following that period, the agent expects the strategic regulator to choose D with a high probability.

In what follows, we restrict attention to histories in which the agent chooses U and denote the resulting filtration by $\mathcal{H}^U_{At}$.

The regulator's reputation does not change if the realization of the agent's completely mixed strategy is T in a given α-NE with $\alpha > 0$. Moreover, the agent plays U with at least α probability in any given private history of the agent. Thus, if there exists an event (denoted by $\omega \in \mathcal{F}$ as in Lemma 9) such that the strategic regulator plays lazy L with some strictly positive probability in every continuation game after any period t, then considering histories in which the agent plays U (on account of being constrained by α-NE) along with R's identification condition, Remark 2, empower us to employ the merging argument of Cripps et al. (2007). In these histories, in which the strategic regulator's behavior does not converge to playing D with a high probability, his true type is going to be revealed to the agent in the long run.[30] But, such histories would be in contradiction with (18) of Lemma 10. So, we obtain:

Lemma 12. *Suppose that (18) given in Lemma 10 is satisfied for all $\omega \in \mathcal{F}$ for a given α-NE $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ arbitrarily small. Then, there exists $\hat{t}_\alpha$ such that the following holds for some $\nu > 0$,*

$$\tilde{Q}\big(\sup_{t'>t} \|\tilde{\sigma}_R - \tilde{E}[\tilde{\sigma}_{Rt'} \mid \mathcal{H}_{At'}]\| < \nu \mid \mathcal{H}^U_{At}\big) > 1 - \nu, \quad \forall t \geq \hat{t}_\alpha. \tag{19}$$

The proof, sketched above, employs the same identification arguments used in the proof of Lemma 11 that parallel the merging argument of Cripps et al. (2007), and hence, is omitted.

D.5 The agent's beliefs about the regulator's future behavior – revisited

To prove Theorem 3 with a contradiction, we suppose that there exists $\mathcal{A} \in \Omega$ with $\tilde{Q}(\mathcal{A}) > 0$ such that for all $\omega \in \mathcal{A}$, $\lim_{t\to\infty} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At} \mid \mathcal{H}_{At}]\| = \alpha$ given an α-NE $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ but arbitrarily small. Then, by Lemma 9, there exists $\mathcal{F} \subset \mathcal{A}$ with $\tilde{Q}(\mathcal{F}) > 0$ such that, there exists $\bar{t}_\alpha$ for which the agent assigns very high probability to the event that all

[30] The agent uses her private history for updating her beliefs about the regulator's type. Her posterior belief at time t is given by the $\mathcal{H}_{At}$ - measurable random variable $\gamma_t \equiv Q(tough \mid \mathcal{H}_{At}) : \Omega \to [0, 1]$. At any equilibrium, γ_t is a bounded martingale with respect to $\{\mathcal{H}_{At}\}_t$ and measure Q. Therefore, γ_t converges Q-almost surely to a random variable γ_∞. Since $\tilde{Q}$ is absolutely continuous with respect to Q, any statement that holds Q-almost surely also holds $\tilde{Q}$ - almost surely. Thus, γ_t also converges to $\tilde{Q}$ - almost surely to a random variable γ_∞.

her continuation strategies from then on are best replies to the commitment strategy of the regulator. For this to be optimal for the agent, Lemmas 10 and 12 establish that the agent believes that the strategic regulator is going to be diligent from $t \geq \max\{\bar{t}_\alpha, \hat{t}_\alpha\}$ on in every continuation game with a very high probability on $\mathcal{F}$. Lemma 11 states that the agent's beliefs about her future behavior will eventually be known to the strategic regulator who is indeed diligent with a high probability (as expected by the agent). But, if the strategic type of the regulator eventually expects the agent to always give a best reply to the commitment strategy of the regulator in every continuation game, then the regulator would like to deviate from the commitment strategy D as it is noncredible, i.e., not a best response to the best response of the agent to the commitment strategy. Now, there seems to be a contradiction with the agent's beliefs about the strategic regulator's behavior (Lemma 12) on $\mathcal{F}$ and the regulator's behavior on a set $\mathcal{G}$ (to be explained below) where the regulator expects to see the agent give a best response to the commitment strategy D and is expected to play D. But, one needs to establish that $\mathcal{G}$ is a subset of $\mathcal{F}$ and measurable for the agent. Instead, following Cripps et al. (2007), we show that $\mathcal{G}$ is close to a $\mathcal{H}^U_{As}$ - measurable set, s sufficiently high and specified below, on which the agent believes that all her future behavior is going to be a best response (restricted by α) to the commitment strategy of the regulator.

