Reputation Building under Uncertain Monitoring

Joyee Deb   Yuhta Ishii*

April 10, 2018

Abstract

We study reputation building under uncertainty: For example, a firm wants to build a reputation for quality; it faces consumers who make purchase decisions based only on public signals of quality like product reviews on a review website, but are uncertain about how exactly to interpret the reviews and link them to the firm's actions. Formally, we study a canonical reputation model with a long-run (LR) player facing a sequence of short-run (SR) opponents, with one difference: The LR player knows the monitoring structure, but SR players are uncertain about it. Can the firm build a reputation under such uncertainty? We show that standard reputation results break down: Even if there is a possibility that firm is a commitment type that plays the Stackelberg action, there exist "bad" equilibria in which the firm gets payoffs no higher than in the one-shot game. We present necessary and sufficient conditions to restore reputation building. In contrast to existing literature, reputation building requires dynamic commitment types that switch between "signaling actions" that help the SR players learn the monitoring structure and "collection actions" that are desirable for payoffs.

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

1 Introduction

Consider a long-run firm building a reputation for producing environmentally friendly products. Such a reputation is valuable for the firm because consumers inherently care about the environmental impact of their purchases and are often willing to pay more for green products. Consumers make purchase decisions based on the presence/absence of "eco-friendly" labeling of the product, but are typically unsure of how to interpret the labels. Indeed, as reported by Ecolabel Index, there are over 400 eco-labels and green certification systems in the market. Many of these labels are genuine certifications with stringent standards, but numerous others have been discredited. As a result, on seeing an eco-label, consumers are uncertain about its informational content, and may not be convinced about the product being environmentally friendly.[1] Faced with such uncertainty among consumers, a firm, even after honest investment in green products and after undergoing reliable certification, may find it difficult to establish a positive reputation. In such a setting, how can a firm build a green reputation and convince consumers that its products are indeed environmentally friendly?

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

To start, consider reputation building in environments in the absence of such uncertainty. Canonical models of reputation (e.g., Fudenberg and Levine (1992)) study the behavior of a long-run agent (say, a firm) who repeatedly interacts with short-run opponents (consumers). There is incomplete information about the firm's type: consumers entertain the possibility that the firm is of a "commitment" type that is committed to playing a particular action at every period of the game. Even when the actions of the firm are noisily observed, the classical reputation building result states that if a sufficiently rich set of commitment types occurs with positive probability, a patient firm can achieve payoffs arbitrarily close to his Stackelberg payoff of the stage game in *every* equilibrium.[2] Intuitively by mimicking a commitment type that always plays the Stackelberg action, a long-run firm can eventually signal to the consumer its intention to play the Stackelberg action in the future and thus obtain high payoffs in *any* equilibrium. Importantly, this result is robust to the introduction of other arbitrary commitment types. This intuition critically relies on the consumer's ability to accurately interpret the noisy signals. But, if monitoring is uncertain, the reputation builder finds it difficult to signal his intentions.

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

[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.

To study reputation building with uncertain monitoring, we consider a canonical model of reputation building, with one key difference. At the beginning of the game, a state of the world, $(\theta, \omega) \in \Theta \times \Omega$ is realized, which determines both the type of the firm, ω, and the monitoring structure, $\pi_\theta : A_1 \to \Delta(Y)$: a mapping from actions taken by the firm to distribution of signals, $\Delta(Y)$, observed by the consumer. We assume for simplicity that the firm knows the state of the world, but the consumer does not.[3]

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

In our main result, we construct a set of commitment types such that when these types occur with positive probability, a sufficiently patient firm obtains payoffs arbitrarily close to the Stackelberg payoff in all equilibria even when the consumers are uncertain about the monitoring environment. Importantly, as in the classical reputation literature, this result is robust to the inclusion of other arbitrary commitment types, and thus is independent of the fine details of the type space. In contrast to the commitment types considered in the previous literature, the commitment types that we construct are committed to *dynamic* strategies (time-dependent but not history dependent) that switch infinitely often between "signaling actions" that help the consumer learn the unknown monitoring state and "collection actions" that are desirable for payoffs (the Stackelberg action). A key contribution of our paper is the construction of these dynamic commitment types that play *periodic* strategies. It turns out that such dynamic commitment types are indeed necessary for reputation building under uncertain monitoring. This is because signaling the unknown monitoring state and Stackelberg payoff collection may require the use of different actions in the stage game.

We can interpret our model as one that represents *subjective* uncertainty that consumers have about the actual monitoring structure and the behavior of the reputation-building firm. We show that the firm can indeed effectively establish a reputation, as long as the consumers assign positive probability to the constructed commitment types and the correct monitoring structure: In particular, our result is a robust reputation theorem in the sense that it holds independently of what beliefs the consumer may have about the behavior of the firm under incorrect monitoring structures.

The proof proceeds in two key steps. The first step is to show that, by mimicking the strategy of the appropriate commitment type, the LR player can ensure that the SR players learn the state at a rate that is *uniform across all equilibria.* We prove this by deriving a new result on robust learning.

It is worth highlighting that the robust learning result is a methodological contribution of this paper, as it is not specific to the reputation context, but is rather applicable to general learning environments, and important in its own right. It provides an easy-to-check sufficient condition on a class of learning

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

environments that guarantees that an observer will indeed learn the validity of an event at a uniform rate across all learning environments in the class, even when there is substantial ambiguity about the true learning environment.[4] More formally, our sufficient condition relates the summability of the supremum of *Hellinger transforms* of learning environments to the uniform learnability of an event across all learning environments in this class.[5] It turns out that this summability condition is is easy to verify and holds in many settings of economic interest.

Next, we apply the robust learning theorem to our reputation model to establish an upper bound on the number of times that the SR player's belief on the true state is low. In other words, the belief, at most times, will be high on the correct state with high probability, in which case action identification is no longer problematic. Finally, we 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. We use merging arguments à la Gossner (2011) to construct our lower bound on payoffs.

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

It is also worth pointing out that in specific applications, once we fix payoffs and the information structure, the full range of dynamic commitment types will not typically be needed for reputation building to be possible. As we show in our example in Section 3, reputation building can be recovered quite easily given a specific setting. The full range of dynamic commitment types is needed to get the canonical reputation result that does not depend on payoffs, prior beliefs, or the fine details of the type space but rather only requires the existence of certain commitment types.

Returning to the eco-labelling example, if consumer purchase decisions depend on product labeling, and the consumer is uncertain about how to interpret eco-labels, a firm cannot build a reputation for environmentally friendly products quality by simply investing effort into producing these products. Effective reputation building requires not only the actual production of environmentally friendly products and reliable certification but also a repeated commitment to credibly convey to the consumer the meaning of the eco-labels. Indeed, we observe this in practice. There are some high-quality eco-labels and stringent certification systems like Energy Star in the US or Blue Angel in Germany. Firms that have invested in producing environmentally friendly products and have certifications from these reliable certification systems will periodically run expensive integrated campaigns aimed to educate the consumer about the environmental practices required

[4] Our robust learning result relates loosely to ideas of uniform learning from Vapnik-Chervonenkis theory used for example in Al-Najjar (2009) and Al-Najjar and Pai (2014). These papers study the uniform learning of a wide class of events given *any* i.i.d process. The main conceptual distinction of our robust learning result is that we study uniform learning of *finitely many* events but allow for any arbitrary stochastic process that may involve arbitrary serial correlations.

[5] The Hellinger transform is a useful concept in the study of statistical experiments. See Section 5.1 for precise statements of our sufficient condition, as well as Torgersen (1991) and Moscarini and Smith (2002) for illustrations of applications of the Hellinger transform.

to obtain these certifications. For instance, TCP a leading seller of energy efficient lighting solutions, ran a widespread campaign with educational events across the US to educate consumers on significance of their Energy Star certification. UPM-Kymmene Corporation, a global forest products company, invested significantly to use recycled and renewable material to obtain certification by the EU ecolabel. Subsequently, the company ran widespread promotion campaigns to increase public awareness and knowledge about what the EU Ecolabel stood for, and the stringent standards the company adhered to.[6]

While this paper is motivated by environments with uncertain monitoring, our results apply more broadly to other types of uncertainty. First, our model allows for both uncertainty regarding monitoring *and* uncertainty about the payoffs of the reputation builder. Our results also extend to environments with symmetric uncertainty about monitoring. For example, consider a firm that is entering a completely new market and is deciding between two different product offerings. Neither the consumer nor the firm initially know which product is better for the consumer. Is it possible for the firm to build a reputation for making the better product? Note that our results apply here. Mimicking a commitment type that both signals and collects is useful to the firm here in two ways: It not only helps the consumer learn about the unknown state of the world, but simultaneously enables the firm to learn the true state of the world. Then, we can interpret the commitment type as one that alternates between learning the state and payoff collection. It is also noteworthy that our reputation results continue to hold even if the firm (long-run player) did not observe the public signals observed by consumers (short-run players).

