Theoretical Economics 20 (2025), 169–208 1555-7561/20250169

Reputation building under uncertain monitoring

Joyee Deb
Department of Economics, New York University

Yuhta Ishii
Department of Economics, Pennsylvania State University

We study the standard reputation model with a long-run (LR) player facing a sequence of short-run (SR) opponents, with one difference: the SR players are uncertain about the monitoring structure, while the LR player knows it. We construct examples where the standard reputation result breaks down: Even if there is a possibility that the LR player is a commitment type who always plays the action to which he wants to commit, there exist "bad" equilibria in which the LR player gets payoffs substantially lower than his commitment payoffs. In contrast, if there is the possibility of dynamic commitment types who switch between "signaling" actions that help the SR players learn the monitoring structure and "collection" actions that are desirable for payoffs, our main theorem shows that a sufficiently patient LR player obtains payoffs of at least the commitment payoffs in each state in every equilibrium.

Keywords. Reputation, monitoring, repeated games, learning.

JEL classification. C73, L14.

1. Introduction

Consider a long-run firm building a reputation for producing environmentally-friendly products. Such a reputation is valuable for the firm when consumers care about the environmental impact of their purchases and are often willing to pay more for green products. Consumers make purchase decisions based on whether products have "eco-friendly" labels, but are typically unsure of how much to trust the labels. Many of these labels are genuine certifications with stringent standards, but numerous others have been discredited as being fake. As a result, on seeing an eco-label, consumers are uncertain about its informational content, and may not be convinced about the product

Joyee Deb: joyee.deb@nyu.edu
Yuhta Ishii: yxi5014@psu.edu
For helpful comments that significantly improved the paper, we thank the three anonymous referees, as well as Dilip Abreu, Heski Bar-Isaac, Martin Cripps, Mehmet Ekmekci, Drew Fudenberg, Johannes Hörner, Michihiro Kandori, Barry Nalebuff, Aniko Öry, Harry Pei, Andy Skrzypacz, Alex Wolitzky, and Jidong Zhou. We also thank Haoning Chen and Kirtivardhan Singh for excellent research assistance. Finally, we are grateful to seminar participants at Brown, Duke, 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.

 https://doi.org/10.3982/TE4758

being environmentally friendly.[1] But if consumers do not trust product labeling, a firm, even after honest investment in green products and after undergoing reliable labeling, may find it difficult to establish a positive reputation and convince consumers that its products are indeed environmentally friendly. This motivates the central question of the paper: Can reputations 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)) consider a long-run (LR) agent (a firm) who repeatedly interacts with short-run (SR) 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 in every period. Even when the actions of the firm are noisily observed, the classical reputation result states that if a sufficiently rich set of commitment types occurs with positive probability, a patient firm can achieve payoffs arbitrarily close to their Stackelberg payoff of the stage game in *every* equilibrium.[2] Intuitively by mimicking a commitment type that always plays the Stackelberg action, a LR 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 remains valid even on 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 may find it difficult to signal his intentions.

To study the effect of uncertain monitoring, we also consider the canonical model of a LR firm facing a sequence of SR consumers, but with one key difference. At the beginning of the game, a persistent state $(\theta, \omega) \in \Theta \times \Omega$ 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 consumers. We assume that the firm knows the state of the world, but the consumer does not.

We first show in a simple example that uncertain monitoring can cause the traditional reputation result to break down: Even if consumers believe that the firm may be a commitment type that plays the Stackelberg action every period, there exist equilibria in which even a patient firm obtains payoffs far below its Stackelberg payoff. Such "bad equilibria" arise due to an identification problem that stems from the uncertainty about monitoring: Good actions in one state cannot be statistically distinguished from a bad action in a different state.

Our simple example with such a bad equilibrium leads us to ask what might restore reputation building under uncertain monitoring in the face of such identification problems. Under an assumption that the action space is sufficiently rich, we construct a set of commitment types such that, if these types occur with positive probability, a sufficiently patient firm obtains payoffs arbitrarily close to the Stackelberg payoff in *all* equilibria,

[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" (Source: E. Wyatt, "FTC Issues Guidelines for Eco-Friendly Labels," *New York Times*, Oct 1, 2012).

[2] The Stackelberg payoff is the payoff that the LR player would get if he could commit to an action in the stage game, and the Stackelberg action is the corresponding commitment action.

even when the consumers are uncertain about the monitoring environment.[3] Importantly, the result holds independent of the fine details of the type space in that it remains valid even if we include other arbitrary commitment types. The commitment types that we construct are committed to *dynamic* (time-dependent) strategies 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 is the construction of these dynamic commitment types that play periodic strategies. As we will discuss later, such dynamic commitment types are generally necessary for reputation building under uncertain monitoring, because signaling the unknown state and Stackelberg payoff collection may require the use of different actions in the stage game.

The proof of the main result involves establishing two properties, which together imply that the LR player can guarantee payoffs close to Stackelberg payoffs in any equilibrium. First, we show that by mimicking any commitment type, the LR player can ensure in any equilibrium with high probability that the SR players' predictions of the public signal distribution are close to the true distribution generated by this commitment type in all but a finite number of periods. This step demonstrates the classic result in the spirit of "merging of opinions," à la Blackwell and Dubins (1962), and is proved using standard arguments from Gossner (2011).[4] In our setting, ensuring accurate predictions of the public signal distribution by the SR players is not sufficient for a reputation result due to potential identification problems across states. Second, we show that by mimicking the appropriate commitment type, the LR player can additionally ensure that the SR players learn the state at a rate that is *uniform across all equilibria*. We prove this by establishing a result on robust learning, which provides an easy-to-check sufficient condition that guarantees that an observer will learn the validity of an event at a uniform rate across a rich class of learning environments. The condition relates the uniform rate at which Hellinger transforms vanish across all learning environments in the class to uniform learnability of an event.[5] To the best of our knowledge, the robust learning theorem is a novel methodological contribution, which applies to general learning environments beyond the specific reputation context of this paper.

A key feature of the constructed dynamic commitment types is that they return to the signaling phase infinitely often. One might reasonably conjecture that the inclusion of a commitment type that begins with a sufficiently long phase of signaling followed by a permanent switch to playing the Stackelberg action for the true state would suffice for reputation building. We show in examples that this is generally *not* sufficient. Also, while this paper is motivated by environments with uncertain monitoring, our model allows for uncertainty both about monitoring *and* about the payoffs of the reputation

[3]We can also 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.

[4]See also the discussion after Lemma 3.

