Reputation Building under Uncertain Monitoring

Joyee Deb Yuhta Ishii

February 21, 2016

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 uncertain state, which determines the monitoring structure in the reputation game. The long-run player plays a stage-game repeatedly against a sequence of short run opponents. We present necessary and sufficient conditions (on the monitoring structure and 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 and highly sensitive to the inclusion of other commitment types. However with the inclusion of appropriate dynamic commitment types, reputation building can again be sustained while maintaining robustness to the inclusion of other arbitrary types.

1 Introduction

Consider a long-run firm that wants to build a reputation as a high-quality producer. Consumers make purchase decisions based on the product review of an individual reviewer but do not know exactly how to interpret these reviews. For instance, consumers face uncertainty to the degree of the bias of the review: a positive review may either signal good or bad quality given the uncertainty regarding the degree of correlation between the reviewer and consumer's tastes. Faced with such uncertainty, the firm, even through honest effort, finds it difficult to establish a positive reputation among the consumers. 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 decide whether to contribute based on information provided by the public media about the candidate's work. Again she faces uncertainty about the degree of bias of the media source and thus cannot tell if positive news is truly indicative of high effort by the representative. In both of these examples, the non-reputation builder faces *persistent uncertainty* (correlation in tastes between reviewer and consumer/bias of the media) regarding the monitoring of the actions of the reputation builder. The central question of this paper is whether reputations can be built in such environments where monitoring suffers from persistent uncertainty.

To start, let us first discuss reputation building in environments without any 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 long-run's player's type / payoffs: consumers entertain the possibility that the firm is of a "commitment" type that is committed to playing a particular action at every stage of the game. Even when the actions of the firm are noisily observed, the classical reputation building result states that if a sufficiently rich set of commitment types occur with positive probability,[1] then, a patient firm can achieve payoffs arbitrarily close

[1] this probability can be arbitrarily small

to his Stackelberg payoff of the stage game in *every* equilibrium.[2] Importantly, this result is robust to the introduction other arbitrary commitment types. The result holds as this incomplete information introduces a connection between the firm's past behavior and the expectations of its future behavior. In particular, by mimicking a commitment type that always plays the Stackelberg action, the firm can signal to the consumer its intention to play the Stackelberg action in the future and thus obtain high payoffs in *any* equilibrium.

Of course this intuition critically relies on the consumer's ability to accurately interpret the noisy signals. On the other hand, as in the examples, if there is persistent uncertainty regarding the mapping from actions to distributions over signals, the reputation builder finds it far more difficult to establish a link between past outcomes and expectations of future behavior. To study this issue, we consider the canonical model of reputation building, with but one key difference. At the start, a state of the world is realized, which determines both the type of the firm and the monitoring structure: a mapping from actions taken by the firm to distribution of signals observed by the consumer. We assume for simplicity that the firm knows the state of the world, but the consumer does not.[3] Since the consumer is uncertain about the monitoring structure, she may not make the correct inferences about the firm's behavior. We first show that uncertain monitoring can cause reputation building to break down. In particular, we present a simple example to show that even if the firm repeatedly takes the desirable Stackelberg action to generate positive signals in one state, there exist equilibria in which these signals become confounded with low signals in another state, leading to consumers' expectations of a different undesirable action. As a result, in equilibrium, the firm obtains low payoffs even if he is arbitrarily patient. The example is quite general, in that reputation building is not effective even if actions are identifiable in each state, and even with a rich set of stationary commitment types. This leads us then to ask what might restore reputation building under such uncertain monitoring.

In the main result of the paper, we provide sufficient conditions on the monitoring structure and a set of commitment types such that reputation can be sustained for sufficiently patient long-run player even when her opponents are uncertain about the monitoring environment. Importantly, this result is robust to the inclusion of other arbitrary commitment types, and thus independent of the fine details of the type space. As observed in the previous paragraph, effective reputation building requires the presence of types that are committed to switching 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, alternating between signaling phases and collection phases *infinitely often*. Without such infinite switches, we show that the reputation result is no longer robust to the inclusion of other arbitrary commitment types.

While the construction is subtle, the broad intuition is that since the uncertainty in monitoring confounds the consumer's ability to interpret the outcomes she observes, reputation building is generally possible only if the firm can teach the consumer about the monitoring state and then play the desirable Stackelberg action. Dynamic commitment types are necessary because "signaling" the state and Stackelberg payoff collection may necessitate the use of different actions. To interpret our result in the context of the motivating examples, if consumer purchase decisions can only be influenced though product reviews and the consumer does not know enough to be able to interpret reviews, a firm cannot build reputation for high quality by simply investing effort into producing high quality products. Rather effective reputation building requires *both* repeated investment in credibly conveying to the consumer the meaning of the product reviews in conjunction with

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

[3] This is in a sense, the easiest such environment in which the firm could hope to establish a reputation. Instead if the firm was also uncertain about the state, then the firm would also have to conduct learning.

the production of high quality products.

How does the presence of these commitment types that both signal and collect enable reputation building? The underlying argument proceeds as follows. First, we show that if there were an appropriate commitment type that could alternate between signaling and collection forever, then by mimicking this type, the long-run player can teach her opponent the true (monitoring) state of the world in all equilibria. Notice that this is not obvious. First we need to establish that the long-run agent can convince her opponent true state via the signaling phases. To do this, we need assumptions on the monitoring structure. However, 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 may be concerned that opponent's learning about the state can be confounded between signaling phases and that the belief on the true state may drop low again during an intervening collection phase. We use Doob's upcrossing inequality for martingales to bound the number of times the belief on the true state can be high and then drop below later. Notice that if the short-run players place high probability on the true state most of the time, then we are done. The long-run player can then play the Stackelberg action for the true state often and earn payoffs arbitrarily close to the Stackelberg payoff. We use the merging arguments à la Gossner (2011) to obtain a lower bound on equilibrium payoffs.

It is worth highlighting an important but subtle issue. Our dynamic commitment type returns to the signaling phase infinitely often. It is intuitive that one of the key ingredients to reputation building under uncertain monitoring is that the short-run player be taught the true state, so that she can make correct inferences from past outcomes about future behavior. Therefore, the presence a commitment type that engages in a signaling phase is not surprising. 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 only at the start (for any arbitrarily long period of time), we construct an example to show that reputation building fails: there exist equilibria in which the long-run player obtains a payoff that is substantially lower than the Stackelberg payoff. In particular, with commitment types whose signaling phases are front-loaded, the lower bound on the long-run player's payoffs is sensitive to the fine details of the distribution of commitment types. As a result, reputation building is no longer robust to the inclusion of other arbitrary commitment types.

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 knows which product is better for the consumer. Is it possible for the firm to build a reputation for making the better product? Note that our results apply here. Mimicking the 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.

Finally, so far we have restricted the discussion to a lower bound on the long-run agent's equilibrium payoff. Of course the immediate question that arises is whether the long-run player can indeed do much better 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.[4] A detailed discussion of an upper bound is outside the scope of this paper. Neverthelss, we provide a joint sufficient condition on the monitoring structure and stage game payoffs that ensure that the lower bound and upper bound coincide for any specification of the type space: Loosely speaking, these are games in which state revelation is desirable.

1.1 Related Literature

There is a vast literature on reputation effects which include 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 also some recent work on uncertainty in payoffs in non-zero sum repeated games by Wiseman (2005), Hörner and Lovo (2009), Hörner, 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 he/she 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 monitoring structure. In contrast, our paper will generally study equilibria where the reputation builder potentially gains drastically from changing the beliefs of the opponent about the monitoring structure. In fact, the possibility of such manipulation will be crucial for our results.

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 arises for very different reasons in the literature on reputation building against long-lived, patient opponents.[5] 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

[4]This is in sharp contrast to the previous papers in the literature where the payoff upper bound is generally independent of the type-space.

[5]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).

leads to high payoffs.[6] In contrast, our non-reputation players are purely myopic and so the threat of punishments has no influence on these players. Rather in our paper, dynamic commitment types are necessary 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 so 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.

2 Model

A long run (LR) player, player 1, faces a sequence of short-lived player 2's. 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 Θ is finite and enumerate $\Theta = \{\theta_1, \ldots, \theta_m\}$, but that Ω may possibly be countably infinite. The realized pair of state and type (θ, ω) is then fixed for the entirety of the game.