Lemma 13. *Let $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ be an α-NE with $\alpha > 0$ and sufficiently small and suppose that $\mathcal{F}$ is the positive probability set of events stated in Lemmas 9 and 10; and $\nu > 0$ be the constant in Lemma 12. For any $\xi > \alpha$ (as in Lemma 9 given α and $\mathcal{F}$) and any $\tau \in \mathbb{N}$, there is an event $\mathcal{G}$ and a time $T(\xi, \tau)$ such that for all $s > T(\xi, \tau)$, there are $C_s \in \mathcal{H}^U_{As}$ and $\phi > 0$ with*

$$\|\hat{\sigma}_R - \tilde{E}[\tilde{\sigma}_{Rs} \mid \mathcal{H}_{As}]\| < \nu, \ \tilde{Q} - a.s. \ \ on \ \ C_s \tag{20}$$

$$\mathcal{G} \cup \mathcal{F} \subset C_s \ \ and \ \ \tilde{Q}(\mathcal{G}) > \tilde{Q}(C_s) - \phi\tilde{Q}(\mathcal{F}), \ and \tag{21}$$

$$on \ \mathcal{G}, \ \tilde{E}[\hat{\sigma}_{As'} \mid \mathcal{H}^D_{Rs}] > 1 - 2\sqrt{\xi}, \ \forall s' = \{s, s+1, ..., s+\tau\}. \tag{22}$$

Proof. We use Lemma 11 and a modified version of Cripps et al. (2007, Lemma 4).

Fix $\tau > 0$. Take $\xi > \alpha$ let $\bar{t}_\alpha$ denote the threshold period stated in Lemma 9 for this ξ and notice that the resulting set of events $\mathcal{F}$ is such that $\tilde{Q}(\mathcal{F})$ tends to one as $t \geq \bar{t}_\alpha$ increases. Thus, there exists $x < 1/5$ such that $\xi \in (\alpha, (x\,\tilde{Q}(\mathcal{F}))^2)$. Regulator's minimum estimate on the probability of truthfulness over periods $s, ..., s+\tau$ when his private history in $\mathcal{H}^D_{Rs}$ can be expressed as $f_s \equiv \min_{s \leq s' \leq s+\tau} \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \mathcal{H}^D_{Rs}]$ where T denotes the truthful action. Note that $f_s > 1 - 2\sqrt{\xi}$ is sufficient to show (22). The first step is to find a lower bound for f_s. For any $t \le s$, the triangle inequality implies $\min_{s \le s' \le s+\tau} \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At})] - k^t_s \le f_s \le 1$ where $k^t_s \equiv \max_{s \le s' \le s+\tau} \left| \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At})] - \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \mathcal{H}^D_{Rs}] \right|$ for $t \le s$. By Lemma 11, $\lim_{s\to\infty} k^t_s = 0$, $\tilde{Q}$ - a.s. Let $\mathcal{G}^0_t \equiv \{\omega : \tilde{\sigma}_{As}(h_{As}) = 1-\alpha, \forall s \ge t\}$. Then, $f_s \ge \tilde{Q}(\mathcal{G}^0_t \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At})) - k^t_s$. The sequence of random variables $\{\tilde{Q}(\mathcal{G}^0_t \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At}))\}_s$ is a martingale with respect to the filtration $\{\mathcal{H}^D_{Rs}\}$, so it converges a.s. to $g^t \equiv \tilde{Q}(\mathcal{G}^0_t \mid \varphi(\mathcal{H}^D_{R\infty}, \mathcal{H}_{At}))$. So,

$$1 \ge f_s \ge g^t - k^t_s - l^t_s \tag{23}$$

where $l^t_s \equiv |g^t - \tilde{Q}(\mathcal{G}^0_t \mid \varphi(\mathcal{H}^D_{Rs}, \mathcal{H}_{At}))|$ and $\lim_{s\to\infty} l^t_s = 0$, $\tilde{Q}$ - a.s.

