Inducing Good Behavior via Reputation*

Ayça Özdoğan† Mehmet Barlo‡

June 19, 2020

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 specific 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 to attain the social optimum in *any* Nash equilibrium. However, with long-lived and patient players, reputation cannot induce permanent good behavior in equilibrium involving sporadic experimentation with bad behavior. The stark contrast between these cases portrays the significance of the longevity of the interaction and provides 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.

*This paper subsumes the result that appeared in a conference proceeding, Özdoğan (2016), under the title "Occurrence of deception under the oversight of a regulator having reputation concerns" and included in the current draft for completeness and comparability. We are grateful to the Editor, Associate Editor and three anonymous referees; 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 the remaining errors are ours.

†**Corresponding Author**: 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. The author gratefully acknowledges financial support from the Scientific and Technological Research Council of Turkey (TUBITAK)–Project No.115K567 for the first part of this study.

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

Contents

1 Introduction

The misbehavior of an agent facing a regulator who spends costly time and resources to audit/investigate is frequently observed. Examples include an investor engaged in fraud by misrepresenting his books to a certified auditor; a bank misreporting the information about its financial health to a regulator; a taxpayer filing false income statements to a tax authority; a construction or mine company neglecting work safety precautions and misreporting its practices; an employee not exerting the promised effort in a business owned by a principal, etc. The instances of financial fraud, corporate deceptions, and work accidents often involve these kinds of misbehavior which are frequently related to the corresponding regulator's reputation for being diligent or lack of it.[1] Indeed, reputation concerns of regulators may prevent or lessen the extent and severity of such socially undesirable outcomes and behavior.[2]

Motivated by instances of "bad" behavior frequently emerging in regulated and repeated environments, this paper aims to unravel whether or not regulators' reputation can induce the "good" behavior permanently when the repeated interaction is neither observable nor contractable. We analyze a dynamic environment where the regulator (he) is responsible to detect deviations from the good behavior via costly auditing yet may not be diligent in doing so because of the associated costs. We propose a dynamic incomplete information game with a stage game possessing a particular payoff and signaling structure played between a long-lived 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 but observe the public history of past play, reputation empowers

[1] For instance, Bernard Madoff was found guilty to several offenses including fraud and making false statements with 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 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. There was a mine accident that 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) told 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*).

the regulator to prevent agents' bad behavior by inducing them to behave (well) in the long-run with no cost (and hence attain the *maximum* payoff which coincides with the social optimum) in *any* Nash equilibrium. To address 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 single myopic (representative) agent.[3] We show that there exists a unique Markov equilibrium 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 (Markov states) at which the agents exhibit good behavior while the regulator incurs no cost. 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.

However, when the regulator faces the same long-lived agent, we show that 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* Nash equilibrium 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]

Hence, our findings display a stark contrast between the myopic agents' case and the case with a long-lived *skeptical* agent who cannot refrain from (costly) testing the regulator's reputation (no matter how high it is) by behaving badly now and then. 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 in the case of perpetual interaction among the participants.[5]

The repeated game between the regulator and the agent(s) involves unobservable actions

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

[4]To ensure that the agent experiments with the bad behavior occasionally, we discuss a setting where she suffers one-period amnesia with some small but positive probability in every period (in which case she hangs on to her sufficiently low initial beliefs that the regulator is tough). Moreover, *perfection* of Selten (1975) implies our notion of robustness as well. Sadly it is too powerful: the regulator (the informed player) would be enforced 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.

[5]We thank an anonymous referee for elegantly outlining and helping us to summarize these observations.

on both sides and incomplete information about the regulator's type. The agent's actions consist of good behavior and bad behavior and the regulator's of firm and lax actions. The behavioral type of the regulator is committed to playing the firm action. The good behavior by the agent generates a good public signal regardless of the regulator's choice. However, if the agent has chosen to behave badly, the regulator can *detect* the misbehavior with some probability (which determines the audit quality) only if he has chosen the costly firm action. The stage game payoffs are so that the agent's best response to believing that the regulator will be *diligent* (i.e., firm) is to be *truthful* (the good behavior), whereas it is to be *untruthful* (the bad behavior) if the agent believes the regulator will be *lazy* (lax). Meanwhile, the regulator's best response is to be lazy when the agent is truthful and to be diligent if the agent is expected to be untruthful. The regulator prefers that the agent is truthful (regulator-preferred action) and the agent prefers that the regulator is lazy (agent-preferred action). The regulator's reputation determines the agent's belief about the regulator's type and choices. If the *tough* type (the Stackelberg type), the regulator is always diligent, while the *strategic* regulator may choose to establish a false reputation by mimicking the tough type.

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 as some instances fit situations where the regulator faces different myopic agents in 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 who play once observing the public history of past play; (2) a long-lived regulator faces a long-lived agent.

In the first, we show that any Nash equilibrium payoff of the strategic regulator converges to the efficient payoff as the discount factor approaches one given any interior initial (common) belief the agents may have about the regulator's type. Moreover, he enjoys a permanent (absorbing) reputation inducing agents' good behavior indefinitely. Then, the strategic regulator refrains from costly auditing and hence attains perpetual social efficiency. In furtherance, we establish that there is a unique Markov equilibrium with a continuous and nondecreasing value function such that the reputation for being diligent becomes permanent whenever it exceeds a threshold: The reputation is above this level implies all the future agents will behave and the regulator will be lax with probability one permanently. Thus, the perpetual social optimum is secured. For low values of the regulator's reputation, players use mixed strategies.

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 beliefs while not considering the information externality that they could initiate and be helpful to future generations.

This is because, when an agent is truthful, Bayesian updating is not called for as there will not be any signal suggesting that the regulator has not been diligent. Thus, the strategic and patient regulator will find it beneficial to make sure that his false reputation eventually reaches a level above which it persists as all the subsequent agents will find it optimal to be truthful thereafter. Hence, the strategic and patient regulator guarantees the maximum payoff (exceeding his Stackelberg payoff) in *any* Nash equilibrium and induce the agents to be truthful permanently in the long-run. Thus, good behavior is attained in perpetuity thanks to the reputation of the patient regulator and short-term incentives of the myopic agents.

The analysis is more complicated if the regulator faces a long-lived and sufficiently patient agent. Both make their choices and update their beliefs according to their private histories while the regulator cannot anticipate the strategy of the agent as her beliefs are private and the long-lived patient agent will not be giving myopic best responses. However still, insisting on a high initial reputation level may block the avenue leading to information externalities, thereby preventing the agent from making critical inferences in the future, by inducing the agent to be truthful with probability one and hence close the gate to Bayesian updating. This, on the other hand, is not robust: Even if the agent were to experiment with untruthfulness very occasionally, the agent will expose the regulator's false reputation in the long-run. So, there is no Nash equilibrium with sufficiently patient players, where the agent sporadically behaves badly and the regulator enjoys a high permanent reputation and hence perpetual social efficiency cannot be sustained in any Nash equilibrium in these cases.

The intuition behind this result 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 date with a very high probability. Thus, the agent must be expecting to see the diligence with a sufficiently high probability on average from then on for sufficiently long periods after observing her private history. In our setup, the "conditional identification of the agent" holds which is formalized in Remark 1 and it implies that the diligent regulator can infer the fixed behavior of the agent from the frequencies of the public signals.[6] Following the footsteps of Cripps, Mailath, and Samuelson (2007) thanks to "conditional identification of the

[6] This enables us to accommodate the techniques of Cripps, Mailath, and Samuelson (2007) suited for a particular set of private histories of the regulator (even though their critical assumptions of full support and full rank do not hold in our model) and surpass the complications due to private beliefs. The basic intuition of Cripps, Mailath, and Samuelson (2004) in establishing impermanency of public reputations is that the informed player knows the beliefs of the uninformed player. Thus, she knows when her deviations from the commitment action have virtually no effect on the opponent's beliefs and continuation payoffs. On the other hand, in proving disappearing private reputations, as the informed player cannot infer the opponent's belief, Cripps, Mailath, and Samuelson (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."

agent", we establish that "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 private history of the regulator (coinciding with the agent's private beliefs about the regulator's future behavior that is induced by the agent's private history). 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, the agent draws the irrefutable inference that the regulator is indeed of the strategic type and is choosing 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, the agent's private beliefs would be updated accordingly which the regulator cannot observe and hence respond to. Thus, there must be a period in which the agent's private beliefs are not compatible with facing a diligent regulator; delivering the desired contradiction.

1.1 Related Literature

Early literature studying the value of reputations focuses on settings where a long-lived player faces a sequence of short-lived players who play once but observe the past play. These studies provide the Stackelberg payoff as the lower bound on the sufficiently patient long-lived player's average limiting payoff given that there is a commitment type who always chooses the Stackelberg strategy.[7] Such one-sided reputation results also arise in settings that involve two long-lived players.[8] Cripps, Mailath, and Samuelson (2004), on the other hand, shows that a long-lived informed player, both against short-lived uninformed opponents and also a long-lived one (with a condition on the equilibrium behavior), can maintain a permanent reputation for playing a commitment strategy in a game with imperfect public monitoring only if that strategy is an equilibrium of the complete information stage game.[9] Thus, the powerful

[7]See Fudenberg and Levine (1989) for games with perfect monitoring, Fudenberg and Levine (1992) for games with imperfect public monitoring and Gossner (2011) for games with any imperfect monitoring.

[8]These studies include: Schmidt (1993a) (conflicting interests with asymmetric discount factors); Celentani, Fudenberg, Levine, and Pesendorfer (1996) and Aoyagi (1996) (with imperfect monitoring and asymmetric discount factors); Cripps, Dekel, and Pesendorfer (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).

[9]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 Markov equilibrium and show that insider traders reveal their true type

results about the lower bounds on the long-lived player's average payoff are short-run reputation effects, where the long-lived informed player's payoff is calculated at the beginning of the game.[10] Cripps, Mailath, and Samuelson (2007) extends their disappearance of noncredible reputations result by allowing for private beliefs in both cases and unrestricted equilibrium strategies for the uninformed long-lived player case. For the asymptotic equilibrium behavior, they show that the continuation play in every Nash equilibrium is a Nash equilibrium of the complete information game when the reputation is public; a correlated equilibrium of the complete information game when the reputation is private (regardless of the longevity of the interaction among the players). So, the selection effects of reputations on the set of equilibrium payoffs should not be assumed to carry over asymptotically. Meanwhile, Cripps, Mailath, and Samuelson (2004) implies non-transient reputations need the incorporation of other mechanisms into the model. One strand of literature attains recurrent reputations by assuming that the type of the player is governed by unobserved replacements and stochastic processes through time such as Gale and Rosenthal (1994), Holmström (1999), Mailath and Samuelson (2001), Phelan (2006), Wiseman (2008) and Ekmekci, Gossner, and Wilson (2012). The others propose mechanisms that rely on the restricted memory of the uninformed player(s) such as the costly discovery of histories (Liu (2011)), rating system (Ekmekci (2011)), bounded memory (Monte (2013)) or limited record-keeping (Liu and Skrzypacz (2014)).[11]

Our findings concerning the asymptotic equilibrium behavior and the permanency of reputation with short-lived uninformed players diverge from those of Cripps, Mailath, and Samuelson (2004) due to the differences in the monitoring structures: While Cripps, Mailath, and Samuelson (2004) assumes full-support and full-rank (any signal is probable after any action profile and identification of each player's constant action is possible following sufficiently many observations), our setting violates both of these assumptions.[12] Our main question is whether or not perpetual good behavior can be sustained in equilibrium via reputation. We provide an affirmative answer with myopic uninformed players and we show that, in the case of a long-lived uninformed player, this conclusion does not hold in any Nash equilibrium involving

asymptotically in any Markov equilibrium. Moreover, Özdoğan (2014) extends the disappearing reputations result to games with two long-lived players with incomplete information on both sides.

[10]The results of Fudenberg and Levine (1992) and Cripps, Mailath, and Samuelson (2004) differ because Fudenberg and Levine (1992) fixes the prior belief of being the commitment type and selects a threshold discount factor depending on this prior above which the player is sufficiently patient for their results to hold; whereas Cripps, Mailath, and Samuelson (2004) fixes the discount factor while allowing the posterior belief to vary which eventually becomes so low that makes the required threshold discount factor (for Fudenberg and Levine (1992)'s result to hold) to exceed way above the fixed discount factor.

[11]See Mailath and Samuelson (2015) for a survey on the reputation literature.

[12]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 agent's untruthfulness is probable only when the regulator is diligent (conditional identification of the agent).

agent's sporadic experimentation with bad behavior.

