# Reputation Building under Uncertain Monitoring

Joyee Deb  Yuhta Ishii*

February 16, 2017

**Abstract**

We study a canonical model of reputation between a long-run player and a sequence of short-run opponents, in which the long-run player is privately informed about an uncertain state, which determines the monitoring structure in the reputation game. We present necessary and sufficient conditions (on the monitoring structure and the type space) to obtain reputation building in this setting. Specifically, in contrast to the previous literature with only stationary commitment types, reputation building is generally not possible when there is uncertainty about monitoring. Moreover, reputation formation is highly sensitive to the inclusion of other commitment types. However, we show that with the inclusion of appropriate dynamic commitment types, reputation formation can again be sustained while maintaining robustness to the inclusion of other arbitrary types.

*Deb: Yale University (joyee.deb@yale.edu), Ishii: Centro de Investigación Económica, ITAM (yuhta.ishii@itam.mx). We are extremely grateful to Dilip Abreu, Heski Bar-Isaac, Martin Cripps, Mehmet Eckmekci, Drew Fudenberg, Johannes Hörner, Michihiro Kandori, Barry Nalebuff, Aniko Öry, Andy Skrzypacz, Alex Wolitzky and Jidong Zhou for very useful comments and suggestions. We also thank seminar participants at Brown, ITAM, Oxford, Queen Mary University of London, University College London, University of Warwick, Yale, and the SITE Summer Workshop 2016 in Dynamic Games, Contracts, and Markets.

# 1 Introduction

Consider a long-run firm building a reputation for producing environmentally friendly products. Such a reputation is valuable for the firm because consumers inherently care about the environmental impact of their purchases and are often willing to pay more for green products. Consumers make purchase decisions based on the presence/absence of "eco-friendly" labeling of the product, but are typically unsure of how to interpret the labels.[1] Many of these labels are genuine certifications with stringent standards, but numerous others have been discredited. As a result, upon seeing an eco-label, consumers are uncertain about its informational content.[2] Faced with such uncertainty among consumers, a firm, even after honest investment in green products and after undergoing reliable certification, may find it difficult to establish a positive reputation among the consumers. In such a setting, how can a firm build a green reputation and convince consumers that its products are indeed environmentally friendly?

Similar examples arise in other contexts. Consumers make purchase decisions based on the product reviews of a particular review site but do not know exactly how to interpret the reviews. For instance, they may face uncertainty about the degree of correlation between their own tastes with those of the reviewer: so, a positive review may signal either good or bad quality. As another example, consider a citizen who must decide whether to contribute to a local political campaign. She wishes to contribute only if she is convinced that the local representative will exert effort to introduce access to universal child-care. She must however decide based on information in public media about the candidate's work and faces uncertainty about the bias of the media source. In all these settings, the audience (consumer or citizen) cannot accurately monitor the actions of the reputation builder, because she is uncertain about how to interpret the signals that she observes (lack of credibility of ecolabel / unknown bias of the reviewer). As a result, she cannot link what she observes directly to the actions of the reputation builder. The central question of this paper is whether reputations can be built in environments with such uncertainty in monitoring.

To start, consider reputation building in environments in the absence of such uncertainty. Canonical models of reputation (e.g., Fudenberg and Levine (1992)) study the behavior of a long-run agent (say, a firm) who repeatedly interacts with short-run opponents (consumers). There is incomplete information about the firm's type: consumers entertain the possibility that the firm is of a "commitment" type that is committed to playing a particular action at every period of the game. Even when the actions of the firm are

[1] Indeed, as reported by Ecolabel Index, there are over 400 eco-labels and green certification systems in the market.

[2] The Federal Trade Commission maintains: "Very few products, if any, have all the attributes consumers seem to perceive from such claims, making these claims nearly impossible to substantiate." According to Scott Poynton, founder of The Forest Trust, "The trouble ... with eco-labels is that some of them are OK, especially when done well. But they can also be highly misleading ... When there are so many cases where a variety of eco-labels are shown up as greenwashing nonsense, how can you have confidence in any of them?"

noisily observed, the classical reputation building result states that if a sufficiently rich set of commitment types occurs with positive probability, a patient firm can achieve payoffs arbitrarily close to his Stackelberg payoff of the stage game in *every* equilibrium.[3] Intuitively by mimicking a commitment type that always plays the Stackelberg action, a long-run firm can eventually signal to the consumer its intention to play the Stackelberg action in the future and thus obtain high payoffs in *any* equilibrium. Importantly, this result is robust to the introduction of other arbitrary commitment types.

Of course this intuition critically relies on the consumer's ability to accurately interpret the noisy signals. On the other hand, if monitoring is uncertain, the reputation builder finds it difficult to signal his intentions. To study reputation building with uncertain monitoring, we consider a canonical model of reputation building, with one key difference. At the beginning of the game, a state of the world, $(\theta, \omega) \in \Theta \times \Omega$ is realized, which determines both the type of the firm, $\omega$, and the monitoring structure, $\pi_\theta : A_1 \to \Delta(Y)$: a mapping from actions taken by the firm to distribution of signals, $\Delta(Y)$, observed by the consumer. We assume for simplicity that the firm knows the state of the world, but the consumer does not.[4]

We first show in a simple example that uncertain monitoring can cause the traditional reputation building result to break down: Even if consumers entertain the possibility that the firm may be a "commitment" type that is committed to playing the Stackelberg action every period, there exist equilibria in which even an arbitrarily patient firm obtains payoffs far below its Stackelberg payoff. In the context of our eco-labeling example, such "bad equilibria" arise because even if the firm mimics the commitment type's strategy of producing green products when being certified by a reliable eco-labelling agency, the consumer may still believe that the firm is using dirty technology and obtaining a certification from an unreliable eco-labeling agency. Such negative examples arise due to an identification problem that arises as a result of the uncertainty about monitoring: Potentially good actions in one state cannot be statistically distinguished from a bad action in a different state. This simple example leads us then to ask what might restore reputation building under uncertain monitoring even in the face of such identification problems.

In our main theorem, we construct a set of commitment types such that when these types occur with positive probability, a sufficiently patient firm obtains payoffs arbitrarily close to the Stackelberg payoff in all equilibria even when the consumers are uncertain about the monitoring environment. Importantly, as in the classical reputation literature, this result is robust to the inclusion of other arbitrary commitment types, and thus is independent of the fine details of the type space. In contrast to the commitment types considered in the previous literature, the commitment types that we construct are committed to *dynamic* strategies

[3] The Stackelberg payoff is the payoff that the long-run player would get if he could commit to an action in the stage game.

[4] We conjecture that our results extend in a straightforward manner to an environment in which the state $\theta$ is also unknown to the reputation builder.

(time-dependent but not history dependent) that switch infinitely often between "signaling actions" that help the consumer learn the unknown monitoring state and "collection actions" that are desirable for payoffs (the Stackelberg action). A key contribution of our paper is the construction of these dynamic commitment types that play *periodic* strategies. It turns out that such dynamic commitment types are indeed necessary for reputation building under uncertain monitoring. This is because signaling the unknown monitoring state and Stackelberg payoff collection may require the use of different actions in the stage game.

Our main reputation theorem admits an additional important interpretation as a robust reputation result. The commitment types and uncertainty regarding the monitoring structure can be interpreted as representations of the *subjective* uncertainty that the consumers possess about both the actual monitoring structure and the behavior of the reputation-building firm. Under this interpretation, our reputation theorem states that the firm can indeed effectively establish a reputation, as long as the consumers assign positive probability to the constructed commitment types and the correct monitoring structure. Importantly our result holds independently of what beliefs the consumer may have about the behavior of the firm under incorrect monitoring structures, highlighting its robustness.

The argument underlying our main theorem proceeds in two key steps. First, under mild assumptions on the information structure, we show that, in type spaces that place positive probability on our constructed types, the firm can, by mimicking these types, guarantee that consumers learn the true (monitoring) state during the signaling phases at a rate that is uniform across all equilibria. The uniformity of the rate of learning is not obvious and we use merging arguments similar to those used in Gossner (2011) and Sorin (1999). Secondly, it may still not suffice to ensure that the opponent's belief on the true monitoring state is high during the signaling phase. Since the commitment type alternates between signaling and collection, we need to ensure that that the consumer's belief on the true state does not dramatically decrease during an intervening collection phase. We use Doob's up-crossing inequality for martingales to produce a uniform (across all equilibria) bound on the number of times such decreases in beliefs can occur in the intervening collection phases. The above arguments establish that by mimicking the constructed commitment type, the consumers' beliefs are guaranteed, in any equilibrium, to be high on the true state with large probability most of the time. Then by using merging arguments à la Gossner (2011) and Sorin (1999), we can show that indeed the firm obtains payoffs arbitrarily close to the Stackelberg payoff.

It is important to highlight a key feature of our dynamic commitment types: They return to the signaling phase infinitely often. Our simple negative example demonstrates the necessity for reputation building of a commitment type that engages in a signaling phase. However, one might conjecture that the inclusion of a commitment type that begins with a sufficiently long phase of signaling followed by a switch to playing the Stackelberg action for the true state would also suffice for reputation building. Importantly this is *not* sufficient, and the recurrent nature of signaling is essential to reputation building. If we restrict commitment types to be able to teach the monitoring state only at the start of the interaction (for any arbitrarily long period of time), we can construct a counterexample to show that reputation building fails: there exist equilibria in which even an arbitrarily patient long-run player obtains a payoff that is substantially lower than the Stackelberg payoff.

Returning to the eco-labelling example, if consumer purchase decisions depend on product labeling, and the consumer is uncertain about how to interpret eco-labels, a firm cannot build a reputation for environmentally friendly products quality by simply investing effort into producing these products. Effective reputation building requires not only the actual production of environmentally friendly products but also a repeated commitment to credibly convey to the consumer the meaning of the eco-labels. Indeed, we observe this in practice. There are some high-quality eco-labels and stringent certification systems like Energy Star in the US or Blue Angel in Germany. Firms that have invested in producing environmentally friendly products typically obtain certifications from these reliable certification systems and, from time to time, run expensive integrated campaigns aimed to educate the consumer about the environmental practices required to obtain these certifications. For instance, TCP a leading seller of energy efficient lighting solutions, ran a widespread campaign with educational events across the US to educate consumers on significance of their Energy Star certification. UPM-Kymmene Corporation, a global forest products company, invested significantly to use recycled and renewable material to obtain certification by the EU ecolabel. Subsequently, the company ran widespread promotion campaigns to increase public awareness and knowledge about what the EU Ecolabel stood for, and the stringent standards the company adhered to.[5] Moreover, we see that firms that have invested in green products typically renew their certifications regularly according to the most recent standards, and employ such integrated awareness campaigns throughout the product lifecycle.

While this paper is motivated by environments with uncertain monitoring, our results apply more broadly to other types of uncertainty. First, our model allows for both uncertainty regarding monitoring *and* uncertainty about the payoffs of the reputation builder. Our results also extend to environments with symmetric uncertainty about monitoring. For example, consider a firm that is entering a completely new market and is deciding between two different product offerings. Neither the consumer nor the firm initially know which product is better for the consumer. Is it possible for the firm to build a reputation for making the better

[5] Curtin (2002) points out that industry analysts maintain that companies that are committed to sound environmental practices also engage in green advertising. As another example, Church & Dwight, maker of Arm & Hammer baking soda products, which has been conservation oriented since its inception in 1888, not only produces environmentally friendly products, but also continues to sponsor green radio and TV broadcasts and produces in-store consumer education displays called Enviro-centers to maintain consumer awareness about their practices and labeling.

product? Note that our results apply here. Mimicking a commitment type that both signals and collects is useful to the firm here in two ways: It not only helps the consumer learn about the unknown state of the world, but simultaneously enables the firm to learn the true state of the world. Then, we can interpret the commitment type as one that alternates between learning the state and payoff collection.

Thus far, we have restricted our discussion to a lower bound on the long-run player's equilibrium payoff. Of course the immediate question that arises is whether the long-run player can possibly obtains payoffs much higher than the Stackelberg payoff: How tight is this lower bound on payoffs? With uncertain monitoring, there may be situations in which a patient long-run player can indeed guarantee himself payoffs that are strictly higher than the Stackelberg payoff of the true state. We present several examples in which this occurs: It turns out that the long-run player does not find it optimal to signal the true state to his opponent, but would rather block learning and attain payoffs that are higher than the Stackelberg payoff in the true state. In general, an upper bound on a patient long-run player's equilibrium payoffs depends on the set of commitment types and the prior distribution over types. Such dependence on the specific details of the game makes a general characterization of an upper bound difficult.[6] A precise characterization of an upper bound is beyond the scope of this paper. Nevertheless, we provide a joint sufficient condition on the monitoring structure and stage game payoffs that ensure that the lower bound and the upper bound coincide for any specification of the type space: Loosely speaking, these are games in which state revelation is desirable for the reputation builder.

## 1.1 Related Literature

There is a vast literature on reputation effects which includes the early contributions of Kreps and Wilson (1982) and Milgrom and Roberts (1982) followed by the canonical models of reputation developed by Fudenberg and Levine (1989), Fudenberg and Levine (1992) and more recent methodological contributions by Gossner (2011). To the best of our knowledge, our paper is the first to consider reputation building in the presence of uncertain monitoring.

Aumann, Maschler, and Stearns (1995) and Mertens, Sorin, and Zamir (2014) study repeated games with uncertainty in both payoffs and monitoring but focus primarily on zero-sum games. In contrast, reputation building matters most in non-zero sum environments where there are large benefits that accrue to the reputation builder from signaling his long-run intentions to the other player. There is some recent work on uncertainty in payoffs in non-zero sum repeated games by Wiseman (2005), Hörner and Lovo (2009), Hörner,

[6] This is in sharp contrast to the previous papers in the literature, where the payoff upper bound is generally independent of the fine details of the type-space such as the relative probabilities of commitment types.

Lovo, and Tomala (2011). In all of these papers, however, the monitoring structure is known to all parties with certainty. Our paper's modeling framework corresponds most closely to Fudenberg and Yamamoto (2010) who study a repeated game model in which there is uncertainty about both monitoring and payoffs. However, Fudenberg and Yamamoto (2010) focus their analysis on an equilibrium concept called perfect public ex-post equilibrium in which players play strategies whose best-responses are independent of any belief that they may have about the unknown state. As a result, in equilibrium, no player has an incentive to affect the beliefs of the opponents about the underlying monitoring structure. In contrast, our paper studies, more generally, equilibria where the reputation builder potentially benefits from affecting the beliefs of the opponent about the monitoring structure. In fact, the possibility of such manipulation is crucial in our setting.

To the best of our knowledge, the construction of the dynamic types necessary to establish a reputation result is novel. The necessity of such dynamic commitment types in our setting is somewhat surprising, and arises for a very different reason than in the literature on reputation building against long-run, patient opponents.[7] In particular, dynamic commitment types arise in Aoyagi (1996), Celentani, Fudenberg, Levine, and Pesendorfer (1996), and Evans and Thomas (1997), since establishing a reputation for carrying through punishments after certain histories potentially leads to high payoffs.[8] In contrast, our non-reputation players are purely myopic and so the threat of punishments has no influence on these players. Dynamic commitment types turn out to be necessary in our setting to resolve a potential conflict between signaling the correct state and Stackelberg payoff collection which are both desirable to the reputation builder: "signaling actions" and "collection actions" discussed in the introduction are generally not the same. As a result, by mimicking such commitment types that switch between signaling and collection actions, the reputation builder, if he wishes, can signal the correct monitoring structure to the non-reputation builders.

The rest of the paper is structured as follows. We describe the model formally in Section 2. In Section 3, we present a simple example to show that reputation building fails due to non-identification issues that arise when there is uncertainty about the monitoring structure. Section 4 contains the main result of the paper, in which we provide sufficient conditions for a positive reputation result to obtain. In this section, we also discuss to what extent our conditions may be necessary. In particular, we explain what features are important for reputation building. The proof of the main result is in Section 5. In Section 6, we discuss potential upper bounds on long-run payoffs. Section 7 concludes.

[7] In this literature, some papers do not require the use of dynamic commitment types by restricting attention to *conflicting interest* games. See, for example, Schmidt (1993) and Cripps, Dekel, and Pesendorfer (2004).

[8] For other papers in this literature that use similar ideas, see e.g., Atakan and Ekmekci (2011), Atakan and Ekmekci (2015), Ghosh (2014).

# 2 Model

A long-run (LR) player, player 1, faces a sequence of short-run (SR) player 2's.[9] Before the interaction begins, a pair $(\theta, \omega) \in \Theta \times \Omega$ of a *state* of nature and *type* of player 1 is drawn independently according to the product measure $\gamma := \mu \times \nu$ with $\nu \in \Delta(\Theta)$, and $\mu \in \Delta(\Omega)$. We assume for simplicity that $\Theta$ is finite and enumerate $\Theta := \{\theta_0, \dots, \theta_{m-1}\}$, but that $\Omega$ may possibly be countably infinite.[10] The realized pair of state and type $(\theta, \omega)$ is then fixed for the entirety of the game.

In each period $t = 0, 1, 2, \dots$, players simultaneously choose actions $a_1^t \in A_1$ and $a_2^t \in A_2$ in their respective action spaces. We assume for simplicity that $A_1$ and $A_2$ are both finite. Let $A$ denote $A_1 \times A_2$. Each period $t \geq 0$, after players have chosen the action profile $a^t$, a public signal $y^t$ is drawn from a finite signal space $Y$ according to the probability $\pi(y^t \mid a_1^t, \theta)$.[11] Note importantly that both the action chosen at time $t$ and the state of the world $\theta$ potentially affect the signal distribution. Accordingly, we interpret $\theta$ as representing the unknown monitoring structure. Denote by $H^t := Y^t$ the set of all $t$-period *public* histories and assume by convention that $H^0 := \emptyset$. Let $H := \bigcup_{t=0}^{\infty} H^t$ denote the set of all *public* histories of the repeated game.