In each period $t = 0, 1, 2, \ldots$, 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. 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 $\pi(y^t \mid \theta, a_1^t)$.[7] Note importantly that both the action chosen at time t and the state of the world θ potentially affect the signal distribution. 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 1 (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 $H_1^t := \Theta \times A_1^t \times Y^t$.[8] Meanwhile the short-lived player 2 at time t observes only the public signals up to time t and so his information coincides exactly with the public history $H_2^t := H^t$.[9] 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 Σ_i. Finally let us denote by $\mathcal{A} = \Delta(A_1)$ the set of mixed actions of player 1 with typical element α_1 and let $\mathcal{B}$ be the set of static state contingent mixed actions, $\mathcal{B} := \mathcal{A}^m$ with typical element β_1.

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

[7] Note that the public signal distribution is only affected by the action of player 1.

[8] We believe that it is a straightforward extension to consider a LR player who must learn the state over time.

[9] Observability or lack thereof of previous SR player's actions do not affect our results.

2.1 Type Space

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

2.2 Payoffs

The payoffs 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 payoffs of the LR opportunistic player 1 is given by:

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

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 θ. The Stackelberg payoff of player 1 in state θ is then given by:

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

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 Distributions over Public Histories

First note that any $\sigma \in \Sigma_1$ and a prior $\nu \in \Delta(\Theta)$ 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 θ.

Furthermore given that type ω^o chooses a strategy $\sigma \in \Sigma_1$, we define $\bar{\sigma} \in \Delta(\Sigma_1)$ to be a mixed strategy that randomizes over the strategies played by the types in Ω according to the respective probabilities:

$$\bar{\sigma}(\sigma) = \mu(\omega^o), \bar{\sigma}(\sigma^\omega) = \mu(\omega) \ \forall \omega \in \Omega^c.$$

$\bar{\sigma}$ is essentially the aggregate strategy of the LR player that the SR players face when the opportunistic type chooses to play σ. Of course, $\bar{\sigma}$ is outcome equivalent to a unique behavioral strategy in Σ_1 and so, with the abuse of notation, henceforth, $\bar{\sigma}$ will refer to this unique behavioral strategy.

Similar probability measures can be defined at any history $h_t \in H$

$$\phi^{\ell}_{\bar{\sigma}}(\cdot \mid h_t) \in \Delta(Y^{\ell})$$

for all *public histories* $h_t \in H^t$, where ϕ^{ℓ} represents the probability distribution over the next ℓ-periods' public signals given the public history h_t, when we aggregate the uncertainty about the state θ and type Ω.

We can define a similar expression conditional on the state being θ:

$$\phi^{\ell}_{\theta,\bar{\sigma}}(\cdot \mid h_t) \in \Delta(Y^{\ell})$$

by conditioning on a certain state. Additionally, given any $\omega \in \Omega^c$, we also obtain the similarly defined measures conditioning on the commitment type ω:

$$\phi^{\ell}_{\omega}(\cdot \mid h_t), \phi^{\ell}_{\theta,\omega}(\cdot \mid h_t) \in \Delta(Y^{\ell})$$

Moreover note that

Finally, given $\bar{\sigma} : \Omega \to \Sigma_1$, the beliefs of the SR player at any public history $h_t \in H_t$ concerning the state and type are pinned down:

$$\mu^{\bar{\sigma}}(\cdot \mid h_t) \in \Delta(\Theta \times \Omega).$$

As with the prior, we denote by $\mu^{\bar{\sigma}}_{\Theta}(\cdot \mid h_t)$ and $\mu^{\bar{\sigma}}_{\Omega}(\cdot \mid h_t)$ the respective marginals.

2.4 Key Definitions Regarding the Signal Structure

Definition 2.1. A signal structure π holds action identification for $(\alpha_1, \theta) \in \Delta A_1 \times \Theta$ if

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

Using the above definition, we impose the following assumptions on the information structure for the remainder of the paper.

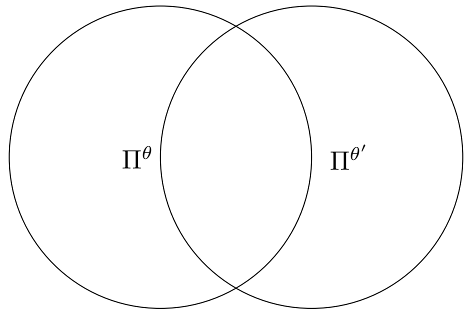
Assumption 2.2. π has action independence for (α_1, θ) for all $\theta \in \Theta$ and some $\alpha_1 \in \alpha_1^*(\theta)$.

Assumption 2.3. For every $\theta, \theta' \in \Theta$ and $\theta' \neq \theta$, there exists some $\alpha_1(\theta, \theta') \in \Delta(A_1)$ such that

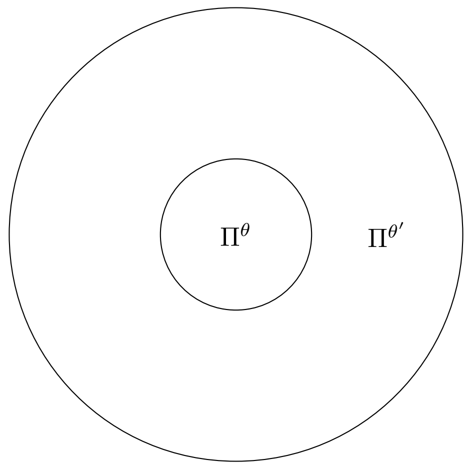
$$\pi\left(\cdot \mid \alpha_1(\theta, \theta'), \theta\right) \neq \pi\left(\cdot \mid \alpha_1, \theta'\right)$$

for all $\alpha_1(\theta') \in \Delta(A_1)$.

The first assumption is a straightforward extension of conditions required of reputation building without uncertainty about the state. The second is new. First note that the above *does not* assume that $\alpha_1^{\theta,\theta'}$ must be the Stackelberg action in state θ. It can be any arbitrary (possibly mixed) action. Graphically, we can represent the assumption above as follows. For each θ, denote by Π^{θ} the set of all probability distributions in $\Delta(Y)$ that are spanned by possibly mixed actions in $\Delta(A_1)$. Note that each point in Π^{θ} is a *probability distribution* and *not* an element of Y. If for each pair, neither $\Pi^{\theta} \subseteq \Pi^{\theta'}$ nor $\Pi^{\theta'} \subseteq \Pi^{\theta}$, then the assumption holds:



The above is violated if there exists a pair of states in which $\Pi^\theta \subseteq \Pi^{\theta'}$ as in the following figure.



Finally 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. Finally our analysis will focus on perfect Bayesian equilibria and to shorten exposition, subsequently we will refer to perfect Bayesian equilibrium as simply equilibrium.

3 An Illustrative Example

We start with a simple example to illustrate that uncertainty in monitoring can have drastic effects on reputation building. Consider the following three action game between a long-ruin player 1 (row player) and a sequence of short-run opponents (the column player). First note that the Stackelberg payoff is 2 and the

	ℓ	r
A	$2, 1$	$-1, 0$
B	$-1, -1$	0,0
C	-100, 0	-100, 0

Figure 1: Stage Game

Stackelberg action is A. Secondly note that in the stage game, ℓ is a best response to player 1's stage game mixed action if and only if $\alpha_1(A) \geq \alpha_1(B)$. Suppose that there are two states $\Theta = \{L, R\}$, which affect the distribution of signals observed by player 2. There are four public signals:

$$Y = \{\bar{c}, \bar{y}, \underline{y}, \underline{c}\}.$$

The information structure is as follows:

$$\pi(\bar{c} \mid L, C) = 1$$
$$\pi(\underline{c} \mid R, C) = 1$$
$$\pi(\bar{y} \mid R, A) = \pi(\bar{y} \mid L, B) = \frac{3}{4}$$
$$\pi(\underline{y} \mid R, A) = \pi(\underline{y} \mid L, B) = \frac{1}{4}$$
$$\pi(\bar{y} \mid R, B) = \pi(\bar{y} \mid L, A) = \frac{1}{4}$$
$$\pi(\underline{y} \mid R, B) = \pi(\underline{y} \mid L, A) = \frac{3}{4}.$$

First note that conditional on any state, actions are always statistically identified. This means that if there were no uncertainty about states, classical reputation results would hold: i.e., if the true state were known, and there was a positive probability that the long-run player could be a commitment type that always plays A, then a patient long-run player would get a payoff arbitrarily close to 2 in *every* equilibrium.