[5]See Section 5.3 for precise statements of our sufficient condition, as well as Torgersen (1991) and Moscarini and Smith (2002) for illustrations of other applications of the Hellinger transform.

builder. Finally, our main result continues to hold even if the signals observed by the SR players are unobserved by the LR player.

While the main result establishes a lower bound on the LR player's equilibrium payoff, a natural question is whether the LR player can obtain payoffs much higher than the Stackelberg payoff. With uncertain monitoring, a patient LR player may be able to obtain payoffs that are strictly higher than the Stackelberg payoff of the true state. The reason is that the LR player may not find it optimal to signal the true state, but would rather block learning to attain payoffs that are higher than the Stackelberg payoff in the true state. Providing a general, sharp characterization of an upper bound on a patient LR player's equilibrium payoffs is difficult, as it depends on the specific set of commitment types and the prior distribution over types.[6] Nevertheless, we provide a joint sufficient condition on the monitoring structure and stage game payoffs that ensures that the lower bound and the upper bound coincide: Loosely speaking, these are games in which state revelation is desirable for the LR player.

1.1 *Related literature*

We contribute to the literature on reputation that started with Kreps and Wilson (1982) and Milgrom and Roberts (1982), and includes the canonical models of Fudenberg and Levine (1989, 1992), and more recent contributions by Gossner (2011). As far as we know, this paper is the first to study reputation under 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 on zero-sum games. Wiseman (2005), Hörner and Lovo (2009), and Hörner, Lovo, and Tomala (2011) study payoff uncertainty in non-zero-sum repeated games, but do not allow uncertainty about the monitoring structure. Our framework is closest to Fudenberg and Yamamoto (2010), who study a repeated game in which there is uncertainty about both monitoring and payoffs. However, Fudenberg and Yamamoto (2010) focus on perfect public ex post equilibrium in which players play strategies whose best responses are independent of any belief about the state. As a result, in equilibrium, no player has an incentive to affect the beliefs of the opponents about the monitoring structure. We study more general equilibria where the LR player may have incentive to affect the beliefs of the SR players about the monitoring structure.

The necessity of dynamic commitment types for reputation building due to identification problems is novel. Dynamic commitment types also arise in reputation building against LR opponents, as in Aoyagi (1996), Celentani, Fudenberg, Levine, and Pesendorfer (1996), and Evans and Thomas (1997), because establishing a reputation for carrying out punishments after certain histories can be beneficial for the reputation builder.[7,8]

[6]This is in 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.

[7]Atakan and Ekmekci (2011, 2015), and Ghosh (2014) also use similar ideas.

[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 (2005).

But in our setting with SR players, the threat of punishments has no bite. Dynamic commitment types turn out to still be necessary to resolve a trade-off between signaling the correct state and collecting the Stackelberg payoff, which are both desirable to the reputation builder.

In a recent paper, Pei (2020) studies reputation with interdependent values. Pei (2020) restricts attention to perfect monitoring and a finite number of stationary commitment types, and studies the conditions under which the repeated game yields a reputation result. In contrast, we study a model where actions are imperfectly observed, but the observed public signals can potentially convey information about the state. We similarly show that reputation building can break down when the type space only consists of stationary commitment types, and further construct *dynamic* commitment types that would restore a reputation result given general type spaces that contain these dynamic commitment types in its support.

Our negative examples demonstrate that reputation building may be fragile in the presence of uncertainty about monitoring, because multiple combinations of state and action lead to the same distribution over observed public signals. Identification problems can also give rise to long-run disagreements between different agents in Acemoglu, Chernozhukov, and Yildiz (2016), and can result in convergence to incorrect beliefs in dynamic games with learning, as in Fudenberg and Levine (1993a, 1993b). The novel question that we address here is whether or not such identification problems can be circumvented by a patient long-lived player in a reputation setting.

Finally, our robust learning theorem also relates to a recent literature that studies rates of learning in decision theoretic settings. Moscarini and Smith (2002) and Mu, Pomatto, Strack, and Tamuz (2021) both provide exact characterizations of the speed of learning in decision theoretic settings, focusing on learning environments where the signals arrive in an independent and identically distributed (i.i.d.) manner conditional on the realized state. On the other hand, our robust learning theorem focuses only on a lower bound on the rate of learning, while allowing for signals that may exhibit arbitrary forms of serial correlation. Our robust learning result also 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 rich 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.

2. Model

2.1 *Notation*

We first introduce some notation that we use throughout the paper. Given a countable set X, let $\Delta(X)$ denote the set of all probability measures on X. Let $\Delta^+(X)$ be the set of full support probability measures on X. For any $B \subseteq X$, we let B^c denote the complement of B.

Given $x, x' \in X$ and some real number $\lambda \in [0,1]$, we let $\lambda x \oplus (1-\lambda)x' \in \Delta(X)$ denote the probability measure that assigns probability λ to x and $1-\lambda$ to x'. If $\nu \in$

$\Delta(X_1 \times \cdots \times X_n)$, then $\mathbf{marg}_{X_j}\nu$ is the marginal distribution of ν on X_j: $\mathbf{marg}_{X_j}\nu(x_j) = \sum_{i\neq j}\nu(x_j, x_{-j})$. Given a probability measure $\nu \in \Delta(X)$ and some function $g : X \to \mathbb{R}$, define $\mathbb{E}_\nu[g(x)]$ to be the expectation of $g(x)$ when x is distributed according to ν.

Given a finite set Y and a countable set X, define $S(Y, X)$ as the set of all possible stochastic processes over Y^∞ with state space X as follows. Formally, an element $s \in S(Y, X)$ is a sequence $s = \{s_t\}_{t=0}^\infty$, where for each t, $s_t \in \Delta(Y^t \times X)$ satisfies the consistency condition $\mathbf{marg}_{Y^{t-1}\times X} s_t = s_{t-1}$. By Kolmogorov's extension theorem, for any $s \in S(Y, X)$, there exists some $s_\infty \in \Delta(Y^\infty \times X)$ such that $\mathbf{marg}_{Y^t\times X} s_\infty = s_t$ for all t. For any $s \in S(Y, X)$ and any subset $C \subseteq X$, we can also define $s^C \in S(Y, X)$ as the corresponding stochastic process conditional on C: $s^C = (s_t(\cdot|C))_{t=0}^\infty$.

We use $\mathbb{N}$ to represent the set of all natural numbers including zero and let $\mathbb{N}_+ := \mathbb{N} \setminus \{0\}$. Finally, we establish the convention that both $\inf\emptyset = \min\emptyset = \infty$ and $\sup\emptyset = \max\emptyset = -\infty$.