Thus far, we have restricted our discussion to a lower bound on the long-run player's equilibrium payoff. Of course the immediate question that arises is whether the long-run player can possibly obtains payoffs much higher than the Stackelberg payoff: How tight is this lower bound on payoffs? With uncertain monitoring, there may be situations in which a patient long-run player can indeed guarantee himself payoffs that are strictly higher than the Stackelberg payoff of the true state. We present several examples in which this occurs: It turns out that the long-run player does not find it optimal to signal the true state to his opponent, but would rather block learning and attain payoffs that are higher than the Stackelberg payoff in the true state. In general, an upper bound on a patient long-run player's equilibrium payoffs depends on the set of commitment types and the prior distribution over types. Such dependence on the specific details of the game makes a general characterization of an upper bound difficult.[7] A precise characterization of an upper bound is beyond the scope of this paper. Nevertheless, we provide a joint sufficient condition on the monitoring structure and stage game payoffs that ensure that the lower bound and the upper bound coincide for any specification of the type space: Loosely speaking, these are games in which state revelation is desirable for the reputation builder.

1.1 Related Literature

There is a vast literature on reputation effects which includes the early contributions of Kreps and Wilson (1982) and Milgrom and Roberts (1982) followed by the canonical models of reputation developed by Fudenberg and Levine (1989), Fudenberg and Levine (1992) and more recent methodological contributions by

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

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

Gossner (2011). To the best of our knowledge, our paper is the first to consider reputation building in the presence of uncertain monitoring.

Aumann, Maschler, and Stearns (1995) and Mertens, Sorin, and Zamir (2014) study repeated games with uncertainty in both payoffs and monitoring but focus primarily on zero-sum games. In contrast, reputation building matters most in non-zero sum environments where there are large benefits that accrue to the reputation builder from signaling his long-run intentions to the other player. There is some recent work on uncertainty in payoffs in non-zero sum repeated games by Wiseman (2005), Hörner and Lovo (2009), Hörner, Lovo, and Tomala (2011). In all of these papers, however, the monitoring structure is known to all parties with certainty. Our paper's modeling framework corresponds most closely to Fudenberg and Yamamoto (2010) who study a repeated game model in which there is uncertainty about both monitoring and payoffs. However, Fudenberg and Yamamoto (2010) focus their analysis on an equilibrium concept called perfect public ex-post equilibrium in which players play strategies whose best-responses are independent of any belief that they may have about the unknown state. As a result, in equilibrium, no player has an incentive to affect the beliefs of the opponents about the underlying monitoring structure. In contrast, our paper studies, more generally, equilibria where the reputation builder potentially benefits from affecting the beliefs of the opponent about the monitoring structure. In fact, the possibility of such manipulation is crucial in our setting.

To the best of our knowledge, the construction of the dynamic types necessary to establish a reputation result is novel. The necessity of such dynamic commitment types in our setting is somewhat surprising, and arises for a very different reason than in the literature on reputation building against long-run, patient opponents.[8] In particular, dynamic commitment types arise in Aoyagi (1996), Celentani, Fudenberg, Levine, and Pesendorfer (1996), and Evans and Thomas (1997), since establishing a reputation for carrying through punishments after certain histories potentially leads to high payoffs.[9] In contrast, our non-reputation players are purely myopic and so the threat of punishments has no influence on these players. Dynamic commitment types turn out to be necessary in our setting to resolve a potential conflict between signaling the correct state and Stackelberg payoff collection which are both desirable to the reputation builder: "signaling actions" and "collection actions" discussed in the introduction are generally not the same. As a result, by mimicking such commitment types that switch between signaling and collection actions, the reputation builder, if he wishes, can signal the correct monitoring structure to the non-reputation builders.

In a concurrent paper, Pei (2017) studies a reputation building model in which there are potential interdependent values. However, there are several main differences. First, Pei (2017) focuses on environments in which monitoring about the action is perfect whereas the focus of our paper is to study general imperfect monitoring environments where non-identification of action and state may potentially cause reputation building problems. Secondly, Pei (2017) restricts attention to a finite number of commitment types that play stationary (possibly state-contingent) actions and asks under what distributional assumptions (over types), a reputation result obtains. We take a complementary approach, and show that the classical reputation result breaks down without any additional distributional assumptions, and then construct commitment types that would restore reputation building given general distributions.

One of the contributions of our paper is to highlight the fragility of reputation building in the presence of

[8] 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).

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

uncertainty about monitoring. This fragility is due to the lack of identification of intended actions because multiple combinations of state and action lead to the same distribution over observed public signals. Such identification problems are at the heart of recent papers featuring failure of learning. For example, Acemoglu, Chernozhukov, and Yildiz (2016) show that multiple Bayesian agents may fail to converge to a common belief if they start with different prior beliefs about the distribution over states. Failure of learning also can arise in settings with misspecified models such as in Heidhues, Köszegi, and Strack (Forthcoming) in which an individual holds wrong beliefs about a fundamental leading to wrong beliefs about other relevant variables. The nature of failure of learning in our negative examples is substantively different in that it arises endogenously, and depends on the strategy of the long-run player, while failure of learning in Acemoglu, Chernozhukov, and Yildiz (2016) and Heidhues, Köszegi, and Strack (Forthcoming) arises due to exogenous assumptions about the information process. Moreover, our main result shows that with appropriate commitment types such identification problems do not pose an obstacle for learning.

The rest of the paper is structured as follows. We describe the model 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 monitoring. We also propose a way to recover reputation building in the example. 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 are necessary. In particular, we explain what features are important for reputation building. The proof of the main result is in Section 5. Finally, in Section 6, we discuss potential upper bounds on long-run payoffs.

2 Model

A long-run (LR) player, player 1, faces a sequence of short-run (SR) player 2's.[10] Before the interaction begins, a pair $(\theta, \omega) \in \Theta \times \Omega$ of a *state* of the world and *type* of player 1 is drawn independently according to the product measure $\gamma := \nu \times \mu$ with $\nu \in \Delta(\Theta)$, and $\mu \in \Delta(\Omega)$. We assume for simplicity that Θ is finite and enumerate $\Theta := \{\theta_0, \dots, \theta_{m-1}\}$, but Ω may possibly be countably infinite.[11] The realized pair of state and type (θ, ω) is then fixed for the entirety of the game.

In each period $t = 0, 1, 2, \dots$, players simultaneously choose actions from their respective action spaces $a_1^t \in A_1$ and $a_2^t \in A_2$. We assume for simplicity that A_1 and A_2 are both finite. Let A denote $A_1 \times A_2$. In each period $t \geq 0$, after players have chosen the action profile a^t, a public signal y^t is drawn from a finite signal space Y according to the probability $\pi(y^t \mid a_1^t, \theta)$.[12] Note importantly that both the action chosen at time t and the state of the world θ potentially affect the signal distribution. We interpret the state of the world θ as representing the unknown monitoring structure. Denote by $H^t := Y^t$ the set of all t-period *public* histories h^t 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 observes the realized state of the world $\theta \in \Theta$ perfectly so that his private history at time t is formally a vector $H_1^t := \Theta \times A_1^t \times Y^t$.[13] Meanwhile the SR player at time t observes only the public signals up to time t and so his information coincides exactly with the public history $H_2^t := H^t$.

[10] In the exposition, we refer to the LR player as male and SR player as female.

[11] The assumption of allowing Ω to be countably infinite is standard in the existing literature (see, for instance, Fudenberg and Levine (1992)) when the Stackelberg action of the stage-game can be mixed.

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

[13] It is worth pointing out that our results hold without any any change even if the LR player did not observe the public signal. Further, we conjecture that it is a straightforward extension to consider a LR player who must learn the state over time.

The observability (or lack of it) of previous SR player's actions does not affect our results. 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.

2.1 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 $\omega \in \Omega^c$ 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 payoff for the SR player 2 at time t is given by:

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

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

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

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

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

Finally, given the stage game payoff u_1, we can define the statewise-Stackelberg payoff of the stage game. First for any $\alpha_1 \in \mathcal{A}$, let us define $B_2(\alpha_1, \theta)$ as the set of best-responses of player 2 when player 2 knows the state to be θ and player 1 plays action α_1. We assume that there exists a mixed action that achieves the Stackelberg payoff. The Stackelberg payoff and actions of player 1 in state θ are given respectively by:

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

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

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

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

[14] Note that α_1^* is generally a correspondence and not a function.

2.3 Information Structure

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

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

Using the above definition, we define the following set.

Definition 2.2. $\Theta^{id} \subseteq \Theta$ is the set of all states $\theta \in \Theta$ such that there exists some $\alpha_1 \in \alpha_1^*(\theta)$ such that information structure π has action identification for (α_1, θ).