The second step involves finding the sets C_s and an intermediate set (to be denoted by $\mathcal{F}^*$) that is used to determine the set $\mathcal{G}$. First, for any $t \ge \max\{\bar{t}_\alpha, \hat{t}_\alpha\} \equiv t_\alpha$ (the critical periods from Lemmas 9 and 10 and Lemma 12), using condition (16) of Lemma 9 implying condition (19) of Lemma 12, we define the associated events

$$\mathcal{K}_t \equiv \{\omega : \tilde{Q}(\mathcal{G}^0_t \mid \mathcal{H}_{At}) > 1-\xi, \|\hat{\sigma}_R - \tilde{E}[\tilde{\sigma}_{Rt} \mid \mathcal{H}_{At}]\| < \nu\} \in \mathcal{H}^U_{At}.$$

Let $\mathcal{F}^s_t \equiv \bigcap_{\tau=t}^s \mathcal{K}_\tau$ and $\mathcal{F}_t \equiv \bigcap_{\tau=t}^\infty \mathcal{K}_\tau$. Note that $\liminf \mathcal{K}_t \equiv \bigcup_{t=t_\alpha}^\infty \bigcap_{\tau=t}^\infty \mathcal{K}_\tau = \bigcup_{t=t_\alpha}^\infty \mathcal{F}_t$. By Lemmas 9, 10, and 12, $\mathcal{F} \subset \mathcal{K}_t$ for all $t \ge t_\alpha$; thus, $\mathcal{F} \subset \mathcal{F}^s_t$, $\mathcal{F} \subset \mathcal{F}_t$ and $\mathcal{F} \subset \liminf \mathcal{K}_t$.

Also define $\mathcal{N}_t \equiv \{\omega : g^t \ge 1 - \sqrt{\xi}\}$, the measure of events that the strategic regulator expects the agent to play T with probability $1-\alpha$ exceeds $1-\sqrt{\xi}$. Set $C_s \equiv F^s_{t_\alpha} \in \mathcal{H}^U_{At}$ and define an intermediate set $\mathcal{F}^*$ by $\mathcal{F}^* \equiv \mathcal{F}_{t_\alpha} \cap \mathcal{N}_{t_\alpha}$. Since $C_s \subset \mathcal{K}_s$, (20) holds (by Lemmas 9 - 12). And, as $\mathcal{F}^* \cup \mathcal{F} \subset C_s$, the first part of (21) holds with $\mathcal{F}^*$ in the role of $\mathcal{G}$. By definition, $\tilde{Q}(C_s) - \tilde{Q}(\mathcal{F}^*) = \tilde{Q}(C_s \cap (\mathcal{F}_{t_\alpha} \cap \mathcal{N}_{t_\alpha})^{\mathrm{C}}) = \tilde{Q}((C_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) \cup (C_s \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}))$ where $(X)^{\mathrm{C}}$ is the complement of X. Since the event $C_s \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}$ is a subset of $\mathcal{K}_{t_\alpha} \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}$, we have

$$\tilde{Q}(C_s) - \tilde{Q}(\mathcal{F}^*) \le \tilde{Q}(C_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) + \tilde{Q}(\mathcal{K}_{t_\alpha} \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}). \tag{24}$$

Next, we find the upper bounds for the two terms on the right-hand side of (24).