2.2 *Setting*

A long-run (LR) player, player 1, faces a sequence of short-run (SR) player 2s. 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 $\gamma_0 := \nu_0 \times \mu_0$ with $\nu_0 \in \Delta^+(\Theta)$ and $\mu_0 \in \Delta^+(\Omega)$. We assume that Θ is finite and enumerate $\Theta := \{\theta_0, \ldots, \theta_{m-1}\}$, but Ω may possibly be countably infinite.[9] The realized pair of state and type (θ, ω) is then fixed for the entirety of the game.

In each period $t = 0, 1, 2, \ldots$, players simultaneously choose actions from their respective action spaces $a_1^t \in A_1$ and $a_2^t \in A_2$. We assume A_1 and A_2 are finite. Let $A = A_1 \times A_2$. Let $\mathcal{A}_i := \Delta(A_i)$ be the set of mixed actions of player i with typical element α_i.

In each period $t \geq 0$, after players have played action profile $a_t \in A$, a public signal y_t is drawn from a finite signal space Y according to the probability measure, $\psi(\cdot|a_t, \theta) \in \Delta(Y)$. Note importantly that both the action profile chosen at time t and the state of the world θ potentially affect the signal distribution. The state of the world θ represents the unknown monitoring structure. Denote by $H^t := Y^t$ the set of all t-period public histories with typical element $h^t = (y_0, \ldots, y_{t-1})$ 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, $h_1^t \in H_1^t := \Theta \times A_1^t \times Y^t$. 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 $H_2^t := H^t$.

A strategy for player i is a map $\sigma_i : \bigcup_{t=0}^\infty H_i^t \to \mathcal{A}_i$. Denote the set of strategies of player i by Σ_i. Finally, let $\mathcal{B}_1$ be the set of static state-contingent mixed actions of player 1, $\mathcal{B}_1 := \{\beta_1 : \Theta \to \mathcal{A}_1\}$ with typical element β_1.

[9]The assumption of allowing Ω to be countably infinite is standard in the existing literature (e.g., Fudenberg and Levine (1992)) when the Stackelberg action of the stage game can be mixed. We do not know whether our arguments can be extended to the setting where $|\Theta|$ is countably infinite. We leave this open for future research.

2.3 *Type space*

We assume that $\Omega = \Omega^{\text{com}} \cup \{\omega^s\}$, where Ω^{com} is the set of commitment types and ω^s is a strategic type. Each commitment type $\omega \in \Omega^{\text{com}}$ is associated with a strategy $\sigma_1^\omega \in \Sigma_1$ such that type ω always plays σ_1^ω. In contrast, type $\omega^s \in \Omega$ is a strategic type who chooses a strategy $\sigma_1 \in \Sigma_1$ to maximize payoffs, which we describe in the next subsection. Thus, a strategy profile, denoted $\sigma = ((\sigma_1(\omega))_{\omega\in\Omega}, \sigma_2)$, is a tuple for which $\sigma_1(\omega) = \sigma_1^\omega$ for all $\omega \in \Omega^{\text{com}}$.

2.4 *Payoffs and equilibrium*

Any strategy profile σ together with the prior γ induces a unique stochastic process, $(\pi_t^\sigma)_{t=0}^\infty \in S(Y \times A, \Omega \times \Theta)$ for all t. By the Kolmogorov extension theorem, there exists some $\pi_\infty^\sigma \in \Delta(H^\infty \times A^\infty \times \Omega \times \Theta)$ such that for all t, $\mathbf{marg}_{H^t\times A^t\times\Omega\times\Theta}\pi_\infty^\sigma = \pi_t^\sigma$.

To study SR players' best responses, it will also be useful to define the following beliefs of the SR players after observing a public signal history:

$$\begin{aligned}
\lambda_t^\sigma\big(\cdot|h^t\big) &:= \mathbf{marg}_{A_1\times\Theta}\pi_t^\sigma\big(\cdot|h^t\big) \in \Delta(A_1 \times \Theta),\\
\gamma_t^\sigma\big(\cdot|h^t\big) &:= \mathbf{marg}_{\Omega\times\Theta}\pi_t^\sigma\big(\cdot|h^t\big) \in \Delta(\Omega \times \Theta),\\
\nu_t^\sigma\big(\cdot|h^t\big) &:= \mathbf{marg}_{\Theta}\pi_t^\sigma\big(\cdot|h^t\big) \in \Delta(\Theta),\\
\mu_t^\sigma\big(\cdot|h^t\big) &:= \mathbf{marg}_{\Omega}\pi_t^\sigma\big(\cdot|h^t\big) \in \Delta(\Omega).
\end{aligned}$$

Then SR players' expected payoffs in any period depend on the belief, $\lambda \in \Delta(A_1 \times \Theta)$:

$$u_2(a_2, \lambda) := \mathbb{E}_\lambda\big[u_2(a_1, a_2, \theta)\big] = \sum_{a_1\in A_1,\,\theta\in\Theta} u_2(a_1, a_2, \theta)\lambda(a_1, \theta).$$

Thus, a strategy profile, σ, yields the expected payoff of $u_2(\sigma_2(h^t), \lambda_t^\sigma(h^t))$ in period t after the public history h^t. Let $B_2(\lambda)$ denote the mixed best responses of player 2, i.e., $B_2(\lambda) := \arg\max_{\alpha_2\in\mathcal{A}_2} u_2(\alpha_2, \lambda)$. With a slight abuse of notation, we write $B_2(\alpha_1, \theta) = B_2(\alpha_1 \times \mathbf{1}_\theta)$, where $\mathbf{1}_\theta$ is the Dirac probability measure that assigns probability 1 to θ, and $B_2(\beta_1, p) = B_2(\lambda_{\beta_1,p})$, where for $\beta_1 \in \mathcal{B}_1$ and $p \in \Delta(\Theta)$, $\lambda_{\beta,p}(a_1, \theta) = p(\theta)\beta_1(a_1|\theta)$.