We assume that the LR player (whichever type he is) observes the realized state of nature $\theta \in \Theta$ fully so that his private history at time $t$ is formally a vector $H_1^t := \Theta \times A_1^t \times Y^t$.[12] Meanwhile the SR player at time $t$ observes only the public signals up to time $t$ and so his information coincides exactly with the public history $H_2^t := H^t$.[13] Then a strategy for player $i$ is a map $\sigma_i : \cup_{t=0}^{\infty} H_i^t \to \Delta(A_i)$. Let us denote the set of strategies of player $i$ by $\Sigma_i$. Finally, let us denote by $\mathcal{A} := \Delta(A_1)$ the set of mixed actions of player 1 with typical element $\alpha_1$ and let $\mathcal{B}$ be the set of static state contingent mixed actions, $\mathcal{B} := \mathcal{A}^m$ with typical element $\beta_1$.

## 2.1 Type Space

We now place more structure on the type space. We assume that $\Omega = \Omega^c \cup \{\omega^o\}$, where $\Omega^c$ is the set of *commitment types* and $\omega^o$ is a *opportunistic* type. For every type $\omega \in \Omega^c$, there exists some strategy $\sigma_\omega \in \Sigma_1$ such that type $\omega$ always plays $\sigma_\omega$. In this sense, every type $\omega \in \Omega^c$ is a commitment type that is committed to playing $\sigma_\omega$ in all scenarios. In contrast, type $\omega^o \in \Omega$ is an *opportunistic type* who is free to choose any strategy $\sigma \in \Sigma_1$.

[9]In the exposition, we refer to the LR player as male and SR player as female.
[10]The assumption of allowing $\Omega$ to be countably infinite is standard in the existing literature (see, for instance, Fudenberg and Levine (1992)) when the Stackelberg action of the stage-game can be mixed.
[11]Note that the public signal distribution is only affected by the action of player 1.
[12]We believe that it is a straightforward extension to consider a LR player who must learn the state over time.
[13]Observability or lack thereof of previous SR player's actions do not affect our results.

## 2.2 Payoffs

The payoff for the SR player 2 at time $t$ is given by:

$$\mathbb{E}\left[u_2(a_1^t, a_2^t, \theta) \mid h^t, \sigma_1, \sigma_2\right].$$

On the other hand, the payoff of the LR opportunistic player 1 in state $\theta$ is given by:

$$U_1(\sigma_1, \sigma_2, \theta) := \mathbb{E}\left[(1-\delta)\sum_{t=0}^{\infty}\delta^t u_1(a_1^t, a_2^t, \theta) \mid \sigma_1, \sigma_2, \theta\right].$$

Then the ex-ante expected payoff of the LR opportunistic player 1 is given by:

$$U_1(\sigma_1, \sigma_2) := \sum_{\theta\in\Theta}\nu(\theta)U_1(\sigma_1, \sigma_2, \theta).$$

Finally, given the stage game payoff $u_1$, we can define the statewise-Stackelberg payoff of the stage game. First for any $\alpha_1 \in \mathcal{A}$, let us define $B_2(\alpha_1, \theta)$ as the set of best-responses of player 2 when player 2 knows the state to be $\theta$ and player 1 plays action $\alpha_1$. The Stackelberg payoff and actions of player 1 in state $\theta$ are given respectively by:

$$u_1^*(\theta) := \max_{\alpha_1\in\mathcal{A}_1}\inf_{\alpha_2\in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta),$$

$$\alpha_1^*(\theta) := \arg\max_{\alpha_1\in\mathcal{A}_1}\inf_{\alpha_2\in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta).$$ [14]

Finally, we define $\mathcal{S}^\varepsilon$ to be the set of state-contingent mixed actions in which the worst best-response of player 2 approximates the Stackelberg payoff up to $\varepsilon > 0$ in every state:

$$\mathcal{S}^\varepsilon := \left\{\beta_1 \in \mathcal{B} : \inf_{\alpha_2\in B_2(\beta_1(\theta),\theta)} u_1(\beta_1(\theta), \alpha_2, \theta) \in (u_1^*(\theta)-\varepsilon, u_1^*(\theta)+\varepsilon)\ \forall\theta\in\Theta\right\}.$$

## 2.3 Information Structure

**Definition 2.1.** A signal structure $\pi$ holds action identification for $(\alpha_1, \theta) \in \mathcal{A}\times\Theta$ if

$$\pi(\cdot \mid \alpha_1, \theta) = \pi(\cdot \mid \alpha_1', \theta) \Longrightarrow \alpha_1 = \alpha_1'.$$

Using the above definition, we define the following set.

[14] Note that $\alpha_1^*$ is generally a correspondence and not a function.

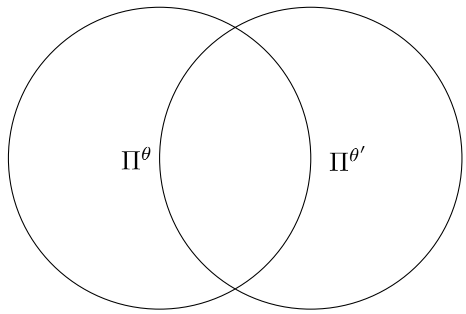


Figure 1: Assumption 2.3 is satisifed.

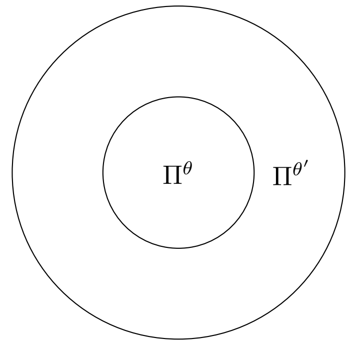


Figure 2: Assumption 2.3 is violated.

**Definition 2.2.** $\Theta^{id} \subseteq \Theta$ is the set of all states $\theta \in \Theta$ such that there exists some $\alpha_1 \in \alpha_1^*(\theta)$ such that information structure $\pi$ has action identification for $(\alpha_1, \theta)$.

In words, a state $\theta \in \Theta^{id}$ if and only if there exists some Stackelberg action such that conditional on the state $\theta$ being common knowledge, the Stackelberg action would be statistically identified from all other actions. Note that this is generally a minimal condition that is required for a LR player to be able to guarantee Stackelberg payoffs in state $\theta$, since if this condition did not hold, reputation building may be impossible even when $\theta$ is common knowledge. Thus our reputation theorem will focus only on reputation building at states $\theta \in \Theta^{id}$. We furthermore make the following assumption for the remainder of the paper.

**Assumption 2.3.** For every $\theta \in \Theta^{id}$ and $\theta' \in \Theta$ such that $\theta' \neq \theta$, there exists some $\alpha_1 \in \mathcal{A}$ such that

$$\pi\left(\cdot \mid \alpha_1, \theta\right) \neq \pi\left(\cdot \mid \alpha_1', \theta'\right)$$

for all $\alpha_1' \in \mathcal{A}$.

Given the above assumption, for any pair of states $\theta \in \Theta^{id}$ and $\theta' \in \Theta$, we denote $\alpha_1(\theta, \theta')$ to be the action defined above. The above assumption is novel. First note that Assumption 2.3 *does not* assume that $\alpha_1(\theta, \theta')$ must be the Stackelberg action in state $\theta$. We can visualize the assumption above as follows. For each $\theta$, denote by $\Pi^\theta$ the set of all probability distributions in $\Delta(Y)$ that are spanned by possibly mixed actions in $\mathcal{A}$ at the state $\theta$:

$$\Pi^\theta = \left\{\pi(\cdot \mid \alpha, \theta) \in \Delta(Y) : \alpha \in \mathcal{A}\right\}.$$

Note that each point in $\Pi^\theta$ is a *probability distribution* over $Y$ and *not* an element of $Y$. If for each pair of states $\theta \neq \theta'$, neither $\Pi^\theta \subseteq \Pi^{\theta'}$ nor $\Pi^{\theta'} \subseteq \Pi^\theta$ holds, then the assumption holds as in Figure 1. On the other hand, Assumption 2.3 is violated if there exists a pair of states in which $\Pi^\theta \subseteq \Pi^{\theta'}$ as in Figure 2. We only impose the condition above pairwise. In fact, even if for some $\theta, \theta', \theta''$, $\Pi^\theta \subseteq \Pi^{\theta'} \cup \Pi^{\theta''}$, the above assumption may still hold. Our analysis will focus on perfect Bayesian equilibria and to shorten the exposition, subsequently we will refer to perfect Bayesian equilibrium as simply equilibrium.

Finally, before we proceed let us establish the following conventions and notation for the remainder of the paper. We will use $\mathbb{N}$ to represent the set of all natural numbers including zero and define $\mathbb{N}_+ := \mathbb{N} \setminus \{0\}$. Whenever the state space $\Theta$ is binary with $\theta \in \Theta$, we will let $-\theta$ denote the state that is complementary to $\theta$ in $\Theta$. Finally, we establish the convention that $\inf \emptyset = \infty$. Given any two actions $a_1, a_1' \in A_1$ and some real number $\lambda \in [0, 1]$, let $\lambda a_1 \oplus (1 - \lambda)a_1'$ denote the mixed strategy that plays $a_1$ with probability $\lambda$ and $a_1'$ with probability $1 - \lambda$.

## 3 Illustrative Example

We begin with a simple example to illustrate that uncertainty in monitoring can have damaging consequences for reputation building. As in the introduction, consider a LR firm (row player) and a sequence of myopic SR consumers who arrive sequentially. At each time period, the LR firm chooses whether to produce a clean product ($C$) or an environmentally dirty product ($D$). Simultaneously, the consumer chooses to buy ($B$) or not buy the product ($N$). The firm has an incentive to protect the environment as long as the consumer is willing to buy the product. Furthermore, the consumer only would like to buy the product if it is a clean product. Otherwise, he would like to abstain from purchasing the product. The stage game is given by:

| | $B$ | $N$ |
|---|---|---|
| $C$ | $\alpha - \gamma + \beta, 1$ | $-\gamma, 0$ |
| $D$ | $\beta, -1$ | $0, 0$ |

Figure 3: Stage Game

We assume that $\alpha, \gamma, \beta > 0$ with $\alpha - \gamma > 0$.[15]

First note that the Stackelberg payoff is $\alpha - \gamma + \beta$ and the Stackelberg action is $C$. Secondly note that in the stage game, $B$ is a best-response to the LR player's stage game mixed action if and only if $\alpha_1(C) \geq \alpha_1(D)$. Suppose now that there are two states $\Theta = \{g, b\}$, which have no effect on payoffs but do affect the distribution of signals observed by the SR players. The state $g$ corresponds to the scenario in which the environmental monitoring agency is indeed good, so that eco-labels and other information on

[15] We can interpret these payoffs as arising from the fact that the long-run firm would like to establish a reputation for producing the "eco-friendly" product due to the consumer's preference for eco-friendly products and its own inherent interests. As a result the firm's utility function given an action profile $(a_1, a_2)$ is:

$$\alpha \mathbf{1}_C(a_1)\mathbf{1}_B(a_2) + \beta \mathbf{1}_B(a_2) - \gamma \mathbf{1}_C(a_1).$$

We assume here that $\alpha - \gamma > 0$ so that the Stackelberg strategy is pure for simplicity. Nevertheless the same negative example with slight modifications survives even if $\alpha - \gamma < 0$.

packaging is informative, while $b$ corresponds to the scenario in which the monitoring agency is bad, in which case information on packaging and labels are completely uninformative.

There are two possible public signals: $Y = \{\bar{y}, \underline{y}\}$. We can interpret these as the signals observed by the consumer (presence of eco-labels or information on packaging). The information structure is given in Figures 4 and 5 and reflect the fact that the signals are informative in one state and not in the other. First

| $\theta = g$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $C$ | $3/4$ | $1/4$ |
| $D$ | $1/4$ | $3/4$ |

Figure 4: Info. Structure under $\theta = g$

| $\theta = b$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $C$ | $3/4$ | $1/4$ |
| $D$ | $3/4$ | $1/4$ |

Figure 5: Info. Structure under $\theta = b$

note that conditional on the state $\theta = g$, actions are identified. Thus if $\theta = g$ were common knowledge, then the classical reputation results would hold: If there was a positive probability that the LR player could be a commitment type that always plays $C$, then a sufficiently patient LR player would achieve a payoff arbitrarily close to $\alpha - \gamma + \beta$ in *every* equilibrium. We will demonstrate below that this observation is no longer true when there is uncertainty about the monitoring states.[16]

## 3.1 Failure of Reputation Building

We will construct an equilibrium in which the LR player gets a payoff close to 0. Suppose that there is uncertainty about the type of the LR player: either he is an opportunistic type, $\omega^o$, who is free to choose any strategy or he is a commitment type, $\omega^c$ that always plays $C$.[17] Consider the following strategy in which types play according to Table 1 at all histories.

| | $\omega^c$ | $\omega^o$ |
|---|---|---|
| $\theta = g$ | $C$ | $D$ |
| $\theta = b$ | $C$ | $D$ |

Table 1: Strategy of Player 1

We show that when $\mu(\omega^c) > 0$ is sufficiently small, there is a perfect Bayesian equilibrium for all $\delta \in (0, 1)$ in which the LR player plays according to the strategy above while the SR player always plays $D$. To simplify notation, we let $\mu(\omega^c) = \xi$ and $\nu(g) = p$. To this end, first consider the Bayesian inference of the SR players. We compute the probability that the SR players assign to the commitment type, which will then be a

[16] In this example, $b \notin \Theta^{id}$, i.e., the Stackelberg action is not identifiable, even conditional on state $b$ being known. This is in keeping with our application that there could be completely uninformative eco-labels. But this feature is inessential for the failure of reputation building. We can construct examples with failure of reputation building even when $\theta \in \Theta^{id}$ for all $\theta$.

[17] This type space mirrors those type spaces studied in the classical reputation literature.

sufficient statistic for his best-response given the candidate equilibrium strategy played by the opportunistic LR player. At any time $t$, conditional on a history $h^t$, let $\mu^t_{a_1\theta}(h^t)$ denote the probability conditional on the public history $h^t$ that the SR player assigns to the event in which the state is $\theta$ and the LR player plays $a_1$.

To analyze these conditional beliefs, consider the following likelihood ratios at any history:

$$\frac{\mu^{t+1}_{Cg}(h^{t+1})}{\mu^{t+1}_{Db}(h^{t+1})} = \frac{\mu^t_{Cg}(h^t)}{\mu^t_{Db}(h^t)}\frac{\pi(y_t \mid C, g)}{\pi(y_t \mid D, b)} = \frac{\mu^t_{Cg}(h^t)}{\mu^t_{Db}(h^t)},$$
$$\frac{\mu^{t+1}_{Cb}(h^{t+1})}{\mu^{t+1}_{Db}(h^{t+1})} = \frac{\mu^t_{Cb}(h^t)}{\mu^t_{Db}(h^t)}\frac{\pi(y_t \mid C, b)}{\pi(y_t \mid D, b)} = \frac{\mu^t_{Cb}(h^t)}{\mu^t_{Db}(h^t)}.$$

Thus the above observation shows that regardless of time $t$ and history $h^t$,

$$\frac{\mu^t_{Cg}(h^t)}{\mu^t_{Db}(h^t)} = \frac{\mu^0_{Cg}(h^0)}{\mu^0_{Db}(h^0)} = \frac{p\xi}{(1-p)(1-\xi)}$$
$$\frac{\mu^t_{Cb}(h^t)}{\mu^t_{Db}(h^t)} = \frac{\mu^0_{Cb}(h^0)}{\mu^0_{Db}(h^0)} = \frac{\xi}{1-\xi}.$$

As a result, we have for all times $t$ and all histories $h^t$,

$$\mu^t(\omega^c \mid h^t) = \mu^t_{Cg}(h^t) + \mu^t_{Cb}(h^t) \leq \left(\frac{p\xi}{(1-p)(1-\xi)} + \frac{\xi}{1-\xi}\right).$$

Then given any $p \in (0,1)$, there exists some $\xi^*$ such that whenever $\xi < \xi^*$, $\mu^t(\omega^c \mid h^t) < \frac{1}{2}$ for all $t$ and all $h^t$.

Given the candidate strategy of the LR player, we have shown that whenever $\xi < \xi^*$, the SR player's best-response is to play $N$ at all histories. Finally given the SR player's strategy, there are no inter-temporal incentives for the LR player, and hence it is also incentive compatible for the opportunistic LR player to always play $D$ (regardless of the discount factor). This gives the LR player a payoff of 0 in this equilibrium, regardless of his discount factor.

Thus this example runs contrary to the reputation results of the classical reputation literature. Such problems for reputation building arise because of the additional problems that non-identification of the Stackelberg action across states poses: $\pi(\cdot \mid C, g) = \pi(\cdot \mid D, b)$. Unlike in the classical reputation models, the opportunistic type here cannot gain by deviating and playing $C$ in state $g$, because by doing so, he will instead, convince the SR player that she is actually facing type $\omega^o$ who always plays $D$ in state $\theta = b$. As a result, the equilibrium renders such deviations unprofitable.

Returning to the motivating example, our observations above imply that if consumer purchase decisions can only be influenced though observed eco-labels and packaging information, and the consumer does not know enough to be able to interpret these labels, a firm cannot build reputation for high quality by simply investing effort into producing environmentally friendly products. In our main theorem to follow, we will show that with additional repeated investment in the form of campaigns promoting awareness about the requirements involved to obtain eco-friendly certification, the firm can once again sustain reputation building. Only after this investment is the firm able to credibly convey to the consumer the meaning of the eco-labels, after which consumers are able to interpret the eco-labels accurately.