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

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

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

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

Given the above assumption, for any pair of states $\theta \in \Theta^{id}$ and $\theta' \in \Theta$, we denote $\alpha_1(\theta, \theta')$ to be the action defined above. The above assumption is novel. First note that Assumption 2.3 *does not* assume that $\alpha_1(\theta, \theta')$ must be the Stackelberg action in state θ. We can visualize 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 $\mathcal{A}$ at the state θ:

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

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

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

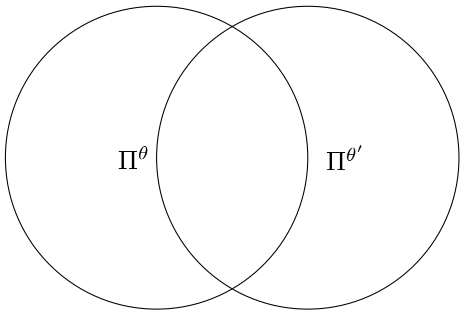

Figure 1: Assumption 2.3 is satisfied.

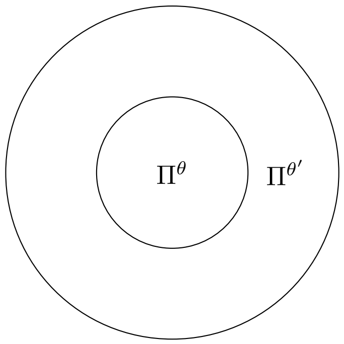

Figure 2: Assumption 2.3 is violated.

3 An Illustrative Example

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

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

Figure 3: Stage Game

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

First note that the Stackelberg payoff is $\alpha - \gamma + \beta$ and the Stackelberg action is C. Secondly note that in the stage game, B is a best-response to the LR player's stage game mixed action if and only if $\alpha_1(C) \geq \alpha_1(D)$.

Suppose now that the consumers do not directly observe the action choice of the firm, but rather observe a signal of the quality of the product: There are two possible public signals: $Y = \{\bar{y}, \underline{y}\}$. Think of these signals as different eco-labels (or the absence of an eco-label) that the consumer observes on the packaging before she make her purchase decision. There are two states of the world $\Theta = \{g, b\}$, which have no effect on payoffs but do affect the distribution of signals. We interpret the state as reflecting whether eco-labels are informative or not. The state g corresponds to the scenario in which the environmental monitoring and certification agency is indeed good, so that eco-labels and other information on packaging is informative,

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

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

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

while b corresponds to the scenario in which the monitoring agency is bad, in which case information on packaging and labels are completely uninformative. The information structure is given in Figures 4 and 5 and reflect the fact that the signals are informative in one state and not in the other. First note that conditional on the state $\theta = g$, actions are identified. Thus if $\theta = g$ were common knowledge, then the classical reputation results would hold: If there was a positive probability that the LR player could be a commitment type that always plays C, then a sufficiently patient LR player would achieve a payoff arbitrarily close to $\alpha - \gamma + \beta$ in *every* equilibrium. We will demonstrate below that this observation is no longer true when there is uncertainty about the monitoring states.

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

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

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

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

3.1 Failure of Reputation Building

We will construct an equilibrium in which the LR player gets a payoff close to 0. Suppose that the LR player is either an opportunistic type, ω^o, who is can choose any strategy or a commitment type, ω^c that always plays C.[16] Consider a strategy profile in which the LR player of type ω^o always plays D regardless of the state of the world θ, and that of type ω^c plays C always. We show that reputation building fails in the sense that when $\mu(\omega^c) > 0$ is sufficiently small, there is a perfect Bayesian equilibrium for all $\delta \in (0, 1)$ in which the LR player plays according to the strategy above while the SR player always plays N. To simplify notation, we let $\mu(\omega^c) = \xi$ and $\nu(g) = p$.

Let us first examine the probability that an SR player assigns to the commitment type, which will then be a sufficient statistic for her best-response given the candidate equilibrium strategy of the opportunistic LR player. At any time t, conditional on a history h^t, let $\mu^t_{a_1\theta}(h^t)$ denote the probability conditional on the public history h^t that the SR player assigns to the event in which the state is θ and the LR player plays a_1.

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

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

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

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

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

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

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

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

Given the candidate strategy of the LR player, if $\xi < \xi^*$, the SR player's best-response is to play N at all histories (since she expects to face action D with probability more than $\frac{1}{2}$. Finally given the SR player's strategy, there are no inter-temporal incentives for the LR player, and hence it is incentive compatible for the opportunistic LR player to always play D. This gives the LR player a payoff of 0 in this equilibrium, regardless of his discount factor.

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

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

3.2 Discussion

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

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

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

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

3.3 Recovering Reputation Building

How can we recover reputation building in this example? Suppose now that the LR player can undertake a costly action to signal the credibility of the eco-labeling agency. Consider the new stage game in Figure 6 below, where the firm can also choose a third action to "inform," denoted by I. In the example highlighted in the introduction, this could take the form of campaigns that promote awareness about the tests involved to obtain eco-friendly certification. The signal structure is given by Figures 7 and 8.

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

Figure 6: Stage Game

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

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

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

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

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

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

4 Main Reputation Theorem

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

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

4.1 Construction of Commitment Types

We first construct the appropriate commitment types. A commitment type ω^{β_1} will be a type that always plays a strategy denoted σ^{β_1} defined as follows. First define the following sequence of natural numbers recursively:

$$n_0 = 0, n_1 = m + 1, n_{k+1} - n_k = m + k + 1.$$

In Assumption 2.3, we have already defined the mixed action $\alpha_1(\theta, \theta')$ for any $\theta \in \Theta^{id}$ and $\theta \neq \theta'$. To simplify notation, let us also choose $\alpha_1(\theta, \theta') \in \mathcal{A}$ arbitrarily if either $\theta = \theta'$ or $\theta \notin \Theta^{id}$. Then for every $\beta_1 \in \mathcal{B}$, we now define the following commitment type, ω^{β_1}, who plays the (possibly dynamic) strategy $\sigma^{\beta_1} \in \Sigma_1$ in every play of the game. We define this strategy σ^{β_1} as follows, which depends only on calendar time:

$$\sigma_\tau^{\beta_1}(\theta) = \begin{cases} \beta_1(\theta) & \text{if } \tau - \max\{n_k : n_k \leq \tau\} \geq m, \\ \alpha_1(\theta, \theta_j) & \text{if } j = \tau - \max\{n_k : n_k \leq \tau\} < m. \end{cases}$$

This commitment type plays a dynamic strategy that depends only on calendar time that consist of blocks that grow in length over time. Essentially, this commitment type starts out in a *signaling phase* that lasts for m periods, trying to convince the SR players of the state. After these first m periods, then this commitment type transitions to a *collection phase* where it plays the action $\beta_1(\theta)$ for one period. Then the commitment type transitions again to a signaling phase for m periods, after which proceeds to a collection phase again, but this time for *two* periods. This commitment type continues along this pattern transitioning between m periods of signaling followed by a collection phase that increases in length by one period after each repetition of the signaling and collection phase. As a result, eventually the times between subsequent signaling phases become longer and longer as t increases. We defer discussion about the important features of this commitment type until after the statement of our main reputation theorem.

4.2 Reputation Theorem

In the main result of the paper, 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 opportunistic LR player will obtain payoffs arbitrarily close to the Stackelberg payoff of the complete information stage game in every equilibrium of the repeated incomplete information game (at all $\theta \in \Theta^{id}$).

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

Before presenting the proof, we first discuss the important features of the constructed commitment types here. Our example in Section 3 already suggested that reputation building may fail with only simple commitment types that are committed to playing the same (possibly mixed) action in every period. The broad intuition is that, since the uncertainty in monitoring confounds the SR player's ability to interpret the outcomes she observes, reputation building is possible only if the LR firm can both teach the SR player about the monitoring state and also the intention to play the desirable Stackelberg action. The commitment types that we constructed above do exactly this: They are committed to playing both "signaling actions" that help the consumer learn the unknown monitoring state and "collection actions" that are desirable for payoffs of the LR player. It is worth highlighting that our commitment types are non-stationary, playing a periodic strategy that alternates between signaling phases and collection phases. A similar reputation theorem can be proved also with stationary commitment types that have access to a public randomization device.[17]

Also, since the dynamic commitment types use a strategy that only depends on calendar time, our results continue to hold even if the LR player did not observe the signals that are observed by the SR players. Furthermore, as we have emphasized previously, our reputation result is robust to the inclusion of other possibly "bad" commitment types. We only require the existence of types ω^{β_1} while placing no restrictions on the existence or absence of other commitment types.

Relatedly, our theorem can also be interpreted as a robust reputation result in the following sense. Suppose that we interpret the state space Θ as a representation of the "subjective" uncertainty of the SR players about the informativeness of the public signals about the actions of the LR player. Then our main theorem states that as long as the SR players place a positive probability on the correct θ and the constructed commitment types, then the LR player can build reputation effectively in the correct state θ, regardless of the SR players' beliefs about play at any other state $\theta' \neq \theta$.

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

4.3 Necessary Characteristics of Commitment Types

Our commitment types ω^{β_1} have two key features: i) They switch play between signaling and collection phases and ii) they do so infinitely often. It turns out that these two features are important and, in a sense, necessary to reputation building. To highlight the necessity of i), we provide an example in which the opportunistic LR player regardless of his discount factor obtains a low equilibrium payoff 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 signaling phases, and again construct equilibria in which the opportunistic LR player gets payoff much below the statewise Stackelberg payoff in all states.