The payoff of the LR strategic type, ω^s, in state θ is given by

$$U_1(\sigma_1, \sigma_2, \theta; \delta) := \mathbb{E}_{\pi_\infty^\sigma}\left[(1-\delta)\sum_{t=0}^\infty \delta^t u_1\big(a_1^t, a_2^t, \theta\big)|\theta, \omega^s\right].$$

Then the ex ante expected payoff of type ω^s is

$$U_1(\sigma_1, \sigma_2; \delta) := \mathbb{E}_{\nu_0}\big[U_1(\sigma_1, \sigma_2, \theta; \delta)\big].$$

Finally, we can define the statewise-Stackelberg payoff of the stage game. The Stackelberg payoff of player 1 in state θ is given by

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

For each $\varepsilon > 0$, let $\mathcal{S}_\theta^\varepsilon$ be the set of ε-Stackelberg actions in state θ, which are the mixed actions that approximate $u_1^*(\theta)$ up to ε in $\theta \in \Theta$:

$$\mathcal{S}_\theta^\varepsilon := \Big\{\alpha_1 \in \mathcal{A}_1 : \inf_{\alpha_2 \in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta) > u_1^*(\theta) - \varepsilon\Big\}.$$

We analogously define $\mathcal{S}^\varepsilon \subseteq \mathcal{B}_1$ as

$$\mathcal{S}^\varepsilon := \big\{\beta_1 \in \mathcal{B}_1 : \beta_1(\theta) \in \mathcal{S}_\theta^\varepsilon \text{ for all } \theta \in \Theta\big\}.$$

Our analysis will focus on Bayes Nash equilibria; to shorten the exposition, subsequently we will refer to Bayes Nash equilibrium simply as equilibrium. We let $\mathbf{BNE}^\delta$ denote the set of all equilibria of the game.[10]

2.5 *Information structure and key assumptions*

We now impose two key assumptions on the information structure, Assumptions 1 and 2, which we maintain for the entirety of the paper. We start with a definition.

Definition 1. A signal structure ψ satisfies *action identification for* $(\alpha_1, \theta) \in \mathcal{A}_1 \times \Theta$ if, for all $\alpha_2 \in \mathcal{A}_2$,

$$\psi(\cdot|\alpha_1, \alpha_2, \theta) = \psi\big(\cdot|\alpha_1', \alpha_2, \theta\big) \Longrightarrow \alpha_1 = \alpha_1'.$$

Let $\mathcal{B}_{id} \subseteq \mathcal{B}_1$ be the set of all $\beta_1 \in \mathcal{B}_1$ such that $(\beta_1(\theta), \theta)$ satisfies action identification for all $\theta \in \Theta$.

Assumption 1. *For every $\varepsilon > 0$, $\mathcal{S}^\varepsilon \cap \mathcal{B}_{id} \neq \emptyset$.*

In words, the above assumption holds if and only if in every state θ, there exists some ε-Stackelberg action in state θ such that this action would be statistically identified from all other actions regardless of the actions played by the SR player. 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 without it, reputation building may be impossible even when θ is common knowledge.

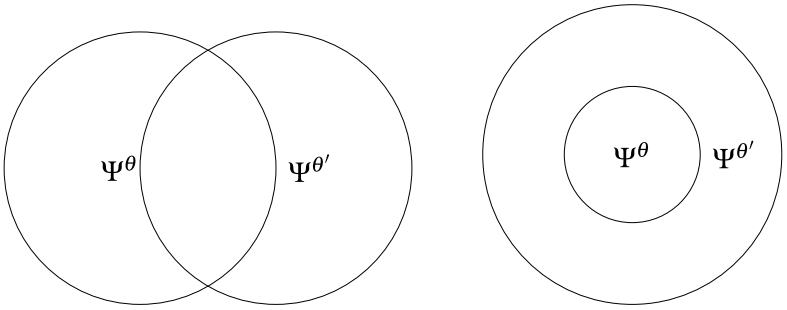
While the above assumption concerns statistical identification of actions for a *fixed* state θ, this is generally not sufficient for a reputation theorem. We furthermore impose the following assumption, which concerns the statistical identification of actions *across states*.

Assumption 2. *For every $\theta' \neq \theta$, there exist some $\alpha_1 \in \mathcal{A}_1$ such that*

$$\psi(\cdot|\alpha_1, \alpha_2, \theta) \neq \psi\big(\cdot|\alpha_1', \alpha_2, \theta'\big)$$

for all $\alpha_1' \in \mathcal{A}_1$ and all $\alpha_2 \in \mathcal{A}_2$.

[10] Our main theorems provide bounds on payoffs across all equilibria. So these bounds also apply even when restricting attention to more stringent solution concepts such as perfect Bayes Nash equilibria.



(a) Assumption 2 satisfied. (b) Assumption 2 violated.

FIGURE 1. Illustration of Assumption 2.

First note that Assumption 2 *does not* assume that α_1 must be a ε-Stackelberg action (for ε small) in state θ. Indeed, in many examples, such as those in Section 3, this assumption will not be satisfied for those ε-Stackelberg actions.

Second, we can also visualize the assumption above as follows. Let us assume for the purposes of illustration that the SR's action does not affect the public signal distribution. Then for each θ, denote by Ψ^θ the set of all probability distributions in $\Delta(Y)$ that are spanned by possibly mixed actions in $\mathcal{A}_1$ at the state θ:

$$\Psi^\theta := \{\psi(\cdot|\alpha_1, \theta) \in \Delta(Y) : \alpha_1 \in \mathcal{A}_1\}.$$

If for each pair of states $\theta \neq \theta'$, neither $\Psi^\theta \subseteq \Psi^{\theta'}$ nor $\Psi^{\theta'} \subseteq \Psi^\theta$ holds, then the assumption holds as in Figure 1(a).[11] On the other hand, Assumption 2 is violated if there exists a pair of states in which $\Psi^\theta \subseteq \Psi^{\theta'}$ as in Figure 1(b).

REMARK. Notice that the model can indeed be recast in the standard reputation framework by interpreting the pair of (θ, ω) as the type of the long-run player. However, in the classical reputation literature, non-identification problems are typically avoided by assuming that the ε-Stackelberg actions (for $\varepsilon > 0$ small) are identified. In our setting, this is no longer true. For an ε-Stackelberg action $\alpha_1 \in \mathcal{S}_\theta^\varepsilon$, our assumptions do not preclude the possibility that there exist some other state θ' and some other action α_1' that generate the same distribution over public signals. We provide a more detailed discussion of the relationship to the previous literature on reputation building with imperfect public monitoring after the presentation of Theorem 1.

3. ILLUSTRATIVE EXAMPLE

We begin with a simple example to illustrate that uncertainty in monitoring can hinder reputation building. Consider a LR player (row player) who faces a sequence of SR opponents (column player). There are two unknown states of the world, $\theta \in \Theta = \{g, b\}$. The state affects both the SR player's payoffs and the monitoring structure. Figure 2 describes the stage game in each state.

[11] Note that we only impose the condition above pairwise. In fact, even if for some θ, θ', and θ'', $\Psi^\theta \subseteq \Psi^{\theta'} \cup \Psi^{\theta''}$, the above assumption may still hold.