The other important work related to our short-lived agents case deals with bad reputations. Building on Ely and Välimäki (2003)'s motorist-mechanic example and bad reputation result, Ely, Fudenberg, and Levine (2008) characterizes a class of games in which the following holds: the short-run uninformed players decide whether to participate in a game with the long-run player while every action of the long-run player that induces the short-run players to participate may generate a signal which could suggest that the long-run player is bad. As a result, any equilibrium payoff of a sufficiently patient long-run player is close to her payoff from the exit decision of the short-run players (with a condition on the size of commitment types) which is assumed to be her minmax payoff. This also suggests a patient long-run player can attain her maximum payoff if the exit action provides her the maximum payoff (exit-maximum). The similarity of results in terms of equilibrium payoff when the participation games have exit-maximum lies in the fact that the public signals in Ely, Fudenberg, and Levine (2008) satisfy our condition of conditional identification of the long-lived informed player and the short-lived players do not find it optimal to experiment and unravel the type of the long-lived player. This leads to a persistent reputation in both studies. Yet, the two signaling structures differ in important ways. In their setup, there are exit signals that occur with probability one if the short-run players choose an exit action, which cannot be observed if the short-run players choose to participate and they 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 the analysis of the two long-lived player case which was left as an open question in Ely, Fudenberg, and Levine (2008).[13, 14]

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.

[13]We would like to note that 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.

[14]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., Dilmé (2019) and Marinovic, Skrzypacz, and Varas (2018) following Board and Meyer-ter Vehn (2013)). While only Halac and Prat (2016) and Marinovic, Skrzypacz, and Varas (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."

2 Model

We model the strategic interaction between the agent and regulator through a simultaneous-move stage game.[15]

At the beginning of the game, the agent (*she*) gets some private information about the state of nature and has the incentive to deceive the uninformed regulator (*he*) by strategically manipulating information through reporting false messages. The agent can either be *truthful* or *untruthful* about the information she has. Thus, the action set of the agent is $A = \{T, U\}$ where $a \in A$. The reporting strategy 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 is supposed to detect deviations from the truthful behavior via costly auditing. He chooses to be *diligent* or *lazy* in auditing the agent. His choice generates different probabilities of eliciting information about the agent's untruthfulness if she is indeed untruthful. The regulator's action set is $R = \{D, L\}$ while his strategy is given by $\sigma_R \in \Delta(R)$. By adopting a slight abuse of notation, we let σ_R also denote the probability of him choosing the diligent action.

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 $(a, r) \in A \times R$:

$$
\begin{aligned}
&\rho(1 \mid U, D) = 1 - \rho(0 \mid U, D) = \beta \qquad && \rho(1 \mid T, D) = 1 - \rho(0 \mid T, D) = 0 \\
&\rho(1 \mid U, L) = 1 - \rho(0 \mid U, L) = 0 \qquad && \rho(1 \mid T, L) = 1 - \rho(0 \mid T, L) = 0
\end{aligned}
$$

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 as $\rho(0|T, r) = 1$ for all $r \in R$.

An action chosen by a player is not observable to the other. Yet, the public signals become commonly observable at the end of the corresponding period (after all have made their choices) and provide information about the chosen actions. More specifically, each player can infer the other's action only when he/she chooses a particular action. In other words, public signals

[15] The stage game considered here is the one presented in Özdoğan (2016). The following is an alternative formulation that parallels 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 performance of the agent is unobservable to the principal. 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 a good signal even though the agent has chosen L.

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 frequencies of signals only when he has been diligent; the agent can identify the fixed action chosen by the regulator from the frequency of the detections only when she has been untruthful. Otherwise, no information about the opponent's behavior is revealed. These are summarized in Remarks 1 and 2.

Remark 1 *The* conditional identification of the agent's actions *is satisfied 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 $\rho(0|T, D) = \rho(0|T, L) = 1$.*

In our model, the full support assumption, typically presumed in many studies in the literature, does not hold.

Independent of the auditing strategy of the regulator, the agent's payoff is normalized to zero if she chooses T. When she is untruthful, 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 the agent's 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 untruthfulness is detected and otherwise his expected loss is f. The regulator incurs a cost 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 $u_R(U, L) = -f$.[16]

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

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

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

[16]Our ex post and ex ante payoff specifications differ from the ones typically assumed in the literature where a player's ex post payoff is presumed to depend on a player's own action and the public signal and the ex ante payoff to equal the expectation of the ex post payoffs taken over the opponents' actions. If we had followed this type of specification, we would have obtained $u_R(U, L) = u_R(T, L)$ for the regulator as he would be choosing the same action and receiving the same signal of no detection with probability one. However then, the forgone societal loss due to the agent choosing U would not be captured. We depart away from the literature in that regard to accommodate the influence of the conditional identification assumptions on ex-ante payoffs by reflecting the effects of the public signals on the ex-post payoffs only when the agent chooses U. Otherwise, the realized payoff of the regulator is independent of the signal and depends only on his own action D.

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

Table 1: Ex ante stage game payoffs under complete information

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.[17] Thus, the regulator would like to convince the agent for being 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 is also maximizing the total welfare.[18]

Given these assumptions, the unique Nash equilibrium is in mixed strategies:

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

Next, we discuss some important properties of the ex-ante stage game payoff structure. First, the *minmax payoffs* (both in pure and mixed strategies) 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 pure Stackelberg action of the regulator is D and D mixed-action minmaxes the agent. Thus, following Schmidt (1993b), the stage game described in the current paper has conflicting

[17]Consider the situation where the ex-post payoffs depend on a player's own action and the public signals. Let $u_R^*(1, D)$ denote the realized payoff following D and signal "detection" and $u_R^*(0, D)$ the realized payoff following action D and signal "no detection". Then, $u_R(T, D) = u_R^*(0, D)$. Thus, $u_R(T, D) > u_R(U, D)$ thanks to Assumption 2 implies $u_R^*(0, D) > u_R^*(1, D)$. This may be puzzling at first sight as the signal "detection" is the rewarding signal for the regulator. However, observing "no detection" is more likely when the agent chooses T (having a probability of $1/(1+\beta)$) compared to observing "no detection" when the agent chooses U (associated with a probability $\beta/(1+\beta)$). Thus, the former intrinsically incorporates that it is less likely for the regulator to incur a forgone societal loss due to untruthfulness. We thank an anonymous referee for pointing out that a justification for $u_R(T, D) > u_R(U, D)$ can be attained by considering a situation where "detection" of untruthfulness has follow-up (disciplinary) procedures requiring the regulator to perform additional tasks (e.g., preparing and presenting the resulting case in a court of law) and hence reducing his ex-ante payoffs.

[18]We thank an anonymous referee for pointing out this observation.

interests. The regulator's preferred 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.[19]

To induce uncertainty about the regulator's preferences and obtain a model with reputation, we incorporate a behavioral type into the game following Harsanyi (1967-68), Kreps and Wilson (1982) and Milgrom and Roberts (1982).[20] In particular, the regulator can be one of two types: *tough* or *strategic*. The tough regulator is committed to being diligent (which is the pure Stackelberg action for the strategic type), whereas the strategic one has the preferences described 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 $\gamma \in (0,1)$. Now, the agent's equilibrium behavior (determined by her belief about the probability of detection) 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 to choose $\sigma_A \in [0,1]$ to maximize

$$(1-\sigma_A)\ [(1-\pi)g - \pi l] \tag{1}$$

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 equilibrium of the incomplete information game is presented in Lemma 1 and its proof is in Appendix A.

Lemma 1 *The following strategy profile (σ_A, σ_R) constitutes a Nash equilibrium,*

1. $\sigma_A = 1$ *and* $\sigma_R = 0$ *if* $\gamma \geq \gamma^*$,
2. $\sigma_A = 1 - \frac{c}{\beta(d+f)}$ *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,

[19] If Assumption 2 is violated and $u_R(T,L) > u_R(U,D) > u_R(T,D) > u_R(U,L)$, then the Stackelberg action of the regulator is still D and (U,D) minmaxes the regulator; however, the Stackelberg action profile gives a payoff strictly less than the minmax payoff. Hence, the reputation of being diligent will have no use to the regulator. So, the strategic regulator will separate himself from a diligent behavioral type. If, on the other hand, $u_R(T,L) > u_R(U,D) > u_R(U,L) > u_R(T,D)$, then (U,D) still minmaxes the regulator but this time L becomes the pure Stackelberg action and the Stackelberg profile (U,L) provides a payoff strictly less than the minmax payoff (U,D) provides. Hence, the resulting model is similar to the one we solve.

[20] If $\beta = 1$, the agent choosing U implies in two possible outcomes: Detection whenever the regulator chooses D; and no detection whenever the regulator has chosen L. The second contingency indicates that he is of the strategic type. Thus, the game becomes one with *perfect monitoring*. Therefore, Fudenberg and Levine (1989) and Cripps, Mailath, and Samuelson (2004) imply that the informed player would not risk additional payoffs but obtain the Stackelberg level and maintain his reputation. Moreover, having $\beta = 1$ contradicts with Assumption 2 as we can no longer have $\beta < \frac{f}{f+d}$.

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 strategy specified in the lemma. Moreover, the equilibrium strategies are monotone in the prior belief.

3 Dynamic game with short-lived agents

The game begins at time $t = 1$ and is infinitely repeated. 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 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 $\gamma_0 \in (0, 1)$.

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^1 stands for the unique null history) consisting of whether or not each of the preceding agents has been detected. We let h_R^t be the private history of the regulator. It is composed of h^t and his past actions up to time t. A strategy for the long-lived player is a sequence of maps $\sigma_{R,t}(h_R^t) \in \Delta(R)$ and a strategy of agent t is $\sigma_{A,t}(h^t) \in \Delta(A)$.

A *Nash equilibrium* is a strategy profile σ such that for all t and positive probability histories h^t and h_R^t, (1) $\sigma_{A,t}(h^t)$ is a best response of agent t against $\sigma_{R,t}(h^t)$; (2) $\sigma_{R,t}(h_R^t)$ is a best response of the strategic regulator against $\sigma_{A,t}(h^t)$. 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 long-lived strategic player action $\sigma_{R,t}(h^t)$ that depends on the public history. Thereby, we restrict attention to public strategies and public equilibria.

The posterior belief of agent t at the beginning of period t is $\gamma_{t-1}(h^t)$ with $\gamma_0(h^1) = \gamma_0$. When the meaning is clear, we shorten $\gamma_{t-1}(h^t)$ to γ_{t-1} As each agent is short-lived, their decision depends only on the updated reputation of the regulator, the strategic regulator's expected behavior, and the resulting detection probability. If $\gamma_{t-1} \geq \gamma^*$, agent t chooses T and the strategic regulator L delivering each a payoff of zero and sustaining efficiency. If $\gamma_{t-1} < \gamma^*$, then agent t chooses U with some probability only if the strategic regulator is diligent with no more than probability $\frac{\pi^* - \gamma_{t-1}\beta}{(1-\gamma_{t-1})\beta}$ as from the perspective of agent t the probability of detection at t is

$$\pi(\gamma_{t-1}, \sigma_{R,t}(h^t)) \equiv \gamma_{t-1}\, \beta + (1 - \gamma_{t-1})\, \sigma_{R,t}(h^t)\, \beta. \tag{2}$$

Bayesian updating is needed at the end of period t if agent t has chosen U (see Remark 2). If agent t is truthful, the reputation of regulator does not change as there is no new information, i.e., $\gamma_t = \gamma_{t-1}$. Then, given that agent t is untruthful, the reputation after the signal $i_d \in \{0,1\}$ is calculated as follows:

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

3.1 Nash equilibrium payoff of the regulator

Özdoğan (2016) shows that when the agents are short-lived, the reputation helps the patient strategic regulator to achieve the maximum attainable payoff (i.e., the socially efficient level) for any prior belief agents may have about the types of the regulator. Specifically, for any prior belief $\gamma_0 > 0$, the average payoff of the strategic regulator across all Nash equilibria converges to zero (the maximum payoff of the regulator) as δ approaches one.

Let $\hat{V}(\gamma_0, \delta)$ be the minimum payoff for the strategic regulator in some Nash equilibrium.

Theorem 1 *For any prior belief $\gamma_0 > 0$ and auditing quality $\beta \in (0,1)$ satisfying Assumptions 1 and 2,* $\lim_{\delta\to 1} \hat{V}(\gamma_0, \delta) = 0$.