4.3.1 Stationary Commitment Types

Below is an example in which we allow for only stationary commitment types (that always plays the same strategy). We show that the mere existence of such types is not sufficient for reputation building. Formally, given a countable set of stationary commitment types, Ω^*, 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 opportunistic LR player obtains payoffs significantly below the statewise Stackelberg payoff.[18]

Consider the stage game described in Figure 9 whose payoffs are state independent. The Stackelberg payoff is 3 and the Stackelberg action is T. Note that L is a best-response in the stage game if and only if $\alpha_1(T) \geq \frac{3}{4}$. The set of states, $\Theta = \{\ell, r\}$, is binary with equal likelihood of both states. The signal space, $Y = \{\bar{y}, \underline{y}\}$ is also binary. The information structure is described in Figures 10 and 11.

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

Figure 9: Stage Game

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

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

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

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

Suppose we are given a set Ω^* of commitment types, each of which is associated with the play of a state-contingent action $\beta \in \mathcal{B}$ at all periods. For each $\omega \in \Omega^*$, let β^ω be the associated state contingent mixed action plan. For any pair of mixed action $\alpha \in \mathcal{A}$ such that $\alpha(T) \geq \frac{3}{4}$ and state $\theta \in \Theta$, let $\bar{\alpha}_{-\theta} \in \mathcal{A}$ be the unique mixed action such that $\pi(\cdot \mid \bar{\alpha}_{-\theta}, -\theta) = \pi(\cdot \mid \alpha, \theta)$.[19] Note that because of the symmetry of the information structure, $\bar{\alpha}_{-\theta}$ does not depend on the state $\theta \in \Theta$ and so we subsequently omit the subscript.

For each ω we construct another type $\bar{\omega}$ who also plays a stationary strategy consisting of the following

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

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

state contingent mixed action at all times:

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

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

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

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

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

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

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

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

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

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

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

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

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

In this example, the co-existence of other "bad" commitment types makes it possible for the LR player to still end up with very low payoffs in equilibrium. In contrast, notice that, Theorem 4.1 does not restrict the type space beyond requiring the existence of the appropriate commitment types.

4.3.2 Finite Type Space with Front-Loaded Signaling

Next, we present an example where we have commitment types that switch between signaling and collection, but not infinitely often: i.e., they can play signaling actions for at most N periods, and then switch to collection forever. In such environments, we show that a reputation theorem does not hold.

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

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

Figure 12: Stage Game

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

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

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

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

For notational simplicity let $\kappa := 4$.[20] Consider the following type space. Let ω^t denote a commitment type that plays B until period $t-1$ and thereafter switches to the action T forever. Let $N \in \mathbb{N}_+$ and consider the following set of types: $\Omega := \{\omega^1, \ldots, \omega^N\} \cup \{\omega^o\}$.

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

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

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

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

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

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

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

But note that given $s < t$, the strategy of ω^s in state θ generates exactly the same distribution of public signals as ω^o in state $-\theta$ at all times τ between s and t. Therefore learning between these two types ceases

[20] 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.

after time s. This allows us to simplify the above expression at any time t and history h^t:

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

But then note that the above implies:

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

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

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

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

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

The reason why reputation building fails here is that when a commitment type can teach for only for up to N periods, then whether the SR players' beliefs about the correct state are high before the switch to the Stackelberg action occurs depends on the probability of that commitment type. If this probability is too small (relative to N), then mimicking that type may not lead to sufficiently high beliefs about the correct state in the future. Thus the relative ratio between κ and the probability of the commitment type crucially matters. As a consequence, the mere existence of such commitment types is again not sufficient for effective reputation building and the fine details of the type space matter.[21]

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

4.3.3 Infinite Type Space with Front-Loaded Signaling

The finiteness of the type spaces in the above example restricts the amount of learning that can be achieved by mimicking the commitment type, and hinders reputation building. However, the reason for failure of reputation building is more subtle. In the example below, we show that even if we allowed inclusion of infinitely many of the above switching types $\{\omega^s\}_{s=t}^{\infty}$ so that the state could be taught to the SR players to any degree of precision, reputation building can still fail if all signaling is front-loaded.

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

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

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

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

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

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

We can show that in the above type space, as long as $\varepsilon > 0$ is sufficiently small, regardless of the discount factor, there always exists an equilibrium in which the opportunistic LR player plays M at all histories and the SR player always plays R. Using arguments similar to those in Subsection 4.3.2, we can show that, at any history at any time t, the SR player never assigns more than $\frac{3}{4}$ probability to the LR player playing T, which means that the SR player's best-response is to play R at all histories. As a result, there are no inter-temporal incentives for the opportunistic LR player and so it is also indeed his best-response to play M always. The interested reader may refer to the appendix for details.

Note that here, reputation building does not fail because the SR player cannot learn the true state. Indeed, as the claim below shows the SR player can learn the true state with an arbitrary level of precision.

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

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

Proof. The proof is a direct consequence of merging arguments that will be illustrated in the next section.[22] □

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

Reputation building cannot be guaranteed in this example because it may be impossible for the LR player to convince the SR player of *both the correct state and the intention to play the Stackelberg action simultaneously.* As the opportunistic LR player mimics any of these commitment types, the SR players' beliefs are converging to the correct state. But, at the same time, the SR players are placing more and more probability on the types that teach the state for longer amounts of time instead of those types that have switched play to the Stackelberg action.

5 Proof of Theorem 4.1

The proof of Theorem 4.1 proceeds in three steps.

1. The first step is to show that in any state $\theta \in \Theta^{id}$, by mimicking the strategy of the appropriate commitment type, the LR player can ensure that the SR players learn the state at a rate that is uniform *across all equilibria.*We prove this by deriving a new theorem on robust learning. This theorem establishes sufficient conditions for uniform learning across a wide class of general learning environments. Such a robust learning result is important in its own right and we establish it in Section 5.1.

2. The robust learning theorem in Section 5.1 allows us to establish an upper bound on the number of times that the SR player's belief on the true state is low. In other words, the belief, at most times, will be high on the correct state with high probability, in which case action identification is no longer problematic.

3. Finally, we 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. Standard merging arguments of Gossner (2011) can be used to construct our lower bound on payoffs.

5.1 A Robust Learning Theorem

Consider the following general model of learning. There is a finite signal space Y. Define the set of all possible stochastic processes over H^∞ as $S(Y)$. Formally, an element $\pi \in S(Y)$ is a sequence $\pi = \{\pi_t\}_{t=1}^\infty$ where for each t, $\pi_t \in \Delta(H^t)$ is a probability measure over H^t that satisfies the following consistency condition:

$$\mathbf{marg}_{H^{t-1}}\pi_t = \pi_{t-1}.$$

This allows for very general stochastic processes that may potentially contain arbitrary forms of serial correlations. By Kolmogorov's extension theorem, for each $\pi \in S(Y)$, there is a unique probability measure $\pi_\infty \in \Delta(H^\infty)$ such that $\mathbf{marg}_{H^t}\pi_\infty = \pi_t$ for all $t = 0, 1, 2, \ldots$.

A **learning environment** is then a pair $\mathcal{E} = \left(\left\{\pi^{\mathcal{E},\xi}\right\}_{\xi\in\mathbb{N}}, \rho^{\mathcal{E}}\right)$ where for each $\xi \in \mathbb{N}$, $\pi^{\mathcal{E},\xi} \in S(Y)$ and $\rho \in \Delta(\mathbb{N})$ is a prior over $\mathbb{N}$. Here $\mathbb{N}$ is interpreted as the set of states of uncertainty.[23] Let $\mathcal{L}(Y)$ be the set of all learning environments with signal space Y. Given any learning environment $\mathcal{E} \in \mathcal{L}(Y)$, we can define the induced stochastic process, $\pi^{\mathcal{E},B} = \{\pi_t^{\mathcal{E},B}\}_{t=0}^\infty$, given any event $B \subseteq \mathbb{N}$ such that $\rho^{\mathcal{E}}(B) > 0$ in the following manner:

$$\pi_t^{\mathcal{E},B}(h^t) := \frac{1}{\rho^{\mathcal{E}}(B)} \sum_{\xi\in B} \rho^{\mathcal{E}}(\xi)\pi_t^{\mathcal{E},\xi}(h^t).$$

[23] To avoid unnecessary measure theoretic complications, we assume throughout that the set of states of uncertainty is countable.

Again by the Kolmogorov extension theorem, this stochastic process also induces a unique probability measure $\pi_\infty^{\mathcal{E},B} \in \Delta(Y^\infty)$.