$\theta = b$	B	N
C	$1,1$	$-1,0$
D	$2,-2$	$0,0$

$\theta = g$	B	N
C	$1,-2$	$-1,3$
D	$2,-2$	$0,3$

FIGURE 2. The stage game in each state.

The SR players do not directly observe the action of the LR player, but rather observe a public signal. There are two possible public signals: $Y = \{\bar{y}, \underline{y}\}$. Figure 3 describes the distribution of signals, which depends only on the LR player's action and the state.

Finally suppose that the LR player can be one of two types: $\omega \in \{\omega^{\text{com}}, \omega^s\}$.[12] The type ω^{com} is the commitment type who always plays C and ω^s is the strategic type who maximizes the sum of discounted payoffs.[13]

First note that if the state θ were commonly known, then all actions would be statistically identified. Thus, if $\theta = b$ were common knowledge, the classical reputation results would imply that, for every $\varepsilon > 0$, for δ sufficiently high, the LR player would get a payoff at least $1 - \varepsilon$ in *every* equilibrium. We show below that this is not true when there is uncertainty about the state.[14]

3.1 *Failure of reputation building*

Consider a strategy profile, σ, in which ω^s always plays D, ω^{com} always plays C, and SR always plays N. We demonstrate a failure of reputation building: when $\mu_0(\omega^{\text{com}}) > 0$ is sufficiently small, the above strategy profile is indeed a perfect Bayesian equilibrium (PBE) for all $\delta \in (0, 1)$ in which the LR player obtains a payoff of 0 in both states.[15]

To see this, suppose that $\mu(\omega^{\text{com}}) > 0$ is sufficiently small so that

$$\mu_0\big(\omega^{\text{com}}\big)\nu_0(b) = \gamma_0\big(\omega^{\text{com}}, b\big) < \frac{2}{3}\gamma_0\big(\omega^s, g\big) = \frac{2}{3}\mu_0\big(\omega^s\big)\nu_0(g).$$

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

$\theta = g$	$\bar{y}$	$\underline{y}$
C	$4/5$	$1/5$
D	$3/4$	$1/4$

FIGURE 3. The monitoring structure in each state.

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

[13] For expositional simplicity, we focus on a setting in which the commitment type plays the *pure* Stackelberg action. With suitable modification of the information structure, the example can easily be extended to settings in which this type plays a mixed action. See Section 4.3.1 for an example along these lines.

[14] In our motivating example of eco-labeling, the state can be interpreted as the accuracy of the eco-label. In that case, it may be more natural to assume that the state affects only the distribution of public signals and not the SR player's payoffs. It is easy to construct a similar example of the failure of reputation building in that setting as well.

[15] We illustrate this example using PBE rather than Bayes Nash equilibrium to emphasize that the example is robust even to standard refinements such as those imposed by perfect Bayesian Nash equilibrium.

Now consider the beliefs $\gamma_t^\sigma(\omega, \theta|h^t)$ that the SR player assigns to the pair (ω, θ) at a public history h^t. Because $\psi(\cdot|C, b) = \psi(\cdot|D, g)$, in equilibrium, at all public histories h^t, the likelihood ratio between (ω^c, b) and (ω^s, g) remains at the prior and, hence,

$$\gamma_t^\sigma\big(\omega^{\text{com}}, b|h^t\big) \leq \frac{\gamma_t^\sigma\big(\omega^c, b|h^t\big)}{\gamma_t^\sigma\big(\omega^s, g|h^t\big)} = \frac{\gamma_0\big(\omega^{\text{com}}, b\big)}{\gamma_0\big(\omega^s, g\big)} < \frac{2}{3}. \tag{1}$$

Moreover, the SR finds it strictly optimal to play N whenever he assigns strictly less than $2/3$ probability to the event that both $\theta = b$ and LR plays C.[16] Since ω^{com} is the only type who plays C in state b, the SR plays N whenever $\gamma_t^\sigma((\omega^{\text{com}}, b)|h^t) < 2/3$, which holds at all public histories, as shown in (1). This implies that the SR's strategy of always playing N is indeed incentive compatible.

Finally given the SR's strategy, there are no intertemporal incentives for LR and, hence, it is optimal for ω^s to always play D. Thus, σ is a PBE and gives the LR a payoff of 0 for all $\delta \in (0, 1)$.

Reputation building fails in this example because of non-identification of the (pure) Stackelberg action across states: $\psi(\cdot|C, b) = \psi(\cdot|D, g)$. Unlike in the classical reputation models, the strategic type cannot gain by deviating to C in state b, because by doing so, he will instead, mistakenly convince the SR player that she is actually facing type ω^s who always plays D in state $\theta = g$. As a result, the equilibrium renders such deviations unprofitable.[17]

3.2 *Recovering reputation building*

How can we recover reputation building in this example? Suppose that it was possible for the LR player to undertake a costly action to signal the state. Consider the new stage game and information structure in Figures 4 and 5. Notice that both the information structure and the stage game payoffs are exactly as before when attention is restricted to action profiles in $\{C, D\} \times \{B, N\}$. The only change is that the firm can play a third action, denoted by I, that can "inform" the SR player about the true state.

In this new game, if the type space is unaltered from the previous example, then we still get a failure of reputation building, i.e., for any discount factor, $\delta \in (0, 1)$, there is a PBE in which ω^s always plays D and obtains a payoff of 0 in both states.

However, suppose now that there exists an additional type of LR that is committed to playing I in period 0 followed by C thereafter. The existence of such a type then would rule out the bad equilibrium constructed above. In equilibrium, a sufficiently patient

[16]To see this, consider any belief distribution λ over $A_1 \times \Theta$. Then

$$u_2(B, \lambda) < u_2(N, \lambda) \Leftrightarrow 3\lambda(C, b) - 2 < 3\big(\lambda(C, g) + \lambda(D, g)\big).$$

The latter inequality holds whenever $\lambda(C, b) < 2/3$.

[17]Unlike in the literature on bad reputation (e.g., Ely and Välimäki (2003) and Ely, Fudenberg, and Levine (2008)), the failure of reputation building in this example does not rely on the existence of bad types. Here, strategic types endogenously play bad actions in equilibrium. 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.

$\theta = b$	B	N
C	$1,1$	$-1,0$
D	$2,-2$	$0,0$
I	$-10,0$	$-10,0$

$\theta = g$	B	N
C	$1,-2$	$-1,3$
D	$2,-2$	$0,3$
I	$-10,0$	$-10,0$

FIGURE 4. The stage game.