What we will show below is that this is no longer true when there is uncertainty about the monitoring states. Reputation building cannot be guaranteed if we only have access to simple commitment types. We will construct an equilibrium in which the long-run player 1 gets a payoff close to 0. Consider the following strategy in which types play according to the following chart at all histories:

	ω^c	ω^o
$\theta = R$	A	B
$\theta = L$	A	B

Table 1: Strategy of Player 1

Furthermore suppose that the SR player always plays r at all histories. We show that when $\mu_\Omega(\omega^c) > 0$ is sufficiently small, the above is a perfect Bayesian equilibrium for all $\delta \in (0,1)$. We let the measure over the type and state to be a product measure in which $\mu_\Omega(\omega^c) = \nu$ and $\mu_\Theta(R) = p$.

To this end, first consider the updating rule of player 2. We must keep track of the probability that player 2 assigns to the commitment type. Note that this is a sufficient statistic for his decision given the candidate strategy played by player 1. At any time t, conditional on a history h_t, player 2 has beliefs given by:

	ω^c	ω^o
$\theta = R$	μ^t_{RA}	μ^t_{RB}
$\theta = L$	μ^t_{LA}	$\mu^t_{\ell LB}$

Table 2: Joint Distribution over (θ, ω) pairs

Consider the following likelihood ratios:

$$\frac{\mu_{RA}^{t+1}}{\mu_{LB}^{t+1}} = \frac{\mu_{RA}^{t}}{\mu_{LB}^{t}} \frac{\pi(y_t \mid R, A)}{\pi(y_t \mid L, B)} = \frac{\mu_{RA}^{t}}{\mu_{LB}^{t}}$$
$$\frac{\mu_{RB}^{t+1}}{\mu_{LA}^{t+1}} = \frac{\mu_{RB}^{t}}{\mu_{LA}^{t}} \frac{\pi(y_t \mid R, B)}{\pi(y_t \mid L, A)} = \frac{\mu_{RB}^{t}}{\mu_{LA}^{t}}.$$

Thus regardless of the history,

$$\alpha := \frac{p\nu}{(1-p)(1-\nu)} = \frac{\mu_{RA}^{0}}{\mu_{LB}^{0}} = \frac{\mu_{RA}^{t}}{\mu_{LB}^{t}},$$
$$\beta := \frac{(1-p)\nu}{p(1-\nu)} = \frac{\mu_{LA}^{0}}{\mu_{RB}^{0}} = \frac{\mu_{LA}^{t}}{\mu_{RB}^{t}}.$$

Define:

$$\mu_s^t = \mu_{RA}^t + \mu_{LB}^t$$
$$\mu_d^t = \mu_{RB}^t + \mu_{LA}^t.$$

Then note that
$$\mu^t(\omega^c) = \frac{\alpha}{1+\alpha}\mu_s^t + \frac{\beta}{1+\beta}\mu_d^t.$$

Now note that given any $p \in (0,1)$, there exists some $\nu^* > 0$ such that for all $\nu < \nu^*$, $\alpha, \beta < \frac{1}{2}$. Then note that for all $\nu < \nu^*$,
$$\mu^t(\omega^c) < \frac{1}{2}(\mu_s^t + \mu_d^t) = \frac{1}{2} \Longrightarrow \mu^t(\omega^c) < \mu^t(\omega^o).$$
Therefore we have shown that for all $\nu < \nu^*$, player 2's best response is to play r at all histories (that are consistent with the play of ω^c and ω^o). Furthermore at any history in which either $\bar{c}$ or $\underline{c}$ arise, we can specify that the SR player places a belief of probability one on the rational type. Thus at any such history, it is a best response for the SR player to play r. Therefore, it is incentive compatible for the SR player to always play r. Since there are no inter-temporal incentives of the LR player, it is also incentive compatible for the LR player to always play B (regardless of the discount factor). Of course, this example runs contrary to the reputation results of the previous literature because of the additional problems that non-identification of the Stackelberg action pose.

3.1 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 R, such examples exist, which seems to suggest a discontinuity at $p = 1$. However this seeming discontinuity arises because $\nu^* > 0$ necessarily becomes vanishingly small as $p \to 1$. This highlights the observation that with access to only simple commitment types, whether one can guarantee Stackelberg payoffs or not depends highly 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 θ. This is in contrast to the previous literature on reputation building where relative probabilities did not matter.

Additionally observe that in the example above, there were no types who played C. However suppose that we included a type that played C for a single period in both states and then subsequently switched to play A forever. The inclusion of such a type then would rule out the "bad equilibrium" constructed above. It is no longer possible for the LR player to play B always since by mimicking this 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 building a reputation to play A. 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 satisfy Assumptions 1 and 2. However the generalization will have to deal with some additional complexities since assumptions 1 and 2 do not rule out information structures in which all signals are possible (full support information structures) in all state, action pairs.

Note first that when the signal structure has full support, learning about the state is never immediate. More importantly however, it is usually not possible to convince the SR players with *certainty* about a state. Therefore there is the difficulty that even after having convinced the SR players to a high level of certainty about the correct state, it is typically possible that the belief about the state dips to a low belief subsequently. We deal with these issues by introducing dynamic commitment types who signal the state in a periodic and recurrent fashion. We present a more detailed discussion of the nuances of the problem at hand in Section 4.

3.1.1 Robustness

Finally one of the main assumptions behind our negative example was that $\pi(\cdot \mid R, A) = \pi(\cdot \mid L, B)$ and $\pi(\cdot \mid R, B) = \pi(\cdot \mid L, A)$. 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 LR rational player equilibrium action (B) in state θ and the action of the commitment type (A) in state $-\theta$.[10]

However such examples generalize if we allow for the inclusion of arbitrary "bad" commitment types (as long as the information structure has two states at which Π_θ and $\Pi_{\theta'}$ overlap). In the following section, we will construct appropriate commitment types to show that the *mere presence* of such types is sufficient to establish effective reputation even if we allow for the inclusion of arbitrary other "bad" commitment types.

4 A Reputation Theorem

Let ω^* be a commitment type that always plays a strategy σ^* and let $\mathcal{G}$ be the set of type spaces (Ω, μ) such that $\omega^* \in \Omega$ and $\mu(\omega^*) > 0$. Virtually all reputation theorems in the existing literature have the following structure. For every $(\Omega, \mu) \in \mathcal{G}$ and every $\varepsilon > 0$, there exists δ^* such that whenever $\delta > \delta^*$, the long run player receives payoffs within ε of the Stackelberg payoff in all Nash equilibria. In short, the fine details of the type space beyond the mere fact that ω^* exists with positive probability in the belief space of the short run players do not matter for reputation building.

Here we ask the following question in the spirit of the reputation theorems proved in the literature thus far: 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 show that with "simple" commitment types this is in many cases impossible. By introducing dynamic

[10]In binary examples, we always denote by $-\theta$, the state that is complementary to state θ.

(time-dependent but not history dependent) commitment types, reputation building is salvaged.[11]

4.1 Statewise Commitment Type

We first construct the appropriate commitment types. For every $k > m-1$ and $\beta_1 \in \mathcal{B}$, we now define the following commitment type, ω^{k,β_1}, who plays the (possibly dynamic) strategy $\sigma^{k,\beta_1} \in \Sigma_1$ in every play of the game. We define this strategy as follows, which depends only on calendar time:

$$\sigma_\tau^{k,\beta_1}(\theta) = \begin{cases} \beta_1(\theta) & \text{if } \tau \mod k > m-1, \\ \alpha_1(\theta, \theta_{\tau \mod k}) & \text{if } \tau \mod k \leq m-1. \end{cases}$$

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, he plays a sequence of mixed actions $\{\alpha_1(\theta, \theta_\ell)\}_{\ell=1}^{m-1}$ which will be used to signal the state θ to the SR players. We call this phase of the commitment type's strategy, the *signaling phase*. Subsequently, the 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.