This theorem says that the slightest uncertainty about the type of the regulator guarantees the patient strategic regulator the best outcome in any Nash equilibrium. This outcome provides more utility than his Stackelberg payoff of $-c$.[21] Thus, the reputation is a useful tool in achieving the regulator-preferred and total welfare-maximizing outcome.

The techniques in Ely and Välimäki (2003) are employed in the proof presented in Appendix B.[22] The intuition of the result follows from the fact that the short-lived agents only care about their own payoffs and give myopic best responses to their updated beliefs about the regulator's type. The myopic agent plays truthfully if and only if her belief about the regulator being diligent is above a threshold. In such cases, there is no learning and the regulator attains his maximum payoff as there is no need to engage in costly auditing (Lemma 2). Thus, whenever the public histories lead to a strong belief about the regulator being tough, all the agents are truthful and the regulator is lazy thereafter. At histories where the agent is not yet strongly convinced of the regulator being diligent, she has an incentive to be untruthful.

[21] When the payoff bound established by Gossner (2011) is applied to the current context, we find that the lower bound of the regulator's payoff in any Nash equilibrium approaches the Stackelberg payoff of $-c$ as he gets arbitrarily patient (since the unique 0-entropy confirming best response of the agent to the Stackelberg action D is T with the associated bound being $-c < 0$).

[22] Unlike ours, the long-lived player's concern for differentiating himself from his bad counterpart results in the loss of all surplus in their bad-reputation model.

This leads to an incentive for the strategic regulator to be diligent with some probability to induce a possible increase in the agent's belief. After the event of subsequent detections, the reputation would eventually reach a level above which it persists as all the agents will find it optimal to be truthful thereafter. Hence, the regulator receives a payoff different than his maximum for a finite number of periods and the permanency of the reputation would enable a sufficiently patient strategic regulator to capture all the surplus (i.e., sustain social efficiency) in the long-run.[23]

When $\gamma_0 < \gamma^*$, a typical equilibrium starts with (totally) mixed actions involving agents experimenting with untruthfulness. Some of these will be detected and this will continue until the posterior belief about the regulator being tough gets strong enough. Thereafter, (T, L) will be played in every period and the reputation for being tough will persist.

3.2 Markov equilibrium

Now, we consider a single short-lived agent who is restricted to use history independent Markov strategies. This situation also corresponds to cases in which agent t is one of a continuum of long-lived (population of) agents, coordination/communication among agents (across periods) is not possible, and all agents observe the same public history.[24]

We characterize the stationary Markov equilibria with the reputation of the regulator being the Markov state variable and the players' strategies of being truthful and diligent, $\sigma_A(\gamma)$ and $\sigma_R(\gamma)$, are functions of only the current reputation level (and not the public history).

The strategic regulator maximizes the normalized discounted sum of expected payoffs and the agents give myopic best responses given the reputation level as to optimize (1) as before. We let $V(\gamma)$ denote the expected life-time payoff to a strategic regulator associated with the

[23] Lemma 4 in Appendix B suggests that every Nash equilibrium continuation path starting from history h^t at which $\gamma_{t-1} < \gamma^*$ must include a play of diligence with some probability. Thus, there is a positive probability of detection in every continuation path. Since for the agent to be untruthful, the regulator's diligence must be lower than $\frac{\pi^* - \gamma_{t-1}\beta}{(1-\gamma_{t-1})\beta}$, the smallest posterior belief after a detection would be $\frac{\gamma\beta}{\pi^*} > \gamma$.

[24] There are many cases where coordination among the agents and the regulator for future punishments/rewards is not plausible (e.g., a large number of taxpayers, agents, are audited by the tax authority, the regulator) if agents do not receive the same signal and the individual signals are not publicly observed by the others. In such cases, it is most innocuous to assume that each agent receives privately observed and independently drawn (i.e., idiosyncratic signals) breaking the coordinated behavior among the agents and the regulator to provide intertemporal incentives (punishment/experimentation) in the continuation games. Mailath and Samuelson (2015) points out that one can construct an equilibrium with coordinated punishments if the idiosyncratic signals are public or they all receive the same signal in a stylized version of Mailath and Samuelson (2001). (Mailath and Samuelson (2006, Remark 18.1.3) provides a construction with idiosyncratic signals in the context of Mailath and Samuelson (2001).) However, the idiosyncrasy of signals causes technical complications (see, for instance, Al-Najjar (1995)) and divert attention away from the reputation effects. Thus, to abstract away from these technical complications and to capture the myopic incentives of the agents due to lack of coordination devices in a large population environment, it suffices to consider a continuum of agents who cannot communicate/coordinate among each other but receive the same signal.

strategy $(\sigma_A, \sigma_R) \equiv (\sigma_A(\gamma), \sigma_R(\gamma))$. The value function $V(\gamma)$ satisfies:

$$\begin{aligned} V(\gamma) &= (1-\delta)\left\{\sigma_R[(1-\sigma_A)(\beta d - (1-\beta)f) - c] - (1-\sigma_R)(1-\sigma_A)f\right\} + \delta\sigma_A V(\gamma) \\ &\quad + \delta(1-\sigma_A)\sigma_R\beta V\left(\frac{\gamma\beta}{\pi(\gamma,\sigma_R)}\right) + \delta(1-\sigma_A)(1-\sigma_R\beta)V\left(\frac{\gamma(1-\beta)}{1-\pi(\gamma,\sigma_R)}\right) \end{aligned} \quad (4)$$

In the complete-information case (when $\gamma = 0$), with history-independent strategies, the only equilibrium is the repetition of the stage-game equilibrium. This is because then

$$V = (1-\delta)\left\{\sigma_R[(1-\sigma_A)(\beta d - (1-\beta)f) - c] - (1-\sigma_R)(1-\sigma_A)f\right\} + \delta V$$

and the agent chooses $\sigma_A = \frac{\beta(d+f)-c}{\beta(d+f)} \in (0,1)$ in order to make the regulator indifferent. In turn, for the agent to use this mixed strategy, the regulator's strategy should satisfy $\sigma_R = \frac{g}{\beta(g+l)}$. The corresponding value is $-\frac{cf}{\beta(d+f)} < 0$ due to Assumption 2.

The (stationary) Markov equilibrium is defined as follows.

Definition 1 *A stationary Markov equilibrium consists of the strategy profile* $(\sigma_A^*(\gamma), \sigma_R^*(\gamma))$ *and the corresponding beliefs such that for all* $\gamma \in [0,1]$*:*

1. *Given the expected probability of detection* $\pi(\gamma, \sigma_R^*)$ *induced by* $\sigma_R^*(\gamma)$, $\sigma_A^*(\gamma)$ *maximizes the agent's problem given in (1);*

2. *Given* $\sigma_A^*(\gamma)$, $\sigma_R^*(\gamma)$ *maximizes the associated value function* $V(\gamma)$ *given in (4);*

3. *Posterior beliefs are determined via Bayes' rule whenever possible (i.e., when* $\sigma_A^* < 1$*) according to (3).*

An equilibrium value function $V : [0,1] \to [-f, 0]$ should satisfy the Bellman equation given by (4). So, it should be a fixed point of the operator T which maps any continuation value function W into $\bar{W}$ in the following way:

$$\begin{aligned} \bar{W}(\gamma) &= TW(\gamma) \\ &= (1-\delta)\left\{\sigma_R[(1-\sigma_A)(\beta d - (1-\beta)f) - c] - (1-\sigma_R)(1-\sigma_A)f\right\} + \delta\sigma_A W(\gamma) \\ &\quad + \delta(1-\sigma_A)\sigma_R\beta W\left(\frac{\gamma\beta}{\pi(\gamma,\sigma_R)}\right) + \delta(1-\sigma_A)(1-\sigma_R\beta)W\left(\frac{\gamma(1-\beta)}{1-\pi(\gamma,\sigma_R)}\right) \end{aligned} \quad (5)$$

where $\sigma_R(\gamma)$ maximizes the right-hand side of the equation (5) given $\sigma_A(\gamma)$ while $\sigma_A(\gamma)$ maximizes (1) given the expected probability of detection $\pi(\gamma, \sigma_R)$ implied by $\sigma_R(\gamma)$.

We focus on the equilibrium that is associated with a value function that is continuous and nondecreasing in the reputation, γ. Let C_+ denote the space of such value functions endowed with the sup norm. We show that there is a unique Markov equilibrium with a continuous and nondecreasing value function where players use mixed strategies for low values of the regulator's reputation, whereas the agents are truthful and the regulator is lazy with probability one above a threshold state.

Theorem 2 *Suppose that Assumptions 1 and 2 hold. Then, there is a unique stationary Markov equilibrium* $(\sigma_A^*(\gamma), \sigma_R^*(\gamma))$ *with a continuous and nondecreasing value function* V *such that* $\gamma \leq \gamma^*$ *implies*

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

and $\gamma \geq \gamma^*$ *implies*

$$\sigma_A^*(\gamma) = 1 \text{ and } \sigma_R^*(\gamma) = 0$$

where $\gamma^* = \frac{g}{\beta(g+l)}$ *and* $\pi^* = \frac{g}{g+l}$. *Moreover, the reputation persists whenever it exceeds* γ^* *with the corresponding value function attaining its maximum value, i.e.,* $V(\gamma) = 0$ *for any* $\gamma \geq \gamma^*$.

The proof, relegated to Appendix C, is built upon a couple of lemmas that follow from 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 (5) given $\sigma_A(\gamma; W)$, and $\sigma_A(\gamma; W)$ maximizes (1) 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., $\bar{W} \in C_+$. Lastly, it is shown that $T : C_+ \to C_+$ is a contraction mapping so that there is a unique fixed point of T in C_+, which we call V.

The stationary Markov equilibrium induces a Nash equilibrium. Note that there is a unique concatenation history after all histories ending with the absorbing state and both equilibria specify the same pure action profile of truthful and lazy thereafter. Given $\gamma < \gamma^*$, Markov equilibrium specifies a completely mixed action profile so that every public history is reached with a strictly positive probability. Let any public history that is induced by the stationary

Markov equilibrium $\sigma^*(\gamma)$ be denoted by $h^t_m \equiv h^t(\sigma^*(\gamma))$. Then, we describe the associated Nash equilibrium as $\sigma(h^t_m) = \sigma^*(\gamma_{t-1})$, for any h^t_m starting with the same belief $\gamma_{t-1}(h^t_m)$. As in any Nash equilibrium, the regulator attains the maximum payoff as δ tends to 1 (Theorem 1), the payoff under the stationary Markov equilibrium also approaches the maximum payoff as δ approaches 1. This theorem says that for $\gamma < \gamma^*$, the regulator and each short-lived agent play mixed actions that result in a positive probability of producing consecutive detections.[25]

We identify an upper bound on the number of consecutive detections that can be 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 reputation at that state which would be permanent thereafter. Corollary 1 establishes this upper bound.[26]

Corollary 1 *Suppose that Assumptions 1 and 2 hold and h^t be a positive probability history involving k consecutive detections starting date τ with respect to the unique Markov equilibrium defined in Theorem 2. Then, k can be at most the smallest integer that is greater than k^*_γ where $k^*_\gamma = \frac{\log(\gamma^*)-\log(\gamma)}{\log(\beta)-\log(\pi^*)}$ for $\gamma \equiv \gamma_{\tau-1}(h^t) < \gamma^*$.*

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 the order that they have been observed. Thus, the concatenation of any history that involves at least $k^*_{\gamma_0}$ more detections than no detections continues with a persistent reputation of at least γ^*.[27]

[25] $\{\gamma_t\}_t$ is a martingale. The details are in Section 4 and in the discussion in Appendix D starting on page 40 and in Footnote 41 which can be verified in the Markovian context by Theorem 2 and Lemma 8 (presented in Appendix C). 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 8, we also observe that the belief at the beginning of tomorrow, γ_t, cannot equal the belief at the beginning of today, γ_{t-1} (even though this must hold for the expected values) whenever the agent has chosen U in t. In fact, γ_t must either be $\gamma^+(\gamma_{t-1})$ or $\gamma^-(\gamma_{t-1})$ with some probabilities specified by Theorem 2 and equation (3).

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

[27] 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 Markov equilibrium requires that the regulator chooses to be diligent with a very small probability and the expected hitting (absorption) time becomes unboundedly large as the transition probability puts higher weight on the lower levels of reputation.