ω^s will no longer find it optimal to play D always in state $\theta = b$. Instead, by mimicking the new commitment type, he can obtain a relatively high payoff 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 Assumptions 1 and 2. The generalization must deal with some additional difficulties, since the information structure may have full support, in which case, learning about the state is not immediate as in our simple example. Moreover, in such circumstances, it is impossible to convince the SR players with certainty about a state in finite time. Therefore, even after having convinced the SR to a high level of certainty about the correct state, the LR cannot be sure that the belief about the correct state will not dip to a low level thereafter. We provide a detailed discussion of these issues after the statement of Theorem 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 $\omega \in \mathcal{C}$. Most reputation theorems in the existing literature have the following structure. There exists a collection of commitment types $\mathcal{C}$ such that 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 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 high payoffs can be sustained in all equilibria for sufficiently patient players? We have already seen an example

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

$\theta = g$	$\hat{g}$	$\bar{y}$	$\underline{y}$	$\hat{b}$
C	0	$4/5$	$1/5$	0
D	0	$3/4$	$1/4$	0
I	0	0	0	1

FIGURE 5. The information structure.

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 now construct the appropriate commitment types. First, by Assumption 2, for each pair $\theta' \neq \theta$, we can choose some $\alpha_1(\theta, \theta') \in \mathcal{A}_1$ such that $\psi(\cdot|\alpha_1(\theta, \theta'), \alpha_2, \theta) \neq \psi(\cdot|\alpha_1', \alpha_2, \theta')$ for all $\alpha_1' \in \mathcal{A}_1$ and all $\alpha_2 \in \mathcal{A}_2$.[18] To simplify exposition, let us also choose an arbitrary action $\alpha_1(\theta, \theta) \in \mathcal{A}_1$ for each $\theta \in \Theta$.

For $\beta_1 \in \mathcal{B}_1$, a commitment type ω^{β_1} will be a type associated with the strategy denoted σ^{β_1} defined as follows. First, recursively define the sequence

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

where $m = |\Theta|$. Then for every $\beta_1 \in \mathcal{B}_1$, we define the 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 and θ: for all private histories h_1^τ at time τ,

$$\sigma_\tau^{\beta_1}(\theta, h_1^\tau) = \begin{cases} \beta_1(\theta) & \text{if } \tau - \max\{n_k : n_k \leq \tau\} \geq m, \\ \alpha_1(\theta, \theta_j) & \text{if } j = \tau - \max\{n_k : n_k \leq \tau\} < m. \end{cases}$$

In state θ, this commitment type plays a dynamic strategy that consist of blocks that grow in length over time. 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 play 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, 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.

We then introduce the following condition on the type space.

DEFINITION 2. We say that the type space (Ω, μ) satisfies *richness* if for every $\varepsilon > 0$, there exists $\beta_1 \in \mathcal{S}^\varepsilon \cap \mathcal{B}_{id}$ such that $\mu(\omega^{\beta_1}) > 0$.

[18] There may be many choices of $\alpha_1(\theta, \theta')$ that satisfy this condition, in which case, we choose this arbitrarily.

4.2 *Reputation theorem*

Our main result shows that our assumptions on the monitoring structure along with richness of the type space suffice for reputation building: A sufficiently patient strategic LR player will obtain (at least) payoffs arbitrarily close to the Stackelberg payoff of the complete information stage game in every equilibrium.

THEOREM 1. *Suppose that* (Ω, μ) *satisfies richness. Then for every* $\rho > 0$*, there exists some* $\delta^* \in (0,1)$ *such that for all* $\delta > \delta^*$ *and all* θ*,* $\inf_{\sigma \in \mathbf{BNE}^\delta} U_1(\sigma_1, \sigma_2, \theta; \delta) \geq u_1^*(\theta) - \rho$.[19]

We present the proof of the theorem in Section 5.[20] Before that, we discuss some important features of our result.

In specific applications, if we fix payoffs and the information structure, the full range of dynamic commitment types required by the richness assumption will not typically be needed for a reputation theorem. The full range of dynamic commitment types is needed to obtain a reputation result that does not depend on the fine details of the prior, μ_0.

To see the relationship of our main result with the classical reputation theorems, consider the special case in which ε-Stackelberg actions are identified in the following strong sense: Suppose that for each $\varepsilon > 0$ and θ, there exists some $\alpha_1 \in \mathcal{S}_\theta^\varepsilon$ such that for every $\theta' \neq \theta$, $\alpha_1' \in \mathcal{A}_1$, and all $\alpha_2 \in \mathcal{A}_2$, $\psi(\cdot|\alpha_1, \alpha_2, \theta) \neq \psi(\cdot|\alpha_1', \alpha_2, \theta')$.[21] Note that this is a stronger requirement than Assumption 2, which does not require α_1 to be an ε-Stackelberg action. In this setting, the strategy of the commitment type ω^{β_1} can be taken to be the stationary strategy that plays $\beta_1 = (\alpha_1(\theta))_{\theta \in \Theta}$ in every period. Thus, under this stronger assumption, the above reputation theorem boils down to the classical reputation theorem with stationary commitment types.[22] The reason is that in such settings, ε-Stackelberg actions themselves can be used to signal the state.

The above reputation theorem is a generalization to environments in which such identification does not hold for the ε-Stackelberg actions. In such environments, 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

[19] Because the commitment types, ω^{β_1}, play time-dependent strategies that do not condition on past public signals, a similar proof shows that Theorem 1 remains true even if the signals, $y_0, y_1, \ldots$, are privately observed only by the SR players.

[20] Our theorem assumes richness, but the requirements that both $\beta_1 \in \mathcal{S}^\varepsilon$ and $\beta_1 \in \mathcal{B}_{id}$ are made for expositional purposes. A reputation theorem with a weaker lower bound can still be established if either of these is relaxed.

[21] For example, in the special case where A_2 does not affect the public signal distribution, this would hold generically if $|Y| > |A_1 \times \Theta| + 1$.

[22] We thank an anonymous referee for making this observation.

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

$\theta = g$	$\bar{y}$	$\underline{y}$
C	$7/8$	$1/8$
D	$1/2$	$1/2$

FIGURE 6. The information structure.

signaling actions that help the consumer learn the unknown monitoring state and collection actions that are desirable for payoffs of the LR player. Because of the necessity to play both types of actions, our commitment types are nonstationary, playing a periodic strategy that alternates between signaling phases and collection phases.[23]

Finally, as we have already emphasized, our reputation result does not depend on specific distributional assumptions on the type space. In particular, it remains valid even if we include other possibly bad commitment types, as richness of the type space (Ω, μ) only requires the existence of types ω^{β_1}, while placing no restrictions on the existence or absence of other commitment types.