4.2 Reputation Theorem

We are now equipped to establish 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: A sufficiently patient long-run player will get equilibrium payoffs arbitrarily close to the maximum possible payoff of the complete information game, in every equilibrium of the incomplete information game.

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

Our illustrative example already suggested that reputation building is impossible in general 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 consumer's ability to interpret the outcomes she observes, reputation building is possible only if the firm can both teach the consumer about the monitoring state and also play the desirable Stackelberg action. To go back to the motivating examples, our result means that if consumer purchase decisions can only be influenced though product reviews and the consumer does not know enough to be able to interpret reviews, a firm cannot build reputation for high quality by simply investing effort into producing high quality products. Rather, it needs to repeatedly invest effort to credibly convey to the consumer the meaning of the product reviews, and then invest effort in producing high quality, so that a subsequent good review convinces the consumer about the type of the firm.

The commitment types that we constructed above are able to do exactly this: They are committed to playing both "signaling actions" that can help the consumer learn the unknown monitoring state and

[11] As pointed out by Johannes Hörner, there is a way to modify all of our constructions using stationary types that use a public randomization device. The distinction under this interpretation is that the public randomization device must be used in order to effectively build reputation using stationary types.

"collection actions" that are desirable for payoffs. 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.

4.3 Examples: Characteristics of Commitment Types

Note that our commitment types ω^{k,β_1} share an important feature: the type switches play between signalling and collection phases infinitely often. In this subsection, we show the importance of both i) the existence of switches between signalling and collection phases in at least some commitment types and ii) the recurrent nature of the signalling phases. To highlight i), we construct an equilibrium in an example in which the long run player regardless of his discount factor obtains low payoffs if all commitment types play stationary strategies. To highlight the importance of ii), we consider type spaces in which all commitment types play strategies that front-load the signalling phases. In such cases, we construct equilibria (for all discount factors) in which the opportunistic long run player gets payoff substantially below the statewise Stackelberg payoff in all states.

4.3.1 Stationary Commitment Types

We show here that in the following stage game that regardless of a given set of commitment types, Ω^I, that is countable (possibly infinite), we can construct a set of commitment types $\Omega^c \supseteq \Omega^*$ and a probability measure μ over $\Omega^c \cup \{\omega^o\}$ such that there exists an equilibrium in which the long run opportunistic type obtains payoffs significantly below the statewise Stackelberg payoff.[12]

Consider the following stage game that is state independent: The Stackelberg payoff is 3 and the Stack-

	L	R
A	$3,1$	$0,0$
B	$0,0$	1,3

Figure 2: Stage Game

elberg action is A. Note that L is a best response in the stage game if and only if $\alpha_1(A) \geq \frac{3}{4}$. The set of states is binary, $\Theta = \{\ell, r\}$ with equal likelihood of both states. The signal space is binary, $Y = \{\overline{y}, \underline{y}\}$, together with the following information structure:

$$\pi(\overline{y} \mid A, \ell) = \frac{1}{3} < \pi(\overline{y} \mid B, \ell) = \frac{5}{6},$$
$$\pi(\overline{y} \mid A, r) = \frac{2}{3} > \pi(\overline{y} \mid B, r) = \frac{1}{6}.$$

Suppose we are given a set $\bar{\Omega}$ of commitment, each of which is associated with the play of a stationary strategy $\beta \in \mathcal{B}$. For each $\omega \in \bar{\Omega}$, let β_ω be the associated state contingent mixed action plan. For any pair of mixed action $\alpha \in \mathcal{A}$ such that $\alpha(A) \geq \frac{3}{4}$ and state $\theta \in \Theta$, let $\bar{\alpha}_{-\theta} \in \mathcal{A}$ be the unique mixed action such that $\pi(\cdot \mid \bar{\alpha}_{-\theta}, -\theta) = \pi(\cdot \mid \alpha, \theta)$.[13] Note that because of the symmetry of the information structure, $\alpha_{-\theta}$

[12]In the public randomization interpretation, these types correspond to types that do not use the public randomization device.
[13]Note that for any $\alpha \in \mathcal{A}$ with $\alpha(A) \geq 3/4$, such an action always exists.

does not depend on the state $\theta \in \Theta$ and so we subsequently omit the subscript.

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

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

Finally let $\bar{\Omega} := \{\bar{\omega} : \omega \in \Omega^*\}$ and let the set of commitment types be $\Omega^c = \bar{\Omega} \cup \Omega^*$. We now show 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:

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

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

Furthermore for any $(\omega, \theta) \in \mathcal{D}_\omega$, 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 $\alpha(A) \geq 3/4$, then

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

Thus we have for all h^t:

$$\begin{aligned} \Pr(A \mid h^t) &= \sum_{(\omega,\theta) \in \Omega \times \Theta} \beta_\omega(\theta)(A)\gamma(\omega, \theta \mid h^t) \\ &= \sum_{(\omega,\theta) \in \Omega^c \times \Theta} \beta_\omega(\theta)(A)\gamma(\omega, \theta \mid h^t) \\ &= \sum_{(\omega,\theta) \in \mathcal{D}} \gamma(\omega, \theta \mid h^t)\beta_\omega(\theta)(A) + \gamma(\bar{\omega}, -\theta \mid h^t)\beta_{\bar{\omega}}(-\theta)(A) + \sum_{(\omega,\theta) \in (\Omega^* \times \Theta) \setminus \mathcal{D}} \gamma(\omega, \theta \mid h^t)\beta_\omega(\theta)(A) \\ &< \sum_{(\omega,\theta) \in \mathcal{D}} \frac{3}{4}\left(\gamma(\omega, \theta \mid h^t) + \gamma(\bar{\omega}, -\theta \mid h^t)\right) + \sum_{(\omega,\theta) \in (\Omega^* \times \Theta) \setminus \mathcal{D}} \frac{3}{4}\gamma(\omega, \theta \mid h^t) < \frac{3}{4}. \end{aligned}$$

As a result, the SR player always plays R and thus it is a best response for the 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 (without the use of a public randomization strategy), then the fine details of the type space matter for a reputation result. More precisely, the example shows that the mere existence of such commitment types is not sufficient for a reputation result, since many other "bad" commitment types may exist in the type space. Our reputation result in contrast does not hinge on the existence (or absence) of such bad commitment types.

4.3.2 Finite Type Space with Front-loaded Signalling

Consider a stage game that augments the one in FIgure 2 by adding a third action C to player 1's action set. In this modified game, the Stackelberg action is again A giving a payoff of 3 to the LR player. Moreover, L

	L	R
A	$3, 1$	$0, 0$
B	$0, 0$	1,3
C	-100, 0	-100, 3

Figure 3: Stage Game

still remains a best response for player 2 if and only if $\alpha_1(A) \geq \frac{3}{2}$.

Let us first describe the signal structure. The public signal space is binary $Y = \{\overline{y}, \underline{y}\}$ and the state space is $\Theta = \{\ell, r\}$. The conditional probabilities are given by the following:

$$\begin{aligned}
\pi\left(\overline{y} \mid C, r\right) &= \frac{4}{5}, \\
\pi\left(\overline{y} \mid A, \ell\right) = \pi\left(\overline{y} \mid B, r\right) &= \frac{3}{5}, \\
\pi\left(\overline{y} \mid B, \ell\right) = \pi\left(\overline{y} \mid A, r\right) &= \frac{2}{5}, \\
\pi\left(\overline{y} \mid C, \ell\right) &= \frac{1}{5}.
\end{aligned}$$

Note that all of our assumptions for the main theorem are satisfied in the signal structure and so the main theorem holds as long as types with recurrent signalling phases exist. For notational simplicity let $\kappa = 4$.[14] We now describe the type space. Let ω^t denote a commitment type that plays C until period $t-1$ and then switches to the action A forever afterward. Let $K \in \mathbb{N}$ and consider the following set of types:

$$\Omega := \{\omega_1, \ldots, \omega_K\} \cup \{\omega_o\}.$$

Both states ℓ and r are equally likely.