In a learning environment, at the beginning of each period $t = 1, 2, \ldots$, an observer updates her beliefs about the true state $\xi \in \mathbb{N}$ according to Bayes' rule upon the realization of a history of signals $h^t = (y_0, \ldots, y_{t-1})$. Let $\rho_t^{\mathcal{E}}(\cdot \mid h^t) \in \Delta(\mathbb{N})$ denote the observer's beliefs at the beginning of period t (before the realization of the t-period signal) about the state after observing h^t. Analytically, due to its tractability, it is oftentimes more convenient to study the likelihood ratios of events $B, C \subseteq \mathbb{N}$ when $\rho^{\mathcal{E}}(C) > 0$:

$$\frac{\rho_t^{\mathcal{E}}(B \mid h^t)}{\rho_t^{\mathcal{E}}(C \mid h^t)} = \frac{\rho^{\mathcal{E}}(B)}{\rho^{\mathcal{E}}(C)} \frac{\pi_t^{\mathcal{E},B}(h^t)}{\pi_t^{\mathcal{E},C}(h^t)}.$$

We now describe formally our definition of **robust learning**.

Definition 5.1. Let $\xi^* \in A \subseteq \mathbb{N}$ and $\mathcal{S} \subseteq \mathcal{L}$. Then we say that an observer **$\mathcal{S}$-robustly learns** A at ξ^* if for every $\nu > 0$, there exists some K such that

$$\inf_{\mathcal{E} \in \mathcal{S}} \pi_\infty^{\mathcal{E},\xi^*} \left(\bigcap_{t=K}^{\infty} \left\{ h^\infty : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} \leq \nu \right\} \right) \geq 1 - \nu.$$

Intuitively, $\mathcal{S}$-robust learning requires an observer to hold high beliefs on A forever after K periods with high probability for all learning environments in $\mathcal{S}$.

Our main theorem in this section establishes a simple sufficient condition on $\mathcal{S}$ that guarantees $\mathcal{S}$-robust learning of A at ξ^*. To state it, we first need a few definitions that are well-known from the theory of experiments. First fix a learning environment $\mathcal{E} \in \mathcal{L}(Y)$, some $\xi^* \in \mathbb{N}$, and $B \subseteq \mathbb{N}$. We now define the function, $\mathcal{H}_t^{\mathcal{E}}(\cdot\,; B, \xi^*) : [0,1] \to \mathbb{R}$, which is also known as the *Hellinger transform* of an information process from the theory of experiments.[24] Formally this function is defined in the following manner. If $\rho^{\mathcal{E}}(B) = 0$, then we simply define it to be the following piecewise function:

$$\mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*) = \begin{cases} 0 & \text{if } \lambda \in (0,1) \\ 1 & \text{if } \lambda \in \{0,1\}. \end{cases}$$

On the other hand if $\rho^{\mathcal{E}}(B) > 0$, we define it as follows:

$$\mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*) := \sum_{h^t \in H^t} \left(\pi_t^{\mathcal{E},B}(h^t)\right)^\lambda \left(\pi_t^{\mathcal{E},\xi^*}(h^t)\right)^{1-\lambda} = \mathbb{E}\left[\left(\frac{\pi_t^{\mathcal{E},B}(h^t)}{\pi_t^{\mathcal{E},\xi^*}(h^t)} \right)^\lambda \mid \xi^* \right].$$

This is essentially the moment generating function of the log-likelihood ratio of the t-period information process where the true stochastic process is given by ξ^*. In Appendix B, we list some important properties of the Hellinger transform. Toward our robust learning result, let us first define the following for any $\mathcal{E} \in \mathcal{L}(Y)$ and $\xi^* \notin B \subseteq \mathbb{N}$:

$$\mathcal{H}_t^{\mathcal{E}}(B, \xi^*) = \inf_{\lambda \in [0,1]} \mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*).$$

The Hellinger transform is useful as it provides a sufficient statistic for the lower bound on the probability

[24] See for example **?**.

of learning after some time K. The following lemma formalizes this lower bound.[25]

Lemma 5.2. *Let $\mathcal{E} \in \mathcal{L}(Y)$, $\xi^* \in \mathbb{N}$, and $A \subseteq \mathbb{N}$ such that $\xi^* \in A$. Suppose that $\sum_{t=1}^{\infty} \mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*) < +\infty$ and that $\rho^{\mathcal{E}}(\xi^*) > 0$. Then for all K and any $\nu > 0$,*

$$\pi_{\infty}^{\mathcal{E},\xi^*}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty} : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} \leq \nu\right\}\right) \geq 1 - \max\{1, 1/(\nu\rho^{\mathcal{E}}(\xi^*))\}\sum_{t=K}^{\infty}\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*).$$

Proof. In the Appendix. □

Note that the above lemma shows that when $\{\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*)\}_{t=0}^{\infty}$ converges to 0 faster, learning is guaranteed to occur at a faster rate since the lower bound increases for all K.[26] Moreover the lower bound of the probability of learning established in the above lemma depend only on four parameters: ν, K, $\rho^{\mathcal{E}}(\xi^*)$, and $\sum_{t=K}^{\infty}\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*)$. In particular other aspects of the learning environment do not influence this lower bound. As a result, this lemma yields the following robust learning theorem.

Theorem 5.3. *Let $\mathcal{S} \subseteq \mathcal{L}(Y)$ and define for every t, $\mathcal{H}_t^{\mathcal{S}}(A^c, \xi^*) = \sup_{\mathcal{E}\in\mathcal{S}}\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*)$. Suppose that $\sum_{t=1}^{\infty}\mathcal{H}_t^{\mathcal{S}}(A^c, \xi^*) < +\infty$ and that $\rho^{\mathcal{E}}(\xi^*) \geq \varepsilon > 0$ for all $\mathcal{E} \in \mathcal{S}$. Then an observer $\mathcal{S}$-robustly learns A at ξ^*.*

Proof. By the previous lemma, for every $\mathcal{E} \in \mathcal{S}$ and every K,

$$\pi_{\infty}^{\mathcal{E},\xi^*}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty} : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} \leq \nu\right\}\right) \geq 1 - \max\{1, 1/(\nu\rho_{\mathcal{E}}(\xi^*))\}\sum_{t=K}^{\infty}\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*).$$

Because $\rho^{\mathcal{E}}(\xi^*) \geq \varepsilon$ and by the definition of $\mathcal{H}_t^{\mathcal{S}}(A^c, \xi^*)$, we have:

$$\pi_{\infty}^{\mathcal{E},\xi^*}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty} : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} \leq \nu\right\}\right) \geq 1 - \max\{1, 1/(\nu\varepsilon)\}\sum_{t=K}^{\infty}\mathcal{H}_t^{\mathcal{S}}(A^c, \xi^*).$$

Because $\sum_{t=0}^{\infty}\mathcal{H}_t^{\mathcal{S}}(A^c, \xi^*) < +\infty$, there exists some K sufficiently large such that $\max\{1, 1/(\nu\varepsilon)\}\sum_{t=K}^{\infty}\mathcal{H}_t^{\mathcal{S}}(A^c, \xi^*) < \nu$. Thus for such a K,

$$\pi_{\infty}^{\mathcal{E},\xi^*}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty} : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} \leq \nu\right\}\right) \geq 1 - \nu.$$

Since the above holds for all $\mathcal{E} \in \mathcal{S}$, this completes the proof. □

The following corollary will be useful: It shows that if we can guarantee $\mathcal{S}$-robust learning of a finite collection of sets at ξ^*, then we can guarantee $\mathcal{S}$-robust learning of the intersection of these sets at ξ^*.

Corollary 5.4. *Let $\xi^* \in A_1, \ldots, A_n \subseteq \mathbb{N}$ and that $\mathcal{S} \subseteq \mathcal{L}(Y)$. Suppose that $\sum_{t=0}^{\infty}\mathcal{H}_t^{\mathcal{S}}(A_\ell, \xi^*) < +\infty$ for all $\ell = 1, 2, \ldots, n$ and that $\rho^{\mathcal{E}}(\xi^*) \geq \varepsilon > 0$ for all $\mathcal{E} \in \mathcal{S}$. Then the observer $\mathcal{S}$-robustly learns $A_1 \cap A_2 \cap \cdots \cap A_n$ at ξ^*.*

Proof. In the Appendix. □

[25] Moscarini and Smith (2002) use the Hellinger transform to compare the informational value of experiments when the experiments are repeated sufficiently many times in an i.i.d. manner. Our learning environments are very different in that they are not derived from just i.i.d experiments; there may be potentially arbitrary serial correlation in public signals.

[26] In this paper, we do not address whether this lower bound is sharp. Whether there exists any learning environments for which the lower bound in the lemma is exactly hit is an open question.

5.2 Uniform Signaling of the State in Reputation Building

Given any equilibrium, σ, we can interpret it as a learning environment along the same lines as in Section 5.1. To this end, for a given equilibrium, σ, and for each $(\omega, \theta) \in \Omega \times \Theta$, let

$$\left\{\pi_t^{\sigma,(\omega,\theta)}\right\}_{t=1}^{\infty} \in \mathcal{S}(Y)$$

denote the stochastic process over public histories induced by the strategy played by type ω in state θ in the equilibrium σ where $\pi_t^{\sigma,(\omega,\theta)} \in \Delta(H^t)$. Note that this is exactly in line with the notation from Section 5.1 where we view σ as the learning environment $\mathcal{E}$ and view $\Omega \times \Theta$ as the set of states of uncertainty $\mathbb{N}$.[27] Again by Kolmogorov's extension theorem, there exists a unique probability measure $\pi_\infty^{\sigma,\omega}$ such that for each t, $\mathbf{marg}_{H^t}\pi_\infty^{\sigma,(\omega,\theta)} = \pi_t^{\sigma,(\omega,\theta)}$.