4.3 *Necessary characteristics of commitment types*

The 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. These two features are important and in some sense also necessary for reputation building, given the possibility of identification problems in the monitoring structure.

Consider again the stage game from Figure 2 and suppose that the information structure is now given by Figure 6. To highlight the importance of (i), we provide an example below in which the strategic LR player regardless of his discount factor obtains a low equilibrium payoff in state $\theta = b$ 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 LR gets a payoff much below the Stackelberg payoff in state b.

4.3.1 *Stationary commitment types* Consider any arbitrary countable set Ω^* of commitment types, each of which is associated with the play of a state-contingent action $\beta \in \mathcal{B}_1$ at all periods. For each $\omega \in \Omega^*$, let β^ω be the associated state-contingent mixed action plan of type ω. Notice that this type space contains only stationary commitment types. We now show that the existence of such types is generally not sufficient for reputation building.

Formally, given any countable set of stationary commitment types, Ω^*, we can construct a set of commitment types $\Omega^{\text{com}} \supseteq \Omega^*$ and a probability measure $\mu \in \Delta^+(\Omega^{\text{com}} \cup \{\omega^s\})$ such that there exists an equilibrium in which the strategic LR player obtains a payoff significantly below the Stackelberg payoff in state b.

[23] A similar reputation theorem can be proved also with stationary commitment types that have access to a public randomization device. In particular, we would need a rich space of stationary commitment types that each signal the state with different probabilities. We thank Johannes Hörner for pointing this out.

To simplify notation, let $\mathcal{A}_b := \{\alpha_1(C) \geq 2/3\}$. Notice that B is a best response to α_1 in state b if and only if $\alpha_1 \in \mathcal{A}_b$. Let $\Omega_b := \{\omega \in \Omega^* : \beta^\omega(b) \in \mathcal{A}_b\}$.

Given the information structure, ψ, for every $\alpha_1 \in \mathcal{A}_b$, there exists a corresponding bad action, $\overline{\alpha}_1$, in state g such that $\psi(\cdot|\alpha_1, b) = \psi(\cdot|\overline{\alpha}_1, g)$. For every $\omega \in \Omega_b$, let $\overline{\omega}$ denote a type who plays $\overline{\beta^\omega(b)}$ in state g and D in state b. Let the type space consist of

$$\Omega = \Omega^* \cup \{\overline{\omega} : \omega \in \Omega_b\} \cup \{\omega^s\}.$$

Claim 1. *Suppose that $\gamma_0(\omega, b) < \frac{2}{3}\gamma_0(\overline{\omega}, g)$ for all $\omega \in \Omega_b$. Then for every $\delta \in (0, 1)$, it is a PBE for the LR to always play D and the SR to always play N. In particular, this PBE yields a payoff of $0 < u_1^*(b) = 4/3$ to the LR in state $\theta = b$.*

Proof. Let σ denote the above strategy profile and consider the belief, $\lambda_t^\sigma((C, b)|h^t)$, that the SR assigns to the event (C, b) at a history h^t. By construction, for any $\omega \in \Omega_b$, $\gamma_t^\sigma((\omega, b)|h^t) = \frac{\gamma_0(\omega,b)}{\gamma_0(\overline{\omega},g)}\gamma_t^\sigma((\overline{\omega}, g)|h^t)$ for any h^t. Therefore,

$$\begin{aligned}\lambda_t^\sigma\big((C, b)|h^t\big) &= \sum_{\omega\in\Omega_b} \gamma_t^\sigma\big((\omega, b)|h^t\big)\beta^\omega(C|b) + \sum_{\omega\notin\Omega_b} \gamma_t^\sigma\big((\omega, b)|h^t\big)\beta^\omega(C|b)\\ &< \sum_{\omega\in\Omega_b} \frac{2}{3}\gamma_t^\sigma\big((\overline{\omega}, g)|h^t\big) + \sum_{\omega\notin\Omega_b} \frac{2}{3}\gamma_t^\sigma\big((\omega, b)|h^t\big) \leq \frac{2}{3}.\end{aligned}$$

Recall that it is a best response to play N at a history if $\lambda_t^\sigma((C, b)|h^t) < 2/3$. Hence, it is a best response for the SR to play N at all histories. Then it is immediate that it is a best response for te LR to play D at all histories. □

If $\nu_0(b) = 1$, as long as the closure of $\mathcal{A}_b \cap \{\beta^\omega(b) : \omega \in \Omega_b\}$ contains $2/3$ (the mixed Stackelberg action), then a sufficiently patient player obtains payoffs close to $4/3$ in any equilibrium, since a deviation to mimicking one of the good commitment types in Ω_b guarantees such a high payoff. Now consider the case when $\nu_0(b) = 1/2$. Consider again a deviation to mimicking a good type in Ω_b. Such a deviation no longer guarantees a high payoff, since there are now also bad commitment types in $\{\bar{\omega} : \omega \in \Omega_b\}$ in state g that replicate exactly the same distribution over public signals as the good commitment types. As a result, SR players are never able to differentiate between these types, and if the prior places relatively higher weight on such types in state g, then the SR players will never become optimistic about the event $\Omega_b \times \{b\}$.

4.3.2 *Type spaces with front-loaded signaling* Next we present an example where each commitment type switches 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 type spaces, we show that a reputation theorem again does not hold generally.

Again consider the same stage game (Figure 2) and information structure (Figure 6) from the previous subsection. Let ω^b be a bad commitment type who always plays $\alpha_1^* = \frac{2}{3}C \oplus \frac{1}{3}D$ in state g and always plays D in state b. Note that $\psi(\cdot|\alpha_1^*, g) = \psi(\cdot|C, b)$. Let

ω^t denote a commitment type who plays D until period t (signaling phase) and thereafter switches to the action C forever after (collection phase).[24] For any $N \in \mathbb{N}_+ \cup \{\infty\}$, consider the set of types $\Omega^N := \{\omega^t : t \in \mathbb{N}_+,\, t \leq N\} \cup \{\omega^s,\, \omega^b\}$.[25]

We now show in the following claim that without further distributional assumptions on the type space, reputation building cannot be guaranteed.