Let us first choose $\mu^* > 0$ such that

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

Consider any type space such that $\mu(\{\omega_1, \ldots, \omega_K\}) < \mu^*$. 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 opportunistic type plays B at all histories and SR players always play R.

To show this, we compute at any history the probability that player 2 assigns to the LR player playing

[14]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.

A (given the proposed candidate strategy profile above):

$$\Pr(A \mid h_t) = \Pr(\{\omega_s : s \leq t\} \mid h_t) = \Pr\left(\{\omega_s : s \leq t\} \cap \ell \mid h_t\right) + \Pr\left(\{\omega_s : s \leq t\} \cap r \mid h_t\right)$$

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

$$\frac{\Pr\left(\{\omega_s : s \leq t\} \cap \theta \mid h_t\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} = \sum_{s=t^*}^{t} \frac{\Pr\left(\{\omega_s\} \cap \theta \mid h_t\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)},$$

where $-\theta$ denotes the state complementary to θ. But note that given $s < t$, ω_s in state θ generates exactly the same distribution of public signals as ω_o in state $-\theta$ at all times $\tau \geq s$. 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:

$$\begin{aligned}
\frac{\Pr(\{\omega_s : s \leq t\} \cap \theta \mid h_t)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} &= \sum_{s=1}^{\min\{t,K\}} \frac{\Pr\left(\omega_s \cap \theta \mid h_t\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} \\
&= \sum_{s=1}^{\min\{t,K\}} \frac{\Pr\left(\omega_s \cap \theta \mid h_s\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_s\right)} \\
&= \sum_{s=1}^{\min\{t,K\}} \frac{\Pr\left(\omega_s \cap \theta \mid h_0\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_0\right)} \frac{\pi(y_0 \mid C, \theta)}{\pi(y_0 \mid B, -\theta)} \cdots \frac{\pi(y_{s-1} \mid C, \theta)}{\pi(y_{s-1} \mid B, -\theta)} \\
&< \sum_{s=1}^{\min\{t,K\}} \frac{\Pr\left(\omega_s \cap \theta \mid h_0\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_0\right)} \kappa^s \\
&\leq \sum_{s=1}^{K} \frac{\Pr\left(\omega_s \cap \theta \mid h_0\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_0\right)} \kappa^s \leq \frac{\Pr(\{\omega_s : s \leq K\} \cap \theta \mid h_t)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} \sum_{s=1}^{K} \kappa^s \\
&= \frac{\Pr(\{\omega_s : s \leq K\} \cap \theta \mid h_t)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} \frac{\kappa^{K+1} - \kappa}{\kappa - 1}.
\end{aligned}$$

Now consider any type space such that

$$0 < \nu(\{\omega_1, \ldots, \omega_K\}) < \nu^*.$$

Using the inequalities derived above, we have for any t and h_t:

$$\begin{aligned}
\Pr(A \mid h_t) = \Pr(\{\omega_s : s \leq t\} \mid h_t) &= \Pr\left(\{\omega_s : s \leq t\} \cap L \mid h_t\right) + \Pr\left(\{\omega_s : s \leq t\} \cap R \mid h_t\right) \\
&< \frac{3}{5} \Pr\left(\{\omega_o\} \cap R \mid h_t\right) + \frac{3}{5} \Pr\left(\{\omega_o\} \cap L \mid h_t\right) = \frac{3}{5} \Pr\left(\{\omega_o\} \mid h_t\right) \leq \frac{3}{5}.
\end{aligned}$$

Then the above shows that the probability that the SR player assigns at any history h_t to the LR player playing A is at most $3/5$. 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 rational LR player to play B at all histories.

Remark. Note that when a type can teach only for up to K periods, then whether 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 K), 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. Therefore, it is not sufficient that such a type merely exists for effective reputation.[15]

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 the source of problems in reputation building does not simply come from limitations in learning the state correctly. In particular, we show more strongly in the example here 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 signalling phases that are front-loaded.

Consider exactly the same game with the same information structure described in the previous subsection 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:

$$\Omega = \{\omega_{t^*}, \omega_{t^*+1}, \dots\} \cup \{\omega_\infty, \omega^o\},$$

where ω_∞ is a type that plays C forever at all histories. Again each state is equally likely and the probability measure over the types is given by $\mu \in \Delta(\Omega)$:

$$\begin{aligned}
\mu(\omega_s) &= \kappa^{-2s}, \\
\mu(\omega_\infty) &= \varepsilon, \\
\mu(\omega_o) &= 1 - \sum_{s=t^*}^{\infty} \kappa^{-2s} = 1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon.
\end{aligned}$$

We will now show that in the above type space, as long as ε is sufficiently small, regardless of the discount factor, there always exists an equilibrium in which the rational player plays B at all histories and player 2 always plays R. Let us now check incentive compatibility at all histories of this proposed candidate strategy profile.

Again as in the previous section, we calculate the probability that the SR player assigns to A 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 $\theta \in \{L, R\}$ in

[15] Of course, if we place more restrictions on the type space, one might conjecture that a reputation theorem might be salvaged. But again such restrictions imply that the fine details of the type space matter beyond just the mere existence of certain types.

the same manner as in the previous example:

$$\begin{aligned}
\frac{\Pr\left(\{\omega_s : s \le t\} \cap \theta \mid h_t\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} &= \sum_{s=t^*}^{t} \frac{\Pr\left(\{\omega_s\} \cap \theta \mid h_t\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_t\right)} = \sum_{s=t^*}^{t} \frac{\Pr\left(\{\omega_s\} \cap \theta \mid h_s\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_s\right)} \\
&= \sum_{s=t^*}^{t} \frac{\Pr\left(\{\omega_s\} \cap \theta \mid h_0\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_0\right)} \frac{\pi(y_0 \mid C, \theta)}{\pi(y_0 \mid B, -\theta)} \cdots \frac{\pi(y_{s-1} \mid C, \theta)}{\pi(y_{s-1} \mid B, -\theta)} \\
&< \sum_{s=t^*}^{t} \frac{\Pr\left(\{\omega_s\} \cap \theta \mid h_0\right)}{\Pr\left(\{\omega_o\} \cap -\theta \mid h_0\right)} \kappa^s \\
&= \sum_{s=t^*}^{t} \frac{\frac{1}{2}\kappa^{-2s}}{\frac{1}{2}\left(1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon\right)} \kappa^s \\
&\le \sum_{s=t^*}^{\infty} \frac{\kappa^{-2s}}{1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon} \kappa^s = \frac{1}{1-\kappa^{-1}} \frac{\kappa^{-t^*}}{1 - \frac{\kappa^{-2t^*}}{1-\kappa^{-2}} - \varepsilon} < \frac{3}{5},
\end{aligned}$$

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

As in the previous example, this again implies that at any history at any time t, the SR player never assigns more than $\frac{3}{5}$ probability to the LR player playing A, which means that the SR player's best response is to play r at all histories. As a result, there are no intertemporal incentives for the opportunistic LR player and so it is also indeed his best response to play B always.

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 there exists some $t > t^*$ such that in any equilibrium,*

$$\Pr(\mu\left(\theta \mid h_t\right) > 1 - \rho \mid \omega_t, \theta) > 1 - \rho.$$

Proof. The proof is a direct consequence of merging arguments that will be illustrated in the next section. □

Remark. One may wonder why we only allow for types ω_s with $s \ge t^*$. 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 ω_∞. The inclusion of this type makes Claim 4.3 very easy to prove. The arguments for the impossibility of reputation building goes through 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 exericse since we are interested in a reputation result that does not depend on what other types are (or are not) included in the type space.

Remark. 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 up to any degree of precision that the LR player wishes. Nevertheless, reputation is not effective 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 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.

5 Proof of Theorem 4.1

Before we proceed to the details of our proof, we provide a brief roadmap for how our arguments will proceed. We first show that by mimicking the strategy of the appropriate commitment type, the long run player can ensure that player 2 learns the state at a rate that is uniform *across all equilibria*.[16] In order to prove this, we first use merging arguments a la Gossner (2011) 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. However note that this does not preclude the possibility that beliefs may drop to low levels outside the signaling phase. To take care of this problem, with the help of the well-known Doob's upcrossing 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). 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.

5.1 Uniform Learning of the State

We first begin by showing that playing the strategy σ^{k,β_1} at state θ leads to uniform learning of the true state θ in all equilibria.