As in Section 5.1, given any event $A \subseteq \Omega \times \Theta$, we can define the stochastic process, $\pi^{\sigma,A} := \{\pi_t^{\sigma,A}\}_{t=1}^{\infty}$ induced over public histories when conditioning on A in the following manner:

$$\pi_t^{\sigma,A}(h^t) = \frac{1}{\gamma(A)} \sum_{(\omega,\theta)\in A} \gamma(\omega,\theta)\pi_t^{\sigma,(\omega,\theta)}(h^t).$$

Finally, given any prior $\nu \in \Delta(\Theta)$ and any equilibrium σ, we can compute the SR players' conditional belief about the state $\theta \in \Theta$ after a history h^t, $\nu^\sigma(\cdot \mid h^t) \in \Delta(\Theta)$, in the following manner:

$$\nu^\sigma(\theta \mid h^t) := \nu(\theta)\frac{\pi_t^{\sigma,\Omega\times\{\theta\}}(h^t)}{\pi_t^{\sigma,\Omega\times\Theta}(h^t)}.$$

Naturally, there is an alternative but equivalent way of viewing the stochastic processes $\{\pi_t^{\sigma,A}\}_{t=0}^{\infty}$. Note that we can represent these processes as simply a sequence of conditional distributions over Y:

$$\pi_t^{\sigma,A}(y \mid h^{t-1}) = \frac{\pi_t^{\sigma,A}((h^{t-1}, y))}{\pi_{t-1}^{\sigma,A}(h^{t-1})}.$$

Thus sometimes we will represent this stochastic process as $\pi^{\sigma,A} = \left\{\{\pi_t^{\sigma,A}(\cdot \mid h^{t-1})\}_{h^{t-1}\in H^{t-1}}\right\}_{t=1}^{\infty}$.

Of course, when we view an equilibrium σ as a learning environment, we can also define the appropriate Hellinger transforms. This will be useful for studying the evolution of beliefs by the SR players. Thus, for any σ an equilibrium and any event $A \subseteq \Omega \times \Theta$, we define the Hellinger transform as follows:

$$\mathcal{H}_t^\sigma(\lambda; A, (\omega, \theta)) = \sum_{h^t \in H^t} \pi_t^{\sigma,A}(h^t)^\lambda \pi_t^{\sigma,(\omega,\theta)}(h^t)^{1-\lambda}.$$

We also define accordingly the following:

$$\mathcal{H}_t^\sigma(A, (\omega, \theta)) = \inf_{\lambda\in[0,1]} \mathcal{H}_t^\sigma(\lambda; A, (\omega, \theta)),$$
$$\mathcal{H}_t^{BE}(A, (\omega, \theta)) = \sup_{\sigma\in BE} \mathcal{H}_t^\sigma(A, (\omega, \theta)).$$

We first bound $\mathcal{H}_t^\sigma(\Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta))$ across all equilibria σ.

[27] Recall that we assumed throughout that $\Omega \times \Theta$ was countable.

Lemma 5.5. *Fix any θ and let $\theta_j \neq \theta$ and let $\lambda \in (0,1)$. Then there exists some $\varepsilon > 0$ such that for any equilibrium σ and any $k \in \mathbb{N}$,*

$$\mathcal{H}^{\sigma}_{j+n_k+1}(\lambda; \Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta)) \leq (1-\varepsilon)^k.$$

As a result, there exists some $\varepsilon > 0$ such that for any $k \in \mathbb{N}$,

$$\mathcal{H}^{BE}_{j+n_k+1}(\Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta)) \leq (1-\varepsilon)^k.$$

The proof is in the Appendix. Given the above lemma, we can apply our robust learning results in the context of our reputation game. This yields the desired uniform learning result across all equilibria.

Proposition 5.6. *Suppose that $\mu(\omega^{\beta_1}) > 0$. Then for every $\kappa > 0$, there exists some $K > 0$ such that for all $\theta \in \Theta$ and all $\sigma \in BE$,*

$$\pi^{\sigma,(\omega^{\beta_1},\theta)}_{\infty}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty} : \frac{\nu^{\sigma}_t(\Theta \setminus \{\theta\} \mid h^t)}{\nu^{\sigma}_t(\theta \mid h^t)} \leq \kappa\right\}\right) \geq 1-\kappa.$$

Proof. Fix some θ and first consider any $\theta_j \neq \theta$. First note that by the previous lemma, there exists some $\varepsilon_j > 0$ such that for all k,

$$\mathcal{H}^{BE}_{j+n_k}(\Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta)) < (1-\varepsilon_j)^k.$$

Then since $\mathcal{H}^{BE}_t(\Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta))$ is non-increasing,

$$\sum_{t=1}^{\infty}\mathcal{H}^{BE}_t(\Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta)) \leq \sum_{k=0}^{\infty}(m+k+1)(1-\varepsilon_j)^k = \frac{m}{\varepsilon_j} + \sum_{k=0}^{\infty}(k+1)(1-\varepsilon_j)^k = \frac{m}{\varepsilon_j} + \frac{1}{\varepsilon_j^2} < +\infty.$$

Thus, we have shown that for each j such that $\theta_j \neq \theta$,

$$\sum_{t=1}^{\infty}\mathcal{H}^{BE}_t(\Omega \times \{\theta_j\}, (\omega^{\beta_1}, \theta)) < +\infty.$$

By Corollary 5.4, for every $\kappa > 0$, there exists some K_θ such that

$$\inf_{\sigma \in BE}\pi^{\sigma,(\omega^{\beta_1},\theta)}_{\infty}\left(\bigcap_{t=K_\theta}^{\infty}\left\{h^{\infty} : \frac{\nu^{\sigma}_t(\Theta \setminus \{\theta\} \mid h^t)}{\nu^{\sigma}_t(\theta \mid h^t)} \leq \kappa\right\}\right) \geq 1-\kappa.$$

Letting $K = \max_{\theta \in \Theta} K_\theta$, we obtain the desired conclusion. □

5.3 Applying Merging

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

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

recall the definition of the Kullback-Leibler divergence of two probability measures: Given two probability measures $P, Q \in \Delta(Y)$,

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

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

Definition 5.7. Let $(\lambda, \varepsilon) \in [0,1]^2$. Then $(\alpha_1, \alpha_2) \in \mathcal{A}_1 \times \mathcal{A}_2$ is a (λ, ε)-confirmed best-response at θ if there exists some $(\beta_1, \nu) \in \mathcal{B} \times \Delta(\Theta)$ such that

- α_2 is a best-response for player 2 to β_1 given belief ν about the state,
- $\frac{1-\nu(\theta)}{\nu(\theta)} < \varepsilon$,
- and $D(\pi(\cdot \mid \alpha_1, \theta)\|\pi(\cdot \mid \beta_1, \nu)) < \lambda$.

Lemma 5.8. *Let $\rho > 0$. Then there exists some $\lambda^* > 0$ and $\varepsilon^* > 0$ such that for all (α_1, α_2) that is a $(\lambda^*, \varepsilon^*)$-confirmed best-response at $\theta \in \Theta$, $u_1(\alpha_1, \alpha_2, \theta) > \inf_{\alpha_2' \in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2', \theta) - \rho$.*

Proof. This lemma is a standard continuity result. □

We first define the following set of histories given an equilibrium σ and a type $\omega \in \Omega$:

$$\mathcal{M}^{\sigma,(\omega,\theta)}(J,\lambda) := \left\{h \in H^\infty : \left|\left\{t : D\left(\pi_t^{\sigma,(\omega,\theta)}(\cdot \mid h^{t-1})\|\pi_t^{\sigma,\Omega\times\Theta}(\cdot \mid h^{t-1})\right) \geq \lambda\right\}\right| < J\right\}.$$

Lemma 5.9. *Let $k > 0$, $\beta_1 \in \mathcal{B}$. Then*

$$\pi_\infty^{\omega^{\beta_1},\theta}\left(\mathcal{M}^{\sigma,(\omega^{\beta_1},\theta)}(J,\lambda)\right) \geq 1 + \frac{\log\left(\gamma(\omega^{\beta_1},\theta)\right)}{J\lambda}.$$

Proof. See Appendix D for the proof. □

Together with Lemma 5.9 and Proposition 5.6, we can now complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Define $\underline{u} := \min_{a\in A}\min_{\theta\in\Theta} u_1(a,\theta)$. Choose any $\theta \in \Theta^{id}$. We will show that there exists some $\delta^* < 1$ such that whenever $\delta > \delta^*$, the LR opportunistic type obtains a payoff of at least $u_1^*(\theta) - \rho$ in every equilibrium. This then proves the theorem, since there are finitely many states $\theta \in \Theta$.