### 3.2 Discussion

Note first that the existence of such an example does not depend on the value of $p$. In fact, even if $p$ becomes arbitrarily close to certainty on state $b$, such examples exist, which seems to suggest a discontinuity at $p = 1$. However, this seeming discontinuity arises because $\xi^* > 0$ necessarily becomes vanishingly small as $p \to 1$. This highlights the observation that when the type space contains only simple commitment types that play the same action every period, whether or not the LR player can guarantee Stackelberg payoffs depends crucially on the fine details of the type space such as the relative probability of the commitment type to the degree of uncertainty about the state $\theta$. This is in contrast to the previous literature on reputation building where such relative probabilities did not matter.

In contrast to Ely, Fudenberg, and Levine (2008), our negative example above does not rely on the existence of bad types. The reason is that in our setting, opportunistic types endogenously play "bad" actions in equilibrium. Meanwhile, in the bad reputation setting of Ely, Fudenberg, and Levine (2008), low payoffs are attainable in equilibrium only if there is sufficiently high probability of bad commitment types.

Finally, one of the assumptions underlying our negative example was that $\pi(\cdot \mid C, g) = \pi(\cdot \mid D, b)$. This may lead one to be suspicious about whether such negative examples are robust, since the example required the *exact equality* of the distribution of the public signal conditional on the opportunistic LR player equilibrium action ($D$) in state $b$ and the public signal distribution conditional on the action of the commitment type ($C$) in state $g$. However, as we will see in Section 4.3.1, negative examples arise even if the information structure does not have this knife-edge characteristic if the type space includes "bad" commitment types.[18]

[18] These examples in Section 4.3.1 show that, in contrast to the existing literature on reputation building, the fine details of the type space matter for a reputation result if the type space only contains the standard stationary commitment types. In this paper, we aim to obtain a *robust* reputation theorem that only requires the existence of certain commitment types but is robust to the addition of other arbitrary types. We show that it is not possible to obtain such a result if the type space only includes the standard stationary commitment types.

### 3.3 Recovering Reputation Building

Suppose now that the LR player can undertake a costly action to signal the credibility of the eco-labeling agency. In the example highlighted in the introduction, this took the form of campaigns that promote awareness about the tests involved to obtain eco-friendly certification. In this setting, the new stage game of Figure 6 includes a third action $I$. The signal structure is given by Figures 7 and 8.

| | $B$ | $N$ |
|---|---|---|
| $C$ | $\alpha - \gamma + \beta, 1$ | $-\gamma, 0$ |
| $D$ | $\beta, -1$ | $0, 0$ |
| $I$ | $-10, 0$ | $-10, 0$ |

Figure 6: Stage Game

| $\theta = g$ | $\hat{g}$ | $\bar{y}$ | $\underline{y}$ | $\hat{b}$ |
|---|---|---|---|---|
| $C$ | 0 | 3/4 | 1/4 | 0 |
| $D$ | 0 | 1/4 | 3/4 | 0 |
| $I$ | 1 | 0 | 0 | 0 |

Figure 7: Info. Structure under $\theta = g$

| $\theta = b$ | $\hat{g}$ | $\bar{y}$ | $\underline{y}$ | $\hat{b}$ |
|---|---|---|---|---|
| $C$ | 0 | 3/4 | 1/4 | 0 |
| $D$ | 0 | 3/4 | 1/4 | 0 |
| $I$ | 0 | 0 | 0 | 1 |

Figure 8: Info. Structure under $\theta = b$

In this new game with the same type space as in the previous example, there still remains a perfect Bayesian equilibrium in which the LR player always plays $D$ and obtains a payoff of 0 in both states. However, suppose we now modify the type space to include a type that plays $I$ in period 0 followed by the play of $C$ thereafter. The inclusion of such a type then would rule out the "bad equilibrium" constructed above. In equilibrium, the LR opportunistic type will no longer find it optimal to play $D$ always in state $\theta = g$, since by mimicking this described commitment type, he could obtain a relatively high payoff (if he is sufficiently patient) by convincing the SR players of the correct state with *certainty* and then subsequently building a reputation to play $C$. Essentially by signaling the state in the initial period, he eliminates all identification problems from future periods.

The remainder of the paper will generalize the construction of such a type to general information structures that satisfies Assumption 2.3. However, the generalization will have to deal with some additional difficulties, since, in general, information structures may have full support, in which case, learning about the state is not immediate, as in our simple illustrative example. Moreover, in such circumstances, it is usually impossible to convince the SR players with *certainty* about a state. Therefore there is an additional difficulty that even after having convinced the SR players to a high level of certainty about the correct state, the LR player cannot necessarily be sure that the belief about the correct state will dip to a low level thereafter. Interestingly, due to such issues, even if the state is persistent as in our model, a *robust* reputation theorem requires the presence of dynamic commitment types that commit to repeated (forever), periodic investment to signaling the state. We present a more detailed discussion of these issues after the statement of Theorem 4.1 in Section 4.

# 4 Main Reputation Theorem

Let $\mathcal{C}$ be a collection of commitment types $\omega$ that always play strategy $\sigma_\omega$ and let $\mathcal{G}_\mathcal{C}$ be the set of type spaces $(\Omega, \mu)$ such that $\mathcal{C} \subseteq \Omega$ and $\mu(\omega) > 0$ for all $\omega \in \mathcal{C}$. Virtually all reputation theorems in the existing literature have the following structure. For every $(\Omega, \mu) \in \mathcal{G}_\mathcal{C}$ and every $\varepsilon > 0$, there exists $\delta^*$ such that whenever $\delta > \delta^*$, the LR player receives payoffs within $\varepsilon$ of the Stackelberg payoff in all equilibria. In short, the fine details of the type space beyond the mere fact that $\omega^*$ exists with positive probability in the belief space of the SR players do not matter for reputation building. In this sense, reputation building is *robust*.

In our model with uncertain monitoring, we ask the following question in this spirit: Is it possible to find a set of commitment types $\mathcal{C}$ such that regardless of the type space in question, as long as all $\omega \in \mathcal{C}$ have positive probability, then reputation can be sustained for sufficiently patient players? We have already seen an example in Section 3 that shows that such a result will generally not hold if $\mathcal{C}$ contains only "simple" commitment types that play the same action every period. By introducing dynamic (time-dependent but not history dependent) commitment types, reputation building is recovered.

## 4.1 Construction of Commitment Types

We first construct the appropriate commitment types. A commitment type of type $(k, \beta_1) \in \mathbb{N}_+ \times \mathcal{B}$ will have a *signaling phase* of length $km$ and a *collection phase* of length $k^2$.[19] The length of a block of this commitment type is then given by $\kappa(k) := km + k^2$.

In Assumption 2.3, we have already defined the mixed action $\alpha_1(\theta, \theta')$ for any $\theta \in \Theta^{id}$ and $\theta \neq \theta'$. To simplify notation, let us also choose $\alpha_1(\theta, \theta') \in \mathcal{A}$ arbitrarily if either $\theta = \theta'$ or $\theta \notin \Theta^{id}$. Then for every $k \in \mathbb{N}_+$ and $\beta_1 \in \mathcal{B}$, we now define the following commitment type, $\omega^{k,\beta_1}$, who plays the (possibly dynamic) strategy $\sigma^{k,\beta_1} \in \Sigma_1$ in every play of the game. We define this strategy $\sigma^{k,\beta_1}$ as follows, which depends only on calendar time:

$$\sigma_\tau^{k,\beta_1}(\theta) = \begin{cases} \beta_1(\theta) & \text{if } \tau \textbf{ mod } \kappa(k) > km, \\ \alpha_1(\theta, \theta_j) & \text{if } \tau \textbf{ mod } \kappa(k) \leq km - 1, j := \lfloor \tau \textbf{ mod } \kappa(k)/k \rfloor. \end{cases}$$

[19] Recall that $m = |\Theta|$.

This commitment type plays a dynamic strategy that depends only on calendar time and is periodic with period $k$. At the beginning of each one of these blocks, the commitment type plays a sequence of mixed actions, $\{\alpha_1(\theta, \theta_0)\}$ $k$-times, followed by $\{\alpha_1(\theta, \theta_1)\}$ $k$-times, and so forth. We call this phase of the commitment type's strategy, the *signaling phase*, which will be used to signal the state $\theta$ to the SR players. After the signaling phase, the commitment type plays a mixed action $\beta_1(\theta)$ until the end of the block, which we call the *collection phase* of the commitment type strategy. The type then returns to the signaling phase and repeats. We defer discussion about the important features of this commitment type until after the statement of our main reputation theorem.

## 4.2 Reputation Theorem

We are now equipped to state the main result of the paper: In Theorem 4.1 below, we show that our assumptions on the monitoring structure along with the existence of the commitment types constructed above is sufficient for reputation building in the following sense. A sufficiently patient opportunistic LR player will obtain payoffs arbitrarily close to the Stackelberg payoff of the complete information stage game in every equilibrium of the repeated incomplete information game (at all states $\theta \in \Theta^{id}$).

**Theorem 4.1.** *Suppose that Assumption 2.3 holds. Furthermore, assume that for every $k \in \mathbb{N}$ and every $\varepsilon > 0$, there exists $\beta_1 \in \mathcal{S}^\varepsilon$ such that $\mu(\omega^{k,\beta_1}) > 0$. Then for every $\rho > 0$ there exists some $\delta^* \in (0,1)$ such that for all $\delta > \delta^*$ and all $\theta \in \Theta^{id}$, the payoff to player 1 in state $\theta$ is at least $u_1^*(\theta) - \rho$ in all equilibria.*

Before proceeding to the proof in Section 5, we first discuss the important features of the constructed commitment types here. Our example in Section 3 already suggested that reputation building is generally impossible with only simple commitment types that are committed to playing the same (possibly mixed) action in every period. The broad intuition is that, since the uncertainty in monitoring confounds the SR player's ability to interpret the outcomes she observes, reputation building is possible only if the LR firm can both teach the SR player about the monitoring state and also the intention to play the desirable Stackelberg action. The commitment types that we constructed above do exactly this: They are committed to playing both "signaling actions" that help the consumer learn the unknown monitoring state and "collection actions" that are desirable for payoffs of the LR player. It is worth highlighting that our commitment types are non-stationary, playing a periodic strategy that alternates between signaling phases and collection phases. A similar reputation theorem can be proved also with stationary commitment types that have access to a public randomization device.[20]

[20]We thank Johannes Hörner for pointing this out.

Furthermore, as we have emphasized previously, our reputation result is robust to the inclusion of other possibly "bad" commitment types. This is seen in our result, since Theorem 4.1 only requires the existence of types $\omega^{k,\beta_1}$ while placing no restrictions on the existence or absence of other commitment types.

Relatedly, our theorem can also be interpreted as a robust reputation result in the following sense. Suppose that we interpret the state space $\Theta$ as a representation of the "subjective" uncertainty of the SR players about the informativeness of the public signals about the actions of the LR player. Then our main theorem states that as long as the SR players place a positive probability on the correct $\theta$ and the constructed commitment types, then the LR player can build reputation effectively in the correct state $\theta$, regardless of the SR players' beliefs about play at any other state $\theta' \neq \theta$. We will show in the remainder of this section that when the type space does not include the constructed commitment types, such a robust reputation result need not hold.

## 4.3 Examples: Necessary Characteristics of Commitment Types

Note that our commitment types $\omega^{k,\beta_1}$ share an important feature: the commitment type switches play between signaling and collection phases infinitely often. In this subsection, we show the importance of both

1. the existence of switches between signalling and collection phases in at least some commitment types

2. and the recurrent nature of the signalling phases.

To highlight 1), we construct an equilibrium in an example in which the opportunistic LR player regardless of his discount factor obtains low payoffs if all commitment types play stationary strategies. To highlight the importance of 2), we consider type spaces in which all commitment types play strategies that front-load the signaling phases. In such cases, we again construct equilibria (for all discount factors) in which the opportunistic LR player gets payoff substantially below the statewise Stackelberg payoff in all states.

### 4.3.1 Stationary Commitment Types

We show here in the example to follow that regardless of a given countable (and possibly infinite) set of commitment types that contains only stationary commitment types, $\Omega^*$, we can construct a set of commitment types $\Omega^c \supseteq \Omega^*$ and a probability measure $\mu$ over $\Omega^c \cup \{\omega^o\}$ such that there exists an equilibrium in which the opportunistic LR player obtains payoffs significantly below the statewise Stackelberg payoff.[21]

Consider the stage game described in Figure 9 whose payoffs are state independent. The Stackelberg payoff is 3 and the Stackelberg action is $T$. Note that $L$ is a best-response in the stage game if and only if

[21] In the public randomization interpretation, these types correspond to types that do not use the public randomization device.

|   | $L$ | $R$ |
|---|---|---|
| $T$ | $3,1$ | $0,0$ |
| $B$ | $0,0$ | $1,3$ |

Figure 9: Stage Game

| $\theta=\ell$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $T$ | $1/3$ | $2/3$ |
| $B$ | $5/6$ | $1/6$ |

Figure 10: Info. Structure under $\theta=\ell$

| $\theta=r$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $T$ | $2/3$ | $1/3$ |
| $B$ | $1/6$ | $5/6$ |

Figure 11: Info. Structure under $\theta=r$

$\alpha_1(T)\geq\frac{3}{4}$. The set of states, $\Theta=\{\ell,r\}$, is binary with equal likelihood of both states. The signal space, $Y=\{\bar{y},\underline{y}\}$ is also binary. The information structure is described in Figures 10 and 11.

Suppose we are given a set $\Omega^*$ of commitment types, each of which is associated with the play of a state-contingent action $\beta\in\mathcal{B}$ at all periods. For each $\omega\in\Omega^*$, let $\beta^\omega$ be the associated state contingent mixed action plan. For any pair of mixed action $\alpha\in\mathcal{A}$ such that $\alpha(T)\geq\frac{3}{4}$ and state $\theta\in\Theta$, let $\bar{\alpha}_{-\theta}\in\mathcal{A}$ be the unique mixed action such that $\pi(\cdot\mid\bar{\alpha}_{-\theta},-\theta)=\pi(\cdot\mid\alpha,\theta)$.[22] Note that because of the symmetry of the information structure, $\bar{\alpha}_{-\theta}$ does not depend on the state $\theta\in\Theta$ and so we subsequently omit the subscript.

For each $\omega$ we construct another type $\bar{\omega}$ who also plays a stationary strategy consisting of the following state contingent mixed action at all times:

$$\beta^{\bar{\omega}}(\theta):=\begin{cases}\overline{\beta^{\omega}}(-\theta) & \text{if } \beta^\omega(-\theta)(T)\geq 3/4,\\ B & \text{otherwise.}\end{cases}$$

Let $\bar{\Omega}:=\{\bar{\omega}:\omega\in\Omega^*\}$ and let the set of commitment types be $\Omega^c=\bar{\Omega}\cup\Omega^*$. We prove the following claim.

**Claim 4.2.** *Consider any $\mu\in\Delta(\Omega)$ such that for all $\omega\in\Omega^*$, $\mu(\omega)\leq\mu(\bar{\omega})$. Then for any $\delta\in(0,1)$, there exists an equilibrium in which the opportunistic type plays $B$ at all histories and states.*

*Proof.* We verify that the candidate strategy profile is indeed an equilibrium. Let us define the following set of type-state pairs:

$$\mathcal{D}:=\left\{(\omega,\theta)\in\Omega^c\times\Theta:\beta^\omega(\theta)(T)\geq\frac{3}{4}\right\}.$$

Let $\mathcal{D}_\Omega$ be the projection of $\mathcal{D}$ onto $\Omega$. Note that $\mathcal{D}_\Omega\subseteq\Omega^*$ by construction.

[22] Note that for any $\alpha\in\mathcal{A}$ with $\alpha(T)\geq 3/4$, such an action always exists.

Furthermore, for any $(\omega, \theta) \in \mathcal{D}$, note that

$$\frac{\gamma(\omega, \theta \mid h^t)}{\gamma(\bar{\omega}, -\theta \mid h^t)} = \frac{\gamma(\omega, \theta)}{\gamma(\bar{\omega}, -\theta)} = \frac{\mu(\omega)}{\mu(\bar{\omega})} \leq 1.$$

Note that by construction, if $\alpha(T) \geq 3/4$, then

$$\frac{1}{2}\alpha(T) + \frac{1}{2}\bar{\alpha}(T) = \frac{2}{3} < 3/4.$$

Thus given the candidate strategy profile, we have for all $h^t$:

$$\begin{aligned}
\mathbb{P}(T \mid h^t) &= \sum_{(\omega,\theta)\in\Omega^c\times\Theta} \beta^\omega(\theta)(T)\gamma(\omega, \theta \mid h^t) \\
&= \sum_{(\omega,\theta)\in\mathcal{D}} \big(\gamma(\omega, \theta \mid h^t)\beta^\omega(\theta)(T) + \gamma(\bar{\omega}, -\theta \mid h^t)\beta^{\bar{\omega}}(-\theta)(T)\big) + \sum_{(\omega,\theta)\in(\Omega^*\times\Theta)\setminus\mathcal{D}} \gamma(\omega, \theta \mid h^t)\beta^\omega(\theta)(T) \\
&< \sum_{(\omega,\theta)\in\mathcal{D}} \frac{3}{4}\big(\gamma(\omega, \theta \mid h^t) + \gamma(\bar{\omega}, -\theta \mid h^t)\big) + \sum_{(\omega,\theta)\in(\Omega^*\times\Theta)\setminus\mathcal{D}} \frac{3}{4}\gamma(\omega, \theta \mid h^t) < \frac{3}{4}.
\end{aligned}$$

As a result, the SR player always plays $R$ and thus it is a best-response for the LR opportunistic type to always play $B$. □

The example above shows that if we only allow for the presence of commitment types that always plays the same strategy, then the reputation result breaks down in the sense that the mere existence of these commitment types is not sufficient for reputation building. The co-existence of other "bad" commitment types makes it possible for the LR player to still end up with very low payoffs in equilibrium. In contrast, notice that, Theorem 4.1 does not restrict the type space beyond requiring the existence of the appropriate commitment types.