First recall the following definition of the relative entropy of probability measures (also often called the Kullback-Leibler divergence): Given two probability measures $P, Q \in \Delta(Y)$,

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

We consider the predictions of the subsequent ℓ periods. Thus a prediction is a measure in $\Delta(Y^\ell)$. Note that there may be correlations between periods and so a prediction is a joint probability distribution over Y^ℓ as opposed to a product measure. At any history and any pair of states and types, define

$$\phi^\ell_{\theta,\sigma}(\cdot \mid h^t) \in \Delta(Y^\ell)$$

to the probability distribution over $\Delta(Y^\ell)$ over the next ℓ periods, given that the state is θ and the strategy is $\sigma \in \Sigma_1$. We extend the definition of the above to

$$\phi^\ell_{\nu,\bar{\sigma}}(\cdot \mid h^t) \in \Delta(Y^\ell)$$

for any strategy $\bar{\sigma} \in \Sigma_1$ and distribution $\nu \in \Delta(\Theta)$.

Lemma 5.1. *There exists some function $\lambda^* : [0,1] \to \mathbb{R}_+$ such that for all $\beta_1 \in \mathcal{B}$, $k \geq m-1$, $t \geq 0$, and all $\bar{\sigma} \in \Sigma_1$,*

$$H\left(\phi^{m-1}_{\theta,\sigma^{k,\beta_1}}\left(\cdot \mid h^{kt}\right) \mid \phi^{m-1}_{\nu,\bar{\sigma}}\left(\cdot \mid h^{kt}\right)\right) \leq \lambda^*(\varepsilon) \Longrightarrow \nu_{\bar{\sigma}}(\theta \mid h^{kt}) \geq 1 - \varepsilon.$$

Proof. This is a straightforward consequence of the identification assumption of $\alpha_1^*(\theta, \theta')$ for $\theta \neq \theta'$. See Appendix A for details. □

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

Define the following sets when given a strategy $\sigma \in \Sigma_1$ played by type ω^o:

$$\mathcal{C}_{\sigma}^{k,\beta_1}(\theta, J, \lambda) := \left\{h \in H^\infty : H\left(\phi_{\theta,\sigma^{k,\beta_1}}^{m-1}\left(\cdot \mid h^{kt}\right) \mid \phi_{\bar{\sigma}}^{m-1}\left(\cdot \mid h^{kt}\right)\right) > \lambda \text{ for at least } J \text{ values of } t\right\}$$
$$\mathcal{C}_{\sigma}^{\beta_1}(\theta, J, \lambda) := \left\{h \in H^\infty : H\left(\phi_{\theta,\sigma^{k,\beta_1}}^{m-1}\left(\cdot \mid h^{t}\right) \mid \phi_{\bar{\sigma}}^{m-1}\left(\cdot \mid h^{t}\right)\right) > \nu \text{ for at least } J \text{ values of } t\right\},$$
$$\mathcal{D}_{\sigma}^{k,\beta_1}(\theta, J, \varepsilon) := \left\{h \in H^\infty : \nu_{\bar{\sigma}}\left(\theta \mid h^{kt}\right) < 1-\varepsilon \text{ for at least } J \text{ values of } t\right\},$$
$$\mathcal{D}_{\sigma}^{\beta_1}(\theta, J, \varepsilon) := \left\{h \in H^\infty : \nu_{\bar{\sigma}}\left(\theta \mid h^{t}\right) < 1-\varepsilon \text{ for at least } J \text{ values of } t\right\}.$$

Note that $\mathcal{D}_{\sigma}^{k,\beta_1}(\theta, J, \varepsilon)$ only concerns times that are multiples of k and so in particular, $\mathcal{D}_{\sigma}^{1,\beta_1} = \mathcal{D}_{\sigma}^{\beta_1}(\theta, J, \varepsilon)$. The following lemma follows from a standard merging argument.

Lemma 5.2. *Let $\lambda, J > 0$. Then given any $\sigma \in \Sigma_1$, an equilibrium strategy of type ω^o,*

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

Proof. See Appendix A for details. □

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

Corollary 5.3. *Let $\varepsilon > 0$ and let $k > m-1$. Then for every equilibrium σ,*

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

Proof. Note that by Lemma 5.1,

$$\mathcal{D}_{\sigma}^{k,\beta_1}(\theta, J, \varepsilon/2) \subseteq \mathcal{C}_{\sigma}^{k,\beta_1}(\theta, J, \lambda^*(\varepsilon/2)).$$

Therefore

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

□

Note however that $\mathcal{D}_{\sigma}^{k,\beta_1}(\theta, J, \varepsilon/2)$ focuses only on beliefs at the beginning of the signaling phases. However for a 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 upcrossing inequality. To use Doob's upcrossing inequality, however note that the stochastic process in question must indeed be a supermartingale (or submartingale). Note that the SR players' beliefs about the state indeed do form a martingale with respect to the measure $\mathbb{P}_{\nu,\bar{\sigma}}$. However these beliefs are generally no longer a martingale with respect to the measure $\mathbb{P}_{\theta,\sigma^{k,\beta_1}}$.[17] Our proof will necessarily take care of these additional issues.

First however, we 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. First define

[17] They may not even be a submartingale or a supermartingale.

the following sequence of times:

$$\tau_0^x := \inf\{t : x_t < a\},$$
$$\tau_1^x := \inf\{t \geq \tau_0^x : x_t > b\}.^{18}$$

Now we define τ_{2k}^x and τ_{2k+1}^x recursively:

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

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

$$U_t^{(a,b)}(x) := \inf\{k \in \mathbb{N}_+ : \tau_{2k-1}^x \leq t\}.^{19}$$

Finally define the number of up-crossings in the whole sequence:

$$U^{(a,b)}(x) := \lim_{t\to\infty} U_t^{(a,b)}(x).$$

Proposition 5.4. *Let $\varepsilon > 0$ and let σ be an equilibrium strategy of type ω^o. Given any $h \in H^\infty$ and $\theta \in \Theta$, we can define the sequence*

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

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

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

As a consequence,

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

Proof. We first show the following:

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

To prove this, note that Doob's up-crossing inequality (see Appendix C) implies that

$$\mathbb{E}_{\nu,\bar{\sigma}}\left[U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu_{\bar{\sigma}}(\theta \mid h))\right] \leq \frac{\varepsilon}{\frac{\varepsilon}{2}} = 2.^{20}$$

But note that

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

[18] We use the convention that $\inf\emptyset = +\infty$.

[19] $\mathbb{N}_+$ is the set of all natural numbers not equal to zero.

[20] See Appendix for the details.

This implies the latter inequality. Then an application of Markov's inequality implies:

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

Finally note that

$$\{h : U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu_{\bar{\sigma}}(\theta \mid h)) \geq J\} = \bigcup_{t=0}^{\infty}\{h : U_t^{(1-\varepsilon,1-\varepsilon/2)}(\nu_{\bar{\sigma}}(\theta \mid h)) \geq J\}$$

and therefore,

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

□

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

Proposition 5.5 (Uniform Learning of True State)**.** *Let σ be an equilibrium strategy of ω^o. Then for every $n \geq 1$,*

$$\mathbb{P}_{\theta,\sigma^{k,\beta_1}}\left(\mathcal{D}_{\sigma}^{\beta_1}(\theta, 2nk, \varepsilon)\right) \leq \frac{1}{n}\left(\frac{2}{\gamma(\theta,\omega^{k,\beta_1})} - \frac{\log\left(\nu(\theta)\mu(\omega^{k,\beta_1})\right)}{\lambda^*(\varepsilon/2)}\right).$$

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

$$\nu_{\bar{\sigma}}\left(\theta \mid h^t\right) < 1 - \varepsilon \text{ for at least } 2nk \text{ values of } t.$$

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

$$\mathcal{D}_{\sigma}^{\beta_1}(\theta, 2nk, \varepsilon) \subseteq \{h : U^{\varepsilon}(\nu_{\bar{\sigma}}(\theta \mid h)) \geq n\} \cup \mathcal{D}_{\sigma}^{k,\beta_1}(\theta, n, \varepsilon/2).$$

As a result,

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

□

Note importantly that the above bound is independent of the equilibrium. Thus we can use such a bound to establish a lower bound on payoffs that is *uniform* across all equilbria.