First, note that $\tilde{Q}(C_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) = \tilde{Q}(\mathcal{F}^s_{t_\alpha}) - \tilde{Q}(\mathcal{F}_{t_\alpha})$ by definition. Since $\lim_{s\to\infty} \tilde{Q}(\mathcal{F}^s_{t_\alpha}) = \tilde{Q}(\mathcal{F}_{t_\alpha})$, there exists $T' \ge t_\alpha$ such that

$$\tilde{Q}(C_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) < \sqrt{\xi} \quad \forall s \ge T'. \tag{25}$$

Also, as $\tilde{Q}(\mathcal{G}^0_t \mid \mathcal{K}_t) > 1-\xi$ and $\mathcal{K}_t \in \mathcal{H}^U_{At}$, iterated expectations imply that $1-\xi < \tilde{Q}(\mathcal{G}^0_t \mid \mathcal{K}_t) = \tilde{E}[g^t \mid \mathcal{K}_t]$. Since $g^t \le 1$, one gets $1-\xi < \tilde{E}[g^t \mid \mathcal{K}_t] \le (1-\sqrt{\xi})\, \tilde{Q}((\mathcal{N}_t)^{\mathrm{C}} \mid \mathcal{K}_t) + \tilde{Q}(\mathcal{N}_t \mid \mathcal{K}_t) = 1 - \sqrt{\xi}\, \tilde{Q}((\mathcal{N}_t)^{\mathrm{C}} \mid \mathcal{K}_t)$. So, $\tilde{Q}((\mathcal{N}_t)^{\mathrm{C}} \mid \mathcal{K}_t) < \sqrt{\xi}$. By taking $t = t_\alpha$, we get

$$\tilde{Q}(\mathcal{K}_{t_\alpha} \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}) < \sqrt{\xi}. \tag{26}$$

Using (25) and (26) in (24), $\tilde{Q}(C_s) - \tilde{Q}(\mathcal{F}^*) < 2\sqrt{\xi}$ for all $s \ge T'$. Given that $\mathcal{F} \subset C_s$ and the bound on ξ, $\tilde{Q}(\mathcal{F}^*) > \tilde{Q}(\mathcal{F}) - 2\sqrt{\xi} > (1-2x)\tilde{Q}(\mathcal{F}) > 0$.

Now, we use the two steps to obtain $\mathcal{G}$ and the bound on f_s. As $\tilde{Q}(\mathcal{F}^*) > 0$ and $k_s^{t_\alpha} + l_s^{t_\alpha}$ converges almost surely to zero; by Egorov's Theorem, there exists $\mathcal{G} \subset \mathcal{F}^*$ such that $\tilde{Q}(\mathcal{F}^* \setminus \mathcal{G}) < \sqrt{\xi}$ and $T'' > t_\alpha$ such that $k_s^{t_\alpha} + l_s^{t_\alpha} < \sqrt{\xi}$ on $\mathcal{G}$ for all $s \geq T''$. Then, letting $T(\xi, \tau) \equiv \max\{T', T''\}$ and noting $\mathcal{G} \cup \mathcal{F} \subset \mathcal{F}^* \cup \mathcal{F} \subset C_s$, we see that the first part of (21) holds. Moreover, $\tilde{Q}(\mathcal{G}) > \tilde{Q}(\mathcal{F}^*) - \sqrt{\xi}$ and as $\tilde{Q}(\mathcal{F}^*) > \tilde{Q}(\mathcal{F}) - 2\sqrt{\xi}$ we see that $\tilde{Q}(\mathcal{G}) > \tilde{Q}(\mathcal{F}) - 3\sqrt{\xi}$ and as $\xi < (x\tilde{Q}(\mathcal{F}))^2$ we obtain $\tilde{Q}(\mathcal{G}) > (1-3x)\tilde{Q}(\mathcal{F})$. Now, as there is ϕ such that $\phi\tilde{Q}(\mathcal{F}) > \tilde{Q}(\mathcal{F}^*)$ with $\phi > 1-2x$ we have that $\tilde{Q}(C_s) - \phi\tilde{Q}(\mathcal{F}) < \tilde{Q}(C_s) - \tilde{Q}(\mathcal{F}^*) < 2\sqrt{\xi} < 2x\tilde{Q}(\mathcal{F})$ and we wish to obtain that $2x\tilde{Q}(\mathcal{F}) < (1-3x)\tilde{Q}(\mathcal{F})$. Thus, because that $x \in (0, 1/5)$ we see that $\tilde{Q}(\mathcal{G}) > (1-3x)\tilde{Q}(\mathcal{F})$ and $\tilde{Q}(C_s) - \phi\tilde{Q}(\mathcal{F}) < 2x\tilde{Q}(\mathcal{F})$ is satisfied. Thus, the second part of (21) holds for all $s > T(\xi, \tau)$. And, notice that $g^{t_\alpha} \geq 1 - \sqrt{\xi}$ on $\mathcal{G}$ since $\mathcal{G} \subset \mathcal{N}_{t_\alpha}$. Thus, on $\mathcal{G}$, $f_s > 1 - 2\sqrt{\xi}$ for all $s > T(\zeta, \tau)$ by (23). This with the bound on ξ gives (22). ■