First choose some $\varepsilon^* > 0$ such that for all $\varepsilon < \varepsilon^*$,

$$(1-2\varepsilon)\left(u_1^*(\theta) - \frac{\rho}{4}\right) + 2\varepsilon\underline{u} > u_1^*(\theta) - \rho.$$

By assumption, we can choose $\beta_1 \in \mathcal{S}^{\rho/8}$ such that $\mu(\omega^{\beta_1}) > 0$. By Lemma 5.8, there exists some $\varepsilon \in (0, \varepsilon^*)$ such that

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

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

By Proposition 5.6, there exists some K such that in any equilibrium σ,

$$\pi_{\infty}^{\sigma,(\omega^{\beta_1},\theta)}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty}:\frac{\nu_t^{\sigma}(\Theta\setminus\{\theta\}\mid h^t)}{\nu_t^{\sigma}(\theta\mid h^t)}\leq\varepsilon\right\}\right)\geq 1-\varepsilon.$$

Therefore, if we choose J sufficiently large such that $-\frac{\log\left(\gamma(\omega^{\beta_1},\theta)\right)}{J\varepsilon}<\varepsilon$, by Lemma 5.9, then in any equilibrium σ,

$$\pi_{\infty}^{\sigma,(\omega^{\beta_1},\theta)}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty}:\frac{\nu_t^{\sigma}(\Theta\setminus\{\theta\}\mid h^t)}{\nu_t^{\sigma}(\theta\mid h^t)}\leq\varepsilon\right\}\cap\mathcal{M}^{\sigma,(\omega^{\beta_1},\theta)}(J,\varepsilon)\right)\geq 1-2\varepsilon.$$

As a result, in any equilibrium, σ, by mimicking the strategy of the commitment type ω^{β_1}, the LR player 1 obtains at least the following payoff:

$$(1-2\varepsilon)\left((1-\delta^{K+J})\underline{u}+\delta^{K+J}\left(u_1^*(\theta)-\frac{\rho}{4}\right)\right)+2\varepsilon\underline{u}.$$

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

$$(1-2\varepsilon)\left((1-\delta^{K+J})\underline{u}+\delta^{K+J}\left(u_1^*(\theta)-\frac{\rho}{4}\right)\right)+2\varepsilon\underline{u}>u_1^*(\theta)-\rho.$$

□

6 Upper Bound on Payoffs

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

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

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

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

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

6.1 Example

The following example shows that the probability of commitment types matters for the upper bound regardless of the discount factor. Consider the quality choice game with the following stage game payoffs: In the repeated game this stage game is repeatedly played and all payoffs are common knowledge. Note that the Stackelberg payoff of the above game is 3/2. Furthermore, note that L is a best-response for the SR player in the stage game if and only if $\alpha_1(T) \geq 1/2$.

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

Figure 15: Quality Choice

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

6.2 Upper Bound Theorem

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

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

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

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

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

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

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

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

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

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

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

6.2.1 Statewise Payoff Bounds and Payoff Uniqueness

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

Corollary 6.6. *Suppose that*

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

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

The proof is in the Appendix. A key distinction between Theorem 6.5 and the above corollary is that the upper bound on payoffs is given in each state. A key step in the proof of this state-wise upper bound in the corollary relies on the assumption that the constructed commitment types exist with positive probability. This assumption is important for the argument as it first allows us to provide a lower bound on payoffs in each state using Theorem 4.1, which then together with the ex-ante payoff upper bound of Theorem 6.5 allows us to establish the upper bound in each state. Thus without the existence of such commitment types, our proof would not go through.[32]

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

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2016): "Fragility of asymptotic agreement under Bayesian learning," *Theoretical Economics*, 11, 187–225.

Al-Najjar, N. I. (2009): "Decision makers as statisticians: Diversity, ambiguity, and learning," *Econometrica*, 77(5), 1371–1401.

Al-Najjar, N. I., and M. M. Pai (2014): "Coarse decision making and overfitting," *Journal of Economic Theory*, 150, 467–486.

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

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

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

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

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

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

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

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

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

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

——— (1992): "Maintaining a Reputation when Strategies are Imperfectly Observed," *Review of Economic Studies*, 59(3), 561–579.

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

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

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

Heidhues, P., B. Köszegi, and P. Strack (Forthcoming): "Unrealistic Expectations and Misguided Learning," *Econometrica*.

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.

Moscarini, G., and L. Smith (2002): "The law of large demand for information," *Econometrica*, 70(6), 2351–2366.

Pei, H. (2017): "Reputation Effects under Interdependent Values," mimeo.

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

Torgersen, E. (1991): *Comparison of statistical experiments*, vol. 36. Cambridge University Press.

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

A Infinite Type Space with Front-loaded Signaling

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

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

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

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

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

B Proofs of Robust Learning

B.1 Properties of the Hellinger Transform

Below, we list some important properties of the Hellinger transform, that we will use later.

Lemma B.1. *Let $\mathcal{E} \in \mathcal{L}(Y)$, $\xi^* \in \mathbb{N}$, $B \subseteq \mathbb{N}$ such that $\xi^* \notin B$. Then $\mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*)$ satisfy the following properties:*

1. *For all t and all $\lambda \in (0,1)$, $0 \leq \mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*) \leq \mathcal{H}_t^{\mathcal{E}}(0; B, \xi^*) = \mathcal{H}_t^{\mathcal{E}}(1; B, \xi^*) = 1$.*
2. *For all t and all $\lambda \in (0,1)$, $\mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*) = 1$ if and only if $\pi_t^{\mathcal{E},B}(h^t) = \pi_t^{\mathcal{E},\xi^*}(h^t)$ for all $h^t \in supp(\pi_t^{\mathcal{E},\xi^*})$.*
3. *$\mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*)$ is differentiable at $\lambda = 0$ and moreover the derivative at $\lambda = 0$ is given by negative of the Kullback-Leibler divergence (relative entropy):*

$$\frac{d}{d\lambda} \mathcal{H}_t^{\mathcal{E}}(0; B, \xi^*) = -\mathbb{E}\left[\log \frac{\pi_t^{\mathcal{E},\xi^*}(h^t)}{\pi_t^{\mathcal{E},B}(h^t)} \mid \xi^*\right] \leq 0.$$

4. $\mathcal{H}_t^{\mathcal{E}}(\lambda; B, \xi^*)$ is weakly decreasing in t for every $\lambda \in [0,1]$.

Proof. Proofs of these claims are standard. □

Note that the above lemma shows that the Hellinger transform indeed satisfies some intuitive properties and moreover contains more information about the learning environment than the Kullback-Leibler divergence since it can be computed from the Hellinger transform.

B.2 Proof of Lemma 5.2

Proof. If $\rho^{\mathcal{E}}(A^c) = 0$, then the lemma holds trivially so let us assume that $\rho^{\mathcal{E}}(A^c) > 0$. First note that for any $\lambda \in [0,1]$,

$$\begin{aligned}
\pi_t^{\mathcal{E},\xi^*}\left(\frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} > \nu\right) &\leq \pi_t^{\mathcal{E},\xi^*}\left(\frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} > \nu\right) \\
&= \pi_t^{\mathcal{E},\xi^*}\left(\left(\frac{\rho_{\mathcal{E}}(A^c)}{\rho_{\mathcal{E}}(\xi^*)}\right)^{\lambda}\left(\frac{\pi_t^{\mathcal{E},A^c}(h^t)}{\pi_t^{\mathcal{E},\xi^*}(h^t)}\right)^{\lambda} > \nu^{\lambda}\right) \\
&\leq \pi_t^{\mathcal{E},\xi^*}\left(\left(\frac{\pi_t^{\mathcal{E},A^c}(h^t)}{\pi_t^{\mathcal{E},\xi^*}(h^t)}\right)^{\lambda} > (\nu\rho_{\mathcal{E}}(\xi^*))^{\lambda}\right) \\
&\leq \frac{1}{(\nu\rho_{\mathcal{E}}(\xi^*))^{\lambda}}\mathbb{E}\left[\left(\frac{\pi_t^{\mathcal{E},A^c}(h^t)}{\pi_t^{\mathcal{E},\xi^*}(h^t)}\right)^{\lambda} \mid \xi^*\right] \leq \max\{1, 1/(\nu\rho_{\mathcal{E}}(\xi^*))\}\mathcal{H}_t^{\mathcal{E}}(\lambda; A^c, \xi^*),
\end{aligned}$$

where the second to last inequality follows from Markov's inequality. Since the above holds for every $\lambda \in [0,1]$, we have:

$$\pi_t^{\mathcal{E},\xi^*}\left(\frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} > \nu\right) \leq \max\{1, 1/(\nu\rho_{\mathcal{E}}(\xi^*))\}\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*).$$