5.2 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. Thus we obtain the relevant reputation theorem. We now extend the notion of ε-confirmed equilibrium of Fudenberg and Levine (1992) and Gossner (2011) to our framework.

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

- α_2 is a best response to β_1' given belief μ about the state,
- $\mu(\theta) > 1 - \varepsilon$,
- and $H(\pi(\cdot \mid \alpha_1, \theta) \mid \pi(\cdot \mid \beta_1', \mu)) < \nu$.

Lemma 5.7. *Let $\rho > 0$. Then there exists some $\varepsilon^* > 0$ and $\nu^* > 0$ such that for all (α_1, α_2) that is a (ν^*, ε^*)-confirmed best response at $\theta \in \Theta$, $u_1(\alpha_1, \alpha_2, \theta) > \inf_{\alpha_2 \in B_2(\alpha_1)} u_1(\alpha_1, \alpha_2, \theta) - \rho$.*

Proof. See Appendix A for details. □

Let us define the set of all histories at which the SR player's predictions about the current period's distribution of public signal is accurate up to ε degree of approximation:

$$\mathcal{M}_\sigma^{\sigma'}(\theta, J, \varepsilon) := \left\{ h \in H^\infty : H\left(\phi^1_{\theta,\sigma'}(\cdot \mid h^t) \mid \phi^1_{\nu,\bar{\sigma}}(\cdot \mid h^t)\right) > \varepsilon \text{ for at least } J \text{ values of } t \right\}.$$

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

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

Proof. See Appendix B for details. □

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

Proof. To simplify notation, let us first define the following:

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

Given $\rho > 0$, let $\beta_1 \in \mathcal{S}^{\rho/8}$. Furthermore choose k such that

$$\frac{m-1}{k}(\overline{u} - \underline{u}) < \frac{\rho}{4}.$$

By Lemma 5.7, we can choose $\varepsilon_\theta > 0$ for every $\theta \in \Theta$, such that

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

for all $(\beta_1(\theta), \alpha_2)$ that is a $(\varepsilon_\theta, \varepsilon_\theta)$-confirmed best response at θ. Then let $\varepsilon = \min_{\theta \in \Theta} \varepsilon_\theta$.

Choose n such that the following two inequalities hold:

$$\frac{\rho}{8(\overline{u}-\underline{u})} > \frac{1}{n}\max_{\theta\in\Theta}\left(\frac{2}{\gamma(\theta,\omega^{k,\beta_1})} - \frac{\log\left(\nu(\theta)\mu(\omega^{k,\beta_1})\right)}{\lambda^*(\varepsilon/2)}\right),$$
$$\frac{\rho}{8(\overline{u}-\underline{u})} > \max_{\theta\in\Theta}\frac{-\log\left(\gamma(\theta,\omega^{k,\beta_1})\right)}{2nk\varepsilon}.$$

Finally choose ℓ such that $(k-m+1)\ell > 4nk$.

Given these chosen parameters, note that for every $\theta\in\Theta$, the following inequalities hold:

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

Then by Proposition 5.5 and Lemma 5.8, note that for every $\theta\in\Theta$,

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

Thus in any equilibrium, by playing the strategy σ^{k,β_1}, the long run player gets at the very least the following payoff in state $\theta\in\Theta$:

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

where

$$\begin{aligned}
g_\delta(\theta) &= (1-\delta^{\ell k})\underline{u} + (1-\delta)\left(\sum_{s=0}^{k-1}\delta^s\right)\sum_{t=\ell}^{\infty}\delta^{tk}\left(\frac{\sum_{s=0}^{m-1}\delta^s\underline{u} + \sum_{s=m}^{k-1}\delta^s\left(\inf_{\alpha_2'\in\mathcal{A}_2}u_1(\beta_1(\theta),\alpha_2',\theta)-\frac{\rho}{8}\right)}{\sum_{s=0}^{k-1}\delta^s}\right)\\
&\ge (1-\delta^{\ell k})\underline{u} + (1-\delta^k)\sum_{t=\ell}^{\infty}\delta^{tk}\left(\frac{\sum_{s=0}^{m-1}\delta^s\underline{u} + \sum_{s=m}^{k-1}\delta^s\left(u_1^*(\theta)-\frac{\rho}{4}\right)}{\sum_{s=0}^{k-1}\delta^s}\right).
\end{aligned}$$

It remains to show that we can find δ^* 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:

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

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

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

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

This concludes our proof. □

6 Upper Bound on Payoffs

Thus far, we have focused our analysis completely on a lower bound reputation theorem. However notice that we have not argued that the lower bound is indeed sharp in any sense. 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 long run opportunistic player does indeed equal the lower bound of Theorem 4.1.

With possible non-identification of actions, there may be scenarios 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 state or over the set of types. 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 long run player is sufficiently patient. This dependence on the initial conditions makes it difficult to provide a general sharp upper bound.

6.1 Example I

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

Table 3: Quality Choice Game

	ℓ	r
A	$1, 1$	$-1, 0$
B	$2, -1$	$0, 0$

repeated game this stage game is repeatedly played and all payoffs are common knowledge.

There are two states $\{L, R\}$ which only affect the signal distribution of the public signal. There are two types in the game, The commitment type, ω, 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.[21] In particular, we assume that the probability of each state is identical and the probability of the commitment type is given by μ.

The information structure is given by the following:

$$\pi(\bar{y} \mid A, L) = \frac{1}{6}, \pi(\bar{y} \mid B, L) = \frac{4}{6}$$
$$\pi(\bar{y} \mid A, R) = \frac{5}{6}, \pi(\bar{y} \mid B, R) = \frac{2}{6}.$$

Note that according to this information structure, the mixed action $\left(\frac{2}{3}A \oplus \frac{1}{3}B, \theta\right)$ is indistinguishable from $(B, -\theta)$:

$$\pi(\bar{y} \mid \frac{2}{3}A \oplus \frac{1}{3}B, L) = \pi(\bar{y} \mid B, R),$$
$$\pi(\bar{y} \mid \frac{2}{3}A \oplus \frac{1}{3}B, R) = \pi(\bar{y} \mid B, L).$$

[21] 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.

Proposition 6.1. *Let $\varepsilon > 0$. Then there exists some μ^* such that for all $\mu > \mu^*$ and any $\delta \in (0,1)$, there exists an equilibrium in which the rational player obtains a payoff of 2 in both states.*

Proof. Consider the candidate equilibrium strategy profile in which the LR player always plays B. Choose $\nu^* = \frac{3}{4}$. Then we will show that when $\nu > \nu^*$, 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 A:

$$\mu(A \mid h_t) = \frac{2}{3}\mu(\omega \mid h_t) = \frac{2}{3}\left(\mu(\{\omega\} \cap L \mid h_t) + \mu(\{\omega\} \cap R \mid h_t)\right)$$

Now let us bound the probability $\mu(\omega \mid h_t)$ from below. To produce this bound, consider the following likelihood ratio:

$$\frac{\mu(\{\omega\} \cap \theta \mid h_t)}{\mu(\{\omega_o\} \cap -\theta \mid h_t)} = \frac{\mu(\{\omega\} \cap \theta \mid h_0)}{\mu(\{\omega_o\} \cap -\theta \mid h_0)} = \frac{\nu}{1-\nu}.$$

This then implies that for all h_t,

$$\mu(\{\omega\} \mid h_t) = \nu, \mu(\{\omega_o\} \mid h_t) = 1 - \nu.$$

Thus, for all h_t,

$$\begin{aligned}\mu(A \mid h_t) &= \frac{2}{3}\frac{\nu}{1-\nu}\left(\mu\left(\{\omega_o\} \cap R \mid h_t\right) + \mu\left(\{\omega_o\} \cap L \mid h_t\right)\right) \\ &= \frac{2}{3}\frac{\nu}{1-\nu}\mu\left(\{\omega_o\} \mid h_t\right) = \frac{2}{3}\nu > \frac{1}{2}.\end{aligned}$$

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 LR player's payoff is 2 in both states, concluding the proof. □

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.