3.3 Possible extensions

Given the seminal result of Cripps, Mailath, and Samuelson (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 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 results will change 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. The new regulator is the behavioral type with probability $\hat{\gamma}$. To simplify exposition, we assume that $\gamma_0 = (1-\lambda)\hat{\gamma}$. Then, when the agent chooses U, the posterior belief that the regulator is the Stackelberg type in period t after the observation of the signal $i_d \in \{0, 1\}$ is:

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

while $\gamma_t = \lambda\gamma_{t-1} + (1-\lambda)\hat{\gamma}$ (***) when the agent is truthful. Then, with replacement we get the following result the proof of which parallels with the proof Theorem 2 and hence is omitted due to space considerations:

Proposition 1 *There is a unique stationary Markov equilibrium 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 sufficiently low (i.e., $\lambda < \gamma^* - \gamma_0$), 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 $\gamma \in (0, \gamma^*)$,*

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

In this case, the resulting dynamic programming problem is very similar to the one in

Appendix C.[28] The condition $\lambda < \gamma^* - \gamma_0$ implies that the posterior beliefs cannot exceed the threshold γ^*.[29] Thus, with a highly probable replacement possibility the regulator cannot attain the maximum and socially efficient payoff of zero and hence his payoffs and efficiency are adversely affected. This is because, in this case, agents cannot get strongly convinced that the regulator is the tough type. Indeed, with such a replacement structure, the agents will never be sure of the true type of the regulator.

Ekmekci, Gossner, and Wilson (2012), on the other hand, shows that the replacement of the long-lived player can generate permanent reputation effects if the replacements are arbitrarily infrequent and the long-lived player is arbitrarily patient.[30] They provide lower bounds (albeit might be quite small) on equilibrium payoffs in *every continuation game*, which coincides with those of Fudenberg and Levine (1989), Fudenberg and Levine (1992) and Gossner (2011), without making any specific assumptions on the stage game payoffs and the monitoring structure as the discount rate goes to one at a faster rate than the replacement rate goes to zero.[31] This payoff bound corresponds to the Stackelberg payoff of the regulator that is $-c$ in our setting. Yet, the regulator could do better as λ approaches 1 in our model. The posterior beliefs that are given by (6) and (***) imply that the absorbing states reemerge and the regulator's payoff approaches zero as in no replacement case of Theorem 2. Thus, we can conclude that unobserved and very rare replacements do not affect our results.

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:[32] Suppose that each agent t is born with the same prior belief γ_0 and can observe only the last k entries of the public history. So, agent t's behavior depends on the k-tail of public history at t (all signals in periods $\tau = t-k, \ldots, t-1$). To eliminate the possibility of deriving complicated inferences with bounded memory (see Barlo, Carmona, and Sabourian (2016, Section 5)), we suppose that agent t's behavior does not depend on the calendar time, t.

When k is strictly less than $k^*_{\gamma_0}$, there is no hope of the regulator to strongly convince

[28]The posterior probabilities stated in (6) and (***) should be substituted into the expressions (12) and (13); δ must be replaced by $\lambda\delta$. Hence, the construction in Appendix C can be replicated and the updating takes the following form: $\gamma^+ = \lambda\frac{\gamma\beta}{\pi^*} + (1-\lambda)\hat{\gamma}$ and $\gamma^- = \lambda\frac{\gamma(1-\beta)}{1-\pi^*} + (1-\lambda)\hat{\gamma}$.

[29]Note that if $\lambda = 0$, we get the repetition of the stage game Bayesian Nash equilibrium.

[30]The reason is that the renewed doubt about the informed player's type provides him the chance to rebuild a reputation when it is damaged and this dominates the adverse effect of the loss in the value of the reputation of the long-lived player because of the possibility of replacement when the limits on the replacement rate and discount rate are taken appropriately.

[31]More precisely, as established by Ekmekci, Gossner, and Wilson (2012, pp.164), "if the discount rate goes to one at a faster rate than the rate at which the logarithm of the replacement rate goes to infinity."

[32]It maybe 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.

the agents that he is tough (i.e., inducing a posterior belief strictly exceeding γ^*). Thus, the persistent reputation effect, desired by the regulator, disappears and he would want to announce the relevant part of the public history to restore the reputation effect.[33]

With bounded but long memory, the analysis becomes more complicated. Liu and Skrzypacz (2014) analyzes a variation of perfect-monitoring product-choice game with limited but long records of the history. They establish that if the record length exceeds a finite lower bound, then an arbitrarily patient long-lived player can attain the Stackelberg payoff as the limiting payoff in any stationary perfect Bayesian equilibria after every history, which gives rise to permanent reputation effects. The equilibria feature recurrent reputation bubbles. Even though the long-lived player's true type is perfectly observed, the informed short-lived players ride up the reputation up to a level after which it is exploited by the long-lived player (hence, revealing his type). 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 mediator (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 payoff after every history implying permanent reputation effects.

We conjecture that with a bounded but long memory, a detailed analysis (which is outside the scope of the current paper) would establish that reputation effects would prevail.

4 Regulator faces a long-lived agent

Now, we assume that the agent is long-lived with the common 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, we have $h_R^t = ((r_0, i_{d0}), (r_1, i_{d1}), ..., (r_{t-1}, i_{dt-1})) \in H_R^t \equiv (R \times I_d)^t$ for the regulator and $h_A^t = ((a_0, i_{d0}), (a_1, i_{d1}), ..., (a_{t-1}, i_{dt-1})) \in H_A^t \equiv (A \times I_d)^t$ for the agent to show the private histories up to period t. As before, $h^t = (i_{d0}, i_{d1}, ..., i_{dt-1}) \in I_d^t$ denotes the public history of signals observed by both players. The set of full histories is then shown by $h_f^t = ((a_0, r_0, i_{d0}), (a_1, r_1, i_{d1}), ..., (a_{t-1}, r_{t-1}, i_{dt-1})) \in H_f^t \equiv (A \times R \times I_d)^t$. 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. We let $K = \{tough, strategic\}$ be the type space for the regulator. 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

[33]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.

type. Agent's strategy, σ_A, is a sequence of maps $\sigma_{At} : H_A^t \to \Delta(A)$. The prior belief γ_0 with strategies $\sigma_R \equiv (\hat{\sigma}_R, \tilde{\sigma}_R)$ and σ_A induce a probability measure Q on the set of states $\Omega \equiv K \times (R \times A \times I_d)^\infty$, which illustrates how the game evolves for an uninformed player.[34] The strategy 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 tough type and strategic type, respectively. The expectation taken with respect to Q is $E \equiv E_{(\sigma_A, \hat{\sigma}_R, \tilde{\sigma}_R)}$ and the expectations associated with $\hat{Q}$ and $\tilde{Q}$ are $\hat{E} \equiv E_{(\sigma_A, \hat{\sigma}_R)}$ and $\tilde{E} \equiv E_{(\sigma_A, \tilde{\sigma}_R)}$, respectively.

A Nash equilibrium is a strategy profile $(\sigma_A, \tilde{\sigma}_R)$ that satisfies

(*i*) $E_{(\sigma_A, \hat{\sigma}_R, \tilde{\sigma}_R)}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_A(a, r)] \geq E_{(\bar{\sigma}_A, \hat{\sigma}_R, \tilde{\sigma}_R)}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_A(a, r)]$, for all $\bar{\sigma}_A$, and

(*ii*) $E_{(\sigma_A, \tilde{\sigma}_R)}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_R(a, r)] \geq E_{(\sigma_A, \bar{\sigma}_R)}[(1-\delta)\sum_{t=0}^{\infty} \delta^t u_R(a, r)]$, for all $\bar{\sigma}_R$.

In the following, as discussed in the introduction, we concentrate on Nash equilibria in which there is costly learning on the agent's behalf: She is bound to experiment with the bad behavior every once in a while. This, in turn, delivers the notion that we call an α-Nash equilibrium with $\alpha > 0$ and arbitrarily small: An *α-Nash equilibrium* is a Nash equilibrium in which the agent is restricted to choose any one of her actions with at least α probability.

Ours is a direct approach. Instead, we could adopt the following formulation involving *one-period amnesia*:[35] 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-1}) 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-1}. 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, Carmona, and Sabourian (2016)), it is natural to 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 unique Markov equilibrium of Theorem 2 hanging on to her initial

[34] The discussion starting on page 40 and in Footnote 40 establishes that $\{\gamma_t\}_t$ is a Martingale.

[35] 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 and observe her private history and chooses 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.

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 $\gamma_0 < \gamma^*$. At the end of period t, she recovers from the amnesia, observes h_A^t along with her period t choice a_t, whether or not there has been a detection in period t (i.e., i_{dt}), performs the Bayesian updating if $i_{dt} = 1$ (i.e., there was a detection in period t), records these as h_A^{t+1} (hence, γ_t), and gets ready for tomorrow. When $\vartheta > 0$ is set to be arbitrarily small, the incentives of the strategic regulator and the agent do not get affected. Thus, a Nash equilibrium of this formulation with $\vartheta > 0$ is an α-Nash equilibrium with $\alpha = \vartheta(1 - \sigma_A^*(\gamma_0)) > 0$.

Regardless of the formulation employed, this notion of robustness does not imply qualitative changes to our results with myopic agents. If the short-lived players were to be restricted to choose any one of their actions with a probability $\alpha > 0$ but sufficiently small, the regulator's best response does not change and calls for the diligent action for histories with $\gamma_{t-1} \leq \gamma^*$ and being lazy otherwise. Hence, the findings presented in Section 3.1 continue to hold with some small modifications to their statements and proofs. Moreover, the same holds for our results concerning Markov equilibrium: 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 $\sigma_R^*(\gamma)$ do not change and $V(\gamma)$ needs a slight alteration.

The following result, Theorem 3, states that in every α-Nash equilibrium with $\alpha > 0$ and arbitrarily small, the sufficiently patient agent plays the regulator-preferred action T with a probability bounded away from $1 - \alpha$ on average indefinitely on a strictly positive probability set of events. That is, the regulator cannot build the absorbing reputation that induces the patient agent to be truthful indefinitely in the long-run in equilibrium whenever learning/experimenting is enforced even with an arbitrarily small probability. In this sense, the permanency of reputation is not robust.

Theorem 3 *Suppose that Assumptions 1 and 2 hold and let $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ be any α-Nash equilibrium with $\alpha > 0$ and arbitrarily small. Then, there is $\delta_\alpha \in (0,1)$ such that there exists no $\mathcal{A} \subset \Omega$ with $\tilde{Q}(\mathcal{A}) > 0$ that is induced by this α-Nash equilibrium 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 all $\delta > \delta_\alpha$.

Appendix D is devoted to the proof of Theorem 3. The technique that employs the identification conditions is borrowed from Cripps, Mailath, and Samuelson (2007) and adapted 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. This is stated formally in Lemma 12 in Appendix D. 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-Nash equilibrium. This is presented formally in Lemma 11 in Appendix D.

The intuition behind the result 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 12 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.

5 Concluding Remarks

This paper analyzes the long-run equilibrium behavior of uninformed players in a repeated game with incomplete information and imperfect public monitoring that aims to model the strategic interaction in a regulatory environment. We ask whether the uninformed players (agents) can be induced to good behavior permanently through the reputation concerns of the informed player (regulator). We provide a positive answer to this question when a patient long-lived regulator faces a sequence of short-lived agents: using his reputation, the regulator prevents agents' bad behavior by inducing them to behave in the long-run with no cost in *any* Nash equilibrium. As a result, reputation secures the perpetual social efficiency. On the other hand, we show that permanent good behavior cannot be an equilibrium outcome when the long-lived agent is sufficiently patient and occasionally experiments with bad behavior. This stark contrast demonstrates the significance of the longevity of the strategic interaction and provides a novel insight into the importance of learning and experimentation in repeated games: While the sufficiently patient strategic regulator may induce permanent social efficiency via sustaining a high reputation with myopic agents, there is no interior reputation level behind which the strategic regulator can hide and induce social efficiency whenever the long-lived sufficiently patient agent tests the regulator's reputation every once in a while.

In repeated games with incomplete information and imperfect public monitoring, identifiability is a critical assumption used in the literature. In our paper, these are not satisfied. The identifiability assumptions we need are the conditional identification properties formalized in Remarks 1 and 2. They suffice and are the key deriving force that results in different long-run equilibrium behavior for the uninformed players. The conditional identification of the informed player results in a sharp and long-lasting equilibrium selection effect (in terms of payoffs and behavior) when the informed player faces a sequence of short-run players.[36] On the other hand, when the uninformed player is also long-lived, the conditional identification of the uninformed player enables to specify whether a particular behavior of the uninformed player is attainable in the long-run by easing the complications due to private beliefs (when the action of the long-lived informed player that makes the agent's behavior is identifiable in Remark 1 coincides with that of the behavioral type). Insisting on an experimentation/learning structure on behalf of the uninformed player by making her go for bad behavior now and then ensures that the conditional identifiability requirements bite. That is, entangling conditional identifiability and this experimentation structure delivers the stark contrast between the case with short-lived agents and the case with a long-lived agent.