### 4.3.2 Finite Type Space with Front-Loaded Signaling

Consider the stage game described in Figure 12 that augments the one in Figure 9 by adding a third action $B$ to the LR player's action set. In this modified game, the Stackelberg action is again $T$ giving a payoff of 3 to the LR player. Moreover, $L$ still remains a best-response for the SR player if and only if $\alpha_1(T) \geq \frac{3}{4}$. The public signal space is binary $Y = \{\overline{y}, \underline{y}\}$ and the state space is $\Theta = \{\ell, r\}$ with each state occurring with equal likelihood. The information structure are described by Figures 13 and 14.

Note that all of our assumptions for the main theorem are satisfied in the information structure presented above and so the main theorem holds as long as types with recurrent signaling phases exist. In contrast,

| | $L$ | $R$ |
|---|---|---|
| $T$ | $3, 1$ | $0, 0$ |
| $M$ | $0, 0$ | $1, 3$ |
| $B$ | $-10, 0$ | $-10, 3$ |

Figure 12: Stage Game

| $\theta = \ell$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $T$ | $3/5$ | $2/5$ |
| $M$ | $2/5$ | $3/5$ |
| $B$ | $1/5$ | $4/5$ |

Figure 13: Info. Structure under $\theta = \ell$

| $\theta = r$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $T$ | $2/5$ | $3/5$ |
| $M$ | $3/5$ | $2/5$ |
| $B$ | $4/5$ | $1/5$ |

Figure 14: Info. Structure under $\theta = r$

we now consider a type space in which such commitment types with recurrent signaling phases do not exist and instead, consider a type space in which all of the commitment types have signaling phases that are front-loaded. In such environments, we will show that a reputation theorem does not hold.

For notational simplicity let $\kappa := 4$.[23] Consider the following type space. Let $\omega^t$ denote a commitment type that plays $B$ until period $t-1$ and thereafter switches to the action $T$ forever. Let $N \in \mathbb{N}_+$ and consider the following set of types: $\Omega := \{\omega^1, \ldots, \omega^N\} \cup \{\omega^o\}$.

To define the measure $\mu$ over the types, first fix some $\mu^* > 0$ such that

$$\frac{\mu^*}{1-\mu^*} \frac{\kappa^{N+1} - \kappa}{\kappa - 1} < \frac{3}{4}.$$

Consider any type space such that $\mu(\{\omega^1, \ldots, \omega^N\}) < \mu^*$. We will now show that for any such type space and any discount factor $\delta \in (0, 1)$, there exists an equilibrium in which the LR opportunistic type plays $M$ at all histories and SR players always play $R$.

To show this, we compute at any history the probability that the SR player assigns to the LR player playing $T$ (given the proposed candidate strategy profile above):

$$\mathbb{P}(T \mid h^t) = \mu(\{\omega^s : s \leq t\} \mid h^t) = \gamma\left(\{\omega^s : s \leq t\}, \ell \mid h^t\right) + \gamma\left(\{\omega^s : s \leq t\}, r \mid h^t\right).$$

Now given state $\theta \in \{\ell, r\}$, we want to bound the following likelihood ratio from above:

$$\frac{\gamma\left(\{\omega^s : s \leq t\}, \theta \mid h^t\right)}{\gamma\left(\{\omega^o\}, -\theta \mid h^t\right)} = \sum_{s=1}^{t} \frac{\gamma\left(\{\omega^s\}, \theta \mid h^t\right)}{\gamma\left(\{\omega^o\}, -\theta \mid h^t\right)}.$$

[23] This corresponds to the maximum likelihood ratio according to the signal structure described above. As the construction proceeds, the reader will see exactly why this is important.

But note that given $s < t$, the strategy of $\omega^s$ in state $\theta$ generates exactly the same distribution of public signals as $\omega^o$ in state $-\theta$ at all times $\tau$ between $s$ and $t$. Therefore learning between these two types ceases after time $s$. This allows us to simplify the above expression at any time $t$ and history $h^t$:

$$\frac{\gamma(\{\omega^s : s \le t\}, \theta \mid h^t)}{\gamma(\{\omega^o\}, -\theta \mid h^t)} = \sum_{s=1}^{\min\{t,N\}} \frac{\gamma(\omega^s, \theta \mid h^t)}{\gamma(\omega^o, -\theta \mid h^t)} = \sum_{s=1}^{\min\{t,N\}} \frac{\gamma(\omega^s, \theta \mid h^s)}{\gamma(\{\omega^o\}, -\theta \mid h^s)}.$$

But then note that the above implies:

$$\begin{aligned}
\frac{\gamma(\{\omega^s : s \le t\}, \theta \mid h^t)}{\gamma(\{\omega^o\}, -\theta \mid h^t)} &= \sum_{s=1}^{\min\{t,N\}} \frac{\gamma\left(\omega^s, \theta \mid h^0\right)}{\gamma(\omega^o, -\theta \mid h^0)} \prod_{\tau=0}^{s-1} \frac{\pi(y_\tau \mid B, \theta)}{\pi(y_\tau \mid M, -\theta)} \\
&< \sum_{s=1}^{\min\{t,N\}} \frac{\gamma\left(\omega^s, \theta \mid h^0\right)}{\gamma(\omega^o, -\theta \mid h^0)} \kappa^s \\
&\le \sum_{s=1}^{N} \frac{\gamma\left(\omega^s, \theta \mid h^0\right)}{\gamma(\omega^o, -\theta \mid h^0)} \kappa^s
\end{aligned}$$

$$\begin{aligned}
&\le \frac{\gamma(\{\omega^s : s \le N\}, \theta \mid h^0)}{\gamma(\{\omega^o\}, -\theta \mid h^0)} \sum_{s=1}^{N} \kappa^s \\
&= \frac{\gamma(\{\omega^s : s \le N\}, \theta \mid h^0)}{\gamma(\{\omega^o\}, -\theta \mid h^0)} \frac{\kappa^{N+1} - \kappa}{\kappa - 1} < \frac{\mu^*}{1-\mu^*} \frac{\kappa^{N+1} - \kappa}{\kappa - 1} \\
&< \frac{3}{4}.
\end{aligned}$$

Using the inequalities derived above, we have for any $t$ and $h^t$:

$$\begin{aligned}
\mathbb{P}(T \mid h^t) = \mu(\{\omega^s : s \le t\} \mid h^t) &= \gamma\left(\{\omega^s : s \le t\}, \ell \mid h^t\right) + \gamma\left(\{\omega^s : s \le t\}, r \mid h^t\right) \\
&< \frac{3}{4}\gamma\left(\omega^o, r \mid h^t\right) + \frac{3}{4}\gamma\left(\omega^o, \ell \mid h^t\right) = \frac{3}{4}\gamma\left(\omega^o \mid h^t\right) \le \frac{3}{4}.
\end{aligned}$$

Then the above shows that the probability that the SR player assigns at any history $h^t$ to the LR player playing $T$ is at most $3/4$. This then implies that the SR player's best-response is to play $R$ at all histories, which in turn means that it is incentive compatible for the opportunistic LR player to play $M$ at all histories.

*Remark.* Note that when a type can teach only for up to $N$ periods, then whether the SR players' beliefs about the correct state are high in the future before the switch to the Stackelberg action occurs depends on the probability of that commitment type. If this probability is too small (relative to $N$), then mimicking that type may not lead to sufficiently large beliefs about the correct state in the future. Thus the relative ratio between $K$ and the probability of the commitment type crucially matters. As a consequence, the existence of such a type is again not sufficient for effective reputation building and the fine details of the type space matter.[24]

#### 4.3.3 Infinite Type Space with Front-Loaded Signaling

Note that the finite nature of the type spaces considered above places restrictions automatically on the amount of learning about the state that can be achieved by mimicking the commitment type. We now argue through an example that problems in reputation building are not caused by such limitations in learning the state correctly. In particular, we show more strongly in the following example that even when there is an infinite type space, and learning about the state can be achieved to any degree of desired precision (by playing $C$ for enough periods), difficulties still persist regarding reputation building if all commitment types have signaling phases that are front-loaded.

Consider exactly the same game with the same information structure described in the Subsection 4.3.2 with the following modification of the type space. First choose $t^* > 0$ such that

$$\frac{\kappa^{-t^*}}{1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}}} \frac{1}{1-\kappa^{-1}} < \frac{3}{4}.$$

Furthermore, we can choose $\varepsilon > 0$ such that

$$\frac{\kappa^{-t^*}}{1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon} \frac{1}{1-\kappa^{-1}} < \frac{3}{4}.$$

The set of types is *infinite* and is given by the following set: $\Omega = \{\omega^{t^*}, \omega^{t^*+1}, \ldots\} \cup \{\omega^\infty, \omega^o\}$, where $\omega^\infty$ is a type that plays $B$ at all histories. Again each state is equally likely and the probability measure over the types is given by $\mu \in \Delta(\Omega)$:

$$\mu(\omega^s) = \kappa^{-2s}, \mu(\omega^\infty) = \varepsilon, \mu(\omega^o) = 1 - \sum_{s=t^*}^{\infty} \kappa^{-2s} = 1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon.$$

We will now show that in the above type space, as long as $\varepsilon > 0$ is sufficiently small, regardless of the discount factor, there always exists an equilibrium in which the opportunistic LR player plays $M$ at all histories and the SR player always plays $R$. Using arguments similar to those in Subsection 4.3.2, we can show that, at any history at any time $t$, the SR player never assigns more than $\frac{3}{4}$ probability to the LR player playing $T$, which means that the SR player's best-response is to play $R$ at all histories. As a result, there

[24]Of course, if we place more restrictions on the measure $\mu$, one might conjecture that a reputation theorem might be salvaged. But again such restrictions imply that the fine details of the pair $(\Omega, \mu)$ matter beyond just positive probability of commitment types.

are no inter-temporal incentives for the opportunistic LR player and so it is also indeed his best-response to play $M$ always. The interested reader may refer to the appendix for details.

To ensure that learning about the correct state is not the source of the problems with reputation building in the above example, we present the following claim.

**Claim 4.3.** *Let $\rho \in (0,1)$. Then for every $\theta = \ell, r$, there exists some $t > t^*$ such that in any equilibrium,*

$$\mathbb{P}(\mu\left(\theta \mid h^t\right) > 1-\rho \mid \omega_t, \theta) > 1-\rho.$$

*Proof.* The proof is a direct consequence of merging arguments that will be illustrated in the next section.[25]

□

One perhaps surprising feature of the above example is that because of the inclusion of infinitely many of these switching types $\{\omega^s\}_{s=t^*}^{\infty}$, the state can be taught to the SR players to any degree of precision that the LR player wishes. Nevertheless, reputation building cannot be guaranteed in this example because it may be impossible for the LR player to convince the SR player of *both the correct state and the intention to play the Stackelberg action simultaneously.* We see this in the example. As the opportunistic LR player mimics any of these commitment types, the SR players' beliefs are converging (with arbitrarily high probability) to the correct state. At the same time however, the SR players are placing more and more probability on the types that teach the state for longer amounts of time instead of those types that have switched play to the Stackelberg action.

# 5 Proof of Theorem 4.1

Before we proceed to the details of our proof, let us provide a brief roadmap for how our arguments will proceed. We first show that in any state $\theta \in \Theta^{id}$, by mimicking the strategy of the appropriate commitment type, the LR player can ensure that the SR players learn the state at a rate that is uniform *across all equilibria*.[26] In order to prove this, we first use merging arguments à la Gossner (2011) and Sorin (1999) to show that at times *within the signaling phases*, player 2's beliefs converge to high beliefs on the correct state at a uniform rate across all equilibria. However, note that this does not preclude the possibility that beliefs

[25] One may wonder why we only allow for types $\omega^s$ with $s \geq t^*$. In fact, the construction can be extended to a setting in which $\omega^0, \ldots, \omega^{t^*-1}$ are all included but with very small probability. We omitted these types to simplify the exposition. Moreover, one may also wonder why we include the type $\omega^\infty$. The inclusion of this type makes Claim 4.3 very easy to prove. The arguments for the impossibility of reputation building proceed without modification even when $\varepsilon = 0$, but it becomes much more difficult to prove a claim of the form above. Nevertheless, the inclusion of such a type does not present issues with the interpretation of the above exercise, since we are mainly interested in a reputation result that does not depend on what other types are (or are not) included in the type space.

[26] In the proof, we will formalize this notion of uniform rate of convergence.

may drop to low levels outside the signaling phase. To take care of this potential difficulty, with the help of the well-known Doob's up-crossing inequalities for martingales, we provide a uniform (across equilibria) bound on the number of times that the belief can rise from a low level outside the signaling phase to a high level in the subsequent signaling phase (see Proposition 5.4 below). This then shows that the belief, at most times, will be high on the correct state with high probability, in which case action identification problems are no longer problematic. We then use the merging arguments of Gossner (2011) again to construct our lower bound on payoffs.

*Remark.* It is worth highlighting that Doob's upcrossing inequality is also an important part of the proof of Fudenberg and Levine (1992)'s reputation result. There are two main differences between our use of the inequality and theirs. First, they use the up-crossing inequality to analyze the SR players' belief process about the types of the LR player while we use the inequality to analyze the evolution about the beliefs of $\theta \in \Theta$, which is absent in Fudenberg and Levine (1992). In particular, they use the up-crossing inequality mainly to show that for any subset $\Omega' \subseteq \Omega$, the number of times that the likelihood ratio, $(1 - \mu(\Omega'))/\mu(\Omega')$, falls dramatically in any equilibrium under the probability measure over histories induced by the play of types in $\Omega'$ must be finite, with large probability.

Secondly and more importantly, the study of the belief process about $\theta \in \Theta$ is inherently different from the analysis of the evolution of the likelihood ratio of types in Fudenberg and Levine (1992). This is because in our analysis, the belief process governing the evolution of $\frac{1-\mu(\theta|h^t)}{\mu(\theta|h^t)}$ is generally neither a super-martingale nor a sub-martingale under the probability measure over histories induced by a commitment type in state $\theta$. Thus Doob's up-crossing inequality cannot immediately be applied to study this process and instead, we study a modified process that admits a natural comparison to the original belief process.[27]

## 5.1 Distributions over Public Histories

Let us first define some notation which will be useful for our proof. First note that any $\sigma \in \Sigma_1$ and a prior $\nu \in \Delta(\Theta)$ together determine an ex-ante probability measure over the set of infinite *public histories*, which we denote $\mathbb{P}_{\nu,\sigma} \in \Delta(H^\infty)$. With a slight abuse of notation, we let $\mathbb{P}_{\theta,\sigma} := \mathbb{P}_{\mathbf{1}_\theta,\sigma}$, where $\mathbf{1}_\theta$ is the Dirac measure that places probability one on state $\theta$.

Furthermore, given that type $\omega^o$ chooses a strategy $\sigma \in \Sigma_1$, we define $\bar{\sigma} \in \Delta(\Sigma_1)$ to be a mixed strategy

[27] See Section 5.2 for a more detailed discussion.

that randomizes over the strategies played by the types in $\Omega$ according to the respective probabilities:

$$\bar{\sigma}(\sigma) = \mu(\omega^o), \bar{\sigma}(\sigma') = \sum_{\{\omega \in \Omega^c : \sigma_\omega = \sigma'\}} \mu(\omega) \ \forall \sigma' \neq \sigma.$$

$\bar{\sigma}$ is essentially the aggregate strategy of the LR player that the SR players face when the opportunistic type chooses to play $\sigma$. Of course, $\bar{\sigma}$ is outcome equivalent to a unique behavioral strategy in $\Sigma_1$ and so, with the abuse of notation, henceforth, $\bar{\sigma}$ will refer to this unique behavioral strategy.

Given any behavioral strategy $\sigma \in \Sigma_1$ and any prior over the states $\nu \in \Delta(\Theta)$, we define at any public history $h^t \in H$ the following probability measure in $\Delta(Y^\ell)$:

$$\phi^\ell_{\nu,\sigma}((y_t, y_{t+1}, \dots, y_{t+\ell-1}) \mid h^t) = \sum_{\theta \in \Theta} \nu(\theta) \frac{\mathbb{P}_{\theta,\sigma}(h^t) \prod_{\tau=t}^{t+\ell-1} \pi(y_\tau \mid \sigma(h^t, y_t, \dots, y_{\tau-1}), \theta)}{\mathbb{P}_{\nu,\sigma}(h^t)}.$$

$\phi^\ell_{\nu,\sigma}$ represents the probability distribution over the next $\ell$-periods' public signals given that the LR player is known to play according to $\sigma$ conditional on the public history $h^t$. As before, we abuse notation slightly to write $\phi^\ell_{\theta,\sigma}$ for $\phi^\ell_{\mathbf{1}_\theta,\sigma}$.

Finally, given any prior $\nu \in \Delta(\Theta)$ and $\sigma \in \Sigma_1$, we can define at any history the conditional probability of a state $\theta \in \Theta$, $\nu^\sigma(\cdot \mid h^t) \in \Delta(\Theta)$ in the following manner:

$$\nu^\sigma(\theta \mid h^t) := \nu(\theta) \frac{\mathbb{P}_{\theta,\sigma}(h^t)}{\mathbb{P}_{\nu,\sigma}(h^t)}.$$

In particular, if $\sigma^e \in \Sigma_1$ is an equilibrium behavioral strategy of type $\omega^o$ then in equilibrium, the SR player's belief about the true state $\theta \in \Theta$ at history $h^t$ is given by $\nu^{\bar{\sigma}^e}(\cdot \mid h^t)$.