D.6 The proof of Theorem 3

Next, we will establish that on the set $\mathcal{G}$, the regulator's strategic type will not find it optimal to play the commitment strategy. This will make the agent's expectation of the strategic regulator's action move away from the commitment strategy on $\mathcal{F}$ through the relations established between the sets $\mathcal{G}$, $\mathcal{F}$ and C_s in Lemma 13. And, this will contradict with the expectations on $\mathcal{H}_{As}^U$ - measurable set $\mathcal{F}$ stated in Lemma 12.

Let $(\varsigma_R, \varsigma_A)$ be the stage game mixed action profile that puts probability one on the commitment action D and the best response to the commitment action T. Note that when the agent uses any strategy sufficiently close to ς_A, say at most $\bar{\nu}$ away from ς_A, playing ς_R is suboptimal by at least some $\mu > 0$. Then, for given δ, there exists a sufficiently large τ such that the loss of μ for one period is larger than any potential gain held off on for τ periods.

Proof of Theorem 3. Suppose that there exists $\mathcal{A} \in \Omega$ with $\tilde{Q}(\mathcal{A}) > 0$ such that for all $\omega \in \mathcal{A}$, $\lim_{t\to\infty} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At} \mid \mathcal{H}_{At}]\| = \alpha$ for the given α-NE $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ and α sufficiently small. Then, fix $\mathcal{F}$ from Lemma 9 with $\xi > \alpha$ and $t_\alpha^* = \max\{\bar{t}_\alpha, \hat{t}_\alpha, T(\xi, \tau)\}$ (as specified in Lemmas 9 – 13). For $\xi < \bar{\nu}$ and τ, let $\mathcal{G}$ and C_s be the events described in Lemma 13 for $s > T(\xi, \tau)$. Consider the period $s > T(\xi, \tau)$ at some state in $\mathcal{G}$. By (22), the regulator, who would be playing D, expects to see a strategy within $2\sqrt{\xi}$ of ς_A for the next τ periods where $\xi < \bar{\nu}$ (by choosing ξ sufficiently small which is feasible as $\alpha > 0$ is sufficiently small, one can ensure that $2\sqrt{\xi} < \bar{\nu}$). Playing ς_R is suboptimal in period s since the most he can gain from playing D is less than playing a best response to ς_A for τ periods. Thus, on $\mathcal{G}$, the strategic type of the regulator would like to play D with zero probability, which essentially is a contradiction. And, to get a contradiction in agent's beliefs, we calculate a lower bound on the difference between ς_R and the agent's beliefs about the strategic type playing action D in period s, $\tilde{E}[\tilde{\sigma}_{Rs}(D) \mid \mathcal{H}^U_{As}]$ on the events in C_s that contains $\mathcal{F}$ due to (21) of Lemma 13:

$$\begin{aligned}
\tilde{E}[\left|\varsigma_R - \tilde{E}[\tilde{\sigma}_{Rs}(D) \mid \mathcal{H}_{As}]\right|\mathbf{1}_{C_s}] &\geq \tilde{E}[(\varsigma_R - \tilde{E}[\tilde{\sigma}_{Rs}(D) \mid \mathcal{H}_{As}])\mathbf{1}_{C_s}] \\
&\geq \tilde{Q}(C_s) - \tilde{E}[\tilde{\sigma}_{Rs}(D)\mathbf{1}_{C_s}] \\
&\geq \tilde{Q}(C_s) - (\tilde{Q}(C_s) - \tilde{Q}(\mathcal{G})) \geq \tilde{Q}(C_s) - \phi\tilde{Q}(\mathcal{F}) \\
&\geq (1-\phi)\tilde{Q}(\mathcal{F}).
\end{aligned}$$

Note that $\mathbf{1}_{C_s}$ is the indicator function on C_s and hence $(\tilde{E}[\tilde{\sigma}_{Rs} \mid \mathcal{H}_{As}]\mathbf{1}_{C_s}) = \tilde{E}[\tilde{\sigma}_{Rs} \mid \mathcal{H}^U_{As}]$. Then, the first inequality is just removing the absolute values. The second inequality applies $\varsigma_R(D) = 1$ and uses the $\mathcal{H}^U_{As}$ - measurability of C_s. The third is the result of $\tilde{\sigma}_{Rs}(D) = 0$ on $\mathcal{G}$ and $\tilde{\sigma}_{Rs}(D) \leq 1$ in the rest of the set with $\mathcal{G} \subset C_s \in \mathcal{H}^U_{As}$ due to (21) of Lemma 13, which also implies the third and fourth inequalities. Finally, the last one is by $\mathcal{F} \subset C_s$. Since $\tilde{E}[\left|\varsigma_R - \tilde{E}[\tilde{\sigma}_{Rs}(D) \mid \mathcal{H}_{As}]\right|\mathbf{1}_{C_s}] > (1-\phi)\tilde{Q}(\mathcal{F})$ for all $s > t^*_\alpha$ and, by Lemmas 12 and 13, on C_s, $\tilde{E}[|\varsigma_R - \tilde{E}[\tilde{\sigma}_{Rs} \mid \mathcal{H}_{As}]|\mathbf{1}_{C_s}] < \nu$, we obtain the desired contradiction. ■

References

Al-Najjar, N. I. (1995). Decomposition and characterization of risk with a continuum of random variables. *Econometrica*, *63*(5), 1195-1224.

Aoyagi, M. (1996). Reputation and dynamic stackelberg leadership in infinitely repeated games. *Journal of Economic Theory*, *71*, 378-393.

Atakan, A. E., & Ekmekci, M. (2012). Reputation in long-run relationships. *The Review of Economic Studies*, *79*(2), 451-480.

Atakan, A. E., & Ekmekci, M. (2015). Reputation in the long-run with imperfect monitoring. *Journal of Economic Theory*, *157*, 553–605.

Barlo, M., Carmona, G., & Sabourian, H. (2016). Bounded memory folk theorem. *Journal of Economic Theory*, *163*, 728–774.

Benabou, R., & Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus and credibility. *Quarterly Journal of Economics*, *107*(3), 921-958.

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

Celentani, M., Fudenberg, D., Levine, D. K., & Pesendorfer, W. (1996). Maintaining a reputation against a long-lived opponent. *Econometrica*, *64*(3), 691-704.

Chan, J. (2000). On the non-existence of reputation effects in two-person infinitely repeated games. *Working paper*.

Chung, K. L. (1974). *A course in probability theory*. New York: Academic Press.

Cripps, M. W., Dekel, E., & Pesendorfer, W. (2005). Reputation with equal discounting in repeated games with strictly conflicting interests. *Journal of Economic Theory*, *121*, 259-272.

Cripps, M. W., Mailath, G. J., & Samuelson, L. (2004). Imperfect monitoring and impermanent reputations. *Econometrica*, *72*(2), 407-432.