[36]We note that Fudenberg, Levine, and Maskin (1994)'s sufficiency conditions on the distribution of public signals that guarantee the Folk theorem result for repeated games with imperfect public monitoring in a complete-information setting are not satisfied in our paper. The set of all equilibrium payoffs of games satisfying some version of conditional identification property is an open question to the best of our knowledge. In our setting, an equilibrium payoff attained in the complete-information game by repeating the stage game Nash equilibrium is eliminated in the incomplete information setting with short-lived uninformed players.

Appendix

A The proof of Lemma 1

The utility of the agent being truthful is $u_A(T, \sigma_R) = 0$; her expected utility when she is untruthful is $u_A(U, \sigma_R) = \gamma[(1-\beta)g - \beta l] + (1-\gamma)\sigma_R[(1-\beta)g - \beta l] + (1-\gamma)(1-\sigma_R)g$. The agent's best response correspondence against the strategic regulator's strategy is:

$$\sigma_A \equiv BR_A(\sigma_R) = \begin{cases} 1 & \text{if} \quad \sigma_R > \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)} \\ [0,1] & \text{if} \quad \sigma_R = \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)} \\ 0 & \text{if} \quad \sigma_R < \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)}. \end{cases} \tag{7}$$

From this, we can deduce the cutoff prior beliefs. The strategy of the regulator that makes the agent indifferent between being truthful and untruthful, $\sigma_R = \frac{g-\gamma\beta(g+l)}{(1-\gamma)\beta(g+l)}$, is greater than 0 if $\gamma < \gamma^* = \frac{g}{\beta(g+l)}$ and equals to 0 if $\gamma = \gamma^*$. If $\gamma > \gamma^*$, then $BR_A(\sigma_R) = 1$ for any value of σ_R, i.e., even if the strategic regulator is lazy for sure.

The expected utility of the regulator choosing to be diligent is $u_R(\sigma_A, D) = (1-\sigma_A)[\beta d - (1-\beta)f] - c$, whereas his expected utility, when he is lazy, is $u_R(\sigma_A, L) = -(1-\sigma_A)f$. Thus, the regulator's best response is given by:

$$\sigma_R \equiv BR_R(\sigma_A) = \begin{cases} 1 & \text{if} \quad \sigma_A < 1 - \frac{c}{\beta(d+f)} \\ [0,1] & \text{if} \quad \sigma_A = 1 - \frac{c}{\beta(d+f)} \\ 0 & \text{if} \quad \sigma_A > 1 - \frac{c}{\beta(d+f)}. \end{cases} \tag{8}$$

The strategy of the agent that makes the regulator indifferent, $\sigma_A = 1 - \frac{c}{\beta(d+f)}$, is greater than 0 if $\beta > \frac{c}{f+d}$.

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 strategy 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 equilibrium strategies 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

Fix an arbitrary Nash equilibrium in public strategies; each positive probability public history and posterior belief that are considered are going to be with respect to this Nash equilibrium. Given an arbitrary Nash equilibrium, for each positive probability public history 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 $\sigma_{A,t}(h^t)$ and D has a positive probability under $\sigma_{R,t}(h^t)$, 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 for $v(h_t; \sigma_{A,t}(h^t), \sigma_{R,t}(h^t))$ is done in the natural way.

The following lemma tells that if there is a positive probability history on which the agent is truthful with probability one at some date, then the regulator must be lazy with probability one on that history and date, too.

Lemma 2 *Suppose that h^t is a positive probability history where $\sigma_{A,t}(h^t) = 1$ (truthful with probability one). Then, $\sigma_{R,t}(h^t) = 0$ (lazy with probability one).*

Proof of Lemma 2. Given that the agent chooses $\sigma_{A,t}(h^t) = 1$ at h^t, choosing diligent or lazy generates the same distribution of public signals so that the continuation payoff $v(h^t, i_d)$ will be the same. Since the one-period utility $u_R(T, L) = 0 > u_R(T, L) = -c$, we conclude that $\sigma_{R,t}(h^t) = 0$ complying with the one-shot deviation principle. ■

This is the consequence of the fact that when the agent is truthful, being lazy and diligent generates the same distribution of public signals and hence the same continuation payoffs $v(h^t, i_d = 0)$. The next lemma suggests that there is no date on a positive probability history on which the regulator is diligent with probability one with respect to a Nash equilibrium.

Lemma 3 *There is no date t and positive probability history h^t at which $\sigma_{R,t}(h^t) = 1$ (diligent with probability one). In other words, $\sigma_{R,t}(h^t) < 1$ for any positive probability h^t.*

This is because as Lemma 2 says that when the agent is truthful, the regulator chooses to be lazy. And, the agent chooses to be untruthful with some probability only if she expects the regulator to be lazy with some probability on a Nash equilibrium.

Proof of Lemma 3. Suppose that there exist a Nash equilibrium with a positive probability history h^t at which $\sigma_{R,t}(h^t) = 1$. If the agent is truthful with probability one at h^t, then the regulator would have a one-shot deviation gain by switching to lazy since the distribution of public signals, the posterior belief, and the continuation payoff would not change regardless of regulator's action when the agent is truthful. The agent chooses to be untruthful with some probability on a Nash equilibrium only when the belief $\gamma_{t-1}(h^t)$ at $t-1$ is less than γ^* and the regulator chooses to be diligent by less than $\bar{\sigma}_R = \frac{\pi^* - \beta\gamma}{\beta(1-\gamma)} < 1$. ■

Lemma 4 *Let h^t be any positive probability history where $\gamma_{t-1}(h^t) < \gamma^*$. Then, there exists some $\tau \geq t$ and h^τ that appends after h^t for which $\sigma_{R,\tau}(h^\tau) > 0$.*

Proof of Lemma 4. Consider an arbitrary public history h^t that is reached with positive probability with respect to some Nash equilibrium. Suppose for a contradiction that for every $\tau \geq t$ and every h^τ that comes after h^t, $\sigma_{R,\tau}(h^\tau) = 0$ (lazy with probability one at every date starting from h^t). Given $\gamma_{t-1}(h^t) < \gamma^*$ and there will not be any detections after h^t for every $\tau \geq t$ and history since $\sigma_{R,\tau}(h^\tau) = 0$ by hypothesis, and thus the expected probability of detection is going to be less than π^* for every $\tau \geq t$ and h^τ. Then, all the myopic agents are untruthful at every date and history following h^t. Thus, the regulator's expected continuation payoff becomes $v(h^\tau) = -f$ from then on which is less than the minmax payoff $-e$, providing us the desired contradiction. ■

This lemma suggests that every Nash equilibrium continuation path starting from history h^t must include the play of diligence with some probability given that $\gamma_{t-1}(h^t) < \gamma^*$.

Lemma 5 *Suppose that h^t is a positive probability history on which detection occurs ($i_d = 1$) at time t given that $\gamma \equiv \gamma_{t-1}(h^t) < \gamma^*$. Then, $0 < \sigma_{R,t}(h^t) \leq \frac{\pi^* - \beta\gamma}{\beta(1-\gamma)}$ and the smallest posterior belief after the detection has been observed denoted by $\Gamma(\gamma) \equiv \gamma_t(h^t, i_d = 1)$ becomes*

$$\Gamma(\gamma) = \frac{\gamma\beta}{\gamma\beta + (1-\gamma)\beta\bar{\sigma}_R} = \frac{\gamma\beta}{\pi^*}$$

where $\bar{\sigma}_R \equiv \bar{\sigma}_R(\gamma) = \frac{\pi^* - \beta\gamma}{\beta(1-\gamma)}$.

Proof of Lemma 5. First, note that in order to observe a detection at time t, the agent must have been untruthful. For the agent to have chosen untruthfulness with some positive probability at h^t, the expected probability of detection must be lower than π^*. Given that the belief at the beginning of time t at h^t is $\gamma \equiv \gamma_{t-1}(h^t) < \gamma^*$, this requires that $\sigma_{R,t}(h^t) \leq \frac{\pi^* - \beta\gamma}{\beta(1-\gamma)}$, which is derived from (2). And, $\sigma_{R,t}(h^t)$ must be greater than zero because otherwise there would not be any detections. Lastly, it is easy to see from expression (3) that the smallest posterior is obtained when $\sigma_{R,t}(h^t) = \bar{\sigma}_R$ and equals to $\Gamma(\gamma) = \frac{\gamma\beta}{\pi^*}$. ■

This result is the consequence of the fact that detection is possible only when the agent is untruthful and the regulator is diligent (with some probability). But, for the agent to be untruthful, the strategic regulator should not be diligent more than some specified probability. Thus, there is an upper bound on the diligence of the regulator for the agent to be untruthful and the regulator must be using a mixed action at that history. The smallest possible posterior after detection occurs is calculated with respect to this upper bound.

Proof of Theorem 1. First, note that if the prior belief satisfies $\gamma_0 \geq \gamma^*$, then there is a unique Nash equilibrium on which the agents are always truthful and the regulator is always lazy. As all the short-lived agents are truthful with probability one, the reputation for being tough persists and the equilibrium payoffs are zero both for the strategic regulator and the short-lived agents.

Now, suppose that $\gamma_0 < \gamma^*$. Again, note that at any positive probability history h^t where $\gamma_{t-1}(h^t) \geq \gamma^*$, the agent is truthful and the regulator is lazy from then on. The reputation of the regulator does not change when the agent is truthful and again it is not optimal for the strategic regulator to be diligent in this situation. Then, the continuation payoff $v(h^t)$ of the regulator is zero on those histories.

Then, we consider a positive probability history h^t for which $\gamma \equiv \gamma_{t-1}(h^t) < \gamma^*$ and the agent is untruthful with some probability (since otherwise, by Lemma 2, the strategic regulator is lazy and the stage game payoffs are all zero in those periods). This time, the reputation (posterior belief about the regulator's type) is updated according to expression (3) as allowed by Remark 2. Recall that the agent is untruthful only if the expected probability of detection is $\pi(\gamma, \sigma_{R,t}(h^t)) \leq \pi^*$, which requires $\sigma_{R,t}(h^t)$ to be less than or equal to $\bar{\sigma}_R(\gamma) \equiv \frac{\pi^* - \gamma\beta}{(1-\gamma)\beta}$. Then, by Lemma 5, we define the smallest posterior probability of a tough regulator after a detection has been observed as: $\Gamma(\gamma) = \frac{\gamma\beta}{\pi^*}$ (where $\pi^* = \frac{g}{g+l}$). By Assumption 1, $\Gamma(\gamma) > \gamma$ for every $\gamma \in (0, \gamma^*)$, i.e., Γ is strictly increasing and it is continuous.

Following the footsteps of Ely and Välimäki (2003), we construct a decreasing sequence of cutoff beliefs p_n such that $p_1 = \gamma^*$ and $p_n = \Gamma^{-1}(p_{n-1})$ for $n > 1$ and use an induction on n to bound the payoffs attained in any Nash equilibria when γ exceeds p_n. For the induction hypothesis, suppose that there exists a lower bound $\hat{V}_n(\delta)$ with $\lim_{\delta\to 1} \hat{V}_n(\delta) = 0$ and $\hat{V}(\gamma, \delta) \geq \hat{V}_n(\delta)$ for all $\gamma > p_n$. Note that this holds for $n = 1$, i.e., for $\gamma > p_1 = \gamma^*$. We assume this holds for n and want to show that holds for $n + 1$. Now, fix $\gamma > p_{n+1}$. We need to prove that $\lim_{\delta\to 1} \hat{V}_{n+1}(\delta) = 0$. As the maximum attainable payoff is zero, this will imply that $\lim_{\delta\to 1} \hat{V}(\gamma, \delta) = 0$.