## 5.2 Uniform Learning of the State

We first begin by showing that playing the strategy $\sigma^{k,\beta_1}$ associated with type $\omega^{k,\beta_1}$ at state $\theta$ leads to uniform learning of the true state $\theta$ in all equilibria. Recall the following definition of the relative entropy of probability measures (also often called the Kullback-Leibler divergence): Given two probability measures $P, Q \in \Delta(Y)$,

$$H(P \mid Q) := \sum_{y \in Y} P(y) \log\left(\frac{P(y)}{Q(y)}\right).$$

Then recall the basic properties of relative entropy that $H(P \mid Q) \geq 0$ for all $P, Q \in \Delta(Y)$ and $H(P \mid Q) = 0$ if and only if $P = Q$.

The following lemma is key to guaranteeing learning of the true state by the SR players.

**Lemma 5.1.** *For every $\varepsilon > 0$, there exists some $k^* \in \mathbb{N}_+$ such that for all $k \geq k^*$ there exists $\lambda > 0$ such that for all $\beta_1 \in \mathcal{B}$, all $\sigma \in \Sigma_1$, all $\theta \in \Theta$, all $\nu \in \Delta(\Theta)$, and all $t$,*

$$H\left(\phi^{km}_{\theta,\sigma^{k,\beta_1}}\left(\cdot \mid h^{\kappa(k)t}\right) \mid \phi^{km}_{\nu,\sigma}\left(\cdot \mid h^{\kappa(k)t}\right)\right) \leq \lambda \Longrightarrow \nu^{\sigma}(\theta \mid h^{\kappa(k)t}) > 1 - \varepsilon.$$

The proof is relegated to the appendix. Define the following sets for an equilibrium strategy $\sigma^e \in \Sigma_1^e$:

$$\begin{aligned}
\mathcal{C}^{k,\beta_1}_{\sigma^e}(\theta, J, \lambda) &:= \left\{h \in H^\infty : H\left(\phi^{km}_{\theta,\sigma^{k,\beta_1}}\left(\cdot \mid h^{\kappa(k)t}\right) \mid \phi^{km}_{\nu,\bar{\sigma}}\left(\cdot \mid h^{\kappa(k)t}\right)\right) > \lambda \text{ for at least } J \text{ values of } t\right\} \\
\mathcal{C}^{\beta_1}_{\sigma^e}(\theta, J, \lambda) &:= \left\{h \in H^\infty : H\left(\phi^{km}_{\theta,\sigma^{k,\beta_1}}\left(\cdot \mid h^{t}\right) \mid \phi^{km}_{\nu,\bar{\sigma}}\left(\cdot \mid h^{t}\right)\right) > \nu \text{ for at least } J \text{ values of } t\right\}, \\
\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, J, \varepsilon) &:= \left\{h \in H^\infty : \nu^{\bar{\sigma}^e}\left(\theta \mid h^{\kappa(k)t}\right) \leq 1 - \varepsilon \text{ for at least } J \text{ values of } t\right\}, \\
\mathcal{D}^{\beta_1}_{\sigma^e}(\theta, J, \varepsilon) &:= \left\{h \in H^\infty : \nu^{\bar{\sigma}^e}\left(\theta \mid h^{t}\right) \leq 1 - \varepsilon \text{ for at least } J \text{ values of } t\right\}.
\end{aligned}$$

Note that $\mathcal{C}^{k,\beta_1}_{\sigma^e}(\theta, J, \lambda)$ and $\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, J, \varepsilon)$ concern only times that are multiples of $\kappa(k)$. The following lemma follows from a standard merging argument.

**Lemma 5.2.** *Let $\lambda, J > 0$. Then for all $\sigma^e \in \Sigma_1^e$ and all $\beta_1 \in \mathcal{B}$,*

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{C}^{\beta_1}_{\sigma^e}(\theta, J, \lambda)\right) \leq \frac{-km \log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{J\lambda}.$$

*Proof.* This is an immediate consequence of Lemma 5.8. □

The next corollary then follows almost immediately from the above two lemmata.

**Corollary 5.3.** *Let $\varepsilon > 0$. Then there exists some $k^* \in \mathbb{N}_+$ such that for every $k \geq k^*$, there exists $\lambda > 0$ such that for every equilibrium strategy $\sigma^e \in \Sigma_1^e$, all $\beta_1 \in \mathcal{B}$, and every $J > 0$,*

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, J, \varepsilon/2)\right) \leq \frac{-km \log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{J\lambda}.$$

*Proof.* Note that by Lemma 5.1, there exists some $k^* \in \mathbb{N}_+$ such that for every $k \geq k^*$, there exists $\lambda > 0$ such that for all $\sigma^e \in \Sigma_1^e$, and all $\beta_1 \in \mathcal{B}$, $\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, J, \varepsilon/2) \subseteq \mathcal{C}^{k,\beta_1}_{\sigma^e}(\theta, J, \lambda)$. Therefore given $k \geq k^*$ and such a $\lambda > 0$, for all $\sigma^e \in \Sigma_1^e$, all $\beta_1 \in \mathcal{B}$, and all $J > 0$,

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, J, \varepsilon/2)\right) \leq \mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{C}^{k,\beta_1}_{\sigma^e}(\theta, J, \lambda)\right) \leq \frac{-km \log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{J\lambda}.$$

□

Note however, that $\mathcal{D}_{\sigma}^{k,\beta_1}(\theta, J, \varepsilon/2)$ focuses only on beliefs at the beginning of the signaling phases. For our reputation theorem, we need the beliefs to be correct outside of the signaling phases, since those are exactly the times of the dynamic game in which the reputation builder actually collects valuable payoffs. To show that with high probability, the beliefs will be correct even outside the signaling phase (for most times), we use Doob's up-crossing inequality. To use Doob's up-crossing inequality, however, note that the stochastic process in question must be either a supermartingale or submartingale. The SR players' beliefs about the state indeed do form a martingale with respect to the measure $\mathbb{P}_{\nu,\bar{\sigma}}$. However, these same beliefs are generally no longer a supermartingale nor a submartingale with respect to the measure $\mathbb{P}_{\theta,\sigma^{k,\beta_1}}$. Our proof will necessarily take care of these additional issues.

First let us introduce some notation. Given a deterministic real-valued sequence $x := \{x_t\}_{t=0}^{\infty}$ and real numbers $a < b$, we can define the up-crossing sequence $U_t^{(a,b)}(x)$ in the following manner. Define the following sequence of times:

$$\tau_0^x := \inf\left\{t : x_t < a\right\},$$
$$\tau_1^x := \inf\left\{t \geq \tau_0^x : x_t > b\right\}.$$

Now we define $\tau_{2k}^x$ and $\tau_{2k+1}^x$ recursively:

$$\tau_{2k}^x := \inf\left\{t \geq \tau_{2k-1}^x : x_t < a\right\},$$
$$\tau_{2k+1}^x := \inf\left\{t \geq \tau_{2k}^x : x_t > b\right\}.$$

Then we can define the number of up-crossings on the interval $(a, b)$ that occur up to time $t$:

$$U_t^{(a,b)}(x) := \inf\{k \in \mathbb{N}_+ : \tau_{2k-1}^x \leq t\}.$$

Finally, since the up-crossing sequence is a non-decreasing sequence, we can define the number of up-crossings in the whole sequence:

$$U_\infty^{(a,b)}(x) := \lim_{t\to\infty} U_t^{(a,b)}(x) \in \mathbb{N} \cup \{\infty\}.$$

**Proposition 5.4.** *Let $\varepsilon > 0$ and let $\sigma^e \in \Sigma^e$. Given any $h \in H^\infty$ and $\theta \in \Theta$, define the sequence*

$$\nu^{\bar{\sigma}^e}(\theta \mid h) := \{\nu^{\bar{\sigma}^e}(\theta \mid h^t)\}_{t=0}^{\infty}$$

*and the corresponding up-crossing sequence* $U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h))$. *Then for all* $t$ *and all* $J > 0$,

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h)) \geq J\right) \leq \frac{\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h))\right]}{J} \leq \frac{2}{\gamma(\theta,\omega^{k,\beta_1})J}.$$

*As a consequence,*

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(U_\infty^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h)) \geq J\right) \leq \frac{2}{\gamma(\theta,\omega^{k,\beta_1})J}.$$

The proof of the proposition is relegated to the Appendix.

We can now use the inequalities proved above together with the previously established observations to bound $\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}_{\sigma^e}^{\beta_1}(\theta, J, \varepsilon)\right)$ uniformly across all equilibrium strategies $\sigma^e \in \Sigma_1^e$.

**Proposition 5.5** (Uniform Learning of True State)**.** *Let* $\varepsilon > 0$. *Then there exists* $k^* \in \mathbb{N}_+$ *such that for all* $k \geq k^*$, *there exists* $\lambda > 0$ *such that for all* $\sigma^e \in \Sigma_1^e$, *all* $\beta_1 \in \mathcal{B}$, *and every* $n \geq 1$,

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}_{\sigma^e}^{\beta_1}(\theta, 2n\kappa(k), \varepsilon)\right) \leq \frac{1}{n}\left(\frac{2}{\gamma(\theta,\omega^{k,\beta_1})} - \frac{km\log\left(\gamma(\theta,\omega^{k,\beta_1})\right)}{\lambda}\right).$$

The proof is in the Appendix. Note importantly that the above bound is independent of the equilibrium, which is useful in establishing a lower bound on payoffs that is *uniform* across all equililbria.

## 5.3 Applying Merging

Having established a bound on the number of times that the belief on the correct state is low, we can then show that at the histories where belief is high on the true state and predictions are correct, the best-response to the Stackelberg action must be chosen. As a result we obtain our main reputation theorem. To this end, we extend the notion of $\varepsilon$-confirmed equilibrium of Gossner (2011) to our framework.[28]

**Definition 5.6.** Let $(\lambda, \varepsilon) \in [0,1]^2$. Then $(\alpha_1, \alpha_2) \in \mathcal{A}_1 \times \mathcal{A}_2$ is a $(\lambda, \varepsilon)$-confirmed best-response at $\theta$ if there exists some $(\beta_1, \nu) \in \mathcal{B} \times \Delta(\Theta)$ such that

- $\alpha_2$ is a best-response for player 2 to $\beta_1$ given belief $\nu$ about the state,
- $\nu(\theta) > 1 - \varepsilon$,
- and $H(\pi(\cdot \mid \alpha_1, \theta) \mid \pi(\cdot \mid \beta_1, \nu)) < \lambda$.

**Lemma 5.7.** *Let* $\rho > 0$. *Then there exists some* $\lambda^* > 0$ *and* $\varepsilon^* > 0$ *such that for all* $(\alpha_1, \alpha_2)$ *that is a* $(\lambda^*, \varepsilon^*)$*-confirmed best-response at* $\theta \in \Theta$, $u_1(\alpha_1, \alpha_2, \theta) > \inf_{\alpha_2' \in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2', \theta) - \rho$.

[28]Fudenberg and Levine (1992) provide a similar definition that uses the notion of total variational distance between probability measures instead of relative entropy.

*Proof.* This lemma is a standard continuity result. ☐

We first define the following set of histories given any two strategies $\sigma, \sigma' \in \Sigma_1$:

$$\mathcal{M}^{\ell}_{\sigma',\sigma}(\theta, J, \lambda) := \left\{h \in H^{\infty} : H\left(\phi^{\ell}_{\theta,\sigma'}(\cdot \mid h^t) \mid \phi^{\ell}_{\nu,\bar{\sigma}}(\cdot \mid h^t)\right) > \lambda \text{ for at least } J \text{ values of } t\right\}.$$

**Lemma 5.8.** *Let $k > 0$, $\beta_1 \in \mathcal{B}$. Then*

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{M}^{\ell}_{\sigma^{k,\beta_1},\sigma}(\theta, J, \lambda)\right) \leq \frac{-\ell \log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{J\lambda}.$$

*Proof.* See Appendix D for the proof. ☐

Together with Lemma 5.8 and Proposition 5.5, we can now complete the proof of Theorem 4.1.

*Proof of Theorem 4.1.* To simplify notation, let us first define the following:

$$\overline{u} := \max_{a \in A} \max_{\theta \in \Theta} u_1(a, \theta),$$
$$\underline{u} := \min_{a \in A} \min_{\theta \in \Theta} u_1(a, \theta).$$

Choose any $\theta \in \Theta$. We will show that there exists some $\delta^* < 1$ such that whenever $\delta > \delta^*$, the LR opportunistic type obtains a payoff of at least $u_1^*(\theta) - \rho$ in every equilibrium. This then proves the theorem, since there are finitely many states $\theta \in \Theta$.

Given $\rho > 0$, choose $k^*$ such that for all $k \geq k^*$, $\frac{km}{\kappa(k)}(\overline{u} - \underline{u}) < \frac{\rho}{4}$. Given these chosen parameters, note that the following inequalities hold for every $k \geq k^*$:

$$\frac{km}{\kappa(k)}\underline{u} + \frac{k^2}{\kappa(k)}\left(u_1^*(\theta) - \frac{\rho}{4}\right) > u_1^*(\theta) - \frac{\rho}{2}, \tag{1}$$
$$\frac{\rho}{4(\overline{u} - \underline{u})}\underline{u} + \left(1 - \frac{\rho}{4(\overline{u} - \underline{u})}\right)\left(u_1^*(\theta) - \frac{\rho}{2}\right) > u_1^*(\theta) - \frac{3}{4}\rho. \tag{2}$$

By Proposition 5.5, there exists some $k > k^*$ for which there exists some $\lambda > 0$ such that for all $\sigma^e \in \Sigma_1^e$, all $\beta_1 \in \mathcal{B}$, and every $n \geq 1$,

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, 2n\kappa(k), \varepsilon)\right) \leq \frac{1}{n}\left(\frac{2}{\gamma(\theta, \omega^{k,\beta_1})} - \frac{km \log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{\lambda}\right).$$

Fix such a $k > k^*$ and $\lambda > 0$ and choose $\beta_1 \in \mathcal{S}^{\rho/8}$ such that $\gamma(\theta, \omega^{k,\beta_1}) > 0$ which exists by assumption.

By Lemma 5.7, there exists some $\varepsilon > 0$ such that

$$u_1(\beta_1(\theta), \alpha_2, \theta) > \inf_{\alpha_2' \in B_2(\beta_1(\theta), \theta)} u_1(\beta_1(\theta), \alpha_2', \theta) - \frac{\rho}{8} \geq u_1^*(\theta) - \frac{\rho}{4}$$

for all $(\beta_1(\theta), \alpha_2)$ that is a $(\varepsilon, \varepsilon)$-confirmed best-response at $\theta$, where the last inequality follows from the assumption that $\beta_1 \in \mathcal{S}^{\rho/8}$.

Given the already fixed $k$, $\lambda > 0$, and $\beta_1$, we can choose $n \in \mathbb{N}$ sufficiently large such that the following two inequalities hold:

$$\frac{\rho}{8(\overline{u} - \underline{u})} > \frac{1}{n}\left(\frac{2}{\gamma(\theta, \omega^{k,\beta_1})} - \frac{km \log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{\lambda}\right),$$
$$\frac{\rho}{8(\overline{u} - \underline{u})} > \frac{-\log\left(\gamma(\theta, \omega^{k,\beta_1})\right)}{2n\kappa(k)\varepsilon}.$$

Then by Proposition 5.5 and Lemma 5.8, for every $\sigma^e \in \Sigma_1^e$,

$$\mathbb{P}_{\theta, \sigma^{k,\beta_1}}\left(\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, 2n\kappa(k), \varepsilon) \cup \mathcal{M}^1_{\sigma^{k,\beta_1}, \sigma^e}(\theta, 2n\kappa(k), \varepsilon)\right) \leq \frac{\rho}{4(\overline{u} - \underline{u})}.$$

Thus in any equilibrium, by playing the strategy $\sigma^{k,\beta_1}$, the LR opportunistic type with discount factor $\delta$ obtains at the very least the following payoff in state $\theta$:

$$\frac{\rho}{4(\overline{u} - \underline{u})}\underline{u} + \left(1 - \frac{\rho}{4(\overline{u} - \underline{u})}\right) g_\delta(\theta),$$

where

$$g_\delta(\theta) = (1 - \delta^{4n\kappa(k)})\underline{u} + (1 - \delta^{4n\kappa(k)})\left(\frac{(1 - \delta^{km})\underline{u} + (\delta^{km} - \delta^{\kappa(k)})\left(u_1^*(\theta) - \frac{\rho}{4}\right)}{1 - \delta^{\kappa(k)}}\right)$$

It remains to show that we can find $\delta^*$ such that for all $\delta > \delta^*$, this lower bound is at least $u_1^*(\theta) - \rho$. To this end, note that as $\delta \to 1$, we have:

$$\begin{aligned}
\frac{\rho}{4(\overline{u} - \underline{u})}\underline{u} + \left(1 - \frac{\rho}{4(\overline{u} - \underline{u})}\right) g_\delta(\theta) &\longrightarrow \frac{\rho}{4(\overline{u} - \underline{u})}\underline{u} + \left(1 - \frac{\rho}{4(\overline{u} - \underline{u})}\right)\left(\frac{km}{\kappa(k)}\underline{u} + \frac{k^2}{\kappa(k)}\left(u_1^*(\theta) - \frac{\rho}{4}\right)\right) \\
&> \frac{\rho}{4(\overline{u} - \underline{u})}\underline{u} + \left(1 - \frac{\rho}{4(\overline{u} - \underline{u})}\right)\left(u_1^*(\theta) - \frac{\rho}{2}\right) \\
&> u_1^*(\theta) - \frac{3}{4}\rho,
\end{aligned}$$

where the last two inequalities follow respectively from Inequalities (1) and (2).