Proposition 6.2. *Let $\varepsilon > 0$. Then there exists some $\nu^* > 0$ such that for all $\nu < \nu^*$, there exists some δ^* such that for all $\delta > \delta^*$, in all equilibria, the (opportunistic) LR player obtains a payoff of at most $3/2 + \varepsilon$.*

Proof. This will be a consequence of a more general upper bound theorem that we will prove later. □

6.2 Example II

The above shows that the size of the commitment type matters. Proposition 6.2 may suggest a conjecture that when the size of the commitment types is sufficiently small, then the upper bound coincides with the statewise Stackelberg payoffs. However the following example shows otherwise.

There are two states $\theta \in \{L, R\}$. The stage games depend on the state in the following manner:

Note that when the state is L, the game resembles a coordination game. In contrast, when the state is R, the game resembles a quality choice game in which the action B is strictly dominant. Furthermore, note that the payoffs of player 2 are independent of the state.

Table 4: Stage Games

$\theta = L$	ℓ	r
A	$3, 1$	$-1, 0$
B	$2, -1$	$0, 0$

$\theta = R$	ℓ	r
A	$1, 1$	$-1, 0$
B	$2, -1$	$0, 0$

Consider the following signal structure with two public signals $Y = \{\bar{y}, \underline{y}\}$:

$$\pi(\bar{y} \mid A, R) = 1,$$
$$\pi(\bar{y} \mid B, R) = \pi(\bar{y} \mid A, L) = \frac{1}{3},$$
$$\pi(\bar{y} \mid B, L) = \frac{2}{3}.$$

Note importantly that (B, R) and (A, L) are indistinguishable from each other when observing the public signals.

There are two states $\{L, R\}$ with the probability of L given by p. Commitment types are given by the following types $\{\omega, \omega_o\}$ where ω is a commitment type that plays the Stackelberg (mixed) action in each state. Let the probability of ω be ν.

We now show that in this environment, there is an equilibrium in which the LR player plays A always in state L and B always in state R. Whenever the prior probability about the state L is sufficiently large, this means that the LR rational player obtains a payoff of 3 in state L and 2 in state R. This is strictly larger than the

Thus in this environment, signalling of the state in such an equilibrium becomes undesirable for the LR player in state $\theta = R$. Since then the best payoff that he can indeed obtain is $\frac{3}{2}$. We now study the question of how

6.3 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).[22] However we will show that this derived upper bound is indeed tight in a class of games, where state revelation is desirable.[23]

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

Definition 6.3. Let $p \in \Delta(\Theta)$. A state-contingent strategy $\beta \in \mathcal{B}$ is called non-revealing at p if for all $\theta, \theta' \in supp(p)$, $\pi(\cdot \mid \beta(\theta), \theta) = \pi(\cdot \mid \beta(\theta'), \theta')$.

In words, this means that if player 1 plays according to a non-revealing strategy at p, then with probability

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

1, player 2's prior will not change regardless of the message he 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 $V(p) = \underline{u}$. Then define $\mathbf{cav}V$ to be the smallest concave function that dominates V.

Theorem 6.4 (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 δ^* such that for all $\delta > \delta^*$, the ex-ante expected payoff of the LR rational player in all equilibria is at most $\mathbf{cav}V(\nu) + \varepsilon$.*

Proof. Here we sketch the proof, relegating the details to the Appendix. Using arguments borrowed from MSZ, we show that the play of $\bar{\sigma}$, which is not necessarily the equilibrium strategy of the opportunistic type, 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 δ^* such that for all $\delta > \delta^*$, any type space, and any equilibrium strategy σ of the opportunistic type, by playing instead $\bar{\sigma}$, the opportunistic type obtains a payoff of at most $\mathbf{cav}V(\nu) + \varepsilon/2$.[24] Now note that for the opportunistic type, the strategy of 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. Thus we must 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 Example I, 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, beliefs in an equilibrium conditional on *the opportunistic type's strategy* is no longer a martingale. In contrast, when the commitment type probabilities are small, the 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.

Finally we now apply the above theorem to a setting in which the type space includes those commitment types constructed in the previous section. 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 long run opportunistic type converge uniquely to the statewise Stackelberg payoffs in every state.

[24]Note that δ^* does not depend on the type space.

Corollary 6.5. *Suppose that V is convex and 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 δ^* such that for all $\delta > \delta^*$ and any state $\theta \in \Theta$, the LR rational player obtains a payoff in the interval $(v_1^\theta - \varepsilon, v_1^\theta + \varepsilon)$ in all equilibria.*

Proof. Because V is convex,

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

□

6.3.1 Returning to Example 1

Let us return to our Example 1. There, for any $q \in (0,1)$, the set of non-revealing strategies $NR(q)$ is given by:

$$NR(q) = \left\{ \left(p, \frac{2}{3} - p\right) : p \in (0, 2/3) \right\}.$$

Given a non-revealing strategy, the SR player believes that A will be played with probability:

$$qp + (1-q)(2/3 - p) = (2q-1)p + \frac{2}{3}(1-q).$$

Assume that $q \geq 1/2$ since the analysis is exactly symmetric for $q \leq 1/2$. Note that there exists some $p \in (0, 2/3)$ such that $(2q-1)p + \frac{2}{3}(1-q) \geq 1/2$ if and only if $q \geq 3/4$. For $q \geq 3/4$, then

$$L \in BR(p, 2/3 - p, q) \Leftrightarrow (2q-1)p + \frac{2}{3}(1-q) \geq 1/2 \Leftrightarrow p \geq \frac{1}{3}\frac{2q - 1/2}{2q-1}.$$

Then the payoff conditional on $L \in BR(p, 2/3 - p, q)$ is

$$1 + \left(1 - qp - (1-q)\left(\frac{2}{3} - p\right)\right) = 1 + \left(1 - \frac{2}{3}(1-q) - p(2q-1)\right),$$

which is decreasing in p. Thus, for any $q \geq 3/4$, we have $V(q) = \frac{3}{2}$ if $q \geq 3/4$. Similarly $V(q) = \frac{3}{2}$ if $q \leq 1/4$. For $q \in (1/4, 3/4)$, the SR player plays R against any non-revealing strategy so that $V(q) \leq 0$. By applying Theorem 6.4, we confirm Proposition 6.2.

A Proofs of Lemma 5.1 and Lemma 5.2

Proof of Lemma 5.1. To be filled. □

Proof of Lemma 5.2. To be filled. □

B Merging and Best Responses

Proof of Lemma 5.8. To be filled. □

The arguments here are analogues of those results provided by Gossner (2011). We modify the arguments and results slightly.

Lemma B.1. *For every $\xi \in \Xi$,*

$$\sum_{t=1}^{\infty} \mathbb{E}_{\xi}\left[\mathbf{Z}_t^{\xi}\right] \leq -\log[\mu(\xi)].$$

Proof. The proof follows easily from an application of the chain rule. For every t,

$$H(P_{\xi}^t \mid P_{\mu}^t) = \sum_{\tau=1}^{t} \mathbb{E}_{\xi}\left[\mathbf{Z}_t^{\xi}\right].$$

From the previous lemma, we have

$$H(P_{\xi}^t \mid P_{\mu}^t) \leq -\log \mu(\xi).$$

Thus we have

$$\begin{aligned}\log \mu(\xi) \geq \lim_{t\to\infty} \sum_{\tau=1}^{t} \mathbb{E}_{\xi}\left[\mathbf{Z}_t^{\xi}\right] &= \lim_{t\to\infty} \mathbb{E}_{\xi}\left[\sum_{\tau=1}^{t} \mathbf{Z}_t^{\xi}\right] \\ &= \mathbb{E}_{\xi}\left[\sum_{\tau=1}^{\infty} \mathbf{Z}_t^{\xi}\right],\end{aligned}$$

where the last equality follows from the monotone convergence theorem. This proves the claim. □

C Martingale Inequality

Here we present the formal statement of the martingale upcrossing inequality.

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

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

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

D Proof of Theorem 6.4

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 lemmas.

Lemma D.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. To be filled. □

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

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

Proof. To be filled. □

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

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

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 $t + 1$:

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

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 (σ, σ_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 6.4. By Lemma D.2, there exists some $\rho > 0$ such that

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

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

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.

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.

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.