It suffices to consider a Nash equilibrium in which the agent is untruthful in the first period with some probability, e.g., $\sigma_{A,t}(h^t) < 1$.[37] Since the agent has been untruthful with some

[37] Take any Nash equilibrium in which the agents have been truthful with probability one until date $s > t$. In these periods, the reputation of the regulator will stay the same regardless of her behavior. Thus, it is a

probability, this implies that the strategic regulator is diligent with no more than $\bar{\sigma}_R(\gamma)$. Then this strategy must be a best-reply for the strategic regulator, i.e., the expected long-run payoff from being lazy should be no less than that of being diligent. Meanwhile, by Lemma 4, for any positive probability history where $\gamma < \gamma^*$, there exists a continuation history at which the regulator is diligent with some probability (but not with probability 1 by Lemma 3). Take h^t to be that history at which $\sigma_{R,t}(h^t) > 0$ (note that this is consistent with $\sigma_A(h^t) < 1$, if the agent were to be truthful with probability one, by Lemma 2, the regulator would be lazy with probability 1). Thus, we obtain the following constraint on continuation values at h^t:

$$(1-\delta)(-e) + \delta Z_D(\gamma) \leq (1-\delta)(-f) + \delta Z_L(\gamma) \equiv \hat{V}_{n+1}(\delta) \tag{9}$$

where $Z_D(\gamma)$ and $Z_L(\gamma)$ are the lower bounds on the expected continuation payoffs from choosing action D and L, respectively.[38] And, we can define $\hat{V}_{n+1}(\delta) := (1-\delta)(-f) + \delta Z_L(\gamma)$.

When there is no detection, in the worst case senario, the posterior probability drops to $\gamma^- = \frac{\gamma(1-\beta)}{\gamma(1-\beta)+(1-\gamma)}$ at h^{t+1} and this happens with probability one when the regulator chooses to be lazy. Let the minimum continuation payoff for γ^- be $\hat{V}(\gamma^-, \delta)$. Then, $Z_L(\gamma) \equiv \hat{V}(\gamma^-, \delta)$. On the other hand, when the strategic regulator chooses to be diligent, there will be detection with probability β and the posterior probability after a detection occurs is at least $\Gamma(\gamma)$, which is at least p_n given $\gamma > p_{n+1}$. Hence, we derive the following lower bound on $Z_D(\gamma)$:

$$\beta\hat{V}_n(\delta) + (1-\beta)\hat{V}(\gamma^-, \delta) \leq Z_D(\gamma). \tag{10}$$

Then, combining (9) and (10) allows us to get:

$$\begin{aligned}
(1-\delta)(-e) + \delta\beta\hat{V}_n(\delta) + \delta(1-\beta)\hat{V}(\gamma^-, \delta) &\leq (1-\delta)(-f) + \delta\hat{V}(\gamma^-, \delta) \\
(1-\delta)(f-e) + \delta\beta\hat{V}_n(\delta) &\leq \delta\beta\hat{V}(\gamma^-, \delta)
\end{aligned}$$

which implies that $\hat{V}_n(\delta) \leq \hat{V}(\gamma^-, \delta)$ since $(f-e) > 0$.

As $\hat{V}_{n+1}(\delta) := (1-\delta)(-f) + \delta Z_L(\gamma) = (1-\delta)(-f) + \delta\hat{V}(\gamma^-, \delta)$ and $\hat{V}_n(\delta) \leq \hat{V}(\gamma^-, \delta)$, we conclude that

$$(1-\delta)(-f) + \delta\hat{V}_n(\delta) \leq \hat{V}_{n+1}(\delta). \tag{11}$$

best response for the regulator to be lazy during these periods. So, her payoff is zero. Then the continuation play starting from date s is a Nash equilibrium with the same prior γ whose payoff can be no more than the original game.

[38]If the agent were to be truthful with probability $\sigma_{A,t}(h^t) = 1$, choosing lazy is superior to diligent. Thus, the constraint involves inequality instead of equality.

By the induction hypothesis, the limit of the left-hand side of (11) is zero as δ approaches to one. Thus, $\lim_{\delta\to 1}\hat{V}_{n+1}(\delta)=0$, which implies $\lim_{\delta\to 1}\hat{V}(\gamma,\delta)=0$ as desired for $\gamma\equiv\gamma_{t-1}(h^t)<\gamma^*$. Then, following h^t, $\lim_{\delta\to 1}\hat{V}(\gamma,\delta)=0$ for any $\gamma\equiv\gamma_{t-1}(h^t)>0$ since the choice of p_n and $\gamma>p_n$ is arbitrary and this holds for every $\gamma\geq\inf_n p_n$ where $(p_n)_{n\in IN}$ is a decreasing sequence that converges to zero. The regulator gets a payoff possible different than zero only for finite number of periods up to h^t. Thus, we can conclude that $\lim_{\delta\to 1}\hat{V}(\gamma_0,\delta)=0$ for any prior belief $\gamma_0>0$. This completes the proof. ■

C The proof of Theorem 2

First, we deal with the short-term game and the temporary equilibrium by taking $W\in C_+$ as given. Then, the value function in the short-term game when the regulator is diligent and lazy, respectively, can be given as:

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

$$W_L(\gamma) = -(1-\delta)(1-\sigma_A)f+\delta\sigma_A W(\gamma)+\delta(1-\sigma_A)W\left(\gamma^-\right). \tag{13}$$

We would like to point out that the regulator is indifferent when $W_D(\gamma)-W_L(\gamma)=0$, which implies that $(1-\sigma_A)\beta\{(1-\delta)(f+d)+\delta[W(\gamma^+)-W(\gamma^-)]\}=(1-\delta)c$. If $W_D(\gamma)>W_L(\gamma)$, this means that the regulator will choose to be diligent, so the expected probability of detection would be $\pi(\gamma)=\beta$ for any γ. If, on the other hand, $W_D(\gamma)<W_L(\gamma)$, the regulator will be lazy for sure, thus the expected probability of detection will be $\pi(\gamma)=\gamma\beta$. Therefore,

Lemma 6 *Given $W\in C_+$, an equilibrium of the short-term game corresponds to 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 (2)), and an implied strategy for the agent $\sigma_A(\pi)$ that maximizes the agent's problem (1) at π, with an associated value function $\bar{W}:[0,1]\to[-f,0]$ such that for any $\gamma\in[0,1]$,*

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

and $\bar{W}(\gamma)=\max\{W_D(\gamma),W_L(\gamma)\}$ where $W_D(\gamma),W_L(\gamma)$ are as in (12) and (13).

Let us now define, for any given γ and W,

$$W_D(\sigma_A(\pi);\gamma,W) = (1-\delta)\{(1-\sigma_A)[\beta d-(1-\beta)f]-c\}+\delta\sigma_A W(\gamma) + \delta(1-\sigma_A)\beta W\left(\frac{\gamma\beta}{\pi}\right)+\delta(1-\sigma_A)(1-\beta)W\left(\frac{\gamma(1-\beta)}{1-\pi}\right) \quad (14)$$

$$W_L(\sigma_A(\pi);\gamma,W) = -(1-\delta)(1-\sigma_A)f+\delta\sigma_A W(\gamma)+\delta(1-\sigma_A)W\left(\frac{\gamma(1-\beta)}{1-\pi}\right) \quad (15)$$

where $\sigma_A(\pi)$ solves the agent's problem defined in (1) at π and let

$$F(\sigma_A(\pi);\gamma,W) \equiv W_D(\sigma_A(\pi);\gamma,W)-W_L(\sigma_A(\pi);\gamma,W) \quad (16)$$
$$= (1-\sigma_A(\pi))\beta\left\{(1-\delta)(f+d)+\delta\left[W\left(\frac{\gamma\beta}{\pi}\right)-W\left(\frac{\gamma(1-\beta)}{1-\pi}\right)\right]\right\}-(1-\delta)c.$$

For the subsequent arguments, we would like to remind that $\gamma^*=\frac{g}{\beta(g+l)}$ and $\pi^*=\frac{g}{g+l}$.

Lemma 7 *For any $\gamma\in[\gamma^*,1]$ and $W\in C_+$, $\pi(\sigma_R,\gamma)\geq\pi^*$ for any value of σ_R. Hence, at the unique temporary equilibrium of the short-term game, $\bar{\sigma}_A(\pi)=1$ solves the agent's problem in (1), $F(\bar{\sigma}_A(\pi);\gamma,W)=-(1-\delta)c<0$ and $\bar{W}(\gamma)=\max\{W_D(\gamma),W_L(\gamma)\}=W_L(\gamma)$ with $\bar{\sigma}_R(\gamma)=0$. Moreover, $\gamma^+(\gamma)=\gamma^-(\gamma)=\gamma$ and $\bar{W}(\gamma)=0$ for any $\gamma\in[\gamma^*,1]$.*

The proof can be easily shown by employing the definitions, γ^*,π^*, and using Lemma 6.

Lemma 8 *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, 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 the agent's problem in (1). 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 (1) 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 F is continuous in (γ,W) and $\bar{\sigma}_A(\pi)$ is unique for a given (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 (1). 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 completely 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. ■

From hereafter, we refer to the temporary equilibrium strategy of the agent for a given W as $\sigma_A^*(\gamma, W)$ and that of the regulator as $\sigma_R^*(\gamma)$ and define $\bar{W}(\gamma, W) := W_D(\sigma_A^*(\gamma, W); \gamma, W) = W_L(\sigma_A^*(\gamma, W); \gamma, W)$ 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.[39] As the value of W changes (in a nondecreasing way) with a change in γ, we should consider both of the arguments of $\bar{W}$ to investigate whether it is nondecreasing in γ or not.

Lemma 9 *$\bar{W}(\gamma, W)$ 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) $\bar{W}(\gamma, W) = W_L(\sigma_A^*(\gamma, W); \gamma, W)$ are true. Note that (ii) implies,

$$\bar{W}(\gamma, W) = -(1-\delta)(1-\sigma_A^*(\gamma, W))f + \delta W(\gamma^-(\gamma)) + \delta\sigma_A^*(\gamma, W)\left[W(\gamma) - W(\gamma^-(\gamma))\right]. \quad (17)$$

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

$$\bar{W}(\gamma, W) = (1-\delta)\big(1-\sigma_A^*(\gamma, W)\big)d - \frac{(1-\delta)c}{\beta} + \delta W(\gamma^+(\gamma)) - \delta\sigma_A^*(\gamma, W)\left[W(\gamma^+(\gamma)) - W(\gamma)\right]. \quad (18)$$

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 (17) 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 (17) 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 (18) will imply that $\bar{W}$

[39] We would like to remind 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.

is nondecreasing. To see that, as this time, $\sigma_A^*(\gamma, W)$ is assumed to be 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 (18) is nondecreasing which implies $\bar{W}(\gamma_2, W_2) \geq \bar{W}(\gamma_1, W_1)$ when $(\gamma_2, W_2) \geq (\gamma_1, W_1)$. Consequently, $\bar{W}$ is nondecreasing in (γ, W) (whether $\sigma_A^*(\gamma_2, W_2) \geq \sigma_A^*(\gamma_1, W_1)$ or $\sigma_A^*(\gamma_2, W_2) < \sigma_A^*(\gamma_1, W_1)$).

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 8, 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 $T(W) : \gamma \in [0,1] \rightarrow \bar{W}(\gamma, W)$. The last step is to argue that T is a contraction that maps continuous and nondecreasing functions on $[0,1]$ into itself. Lemma 9 shows that the operator T defined above maps C_+ into C_+. 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 V. Finally, the equilibrium strategies and the value function V of Theorem 2 follow from Lemmas 7 and 8.

C.1 The proof of Corollary 1

Let the belief at the beginning of τ be $\gamma \equiv \gamma_{\tau-1}$, the posterior probability when the regulator chooses $\sigma_R^*(\gamma) = \frac{\pi^* - \gamma\beta}{(1-\gamma)\beta}$ is derived from (3) and equals to $\frac{\gamma\beta}{\pi^*} > \gamma$ upon observing a detection. After k consecutive detections starting at τ, we obtain $\gamma_{\tau-1+k} = \gamma_{\tau-1}(\frac{\beta}{\pi^*})^k$. Since detection is possible only when agent chooses U, which needs the posterior beliefs to be less than γ^*:

$$\gamma_{\tau-1+k} \leq \gamma^* \Rightarrow \gamma_{\tau-1}(\frac{\beta}{\pi^*})^k \leq \gamma^* \text{ which implies } k \leq \frac{\log(\gamma^*) - \log(\gamma)}{\log(\beta) - \log(\pi^*)}.$$

D The proof of Theorem 3

The commitment strategy of the tough regulator, $\hat{\sigma}_R$, is simple (it is a constant function) and thus it is public.[40] Also, note that the agent's best response to $\hat{\sigma}_R$, denoted by $\hat{\sigma}_A$, is the regulator-preferred strategy as it involves a play of truthful action T with probability one in every period after every history (hence, it is also public). On the other hand, for the strategic regulator, $\hat{\sigma}_R$ is not a best response to the agent's best response to the commitment strategy $\hat{\sigma}_A$; thus, it is not credible. The strategies $\hat{\sigma}_R$ and $\hat{\sigma}_A$ satisfy the conditions in Definition 5.1 and 5.2 of Cripps, Mailath, and Samuelson (2007), which state that the commitment strategy of the regulator has no long-run credibility and long-lived agent's best response to the commitment strategy is unique on the equilibrium path.