Thus we can find $\delta^* \in (0,1)$ such that for all $\delta > \delta^*$,

$$\frac{\rho}{4(\overline{u}-\underline{u})}\underline{u} + \left(1 - \frac{\rho}{4(\overline{u}-\underline{u})}\right) g_\delta(\theta) > u_1^*(\theta) - \rho.$$

This concludes our proof. □

# 6 Upper Bound on Payoffs

Thus far, we have focused our analysis completely on a lower bound reputation theorem and the sharpness of the lower bound remains to be investigated. This section studies whether and when the lower bound previously established is indeed tight. To this end, we study when an upper bound on payoffs of the opportunistic LR player does indeed equal the lower bound of Theorem 4.1.

Let us impose the following assumption for the remainder of the section.

**Assumption 6.1.** $\Theta^{id} = \Theta$.

We impose this assumption for simplicity, since when $\theta \notin \Theta^{id}$, even when $\theta$ is common knowledge, a precise upper bound is difficult to obtain. This is because when $\theta \notin \Theta^{id}$, there is an action other than the Stackelberg action that generates exactly the same distribution over public signals as the Stackelberg action. Thus, such actions are statistically indistinguishable from the Stackelberg actions. As a result, it may be possible in equilibrium for the LR player to achieve payoffs strictly above the Stackelberg payoff.

Even when $\Theta^{id} = \Theta$, as we assume, because of possible non-identification of actions *across different states*, there may be equilibria in which the LR player obtains payoffs strictly above the Stackelberg payoff. In fact, the upper bound (even for very patient players) typically depends on the initial conditions of the game such as the probability distribution over the states $\Theta$ or over the set of types, $\Omega$. In contrast, in reputation games without any uncertainty about the monitoring structure (and with suitable action identification assumptions), the upper bound on payoffs is independent of these initial conditions as long as the LR player is sufficiently patient. This dependence on the initial conditions makes it difficult to provide a general sharp upper bound as the following example will illustrate.

## 6.1 Example

The following example shows that the probability of commitment types matters for the upper bound regardless of the discount factor. Consider the quality choice game with the following stage game payoffs: In the

| | $L$ | $R$ |
|---|---|---|
| $T$ | $1, 1$ | $-1, 0$ |
| $B$ | $2, -1$ | $0, 0$ |

Figure 15: Quality Choice

repeated game this stage game is repeatedly played and all payoffs are common knowledge. Note that the Stackelberg payoff of the above game is 3/2. Furthermore, note that $L$ is a best-response for the SR player in the stage game if and only if $\alpha_1(T) \geq 1/2$.

There are two states $\Theta = \{\ell, r\}$ which only affect the signal distribution of the public signal. There are two types in the game, $\Omega = \{\omega^c, \omega^o\}$. The commitment type, $\omega^c$, in this game is a type that always plays the mixed action $\frac{2}{3}A \oplus \frac{1}{3}B$ regardless of the state.[29] In particular, we assume that the probability of each state is identical and the probability of the commitment type is given by $\mu$.

The signal space is binary, $Y = \{\bar{y}, \underline{y}\}$, and the information structure is given by the following figures:

| $\theta = \ell$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $T$ | $1/6$ | $5/6$ |
| $B$ | $4/6$ | $2/6$ |

Figure 16: Info. Structure under $\theta = \ell$

| $\theta = r$ | $\bar{y}$ | $\underline{y}$ |
|---|---|---|
| $T$ | $5/6$ | $1/6$ |
| $B$ | $2/6$ | $4/6$ |

Figure 17: Info. Structure under $\theta = r$

Note that according to this information structure, the mixed action $\left(\frac{2}{3}T \oplus \frac{1}{3}B, \theta\right)$ is statistically indistinguishable from $(B, -\theta)$: $\pi\left(\bar{y} \mid \frac{2}{3}T \oplus \frac{1}{3}B, \theta\right) = \pi(\bar{y} \mid B, -\theta)$. In this example, we have the following observation.

**Claim 6.2.** *Let $\varepsilon > 0$. Then there exists some $\mu^*$ such that for all $\mu > \mu^*$ and any $\delta \in (0, 1)$, there exists an equilibrium in which the opportunistic player obtains a payoff of 2 in both states.*

*Proof.* Consider the candidate equilibrium strategy profile in which the opportunistic LR player always plays $B$. Choose $\mu^* = \frac{3}{4}$. Then we will show that when $\mu > \mu^*$, this strategy profile is indeed an equilibrium for any $\delta \in (0, 1)$.

Consider the incentives of the SR player. To study this, we want to compute the probability that the SR player assigns to action $T$ given the candidate equilibrium strategy of the LR player:

$$\mathbb{P}(T \mid h^t) = \frac{2}{3}\mu(\omega^c \mid h^t) = \frac{2}{3}\left(\gamma(\omega^c, \ell \mid h^t) + \gamma(\{\omega^c, r \mid h^t)\right)$$

Now let us compute the probability $\mu(\omega^c \mid h^t)$ from below. To produce this bound, consider the following

[29] Note that this is in reality not the mixed Stackelberg action. However, the example goes through without modification as long as the commitment type plays $A$ with any probability between 1/3 and 1/2.

likelihood ratio:

$$\frac{\gamma(\omega^c, \theta \mid h^t)}{\gamma(\omega^o, -\theta \mid h^t)} = \frac{\gamma(\omega^c, \theta \mid h^0)}{\gamma(\omega^o, -\theta \mid h^0)} = \frac{\mu}{1-\mu}.$$

This then implies that for all $h^t$, $\mu(\omega^c \mid h^t) = \mu, \mu(\omega^o \mid h^t) = 1 - \mu$. Thus, for all $h^t$ and all $\mu > \mu^*$,

$$\mathbb{P}(T \mid h^t) = \frac{2}{3}\mu > \frac{1}{2}.$$

This then implies that for all $h^t$, the SR player's best-response is to play $L$. Furthermore, because the SR player is playing the same action at all histories, the opportunistic LR player's best-response is to play $B$ at all histories. Thus the proposed strategy profile is indeed an equilibrium. Furthermore, according to this strategy profile, the opportunistic LR player's payoff is 2 in both states, concluding the proof. □

The above shows that even an arbitrarily patient opportunistic LR player obtains a payoff strictly greater than the Stackelberg payoff in equilibrium. The problem with the above example is that the commitment type probability is rather large. Therefore, it is instructive to examine an upper bound for the case in which the commitment type probability is indeed small, which we see in the following claim.

**Claim 6.3.** *Let $\varepsilon > 0$. Then there exists some $\mu^* > 0$ such that for all $\mu < \mu^*$, there exists some $\delta^*$ such that for all $\delta > \delta^*$, in all equilibria, the (opportunistic) LR player obtains an ex-ante payoff of at most $3/2 + \varepsilon$.*

*Proof.* This will be a consequence of Theorem 6.5 to be presented in the next subsection. See Appendix for the details. □

## 6.2 Upper Bound Theorem

Here we provide sufficient conditions for when the lower bound and upper bound coincide. In the process, we will provide a general upper bound theorem, with the caveat that generally this upper bound may not be tight (even for patient players).[30] However, we will show that this derived upper bound is indeed tight in a class of games, where state revelation is desirable.[31]

We first provide some definitions that will be useful for constructing our upper bound. The methods presented here follow closely the analysis conducted by Aumann, Maschler, and Stearns (1995) as well as Mertens, Sorin, and Zamir (2014) Chapter V.3 [MSZ].

**Definition 6.4.** Let $p \in \Delta(\Theta)$. A state-contingent strategy $\beta \in \mathcal{B}$ is called non-revealing at $p$ if for all $\theta, \theta' \in supp(p)$, $\pi(\cdot \mid \beta(\theta), \theta) = \pi(\cdot \mid \beta(\theta'), \theta')$.

[30] The previous example should suggest that a general tight upper bound is very difficult to obtain.
[31] We will formalize this informal statement in the following discussion.

In words, this means that if player 1 plays according to a non-revealing strategy at $p$, then with probability 1, player 2's prior will not change regardless of the public signal she sees. For any $p \in \Delta(\Theta)$, define:

$$NR(p) := \{\beta \in \mathcal{B} : \beta \text{ is non-revealing at } p\}.$$

We can define the value function as follows if $NR(p) \neq \emptyset$:

$$V(p) := \max_{\beta \in NR(p)} \max_{\alpha_2 \in B_2(\beta,p)} \sum_{\theta \in \Theta} p(\theta) u_1(\beta(\theta), \alpha_2, \theta).$$

If $NR(p) = \emptyset$, let us define $V(p) = \underline{u}$. Define $\mathbf{cav}V$ to be the smallest concave function that is weakly greater than $V$ pointwise.

**Theorem 6.5** (Upper Bound Theorem). *Let $\varepsilon > 0$ and suppose that the initial prior on the states is given by $\nu \in \Delta(\Theta)$. Then there exists some $\rho^* > 0$ such that whenever $\mu(\Omega^c) < \rho^*$, there exists some $\delta^*$ such that for all $\delta > \delta^*$, the ex-ante expected payoff of the opportunistic LR player in all equilibria is at most $\mathbf{cav}V(\nu) + \varepsilon$.*

Here we sketch the proof, relegating the details to the Appendix. Recall that $\bar{\sigma}$ was defined to be the aggregate strategy that randomizes over the strategies played all types in $\Omega$ according to the respective probabilities.[32] To obtain the desired payoff upper bound, we first obtain a payoff upper bound in any equilibrium if the opportunistic type deviates and plays $\bar{\sigma}$. To establish this payoff upper bound, we use arguments borrowed from MSZ. In particular, we show that the play of $\bar{\sigma}$ leads to the play of "almost" non-revealing strategies at most time periods. Furthermore, at any history $h^t$, the SR player is indeed playing a best-response to the state-contingent mixed action $\bar{\sigma}(h^t)$. Given these two observations, there exists some $\delta^*$ such that for all $\delta > \delta^*$, any type space, and any equilibrium strategy $\sigma$ of the opportunistic type, by playing instead $\bar{\sigma}$, the opportunistic type obtains a payoff of at most $\mathbf{cav}V(\nu) + \varepsilon/2$.[33]

Having established a payoff upper bound to playing $\bar{\sigma}$, we then establish an upper bound on the ex-ante expected payoff to the opportunistic LR player in equilibrium. To this end, note first that for the opportunistic type, the strategy of deviating and playing $\bar{\sigma}$ gives a payoff of at least:

$$\mu(\Omega^c)\underline{u} + \mu(\omega^o)U,$$

where $U$ is the opportunistic type's equilibrium payoff. Therefore with upper bound to playing $\bar{\sigma}$, we must

[32] See Section 5.1 for the precise definition.
[33] Note that $\delta^*$ does not depend on the type space.

have:

$$U \leq \frac{1}{\mu(\omega^o)} \left(\mathbf{cav}V(\nu) + \varepsilon/2 - \mu(\Omega^c)\underline{u}\right).$$

Then by taking $\rho^* > 0$ sufficiently small, we must have $U < \mathbf{cav}V(\nu) + \varepsilon$.

*Remark.* One should note that the above theorem crucially places a requirement on the probability of the commitment types. In the example of Subsection 6.1, we saw that when commitment types are large in probability, the bound provided here does not apply. The reason for the discrepancy when commitment type probabilities are large is that, the SR player's beliefs about the true state $\theta \in \Theta$ in an equilibrium conditional on *the opportunistic type's strategy* is no longer a martingale. In contrast, when the commitment type probabilities are small, these beliefs conditional on the opportunistic type's strategy follow a stochastic process that "almost" resembles a martingale, in which case $\mathbf{cav}V$ provides an approximate upper bound.

### 6.2.1 Statewise Payoff Bounds and Payoff Uniqueness

Finally, we apply Theorem 6.5 to a setting in which the type space includes those commitment types constructed in Section 4. It is easy to see in this scenario that when $V$ is indeed convex, the lower bound and upper bound coincide for patient players and the payoffs of the opportunistic LR player converge uniquely to the statewise Stackelberg payoffs in every state as he becomes arbitrarily patient.

**Corollary 6.6.** *Suppose that*

$$V(\nu) \leq \sum_{\theta \in \Theta} \nu(\theta) u_1^*(\theta)$$

*for all $\nu \in \Delta(\Theta)$. Furthermore, assume that for every $k > m - 1$ and every $\varepsilon > 0$, there exists $\beta_1 \in \mathcal{S}^\varepsilon$ such that $\mu(\omega^{k,\beta_1}) > 0$. Let $\varepsilon > 0$. Then there exists some $\rho^* > 0$ such that whenever $\mu(\Omega^c) < \rho^*$, there exists $\delta^* < 1$ such that for all $\delta > \delta^*$ and any state $\theta \in \Theta$, the opportunistic LR player obtains a payoff in the interval $(u_1^*(\theta) - \varepsilon, u_1^*(\theta) + \varepsilon)$ in all equilibria.*

The proof is in the Appendix. It is worth highlighting a slightly subtle aspect of the corollary. Note that a key distinction between the statement of Theorem 6.5 and the above corollary is that the upper bound on payoffs is given in each state. A key step in the proof of this state-wise upper bound in the corollary relies on the assumption that the constructed commitment types exist with positive probability. This assumption is important for the argument as it first allows us to provide a lower bound on payoffs in each state using Theorem 4.1, which then together with the ex-ante payoff upper bound of Theorem 6.5 allows us to establish the upper bound in each state. Thus without the existence of such commitment types, our proof would not go through.[34]

## 7 Conclusion

We study reputation building by a long-run agent in environments in which there is uncertainty about how the agent's actions relate to observed outcomes. A leading example is that of a long-run firm that wants to build a reputation for making green products. Its consumers make purchase decisions based on eco-labels and information on packaging, but do not understand how exactly to interpret these eco-labels. In particular, they may not know if these eco-labels reliably indicate an environmentally friendly product, or are uninformative. The central question we ask is whether reputations can be built effectively in such settings: Can the firm build a reputation for being environmentally friendly even when consumers do not necessarily trust the green certification or packaging information on their products?

Formally, we study a canonical model of reputation building between a long-run player and a sequence of short-run opponents, in which the long-run player is privately informed about an uncertain state, which determines the monitoring structure, and ask if the classical reputation building result holds: If there is a small positive probability that there is a firm type that is committed to playing the Stackelberg action, then, can a patient firm achieve payoffs arbitrarily close to the Stackelberg payoff of the stage game in every equilibrium?

We first show through a simple example that uncertainty in monitoring can cause reputation building to break down. Specifically, even when consumers entertain the possibility of a commitment type that plays the Stackelberg action, there are equilibria in which the long-run agent gets payoffs that are much below the Stackelberg payoff. Due to the uncertainty in monitoring, the long-run agent cannot convince her opponents about her intention to play the Stackelberg action in the future.

We then present necessary and sufficient conditions on the monitoring structure and type space to restore reputation building in this setting. In contrast to the previous literature, reputation building requires the inclusion of appropriate dynamic commitment types: commitment types that switch infinitely often between "signaling actions" that help the consumer learn the unknown monitoring state and "collection actions" that are desirable for payoffs (the Stackelberg action). A key novelty of our paper is the construction of these dynamic commitment types, and to establish the somewhat surprising fact that not only do we need types that signal the true state, but we need them to signal the state recurrently forever.

While we were motivated by an application of one-sided uncertainty about monitoring, our results gen-

[34] We however, do not know if there exist equilibria in which the state-wise upper bounds fail when such commitment types occur with zero probability.

eralize to both settings in which neither player know the state, and settings in which the uncertainty is about payoffs rather than the monitoring structure. An interesting question is whether reputation building is possible when the monitoring structure is not just uncertain but also changing over time. This is the subject of future research.

# A Infinite Type Space with Front-loaded Signaling

In the main text, we omitted the formal proof that the SR player never assigns more than $\frac{3}{4}$ probability to the LR player playing $T$. We provide the proof below.

As in Subsection 4.3.2, we calculate the probability that the SR player assigns to $T$ being played at history $h^t$ (given the proposed strategy profile). Note that for any $t < t^*$, the above is 0 regardless of the history. So consider $t \geq t^*$. Then we calculate the following likelihood ratio given any state $\theta \in \{\ell, r\}$ in the same manner as in the example in Subsection 4.3.2 by first bounding the following likelihood ratio:

$$\begin{aligned}
\frac{\gamma\left(\{\omega^s : s \leq t\}, \theta \mid h^t\right)}{\gamma\left(\{\omega^o\}, -\theta \mid h^t\right)} &= \sum_{s=t^*}^{t} \frac{\gamma\left(\{\omega^s\}, \theta \mid h^t\right)}{\gamma\left(\{\omega^o\}, -\theta \mid h^t\right)} = \sum_{s=t^*}^{t} \frac{\gamma\left(\omega^s, \theta \mid h^s\right)}{\gamma\left(\omega^o, -\theta \mid h^s\right)} \\
&= \sum_{s=t^*}^{t} \frac{\gamma\left(\omega^s, \theta \mid h^0\right)}{\gamma\left(\omega^o, -\theta \mid h^0\right)} \prod_{\tau=0}^{s-1} \frac{\pi(y_\tau \mid B, \theta)}{\pi(y_\tau \mid M, -\theta)} \\
&< \sum_{s=t^*}^{t} \frac{\gamma\left(\omega^s, \theta \mid h^0\right)}{\gamma\left(\omega^o, -\theta \mid h^0\right)} \kappa^s \\
&= \sum_{s=t^*}^{t} \frac{\frac{1}{2}\kappa^{-2s}}{\frac{1}{2}\left(1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon\right)} \kappa^s \\
&\leq \sum_{s=t^*}^{\infty} \frac{\kappa^{-2s}}{1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon} \kappa^s \\
&= \frac{1}{1-\kappa^{-1}} \frac{\kappa^{-t^*}}{1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon} \\
&< \frac{3}{4},
\end{aligned}$$

where the last inequality was due to our particular choice of $t^*$ and $\varepsilon$.