Then we have:

$$\begin{aligned}
\pi_\infty^{\mathcal{E},\xi^*}\left(\bigcap_{t=K}^{\infty}\left\{h^\infty : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} \leq \nu\right\}\right) &= 1 - \pi_\infty^{\mathcal{E},\xi^*}\left(\bigcup_{t=K}^{\infty}\left\{h^\infty : \frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} > \nu\right\}\right) \\
&\geq 1 - \sum_{t=K}^{\infty}\pi_t^{\mathcal{E},\xi^*}\left(\frac{\rho_t^{\mathcal{E}}(A^c \mid h^t)}{\rho_t^{\mathcal{E}}(A \mid h^t)} > \nu\right) \\
&\geq 1 - \max\{1, 1/(\nu\rho_{\mathcal{E}}(\xi^*))\}\sum_{t=K}^{\infty}\mathcal{H}_t^{\mathcal{E}}(A^c, \xi^*).
\end{aligned}$$

□

B.3 Proof of Corollary 5.4

Proof. Let $\nu > 0$. By studying the proofs of Lemma 5.2 and Theorem 5.3, it is clear that we can easily modify the arguments to conclude more strongly that for each ℓ, there exists some K_ℓ such that

$$\inf_{\mathcal{E}\in\mathcal{S}}\pi_\infty^{\mathcal{E},\xi^*}\left(\bigcap_{t=K_\ell}^{\infty}\left\{h^\infty : \frac{\rho_t^{\mathcal{E}}(A_\ell^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} \leq \frac{\nu}{n}\right\} \middle| \xi^*\right) \geq 1 - \frac{\nu}{n}.$$

Let $K = \max\{K_1, \dots, K_\ell\}$ and note that

$$\sup_{\mathcal{E} \in \mathcal{S}} \pi_\infty^{\mathcal{E}, \xi^*} \left(\bigcup_{\ell=1}^{n} \bigcup_{t=K}^{\infty} \left\{ y^\infty : \frac{\rho_t^{\mathcal{E}}(A_\ell^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} > \frac{\nu}{n} \right\} \Bigg| \ \xi^* \right) \leq \nu.$$

Then we have:

$$\begin{aligned}
\inf_{\mathcal{E} \in \mathcal{S}} \pi_\infty^{\mathcal{E}, \xi^*} \left(\bigcap_{t=K}^{\infty} \left\{ h^\infty : \frac{\rho_t^{\mathcal{E}}((\bigcap_{\ell=1}^n A_\ell)^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} \leq \nu \right\} \right) &= \inf_{\mathcal{E} \in \mathcal{S}} \pi_\infty^{\mathcal{E}, \xi^*} \left(\bigcap_{t=K}^{\infty} \left\{ h^\infty : \frac{\rho_t^{\mathcal{E}}(\bigcup_{\ell=1}^n A_\ell^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} \leq \nu \right\} \right) \\
&\geq \inf_{\mathcal{E} \in \mathcal{S}} \pi_\infty^{\mathcal{E}, \xi^*} \left(\bigcap_{t=K}^{\infty} \left\{ h^\infty : \sum_{\ell=1}^{n} \frac{\rho_t^{\mathcal{E}}(A_\ell^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} \leq \nu \right\} \right) \\
&\geq \inf_{\mathcal{E} \in \mathcal{S}} \pi_\infty^{\mathcal{E}, \xi^*} \left(\bigcap_{\ell=1}^{n} \bigcap_{t=K}^{\infty} \left\{ h^\infty : \frac{\rho_t^{\mathcal{E}}(A_\ell^c \mid h^t)}{\rho_t^{\mathcal{E}}(\xi^* \mid h^t)} \leq \frac{\nu}{n} \right\} \right) \\
&\geq 1 - \nu.
\end{aligned}$$

By definition of $\mathcal{S}$-robust learning, this proves the claim. □

C Uniform Learning across All Equilibria

Proof of Lemma 5.5. Fix any $\lambda \in (0,1)$. Since by construction, $\pi(\cdot \mid \alpha_1, \theta_j) \neq \pi(\cdot \mid \alpha_1(\theta, \theta_j), \theta)$ for all $\alpha_1 \in \mathcal{A}$, by Lemma B.1, there exists some $\varepsilon > 0$ such that

$$\sup_{\alpha_1 \in \mathcal{A}} \sum_{y \in Y} \pi(y \mid \alpha_1, \theta_j)^\lambda \pi(y \mid, \alpha_1(\theta, \theta_j), \theta)^{1-\lambda} \leq 1 - \varepsilon.$$

Note that this chosen ε only depends on the information structure π and is independent of the chosen equilibrium, commitment types, etc.

First the claim holds trivially for $k = 0$. By induction, suppose that the claim holds for $t' = n_{k-1} + j + 1$ and consider the claim for $t = n_k + j + 1$. For any equilibrium, let us define $\bar{\sigma}(h^t, \theta) \in \mathcal{A}_1$ to be the expected action distribution of the LR player after history h^t when conditioning on the state θ:

$$\bar{\sigma}(h^t, \theta)(a_1) = \sum_{\omega \in \Omega} \mu(\omega \mid h^t, \theta) \sigma(a_1 \mid h^t, \theta, \omega).$$

Then by the law of iterated expectations, note that

$$\begin{aligned}
\mathcal{H}_t(\lambda; \Omega \times \theta_j, (\omega_1^\beta, \theta)) &= \mathbb{E}\left[\left(\frac{\pi_t^{\sigma, \Omega \times \{\theta_j\}}(h^t)}{\pi_t^{\sigma, (\omega^{\beta_1}, \theta)}(h^t)} \right)^\lambda \mid (\omega^{\beta_1}, \theta) \right] \\
&= \mathbb{E}\left[\left(\frac{\pi_t^{\sigma, \Omega \times \{\theta_j\}}(h^{t-1})}{\pi_t^{\sigma, (\omega^{\beta_1}, \theta)}(h^{t-1})} \right)^\lambda \mathbb{E}\left[\left(\frac{\pi(y_t \mid \bar{\sigma}(h^{t-1}, \theta_j), \theta_j)}{\pi(y_t \mid \alpha_1(\theta, \theta_j), \theta)} \right)^\lambda \mid (\omega^{\beta_1}, \theta), h^{t-1} \right] \mid (\omega^{\beta_1}, \theta) \right] \\
&\leq (1 - \varepsilon) \mathbb{E}\left[\left(\frac{\pi_t^{\sigma, \Omega \times \{\theta_j\}}(h^{t-1})}{\pi_t^{\sigma, (\omega^{\beta_1}, \theta)}(h^{t-1})} \right)^\lambda \mid (\omega^{\beta_1}, \theta) \right] = (1 - \varepsilon) \mathcal{H}_{t-1}(\lambda; \Omega \times \theta_j, (\omega_1^\beta, \theta)).
\end{aligned}$$

Again by Lemma B.1, since $\mathcal{H}_t$ is a non-increasing sequence, we have:

$$\mathcal{H}_t(\lambda; \Omega \times \theta_j, (\omega_1^\beta, \theta)) \leq (1 - \varepsilon) \mathcal{H}_{t'}(\lambda; \Omega \times \theta_j, (\omega_1^\beta, \theta)) \leq (1 - \varepsilon)^k.$$

Since the above holds for fixed $\lambda > 0$, the claim also holds for the infimum over $\lambda \in [0, 1]$. □

D Merging and Confirmed Best-Responses

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

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

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

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

With this, we can prove Lemma 5.9.

Proof of Lemma 5.9. Note that by the chain rule for relative entropy,

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

For every T,

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

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

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

Then by Markov's inequality,

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

□

E Proof of Theorem 6.5

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

$$\nu_{\bar{\sigma}}(h^t) := \left(\nu_{\bar{\sigma}}(\theta_0 \mid h^t), \nu_{\bar{\sigma}}(\theta_1 \mid h^t), \ldots, \nu_{\bar{\sigma}}(\theta_{m-1} \mid h^t)\right).$$

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

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

We begin with a couple lemmata.

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

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

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

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

Then define the function

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

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

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

G_ε is continuous in ε and this proves the claim. □

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

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

Proof. Consider the following function:

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

Then G_ρ is continuous in ρ and thus proves the claim. □

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

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

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

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

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

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

This then implies the desired conclusion. □

All that remains is to show that for every $\varepsilon > 0$, there exists some $\delta^* < 1$ such that for all $\delta > \delta^*$ and any equilibrium (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 6.5. By Lemma E.2, there exists some $\rho > 0$ such that

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

Choose $n \in \mathbb{N}$ such that $\frac{1}{n}(\overline{u} - \underline{u}) < \varepsilon/4$. By Lemma E.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 by Markov's inequality,

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

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. □

F Proof of Claim 6.3

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

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

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

We want to bound $V(\nu)$. To do this, we consider four cases.
Case 1: $\nu \geq 3/4$

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

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

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

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

Therefore,

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

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

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

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

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

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

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

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

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

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

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

Therefore

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

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

G Proof of Corollary 6.6

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

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

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

By Theorem 6.5, there exists some $\rho^* > 0$ such that whenever $\mu(\Omega^c) < \rho^*$, there exists some $\delta^* < 1$ such that for all $\delta > \delta^*$ and all equilibrium strategy profiles (σ_1, σ_2),

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

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

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

This then implies that for all n,

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

Then for each n, we can find some θ_n such that

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

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

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

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