[40] A behavior strategy is public if it is measurable with respect to the filtration induced by the public signals, $\{\mathcal{H}_t\}_{t=0}^{\infty}$.

Theorem 3 involves Nash equilibria constrained as follows: a given an arbitrarily small $\alpha > 0$, an α-Nash equilibrium is a Nash equilibrium with $\min\{1 - \tilde{\sigma}_{At}, \tilde{\sigma}_{At}\} \geq \alpha$ for all $t \in \mathbb{N}$.

The agent's beliefs about her own future behavior

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

Lemma 10 *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-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$, 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*

$$\tilde{E}\big[\sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\| \mid \mathcal{H}_{At}\big] < \xi, \quad \forall t \geq \bar{t}_\alpha \tag{19}$$

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

$$\tilde{Q}\big(\alpha \leq \sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\| < \psi \mid \mathcal{H}_{At}\big) \to 1, \quad \forall t \geq \bar{t}_\alpha \tag{20}$$

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 $Z_t := \sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\|$ 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 $\mathcal{F}$,

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

Finally, the last expression in Lemma 10 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 (20). ■

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 truthfully with very high probability on average in every continuation game after some period (which is the unique best reply to the commitment strategy), then she must be convinced to see the commitment action D in every continuation game from then on with a high probability. The intuition for this relies on the uniqueness of the best response of the agent against the regulator's repeated strategy of playing D (i.e., the commitment strategy).

Lemma 11 *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 (19) and (20) stated in Lemma 10 hold for all $\omega \in \mathcal{F}$ for a given α-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ arbitrarily small. Then, for some $\zeta_\alpha > 0$,*

$$\tilde{Q}\big(\sup_{t'>t} \|\hat{\sigma}_R - E[\tilde{\sigma}_{Rt'} \mid \mathcal{H}_{At'}]\| < \zeta_\alpha \mid \mathcal{H}_{At}\big) > 1 - \zeta_\alpha, \quad \forall t \geq \bar{t}_\alpha \tag{21}$$

for all $\omega \in \mathcal{F}$.

Proof. Let $Z_t := \sup_{t'>t} \|\hat{\sigma}_A - \tilde{E}[\tilde{\sigma}_{At'} \mid \mathcal{H}_{At'}]\|$ and (19) and (20) hold by the hypothesis. For any $\epsilon > 0$ (so that $\xi = \epsilon\psi$ as given in Lemma 10), $\tilde{Q}(Z_t < \psi \mid \mathcal{H}_{At}) > 1 - \epsilon$ 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-Nash equilibrium.

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, one deduces

$$\tilde{Q}\big(\sup_{t'>t} \|\hat{\sigma}_R - E[\tilde{\sigma}_{Rt'} \mid \mathcal{H}_{At'}]\| < \zeta \mid \mathcal{H}_{At}\big) > 1 - \zeta, \quad \forall t \geq \bar{t}_\alpha. \tag{22}$$

■

The regulator's beliefs about the agent's behavior (given a particular history)

Next step shows that the regulator eventually becomes convinced that the agent plays a best response to the commitment strategy (i.e., T) in the continuation game on a class of private histories that involve the successive play 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, Mailath, and Samuelson (2007, Lemma 3); let the σ-algebra of the regulator who has played D up to (and not including) period s be $\hat{\mathcal{H}}_{Rs}$. 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 $\varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At})$, the smallest σ-algebra containing the σ-algebras $\hat{\mathcal{H}}_{Rs}$ and $\mathcal{H}_{At}$.

Lemma 12 *For any given α-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ arbitrarily small and for any $t > 0$ and $\tau \geq 0$,*

$$\lim_{s\to\infty} \left\| \tilde{E}[\tilde{\sigma}_{A,s+\tau} \mid \varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At})] - \tilde{E}[\tilde{\sigma}_{A,s+\tau} \mid \hat{\mathcal{H}}_{Rs}] \right\| = 0, \quad \tilde{Q} - a.s.$$

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, Mailath, and Samuelson (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:

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

Now, we subtract $\tilde{Q}[K \mid \hat{h}_{Rs}]$ from both sides of the above equation to get

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

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 any given $i_d \in I_d$. Hence,

$$\begin{aligned}\left|\tilde{Q}[K \mid \hat{h}_{R,s+1}]\right. & - \left.\tilde{Q}[K \mid \hat{h}_{Rs}]\right| \\ & \geq \tilde{Q}[K \mid \hat{h}_{Rs}]\left|\sum_{a\in A}\rho(i_d \mid a, D)\Big(\tilde{E}[\tilde{\sigma}^a_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{E}[\tilde{\sigma}^a_A(h_{As}) \mid \hat{h}_{Rs}]\Big)\right|.\end{aligned}$$

As each of the random variables $\{\tilde{Q}[K \mid \hat{\mathcal{H}}_{Rs}]\}_s$ is a martingale with respect to $(\{\hat{\mathcal{H}}_{Rs}\}_s, \tilde{Q})$, 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

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

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

to rewrite the RHS as

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

As there exists a strictly positive constant x such that

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

we get

$$\lim_{s\to\infty}\tilde{Q}[K \mid \hat{\mathcal{H}}_{Rs}]\left\|\tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid K, \hat{h}_{Rs}] - \tilde{\mathbf{E}}[\tilde{\sigma}_A(h_{As}) \mid \hat{h}_{Rs}]\right\| = 0$$

$\tilde{Q}$-a.s on K, where $\varphi(\hat{\mathcal{H}}_{Rs}, K)$ is the smallest σ-algebra containing the σ-algebra $\hat{\mathcal{H}}_{Rs}$ and event K. Since, $\tilde{Q}[K \mid \hat{\mathcal{H}}_{R\infty}](\omega) \geq \tilde{Q}[K \mid \hat{\mathcal{H}}_{Rs}](\omega) > 0$ for all $s > t$ for $\tilde{Q}$-almost all $\omega \in K$,

$$\lim_{s\to\infty}\left\|\tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \varphi(\hat{\mathcal{H}}_{Rs}, K)] - \tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \hat{\mathcal{H}}_{Rs}]\right\| = 0$$

$\tilde{Q}$-a.s on K. Since this is true for all $K \in \mathcal{H}_{At}$, we obtain

$$\lim_{s\to\infty}\left\|\tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At})] - \tilde{\mathbf{E}}[\tilde{\sigma}_{As} \mid \hat{\mathcal{H}}_{Rs}]\right\| = 0$$

$\tilde{Q}$ − a.s., proving the desired result for $\tau = 0$. ■

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

The next result establishes that for any given α-Nash equilibrium with $\alpha > 0$, there exists a time period $\hat{t}_\alpha$ such that in any history following that period, the agent expects the strategic regulator to be diligent (i.e., choose D) with a high probability.

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

Note that the reputation of the regulator does not change if the realization of the agent's completely mixed strategy is T in a given α-Nash equilibrium 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 10) 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-Nash concerns) along with R's identification condition, Remark 2, empower us to employ the merging argument of Cripps, Mailath, and Samuelson (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.[41] But, such histories would be in contradiction with (21) of Lemma 11. And this makes us obtain the following:

Lemma 13 *Suppose that (21) given in Lemma 11 is satisfied for all $\omega \in \mathcal{F}$ for a given α-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ arbitrarily small. Then, there exists $\hat{t}_\alpha$ such that the following holds for some $\nu > 0$,*

$$\tilde{Q}\big(\sup_{t'>t} \|\hat{\sigma}_R - \tilde{E}[\tilde{\sigma}_{Rt'} \mid \mathcal{H}_{At'}]\| < \nu \mid \mathcal{H}_{At}\big) > 1 - \nu, \quad \forall t \geq \hat{t}_\alpha. \tag{23}$$

The proof of this result, sketched above, employs the same arguments used for purposes of identification as in the proof of Lemma 12 in line with the merging argument of Cripps, Mailath, and Samuelson (2007), and hence, is omitted.

By Lemma 10, for any given α-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ there is a time $\bar{t}_\alpha$ after which the agent assigns very high probability to the event that all her continuation

[41] Note that the agent uses the information gathered along her private history while updating her beliefs about the regulator's type. The agent's posterior belief at time t (that the regulator is tough) 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_∞.

strategies from then on are best replies to the commitment strategy of the regulator (given that the conditions of the lemma are met). And thus, by Lemmas 11 and 13, the agent must be believing that she is going to see the diligent behavior D with a high probability from then on in every continuation game. Lemma 12 states that these beliefs of the agent about her future behavior will eventually be known by the regulator (even if he cannot observe the agent's private history) given that he has indeed been playing diligently with a high probability (as expected by the agent).

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-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ but arbitrarily small. Then, by Lemma 10, 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 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 11 and 13 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 12 states that these beliefs of the agent 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 agent's beliefs about the strategic regulator's behavior (Lemma 13) 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 also measurable for the agent. Instead, following Cripps, Mailath, and Samuelson (2007), we show that $\mathcal{G}$ is close to a $\mathcal{H}_{As}$ - measurable set 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 14 *Let $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ be any given α-Nash equilibrium with $\alpha > 0$ and sufficiently small and suppose that $\mathcal{F}$ is the positive probability set of events stated in Lemmas 10 and 11; and $\nu > 0$ be the constant in Lemma 13. For any $\xi > \alpha$ (as in Lemma 10 given α and $\mathcal{F}$) and number of periods τ, there exists an event $\mathcal{G}$ and a time $T(\xi, \tau)$ such that for all $s > T(\xi, \tau)$,* *there exists* $\mathcal{C}_s \in \mathcal{H}_{As}$ *and* $\phi > 0$ *with*

$$\|\hat{\sigma}_R - \tilde{E}[\tilde{\sigma}_{Rs} \mid \mathcal{H}_{As}]\| < \nu, \quad \tilde{Q} - a.s. \quad on \quad \mathcal{C}_s \tag{24}$$

$$\mathcal{G} \cup \mathcal{F} \subset \mathcal{C}_s \quad and \quad \tilde{Q}(\mathcal{G}) > \tilde{Q}(\mathcal{C}_s) - \phi\tilde{Q}(\mathcal{F}), \tag{25}$$

and, on $\mathcal{G}$*, for any* $s' = \{s, s+1, ..., s+\tau\}$

$$\tilde{E}[\hat{\sigma}_{As'} \mid \hat{\mathcal{H}}_{Rs}] > 1 - 2\sqrt{\xi}. \tag{26}$$

Proof. The proof uses arguments in Lemma 12 and modified version of Lemma 4 of Cripps, Mailath, and Samuelson (2007).

Fix $\tau > 0$. Take $\xi > \alpha$ let $\bar{t}_\alpha$ denote the threshold period stated in Lemma 10 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 indicates $\hat{\mathcal{H}}_{Rs}$ can be expressed as $f_s \equiv \min_{s \leq s' \leq s+\tau} \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \hat{\mathcal{H}}_{Rs}]$ where T denotes the truthful action. Note that $f_s > 1 - 2\sqrt{\xi}$ is sufficient to show (26). The first step is to find a lower bound for f_s. For any $t \leq s$, the triangle inequality implies

$$\min_{s \leq s' \leq s+\tau} \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At})] - k_s^t \leq f_s \leq 1$$

where $k_s^t \equiv \max_{s \leq s' \leq s+\tau} \left|\tilde{E}[\tilde{\sigma}_{As'}(T) \mid \varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At})] - \tilde{E}[\tilde{\sigma}_{As'}(T) \mid \hat{\mathcal{H}}_{Rs}]\right|$ for $t \leq s$. By Lemma 12, $\lim_{s\to\infty} k_s^t = 0$, $\tilde{Q}$ - a.s. Let $\mathcal{G}_t^0 \equiv \{\omega : \tilde{\sigma}_{As}(h_{As}) = 1 - \alpha, \forall s \geq t\}$. Then,

$$f_s \geq \tilde{Q}\big(\mathcal{G}_t^0 \mid \varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At})\big) - k_s^t.$$

The sequence of random variables $\{\tilde{Q}(\mathcal{G}_t^0 \mid \varphi(\hat{\mathcal{H}}_{Rs}, \mathcal{H}_{At}))\}_s$ is a martingale with respect to the filtration $\{\hat{\mathcal{H}}_{Rs}\}$, so it converges almost surely to $g^t \equiv \tilde{Q}(\mathcal{G}_t^0 \mid \varphi(\hat{\mathcal{H}}_{R\infty}, \mathcal{H}_{At}))$. Hence,

$$1 \geq f_s \geq g^t - k_s^t - l_s^t \tag{27}$$

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

The second step involves finding the sets $\mathcal{C}_s$ and an intermediate set (to be denoted by $\mathcal{F}^*$) that is used to determine the set $\mathcal{G}$. First, for any $t \geq \max\{\bar{t}_\alpha, \hat{t}_\alpha\} \equiv t_\alpha$ (the critical periods from Lemmas 10 and 11 and Lemma 13), using condition (19) of Lemma 10 implying condition (23) of Lemma 13, we define the associated events

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

Let $\mathcal{F}_t^s \equiv \bigcap_{\tau=t}^{s} \mathcal{K}_\tau$ 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 10, 11, and 13, $\mathcal{F} \subset \mathcal{K}_t$ for all $t \geq t_\alpha$; thus, $\mathcal{F} \subset \mathcal{F}_t^s$, $\mathcal{F} \subset \mathcal{F}_t$ and $\mathcal{F} \subset \liminf \mathcal{K}_t$.