As in the example of Subsection 4.3.2, this again implies that at any history at any time $t$, the SR player never assigns more than $\frac{3}{4}$ probability to the LR player playing $T$, which means that the SR player's best-response is to play $R$ at all histories. As a result, there are no inter-temporal incentives for the opportunistic LR player and so it is also indeed his best-response to play $M$ always.

# B Proof of Lemma 5.1

*Proof.* Note that it is sufficient to prove the above claim for the history $h^0$. By Lemma C.1 in the Appendix, there exist a collection of compact sets $\{D_{\theta',\theta}\}_{\theta' \neq \theta}$ and $\rho^* > 0$ such that $\Pi_{\theta'} \subseteq D_{\theta',\theta}$ for every $\theta' \neq \theta$, $\alpha_1^{\theta,\theta'} \notin D_{\theta',\theta}$ and for every $T$ and every $\sigma \in \Sigma_1$,

$$\mathbb{P}_{\theta',\sigma}\left(\frac{1}{T+1}\sum_{t=0}^{T} \mathbf{1}_{Y_t} \in D_{\theta',\theta}\right) > \rho^*.$$

By the weak law of large numbers for iid random variables, we can choose $k^* \in \mathbb{N}_+$ such that for all $k \geq k^*$,

$$\max_{\theta\in\Theta}\left\{\max_{\theta'\neq\theta}\left\{\mathbb{P}_{\theta,\alpha_1^{\theta,\theta'}}\left(\frac{1}{k+1}\sum_{t=0}^{k}\mathbf{1}_{Y_t} \in D_{\theta',\theta}\right)\right\}\right\} < \frac{\rho^*\varepsilon}{2(m-1)}.$$

If for some $\theta \in \Theta$, $\nu \in \Delta(\Theta)$, $\sigma \in \Sigma_1$, and $k \geq k^*$, $H(\phi^{km}_{\theta,\sigma^k,\beta_1} \mid \phi^{km}_{\nu,\sigma}) = 0$, we must have $\phi^{km}_{\theta,\sigma^k,\beta_1} = \phi^{km}_{\nu,\sigma}$

in which case,

$$
\begin{aligned}
0 &= H\left(\phi^k_{\theta,\alpha_1^{\theta,\theta_0}}\left(\cdot \mid h^0\right) \mid \phi^k_{\nu,\sigma}\left(\cdot \mid h^0\right)\right), \\
0 &= H\left(\phi^k_{\theta,\alpha_1^{\theta,\theta_1}}\left(\cdot \mid h^k\right) \mid \phi^k_{\nu,\sigma}\left(\cdot \mid h^k\right)\right) \ \forall h^k, \\
\vdots & \ \ \vdots \\
0 &= H\left(\phi^k_{\theta,\alpha_1^{\theta,\theta_{m-1}}}\left(\cdot \mid h^{(m-1)k}\right) \mid \phi^k_{\nu,\sigma}\left(\cdot \mid h^{(m-1)k}\right)\right) \ \forall h^{(m-1)k}.
\end{aligned}
$$

By construction this means that for every $\tau$ such that $\theta_\tau \neq \theta$, and every $h^{\tau k}$, $\nu^\sigma(\theta_\tau \mid h^{\tau k}) < \varepsilon/2(m-1)$. Moreover, due to the martingale property of beliefs, we must have $\nu^\sigma(\theta' \mid h^0) < \varepsilon/2(m-1)$ for all $\theta' \neq \theta$ which then implies that $\nu(\theta) > 1 - \varepsilon/2$.

Thus because the relative entropy function is continuous, for every $k \geq k^*$,

$$
\min_{\theta \in \Theta} \left\{ \inf \left\{ H(\phi^{km}_{\theta,\sigma^{k,\beta_1}}(\cdot \mid h^0) \mid \phi^{km}_{\nu,\sigma}(\cdot \mid h^0)) : \nu(\theta) \leq 1-\varepsilon, \sigma \in \Sigma_1 \right\} \right\} > 0.
$$

Therefore we can choose $\lambda > 0$ such that

$$
\min_{\theta \in \Theta} \left\{ \inf \left\{ H(\phi^{km}_{\theta,\sigma^{k,\beta_1}}(\cdot \mid h^0) \mid \phi^{km}_{\nu,\sigma}(\cdot \mid h^0)) : \nu(\theta) \leq 1-\varepsilon, \sigma \in \Sigma_1 \right\} \right\} > \lambda > 0.
$$

Now given any $k \geq k^*$ we have found $\lambda > 0$ to prove the claim. For any $\theta \in \Theta$, $\nu \in \Delta(\Theta)$, and $\sigma \in \Sigma_1$,

$$
H\left(\phi^{km}_{\theta,\sigma^{k,\beta_1}}(\cdot \mid h^0) \mid \phi^{km}_{\nu,\sigma}(\cdot \mid h^0)\right) \leq \lambda \Longrightarrow \nu(\theta) > 1 - \varepsilon.
$$

□

# C Proof of Missing Step in Lemma 5.1

For a random variable $X$ taking values in a finite set $Y$, define a random empirical vector

$$
\mathbf{1}_X := \left(\mathbf{1}_y(X)\right)_{y \in Y}.
$$

To begin, we prove the following. Let us fix a compact, convex set $C \subseteq \Delta(Y)$. We are interested in a set of stochastic processes with conditional probability vectors lying in $C$ at all times almost-surely. Fix a probability space $(X, \pi, \mathcal{F})$. Then let the space of $Y$-valued random variables be denote by $L(X, Y)$.

Define

$$
\mathcal{S}^T(C) := \{(X_0, X_1, \ldots, X_T) \in L(X,Y)^{T+1} : \forall t, \forall y_0, \ldots, y_{t-1}, \left(\pi\left(X_t = y \mid X_0 = y_0, \ldots, X_{t-1} = y_{t-1}\right)\right)_{y \in Y} \in C\}.
$$

We now prove the following lemma. Note that the law of large numbers does not immediately apply, since the sequence of random variables under consideration may be dependent.

**Lemma C.1.** *Suppose that $C \subseteq \Delta(Y)$ is a compact, convex set and let $a \in \Delta(Y)$ be such that $a \notin C$. Then there exists some $\rho^* > 0$ and some compact set $D \subseteq \Delta(Y)$ such that $D \supseteq C$, $a \notin D$, and such that for every $T$ and for every $(X_0, X_1, \ldots, X_T) \in \mathcal{S}^T(C)$,*

$$
\pi\left(\frac{1}{T+1}\sum_{t=0}^{T} \mathbf{1}_{X_t} \in D\right) \geq \rho^*.
$$

*Proof.* Because $a \notin C$, by the (strong) separating hyperplane theorem, there exists some $\lambda' \neq 0$ such that

$$\sup_{c \in C} \lambda' \cdot c < \lambda' \cdot a.$$

Furthermore, because $C \subseteq \Delta(Y)$ and $a \in \Delta(Y)$, for any $\xi > 0$ and the vector $\mathbf{1} = (1, 1, \ldots, 1)$, we also have:

$$\sup_{c \in C} (\lambda' + \xi \mathbf{1}) \cdot c = \xi + \sup_{c \in C} \lambda' \cdot c < \xi + \lambda' \cdot a = (\lambda' + \xi \mathbf{1}) \cdot a.$$

Choose $\xi > 0$ sufficiently large such that $\lambda' + \xi \mathbf{1} \geq 0$ and let $\lambda = \lambda' + \xi \mathbf{1} \geq 0$. Then we have:

$$g := \sup_{c \in C} \lambda \cdot c < \lambda \cdot a.$$

Now choose $\varepsilon > 0$ and $\rho^*$ such that $\rho^* := 1 - \frac{g}{\lambda \cdot a - \varepsilon} > 0$. Define $D$ to be the following set: $D = \{x \in \Delta(Y) : \lambda \cdot x \leq \lambda \cdot a - \varepsilon\}$. Clearly, $D$ is compact, $C \subseteq D$ and $a \notin D$. Moreover, for any $T$ and any $(X_0, X_1, \ldots, X_T) \in \mathcal{S}^T(C)$, since $\mathbb{E}\left[\frac{1}{T+1} \sum_{t=0}^{T} \mathbf{1}_{X_t}\right] \in C$, $\lambda \cdot \mathbb{E}\left[\frac{1}{T+1} \sum_{t=0}^{T} \mathbf{1}_{X_t}\right] \leq g$. Therefore, by Markov's inequality,

$$\pi\left(\lambda \cdot \left(\frac{1}{T+1} \sum_{t=0}^{T} \mathbf{1}_{X_t}\right) > \lambda \cdot a - \varepsilon\right) \leq \frac{g}{\lambda \cdot a - \varepsilon}.$$

This then implies that for every $T$,

$$\pi\left(\frac{1}{T+1} \sum_{t=0}^{T} \mathbf{1}_{X_t} \in D\right) \geq \pi\left(\lambda \cdot \left(\frac{1}{T+1} \sum_{t=0}^{T} \mathbf{1}_{X_t}\right) \leq \lambda \cdot a - \varepsilon\right) \geq 1 - \frac{g}{\lambda \cdot a - \varepsilon} = \rho^*.$$

□

# D Merging and best-responses

The arguments in this section are analogues of those results proved by Gossner (2011). We modify the arguments and notation slightly.

**Lemma D.1.** *Let $\varepsilon \in (0, 1)$ and suppose that $Q = \varepsilon P + (1 - \varepsilon) P'$. Then*

$$H(P \mid Q) \leq -\log \varepsilon.$$

*Proof.* See Lemma 3 of Gossner (2011) for the proof. □

With this, we can prove Lemma 5.8.

*Proof of Lemma 5.8.* Note that by the chain rule for relative entropy,

$$\begin{aligned}
&\mathbb{E}_{\theta, \sigma^{k, \beta_1}}\left[\sum_{t=0}^{\infty} H(\phi^{\ell}_{\theta, \sigma^{k, \beta_1}}(\cdot \mid h^t) \mid \phi^{\ell}_{\nu, \bar{\sigma}}(\cdot \mid h^t))\right] \\
&= \mathbb{E}_{\theta, \sigma^{k, \beta_1}}\left[\sum_{t=0}^{\infty} \sum_{\tau=0}^{\ell-1} \mathbb{E}_{\theta, \sigma^{k, \beta_1}}\left[H(\phi^{1}_{\theta, \sigma^{k, \beta_1}}(\cdot \mid h^{t+\tau}) \mid \phi^{1}_{\nu, \bar{\sigma}}(\cdot \mid h^{t+\tau})) \mid h^t\right]\right].
\end{aligned}$$

For every $T$,

$$
\begin{aligned}
&\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[\sum_{t=0}^{T}\sum_{\tau=0}^{\ell-1}\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[H(\phi^1_{\theta,\sigma^{k,\beta_1}}(\cdot\mid h^{t+\tau})\mid \phi^1_{\nu,\bar\sigma}(\cdot\mid h^{t+\tau}))\mid h^t\right]\right]\\
&=\sum_{t=0}^{T}\sum_{\tau=0}^{\ell-1}\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[H(\phi^1_{\theta,\sigma^{k,\beta_1}}(\cdot\mid h^{t+\tau})\mid \phi^1_{\nu,\bar\sigma}(\cdot\mid h^{t+\tau}))\right]\\
&\le \ell\sum_{t=0}^{T}\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[H(\phi^1_{\theta,\sigma^{k,\beta_1}}(\cdot\mid h^{t})\mid \phi^1_{\nu,\bar\sigma}(\cdot\mid h^{t}))\right]\\
&=\ell H(\phi^T_{\theta,\sigma^{k,\beta_1}}(\cdot\mid h^0)\mid \phi^T_{\nu,\bar\sigma}(\cdot\mid h^0))\le -\ell\log\left(\gamma(\theta,\omega^{k,\beta_1})\right),
\end{aligned}
$$

where the last inequality comes from the previous lemma. Therefore by monotone convergence,

$$
\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[\sum_{t=0}^{\infty}H(\phi^\ell(\cdot\mid h^t)\mid \phi^\ell(\cdot\mid h^t))\right]\le -\ell\log\left(\gamma(\theta,\omega^{k,\beta_1})\right).
$$

Then by Markov's inequality,

$$
\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{M}^\ell_{\sigma^{k,\beta_1},\sigma}(\theta,J,\lambda)\right)\le -\frac{\ell\log\left(\gamma(\theta,\omega^{k,\beta_1})\right)}{\lambda J}.
$$

□

# E Doob's Up-Crossing Inequality

Here we present the formal statement of the martingale up-crossing inequality.

**Theorem E.1** (Doob's Up-Crossing Inequality)**.** *Let $X := \{X_t\}_{t=0}^{\infty}$ be a submartingale defined on a probability space $(\Xi,\mathbb{P},\mathcal{F})$ and let $a<b$. Then*

$$
\mathbb{E}\left[U_t^{(a,b)}(X(\xi))\right]\le \frac{\mathbb{E}\left[(X_t(\xi)-a)^+\right]-\mathbb{E}\left[(X_0-a)^+\right]}{(b-a)}.
$$

See, for example, Shiryaev (1996) or Stroock (2010) for a more detailed treatment of the Doob's up-crossing inequality and its proof.

# F Proof of Proposition 5.4

*Proof.* Note that Doob's up-crossing inequality above implies that

$$
\mathbb{E}_{\nu,\bar\sigma^e}\left[U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar\sigma^e}(\theta\mid h))\right]\le\frac{\varepsilon}{\frac{\varepsilon}{2}}=2.
$$

But note that

$$
\gamma(\theta,\omega^{k,\beta_1})\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar\sigma^e}(\theta\mid h))\right]\le \mathbb{E}_{\nu,\bar\sigma^e}\left[U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar\sigma^e}(\theta\mid h))\right]\le 2.
$$

Then for every $t$, an application of Markov's inequality implies:

$$
\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar\sigma^e}(\theta\mid h))\ge J\right)\le\frac{1}{\gamma(\theta,\omega^{k,\beta_1})J}\mathbb{E}_{\theta,\sigma^{k,\beta_1}}\left[U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar\sigma^e}(\theta\mid h))\right]\le\frac{2}{\gamma(\theta,\omega^{k,\beta_1})J}.
$$

Finally,

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(U_\infty^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h)) \geq J\right) = \lim_{t\to\infty} \mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h)) \geq J\right) \leq \frac{2}{\gamma(\theta,\omega^{k,\beta_1})J}.$$

□

# G Proof of Proposition 5.5

*Proof.* Choose $h \in \mathcal{D}^{\beta_1}_{\sigma^e}(\theta, 2n\kappa(k), \varepsilon)$. Then by definition,

$$\nu^{\bar{\sigma}^e}\left(\theta \mid h^t\right) \leq 1-\varepsilon \text{ for at least } 2n\kappa(k) \text{ values of } t.$$

Suppose that $h \notin \mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, n, \varepsilon/2)$. Then by the pigeon-hole principle, there must be at least $n$ up-crossings of the belief $\nu^{\bar{\sigma}^e}(\theta \mid h^t)$ from $1-\varepsilon$ to $1-\varepsilon/2$. Therefore,

$$\mathcal{D}^{\beta_1}_{\sigma^e}(\theta, 2n\kappa(k), \varepsilon) \subseteq \left\{h : U^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h)) \geq n\right\} \cup \mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, n, \varepsilon/2).$$

By Corollary 5.3, there exist $k^* \in \mathbb{N}_+$ such that for all $k \geq k^*$, there exists $\lambda > 0$ such that for all $\beta_1 \in \mathcal{B}$ and all $n$,

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, n, \varepsilon/2)\right) \leq -\frac{km \log\left(\gamma(\theta,\omega^{k,\beta_1})\right)}{n\lambda}.$$

This together with Proposition 5.4 implies:

$$\begin{aligned}\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}^{\beta_1}_{\sigma^e}(\theta, 2n\kappa(k), \varepsilon)\right) &\leq \mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(U^{(1-\varepsilon,1-\varepsilon/2)}(\nu^{\bar{\sigma}^e}(\theta \mid h)) \geq n\right) + \mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}^{k,\beta_1}_{\sigma^e}(\theta, n, \varepsilon/2)\right)\\ &\leq \frac{1}{n}\left(\frac{2}{\gamma(\theta,\omega^{k,\beta_1})} - \frac{km\log\left(\gamma(\theta,\omega^{k,\beta_1})\right)}{\lambda}\right).\end{aligned}$$

□

# H Proof of Theorem 6.5

Let us denote the vector of beliefs over all states $\theta \in \Theta$ at time $t$ and history $h^t$ by the following:

$$\nu_{\bar{\sigma}}(h^t) := \left(\nu_{\bar{\sigma}}(\theta_1 \mid h^t), \nu_{\bar{\sigma}}(\theta_2 \mid h^t), \ldots, \nu_{\bar{\sigma}}(\theta_m \mid h^t)\right).$$

Given any vector $x \in \mathbb{R}^m$, let $\|x\|$ denote the Euclidean norm:

$$\|x\|^2 = \sum_{k=1}^m x_k^2.$$

We begin with a couple lemmata.