Cripps, M. W., Mailath, G. J., & Samuelson, L. (2007). Disappearing private reputations in long-run relationships. *Journal of Economic Theory*, *134*(1), 287–316.

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

Dorigo, M. (1992). Optimization, learning and natural algorithms. *PhD Thesis, Politecnico di Milano*.

Ekmekci, M. (2011). Sustainable reputations with rating systems. *Journal of Economic Theory*, *146*, 479-503.

Ekmekci, M., Gossner, O., & Wilson, A. (2012). Impermanent types and permanent reputations. *Journal of Economic Theory*, *147*, 162 - 178.

Ely, J. C., Fudenberg, D., & Levine, D. K. (2008). When is reputation bad? *Games and Economic Behavior*, *63*, 498 - 526.

Ely, J. C., & Välimäki, J. (2003). Bad reputation. *The Quarterly Journal of Economics*, *118*(3), 785 - 814.

Fudenberg, D., & Levine, D. K. (1989). Reputation and equilibrium selection in games with a patient player. *Econometrica*, *57*(4), 759-778.

Fudenberg, D., & Levine, D. K. (1992). Maintaining a reputation when strategies are imperfectly observed. *Review of Economic Studies*, *59*, 561-579.

Gale, D., & Rosenthal, R. (1994). Price and quality cycles for experience goods. *The RAND Journal of Economics*, *25*, 590-607.

Gossner, O. (2011). Simple bounds on the value of reputation. *Econometrica*, *79*(5), 1627-1641.

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

Harsanyi, J. C. (1967-68). Games of incomplete information played by bayesian players, parts i-iii. *Management Science*, *14*(3), 159-182.

Hart, S. (1985). Nonzero-sum two person repeated games with incomplete information.

Mathematics of Operations Research, *10*, 117-153.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, *66*, 169-182.

Kreps, D., & Wilson, R. (1982). Sequential equilibria. *Econometrica*, *50*, 863 - 894.

Liu, Q. (2011). Information acquisition and reputation dynamics. *Review of Economic Studies*, *78*, 1400-1425.

Liu, Q., & Skrzypacz, A. (2014). Limited records and reputation bubbles. *Journal of Economic Theory*, *151*, 2–29.

Mailath, G. J., & Samuelson, L. (2001). Who wants a good reputation? *Review of Economic Studies*, *68*, 415-441.

Mailath, G. J., & Samuelson, L. (2006). *Repeated games and reputations: Long-run relationships*. Oxford: Oxford University Press.

Mailath, G. J., & Samuelson, L. (2015). Chapter 4: Reputations in repeated games. In P. Young & S. Zamir (Eds.), *Handbook of game theory, volume 4*. Elsevier Science Publishers.

Marinovic, I., Skrzypacz, A., & Varas, F. (2018). Dynamic certification and reputation for quality. *American Economic Journal: Microeconomics*, *10*(2), 58–82.

Milgrom, P., & Roberts, J. (1982). Predation, reputation and entry deterrence. *Journal of Economic Theory*, *27*, 280 - 312.

Özdoğan, A. (2014). Disappearance of reputations in two-sided incomplete-information games. *Games and Economic Behavior*, *88*, 211–220.

Özdoğan, A. (2016). Occurrence of deception under the oversight of a regulator having reputation concerns. In L. A. Petrosyan & V. V. Mazalov (Eds.), *Recent advances in game theory and applications, european meeting on game theory, saint petersburg, russia, 2015 and networking games and management, petrozavodsk, russia, 2015.* Birkhauser.

Phelan, C. (2006). Public trust and government betrayal. *Journal of Economic Theory*, *130*, 27-43.

Schmidt, K. M. (1993a). Commitment through incomplete information in a simple repeated bargaining game. *Journal of Economic Theory*, *60*, 114-139.

Schmidt, K. M. (1993b). Reputation and equilibrium characterization in repeated games of conflicting interests. *Econometrica*, *61*(2), 325-351.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*, 25-55.

Wiseman, T. (2008). Reputation and impermanent types. *Games and Economic Behavior*, *62*(1), 190-210.