Also define $\mathcal{N}_t \equiv \{\omega : g^t \geq 1-\sqrt{\xi}\}$, the measure of events that the strategic regulator expects the agent to play T with probability $1-\alpha$ exceeds $1-\sqrt{\xi}$. Set $\mathcal{C}_s \equiv F_{t_\alpha}^s \in \mathcal{H}_{At}$ and define an intermediate set $\mathcal{F}^*$ by $\mathcal{F}^* \equiv \mathcal{F}_{t_\alpha} \cap \mathcal{N}_{t_\alpha}$. Since $\mathcal{C}_s \subset \mathcal{K}_s$, (24) holds (by Lemmas 10 - 13). And, as $\mathcal{F}^* \cup \mathcal{F} \subset \mathcal{C}_s$, the first part of (25) holds with $\mathcal{F}^*$ in the role of $\mathcal{G}$. By definition,

$$\tilde{Q}(\mathcal{C}_s) - \tilde{Q}(\mathcal{F}^*) = \tilde{Q}(\mathcal{C}_s \cap (\mathcal{F}_{t_\alpha} \cap \mathcal{N}_{t_\alpha})^{\mathrm{C}}) = \tilde{Q}\big((\mathcal{C}_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) \cup (\mathcal{C}_s \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}})\big)$$

where the complement of a set X is denoted by $(X)^{\mathrm{C}}$. Since the event $\mathcal{C}_s \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}$ is a subset of $\mathcal{K}_{t_\alpha} \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}$, we have

$$\tilde{Q}(\mathcal{C}_s) - \tilde{Q}(\mathcal{F}^*) \leq \tilde{Q}(\mathcal{C}_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) + \tilde{Q}(\mathcal{K}_{t_\alpha} \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}). \tag{28}$$

Next, we find the upper bounds for the two terms on the right hand side of (28).

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

$$\tilde{Q}(\mathcal{C}_s \cap (\mathcal{F}_{t_\alpha})^{\mathrm{C}}) < \sqrt{\xi} \quad \forall s \geq T'. \tag{29}$$

Also, as $\tilde{Q}(\mathcal{G}_t^0 \mid \mathcal{K}_t) > 1-\xi$ and $\mathcal{K}_t \in \mathcal{H}_{At}$, iterated expectations imply that $1-\xi < \tilde{Q}(\mathcal{G}_t^0 \mid \mathcal{K}_t) = \tilde{E}[g^t \mid \mathcal{K}_t]$. Since $g^t \leq 1$, one gets

$$\begin{aligned} 1-\xi < \tilde{E}[g^t \mid \mathcal{K}_t] \;&\leq\; (1-\sqrt{\xi})\, \tilde{Q}((\mathcal{N}_t)^{\mathrm{C}} \mid \mathcal{K}_t) \;+\; \tilde{Q}(\mathcal{N}_t \mid \mathcal{K}_t) \\ &=\; 1-\sqrt{\xi}\, \tilde{Q}((\mathcal{N}_t)^{\mathrm{C}} \mid \mathcal{K}_t). \end{aligned}$$

This implies that $\tilde{Q}((\mathcal{N}_t)^{\mathrm{C}} \mid \mathcal{K}_t) < \sqrt{\xi}$. By taking $t = t_\alpha$, we get

$$\tilde{Q}(\mathcal{K}_{t_\alpha} \cap (\mathcal{N}_{t_\alpha})^{\mathrm{C}}) < \sqrt{\xi}. \tag{30}$$

Using (29) and (30) in (28), $\tilde{Q}(\mathcal{C}_s) - \tilde{Q}(\mathcal{F}^*) < 2\sqrt{\xi}$ for all $s \geq T'$. Given that $\mathcal{F} \subset \mathcal{C}_s$ and the bound on ξ, $\tilde{Q}(\mathcal{F}^*) > \tilde{Q}(\mathcal{F}) - 2\sqrt{\xi} > (1-2x)\tilde{Q}(\mathcal{F}) > 0$.

Let $T(\xi, \tau) \equiv \max\{T', T''\}$. Since $\mathcal{G} \cup \mathcal{F} \subset \mathcal{F}^* \cup \mathcal{F} \subset \mathcal{C}_s$, the first part of (25) holds.

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 $s \geq T''$. Therefore, $\tilde{Q}(\mathcal{G}) > \tilde{Q}(\mathcal{F}^*) - \sqrt{\xi}$ 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 $\tilde{Q}(\mathcal{C}_s) - \phi\tilde{Q}(\mathcal{F}) < \tilde{Q}(\mathcal{C}_s) - \tilde{Q}(\mathcal{F}^*) < 2\sqrt{\xi} < 2x\tilde{Q}(\mathcal{F})$ 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 $\tilde{Q}(\mathcal{C}_s) - \phi\tilde{Q}(\mathcal{F}) < 2x\tilde{Q}(\mathcal{F})$ is satisfied. Thus, the second part of (25) 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 (27). This with the bound on ξ gives (26) and completes the proof. ■

Next, we will establish that on the set $\mathcal{G}$ the strategic type of the regulator 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 $\mathcal{C}_s$ in Lemma 14. And, this will contradict with the expectations on $\mathcal{H}_{As}$ - measurable set $\mathcal{F}$ stated in Lemmas 10, 11, and 13.

The proof of Theorem 3

Let $(\varsigma_R, \varsigma_A)$ be the stage game strategy 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 that sufficiently close to ς_A, say at most $\bar{v}$ distant to ς_A, playing ς_R is suboptimal by at least $\mu > 0$. Then, for a given discount factor δ, there exists a sufficiently large τ such that the loss of μ for one period is larger than any potential gain hold 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-Nash equilibrium $(\tilde{\sigma}_A, \tilde{\sigma}_R)$ with $\alpha > 0$ and α sufficiently small. Then, fix $\mathcal{F}$ from Lemma 10 with $\xi > \alpha$ and $t_\alpha^* = \max\{\bar{t}_\alpha, \hat{t}_\alpha, T(\xi, \tau)\}$ (as specified in Lemmas 10 – 14). For $\xi < \bar{v}$ and τ, let $\mathcal{G}$ and $\mathcal{C}_s$ be the events described in Lemma 14 for $s > T(\xi, \tau)$. Consider the period $s > T(\xi, \tau)$ at some state in $\mathcal{G}$. By (26), 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 $\xi < \bar{v}$ (by choosing ξ sufficiently small which is feasible as $\alpha > 0$ is sufficiently small, one can ensure that $2\sqrt{\xi} < \bar{v}$). 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, $\tilde{E}[\tilde{\sigma}_{Rs}(D) \mid \mathcal{H}_{As}]$ on the events in $\mathcal{C}_s$ (that contains $\mathcal{F}$ where the agent is truthful after $t > t^*_\alpha$):

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

The first inequality is just removing the absolute values. The second inequality applies $\varsigma_R(D) = 1$ and uses the $\mathcal{H}_{As}$ - measurability of $\mathcal{C}_s$. 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 $\mathcal{G} \subset \mathcal{C}_s \in \mathcal{H}_{As}$. The third and fourth follows from (25) in Lemma 14. Finally, the last one is by $\mathcal{F} \subset \mathcal{C}_s$. Since $\tilde{E}\big[\big|\varsigma_R - \tilde{E}[\tilde{\sigma}_{Rs}(D) \mid \mathcal{H}_{As}]\big|\mathbf{1}_{\mathcal{C}_s}\big] > (1-\phi)\tilde{Q}(\mathcal{F})$ for all $s > t^*_\alpha$ and, by Lemma 13, on $\mathcal{C}_s$, $\tilde{E}[|\varsigma_R - \tilde{E}[\tilde{\sigma}_{Rs} \mid \mathcal{H}_{As}]|\mathbf{1}_{\mathcal{C}_s}] < \nu$, we obtain the desired contradiction and hence concluding the proof. ■

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., and M. Ekmekci (2012): "Reputation in Long-Run Relationships," *The Review of Economic Studies*, 79(2), 451–480.

——— (2015): "Reputation in the Long-Run with Imperfect Monitoring," *Journal of Economic Theory*, 157, 553–605.

Barlo, M., G. Carmona, and H. Sabourian (2016): "Bounded Memory Folk Theorem," *Journal of Economic Theory*, 163, 728–774.

Benabou, R., and G. Laroque (1992): "Using Privileged Information to Manipulate Markets: Insiders, Gurus and Credibility," *Quarterly Journal of Economics*, 107(3), 921–958.

Board, S., and M. Meyer-ter Vehn (2013): "Reputation for Quality," *Econometrica*, 81(6), 2381–2462.

Celentani, M., D. Fudenberg, D. K. Levine, and W. Pesendorfer (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., E. Dekel, and W. Pesendorfer (2005): "Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests," *Journal of Economic Theory*, 121, 259–272.

Cripps, M. W., G. J. Mailath, and L. Samuelson (2004): "Imperfect Monitoring and Impermanent Reputations," *Econometrica*, 72(2), 407–432.

——— (2007): "Disappering Private Reputations in Long-Run Relationships," *Journal of Economic Theory*, 134, 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., O. Gossner, and A. Wilson (2012): "Impermanent Types and Permanent Reputations," *Journal of Economic Theory*, 147, 162 – 178.

Ely, J. C., D. Fudenberg, and D. K. Levine (2008): "When is Reputation Bad?," *Games and Economic Behavior*, 63, 498 – 526.

Ely, J. C., and J. Välimäki (2003): "Bad Reputation," *The Quarterly Journal of Economics*, 118(3), 785 – 814.

Fudenberg, D., and D. K. Levine (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57(4), 759–778.

——— (1992): "Maintaining a Reputation when Strategies are Imperfectly Observed," *Review of Economic Studies*, 59, 561–579.

Fudenberg, D., D. K. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62(5), 997–1039.

Gale, D., and R. Rosenthal (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., and A. Prat (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., and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863 – 894.

Liu, Q. (2011): "Information Acquisition and Reputation Dynamics," *Review of Economic Studies*, 78, 1400–1425.

Liu, Q., and A. Skrzypacz (2014): "Limited Records and Reputation Bubbles," *Journal of Economic Theory*, 151, 2–29.

Mailath, G. J., and L. Samuelson (2001): "Who Wants a Good Reputation?," *Review of Economic Studies*, 68, 415–441.

——— (2006): *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, Oxford.

——— (2015): "Chapter 4: Reputations in Repeated Games," in *Handbook of Game Theory, Volume 4*, ed. by P. Young, and S. Zamir. Elsevier Science Publishers.

Marinovic, I., A. Skrzypacz, and F. Varas (2018): "Dynamic Certification and Reputation for Quality," *American Economic Journal: Microeconomics*, 10(2), 58–82.

Milgrom, P., and J. Roberts (1982): "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, 27, 280 – 312.

Monte, D. (2013): "Bounded Memory and Permanent Reputations," *Journal of Mathematical Economics*, 49, 345–354.

Özdoĝan, A. (2014): "Disappearance of Reputations in Two-Sided Incomplete-Information Games," *Games and Economic Behavior*, 88, 211–220.

——— (2016): "Occurrence of Deception under the Oversight of a Regulator Having Reputation Concerns," in *Recent Advances in Game Theory and Applications, European Meeting on Game Theory, Saint Petersburg, Russia, 2015 and Networking Games and Management, Petrozavodsk, Russia, 2015*, ed. by L. A. Petrosyan, and V. V. Mazalov. 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.

——— (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.