**Lemma H.1.** *Let $\rho > 0$. Then there exists some $\varepsilon > 0$ such that for all $t$ and $h^t \in H^t$,*

$$\mathbb{E}_{\nu,\bar{\sigma}}\left[\|\nu_{\bar{\sigma}}(h^{t+1}) - \nu_{\bar{\sigma}}(h^t)\| \mid h^t\right] \leq \varepsilon \Longrightarrow \inf_{\beta \in NR(\nu_{\bar{\sigma}}(h^t))} \|\bar{\sigma}(h^t) - \beta\| \leq \rho.$$

*Proof.* Given $\beta \in \mathcal{B}$ and $\nu \in \Delta(\Theta)$, the updated belief after observing $y \in Y$ is given by:

$$\nu_{\beta,\nu}(\cdot \mid y) = \left(\frac{\nu(\theta)\pi(y \mid \theta, \beta(\theta))}{\sum_{\theta' \in \Theta} \nu(\theta')\pi(y \mid \theta', \beta(\theta'))}\right)_{\theta \in \Theta}.$$

Then define the function

$$F(\beta,\nu) := \mathbb{E}_{\nu,\beta}\left[\|\nu_{\beta,\nu}(\cdot \mid y) - \nu\|\right].$$

First note that if $F(\beta,\nu) = 0$ then $\beta \in NR(\nu)$. Now given any $\varepsilon \geq 0$, the set $F(\beta,\nu) \leq \varepsilon$ is compact. Then note that if we define

$$G_\varepsilon := \max_{\{(\beta,\nu): F(\beta,\nu)\leq\varepsilon\}} \|\beta - NR(\nu)\|,$$

$G_\varepsilon$ is continuous in $\varepsilon$ and this proves the claim. □

**Lemma H.2.** *For every $\varepsilon > 0$ there exists $\rho > 0$ such that for all $\nu \in \Delta(\Theta)$ and $\beta \in \mathcal{B}$,*

$$\inf_{\beta' \in NR(\nu)} \|\beta - \beta'\| \leq \rho \Longrightarrow \max_{\alpha_2 \in B_2(\beta,\nu)} \sum_{\theta\in\Theta} \nu(\theta) u_1(\beta(\theta), \alpha_2, \theta) < V(\nu) + \varepsilon.$$

*Proof.* Consider the following function:

$$G_\rho := \max_{\{(\beta,\nu): \|\beta - NR(\nu)\|\leq\rho\}} \max_{\alpha_2\in B_2(\beta,\nu)} \sum_{\theta\in\Theta} \nu(\theta) u_1(\beta(\theta), \alpha_2, \theta).$$

Then $G_\rho$ is continuous in $\rho$ and thus proves the claim. □

**Lemma H.3.** *Let $\varepsilon > 0$. Then given any equilibrium strategy $\sigma$ of the opportunistic type, there exists at most $m/\varepsilon$ times $t$ at which*

$$\mathbb{E}_{\nu,\bar\sigma}\left[\|\nu^{\bar\sigma}(h^{t+1}) - \nu^{\bar\sigma}(h^t)\|^2\right] \geq \varepsilon.$$

*Proof.* First consider any joint random variable $(X, Z)$ such that $\mathbb{E}[X \mid Z] = Z$. Then

$$\begin{aligned}
\mathbb{E}\left[\|X - Z\|^2\right] &= \mathbb{E}\left[\|X\|^2 + \|Z\|^2\right] - 2\mathbb{E}\left[\langle X, Z\rangle\right] \\
&= \mathbb{E}\left[\|X\|^2 + \|Z\|^2\right] - 2\sum_z \mathbb{P}(Z = z)\, \mathbb{E}\left[\langle X, Z\rangle \mid Z = z\right] \\
&= \mathbb{E}\left[\|X\|^2 + \|Z\|^2\right] - 2\sum_z \mathbb{P}(Z = z)\, \|z\|^2 \\
&= \mathbb{E}\left[\|X\|^2 - \|Z\|^2\right].
\end{aligned}$$

The using the above, consider the beliefs at any time $t + 1$:

$$\begin{aligned}
m \geq \mathbb{E}_{\nu,\bar\sigma}\left[\|\nu_{\bar\sigma}(h^{t+1}) - \nu\|^2\right] &= \mathbb{E}_{\nu,\bar\sigma}\left[\|\nu_{\bar\sigma}(h^{t+1})\|^2 - \|\nu\|^2\right] \\
&= \sum_{\tau=0}^{t} \mathbb{E}_{\nu,\bar\sigma}\left[\|\nu_{\bar\sigma}(h^{\tau+1})\|^2 - \|\nu_{\bar\sigma}(h^\tau)\|^2\right] \\
&= \sum_{\tau=0}^{t} \mathbb{E}_{\nu,\bar\sigma}\left[\|\nu_{\bar\sigma}(h^{\tau+1}) - \nu_{\bar\sigma}(h^\tau)\|^2\right].
\end{aligned}$$

This then implies the desired conclusion. □

All that remains is to show that for every $\varepsilon > 0$, there exists some $\delta^* < 1$ such that for all $\delta > \delta^*$ and any equilibrium $(\sigma, \sigma_2)$, the payoff to playing $\bar\sigma$ for the opportunistic type is at most $\mathbf{cav}V(\nu) + \varepsilon$. We demonstrate in the following proof.

*Proof of Theorem 6.5.* By Lemma H.2, there exists some $\rho > 0$ such that

$$\inf_{\beta\in NR(\nu_{\bar\sigma}(h^t))} \|\bar\sigma(h^t) - \beta\| \leq \rho \Longrightarrow \max_{\alpha_2\in B_2(\bar\sigma(h^t),\nu_{\bar\sigma}(h^t))} \sum_{\theta\in\Theta} \nu_{\bar\sigma}(\theta) u_1(\bar\sigma(h^t), \alpha_2, \theta) < V(\nu_{\bar\sigma}(h^t)) + \frac{\varepsilon}{4}.$$

Choose $n \in \mathbb{N}$ such that $\frac{1}{n}(\overline{u} - \underline{u}) < \varepsilon/4$. By Lemma H.3, there are at most $nm/\rho$ times at which

$$\mathbb{E}_{\nu,\bar{\sigma}}\left[\|\nu_{\bar{\sigma}}(h^{t+1}) - \nu_{\bar{\sigma}}(h^t)\|^2\right] \geq \frac{\rho}{n}.$$

Note that for all times $t$ such that $\mathbb{E}_{\nu,\bar{\sigma}}\left[\|\nu_{\bar{\sigma}}(h^{t+1}) - \nu_{\bar{\sigma}}(h^t)\|^2\right] < \frac{\rho}{n}$, then

$$\mathbb{P}_{\nu,\bar{\sigma}}\left[\|\nu_{\bar{\sigma}}(h^{t+1}) - \nu_{\bar{\sigma}}(h^t)\|^2\right] < \frac{1}{n}.$$

Thus at all such times, the expected payoff is at most

$$\frac{1}{n}(\overline{u} - \underline{u}) + \mathbb{E}_{\nu,\bar{\sigma}}\left[V(\nu_{\bar{\sigma}}(h^t)) + \frac{\varepsilon}{4}\right] \leq \mathbf{cav}V(\nu) + \frac{\varepsilon}{2}.$$

Thus the most that a player could obtain from playing $\bar{\sigma}$ is:

$$\left(1 - \delta^{\frac{nm}{\rho}}\right)\overline{u} + \delta^{\frac{nm}{\rho}}\left(\mathbf{cav}V(\nu) + \frac{\varepsilon}{2}\right).$$

Then we can choose $\delta^* < 1$ such that for all $\delta > \delta^*$,

$$\left(1 - \delta^{\frac{nm}{\rho}}\right)\overline{u} + \delta^{\frac{nm}{\rho}}\left(\mathbf{cav}V(\nu) + \frac{\varepsilon}{2}\right) < \mathbf{cav}V(\nu) + \varepsilon.$$

This concludes the proof. □

# I Proof of Claim 6.3

*Proof.* To simplify notation, let us denote by $(x, y) \in [0,1]^2$ the state-contingent strategy $\beta \in \mathcal{B}$ in which $T$ is played with probability $x$ in state $\ell$ and $T$ is played with probability $y$ in state $r$.

Given the information structure in that example, for any $\nu \in (0,1)$ representing the probability distribution over states in which $\ell$ occurs with probability $\nu$, the set of non-revealing strategies $NR(\nu)$ is given by:

$$NR(\nu) = \left\{\left(x, \frac{2}{3} - x\right) : x \in [0, 2/3]\right\}.$$

We want to bound $V(\nu)$. To do this, we consider four cases.

**Case 1:** $\nu \geq 3/4$

Given a non-revealing strategy $(x, 2/3 - x) \in NR(\nu)$, the SR player believes that $T$ will be played with probability:

$$\nu x + (1-\nu)(2/3 - x) = (2\nu - 1)x + \frac{2}{3}(1-\nu).$$

Thus $L$ is a SR player best-response against $(x, 2/3 - x)$ given belief $\nu$ if and only if

$$(2\nu - 1)x + \frac{2}{3}(1-\nu) \geq 1/2 \Leftrightarrow x \geq \frac{1}{3}\frac{2\nu - 1/2}{2\nu - 1}.$$

Therefore,

$$\begin{aligned} V(\nu) &\leq \max\left\{\max_{x \geq \frac{1}{3}\frac{2\nu-1/2}{2\nu-1}}\left(\nu x + (1-\nu)\left(\frac{2}{3} - x\right)\right) + 2\left(1 - \nu x - (1-\nu)\left(\frac{2}{3} - x\right)\right), 0\right\} \\ &= \max\left\{\max_{x \in \left[\frac{1}{3}\frac{2\nu-1/2}{2\nu-1}, \frac{2}{3}\right]} 2 + (1-2\nu)x - (1-\nu)\frac{2}{3}, 0\right\} = \frac{3}{2}. \end{aligned}$$

**Case 2:** $x \in (1/4, 3/4)$

If $\nu \in [1/2, 3/4)$, $L$ is a SR player best-response against $(x, 2/3 - x)$ given prior $\nu$ if and only if

$$(2\nu - 1)x + \frac{2}{3}(1 - \nu) \geq 1/2 \Leftrightarrow x \geq \frac{1}{3}\frac{2\nu - 1/2}{2\nu - 1}.$$

But the latter is strictly greater than 2/3 for all $\nu \in [1/2, 3/4)$. Thus for all $(x, 2/3 - x) \in NR(\nu)$, the SR player's best response at belief $\nu \in [1/2, 3/4)$ is $R$.

On the other hand, if $\nu \in (1/4, 1/2)$, $L$ is a SR player best-response against $(x, 2/3 - x)$ given prior $\nu$ if and only if

$$(2\nu - 1)x + \frac{2}{3}(1 - \nu) \geq 1/2 \Leftrightarrow x \leq \frac{1}{3}\frac{2\nu - 1/2}{2\nu - 1}.$$

Again the last term is strictly negative for all $\nu \in (1/4, 1/2)$ and hence, for all $(x, 2/3 - x) \in NR(\nu)$, the SR player's best-response at belief $\nu \in (1/2, 3/4)$ is again $R$. This then shows that for all $\nu \in (1/4, 3/4)$, $V(\nu) \leq 0$.

**Case 3:** $\nu \leq 1/4$

This case is symmetric to Case 1. Note that $L$ is a SR player best-response against $(x, 2/3 - x)$ given belief $\nu$ if and only if

$$(2\nu - 1)x + \frac{2}{3}(1 - \nu) \geq \frac{1}{2} \Leftrightarrow x \leq \frac{1}{3}\frac{2\nu - 1/2}{2\nu - 1}.$$

Therefore

$$\begin{aligned} V(\nu) &\leq \max\left\{\max_{0 \leq x \leq \frac{1}{3}\frac{2\nu - 1/2}{2\nu - 1}} \left(\nu x + (1 - \nu)\left(\frac{2}{3} - x\right)\right) + 2\left(1 - \nu x - (1 - \nu)\left(\frac{2}{3} - x\right)\right), 0\right\} \\ &= \max\left\{\max_{x \in \left[0, \frac{1}{3}\frac{2\nu - 1/2}{2\nu - 1}\right]} 2 + (1 - 2\nu)x - (1 - \nu)\frac{2}{3}, 0\right\} = \frac{3}{2}. \end{aligned}$$

Given the above bounds for $V$, we arrive at the conclusion of Claim 6.3 by applying Theorem 6.5: for every $\varepsilon > 0$, there exists $\rho^* > 0$ such that for all $\mu < \rho^*$, there exists some $\delta^*$ such that for all $\delta > \delta^*$, in all equilibria, the (opportunistic) LR player obtains an ex-ante payoff of at most $3/2 + \varepsilon$. □

# J Proof of Corollary 6.6

*Proof.* The lower bound is a consequence of Theorem 4.1. Let us now show the upper bound. Note that by assumption,

$$\mathbf{cav}V(\nu) = \sum_{\theta \in \Theta} \nu(\theta) u_1^*(\theta).$$

Let $\underline{\nu} = \min_{\theta \in \Theta} \nu(\theta)$.

By Theorem 6.5, there exists some $\rho^* > 0$ such that whenever $\mu(\Omega^c) < \rho^*$, there exists some $\delta^* < 1$ such that for all $\delta > \delta^*$ and all equilibrium strategy profiles $(\sigma_1, \sigma_2)$,

$$\sum_{\theta \in \Theta} \nu(\theta) U_1(\sigma_1, \sigma_2, \theta, \delta) < \mathbf{cav}V(\nu) + \frac{\underline{\nu}}{2}\varepsilon.$$

Suppose by way of contradiction that there exists some state $\theta^* \in \Theta$ and some sequence $\delta_n \to 1$ such that there exists some $(\sigma_1^n, \sigma_2^n) \in \Sigma_1^e \times \Sigma_2^e$ such that for all $n$, $U_1(\sigma_1^n, \sigma_2^n, \theta^*, \delta_n) \geq u_1^*(\theta^*) + \varepsilon$. Then note that for all $n$,

$$\nu(\theta^*)(u_1^*(\theta^*) + \varepsilon) + \sum_{\theta \neq \theta^*} \nu(\theta) U_1(\sigma_1^n, \sigma_2^n, \theta, \delta_n) < \sum_{\theta \in \Theta} \nu(\theta) u_1^*(\theta) + \frac{\underline{\nu}}{2}\varepsilon.$$

This then implies that for all $n$,

$$\sum_{\theta \neq \theta^*} \nu(\theta) U_1(\sigma_1^n, \sigma_2^n, \theta, \delta_n) < \sum_{\theta \neq \theta^*} \nu(\theta) u_1^*(\theta) - \left(\nu(\theta^*) - \frac{\nu}{2}\right) \varepsilon < \sum_{\theta \neq \theta^*} \nu(\theta) \left( u_1^*(\theta) - \frac{\nu}{2\nu(\theta)(m-1)} \varepsilon \right).$$

Then for each $n$, we can find some $\theta_n$ such that

$$U_1(\sigma_1^n, \sigma_2^n, \theta_n, \delta_n) < u_1^*(\theta_n) - \frac{\nu}{2\nu(\theta_n)(m-1)} \varepsilon.$$

Because there are only finitely many states $\theta \in \Theta$, there exists some $\theta \neq \theta^*$ and a subsequence $n_k$ such that for all $k$,

$$U_1(\sigma_1^{n_k}, \sigma_2^{n_k}, \theta, \delta_{n_k}) < u_1^*(\theta) - \frac{\nu}{2\nu(\theta)(m-1)} \varepsilon.$$

This contradicts the lower bound theorem, concluding the proof. □

# References


Aoyagi, M. (1996): "Reputation and Dynamic Stackelberg Leadership in Infinitely Repeated Games," *Journal of Economic Theory*, 71(2), 378–393.

Atakan, A. E., and M. Ekmekci (2011): "Reputation in Long-Run Relationships," *The Review of Economic Studies*, p. rdr037.

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

Aumann, R. J., M. Maschler, and R. E. Stearns (1995): *Repeated Games with Incomplete Information*. MIT press.

Celentani, M., D. Fudenberg, D. K. Levine, and W. Pesendorfer (1996): "Maintaining a Reputation against a Long-Lived Opponent," *Econometrica*, 64(3), 691–704.

Cripps, M. W., E. Dekel, and W. Pesendorfer (2004): "Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests," *Journal of Economic Theory*, 121(2), 259–272.

Curtin, P. A. (2002): *The Advertising Age Encyclopedia of Advertising*chap. Environmental Movement. Fitzroy Dearborn Publishers.

Ely, J. C., D. Fudenberg, and D. K. Levine (2008): "When is Reputation Bad?," *Games and Economic Behaivor*, 63(2), 498–526, Northwestern University, Harvard and University of California at Los Angeles.

Evans, R., and J. P. Thomas (1997): "Reputation and Experimentation in Repeated Games with Two Long-Run Players," *Econometrica*, 65(5), 1153–1173.

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(3), 561–579.

Fudenberg, D., and Y. Yamamoto (2010): "Repeated Games where the Payoffs and Monitoring Structure are Unknown," *Econometrica*, 78(5), 1673–1710.

Ghosh, S. (2014): "Multiple Long-lived Opponents and the Limits of Reputation," Mimeo.

Gossner, O. (2011): "Simple Bounds on the Value of a Reputation," *Econometrica*, 79(5), 1627–1641.

Hörner, J., and S. Lovo (2009): "Belief-Free Equilibria in Games With Incomplete Information," *Econometrica*, 77(2), 453–487.

Hörner, J., S. Lovo, and T. Tomala (2011): "Belief-free equilibria in games with incomplete information: Characterization and existence," *Journal of Economic Theory*, 146(5), 1770–1795.

Kreps, D. M., and R. J. Wilson (1982): "Reputation and Imperfect Information," *Journal of Economic Theory*, 27(2), 253–279.

Mertens, J.-F., S. Sorin, and S. Zamir (2014): *Repeated Games*, vol. 55. Cambridge University Press.

Milgrom, P. R., and J. Roberts (1982): "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, 27(2), 280–312.

Schmidt, K. M. (1993): "Reputation and Equilibrium Characterization in Repeated Games of Conflicting Interests," *Econometrica*, 61(2), 325–351.

Shiryaev, A. N. (1996): *Probability*. Springer-Verlag, New York, second edn.

Sorin, S. (1999): "Merging, Reputation, and Repeated Games with Incomplete Information," *Games and Economic Behavior*, 29(1/2), 274–308.

Stroock, D. W. (2010): *Probability Theory: an Analytic View*. Cambridge university press.

Wiseman, T. (2005): "A Partial Folk Theorem for Games with Unknown Payoff Distributions," *Econometrica*, 73(2), 629–645.