CLAIM 2. *Let $N \in \mathbb{N}_+ \cup \{\infty\}$ and $\nu_0(g) = \nu_0(b) = 1/2$. Then there exists some $\mu_0 \in \Delta^+(\Omega^N)$ such that for any $\delta \in (0, 1)$, it is a PBE for the LR to always play D and the SR to always play N. Moreover, this PBE yields a payoff of $0 < 4/3 = u_1^*(b)$ to ω^s for all discount factors in state b.*

PROOF. Consider the probability distribution over types given by

$$\mu_0\big(\omega^t\big) = \kappa^t \varepsilon,\ \mu_0\big(\omega^b\big) = \varepsilon,\ \mu_0\big(\omega^s\big) = 1 - \sum_{\tau=0}^{N} \kappa^\tau \varepsilon.$$

We assume that $\kappa \in (0, 1/4]$ and $\varepsilon \in (0, 1/2)$, in which case, μ_0 is a valid probability measure since $\sum_{\tau=0}^{N} \kappa^\tau \varepsilon < 1$.

Let σ denote the above strategy profile. Consider the probability, $\lambda_t^\sigma((C, b)|h^t)$. Since only types $\{\omega^1, \ldots, \omega^t\}$ play C in state b at such a history,

$$\lambda_t^\sigma\big((C, b)|h^t\big) = \gamma_t^\sigma\big(\big\{\omega^1, \ldots, \omega^{t-1}\big\} \times \{b\}|h^t\big),$$

but for each τ, the likelihood ratio between (ω^τ, b) and (ω^b, g) is given by

$$\frac{\gamma_t^\sigma\big(\omega^\tau, b|h^t\big)}{\gamma_t^\sigma\big(\omega^b, g|h^t\big)} = \frac{\mu_0\big(\omega^\tau\big)}{\mu_0\big(\omega^b\big)} \prod_{\tau'=1}^{\tau} \frac{\psi(y_{\tau'}|D, b)}{\psi(y_{\tau'}|\alpha_1, g)} \leq \frac{\mu_0\big(\omega^\tau\big)}{\mu_0\big(\omega^b\big)} \left(\frac{3}{2}\right)^\tau = \left(\frac{3}{2}\kappa\right)^\tau$$

Therefore,

$$\lambda_t^\sigma\big((C, g)|h^t\big) \leq \sum_{\tau=1}^{t-1} \left(\frac{3}{2}\kappa\right)^\tau \gamma_t^\sigma\big(\omega^b, g|h^t\big) \leq \sum_{\tau=1}^{\infty} \left(\frac{3}{8}\right)^\tau < \frac{2}{3}.$$

Again recall that whenever $\lambda_t^\sigma((C, g)|h^t) < \frac{2}{3}$, the SR has a strict incentive to play N. Therefore, this shows that it is indeed optimal for the SR to play N at all histories. Given this, it is immediate that it is a best response for ω^s to always play D. □

Reputation building fails in this example because all signaling is front-loaded by all commitment types. To see the basic idea, consider again the deviation to a strategy of mimicking ω^t. The hope under such a deviation for the LR is that the initial t periods of signaling would be sufficient to convince the SR players that the state is b to a sufficient

[24]For expositional simplicity, we focus only on those commitment types who play the pure Stackelberg action in the collection phase. The example can be easily extended to settings where such types play mixed actions in the collection phase.

[25]When $N = \infty$, $\Omega^N = \{\omega^t : t \in \mathbb{N}_+\} \cup \{\omega^s, \omega^b\}$.

degree of confidence that it eliminates identification problems across states. However, the claim above shows that this is infeasible for the LR. The problem is that under the constructed belief, $\mu_0 \in \Delta^+(\Omega)$, the likelihood of (ω^t, b) is much smaller than the likelihood of (ω^b, g), so that even after t periods of signaling, the SR still maintains high probability on (ω^b, g).

REMARK. If instead $\gamma_0(\omega^t, b)$ were sufficiently large relative to t for every t, then a reputation result would hold in the example. However, as previously emphasized, this illustrates the dependence of reputation building on the fine details of the prior distribution over types (beyond just the support of the prior distribution) when signaling is front-loaded.

Both of these issues highlighted in the above examples are no longer problematic given the commitment types constructed in the main theorem. First, because the commitment types enter signaling phases many times, there are no bad types in other states that can replicate similar distributions over public signals during these signaling phases for long periods of time. Second, the commitment types signal the state indefinitely so that a LR player who mimics such a commitment type can ensure eventual correct learning of the state even if the probability of such a commitment type is initially very small.

Finally, while the above analysis demonstrates the necessity of dynamic commitment types in particular examples, there are many specific settings where either stationary commitment types or commitment types with front-loaded signaling suffice.[26] An exact characterization in general games of when such simpler types suffice is beyond the scope of this paper.[27] Despite this, we emphasize again that Theorem 1 holds as long as richness is satisfied without any restrictions on what other types are or are not present.

5. PROVING THEOREM 1

We now return to prove Theorem 1. The overall structure of the proof follows the standard approach in the reputation literature. We show that for $\beta_1 \in \mathcal{S}^\varepsilon$, a sufficiently patient LR player, by playing the strategy, σ^{β_1}, associated with type ω^{β_1}, can obtain payoffs at least $u_1^*(\theta) - \rho$ in any equilibrium.

To show this, we prove two key properties that hold uniformly across all equilibria: For every $\varepsilon > 0$, there exists some J (that can be chosen independent of the choice of equilibrium) such that by deviating to play σ^{β_1} in any equilibrium, the following statements hold:

P1. The SR players' predictions of the current period's public signal distribution are approximately correct in all but J periods with probability at least $1 - \varepsilon$ (see Lemma 3 for details).[28]

[26] See the discussion after Theorem 1 and the remark above.

[27] The main obstacle is the difficulty of explicit construction of equilibria in general reputation games.

[28] By "approximately correct," we mean that the SR players' predictions will be close to the actual public signal distribution under σ^{β_1} and state θ.

P2. The SR players' beliefs assign probability at least $1-\varepsilon$ on the correct state θ in all but J periods with probability at least $1-\varepsilon$ (see Lemma 4 for details).

Property P1 holds in previous papers studying reputation building under imperfect monitoring such as Fudenberg and Levine (1986) and Gossner (2011), and its proof follows these standard arguments. In those environments, with appropriate identification assumptions, P1 implies that with high probability, the SR players' best responses will be approximately correct in all but J_1 periods. However, this property alone is inadequate for a reputation theorem in our environment. Even if the SR players' predictions of today's public signal distribution is exactly the same in all periods as that of σ^{β_1} in state θ, because of identification problems across states, this does not necessarily imply that the SR players' beliefs are concentrated on the correct state θ.

Property P2 addresses this issue. To prove it, we prove a theorem on robust learning (Theorem 2) that establishes a simple-to-check sufficient condition to ensure that an observer learns the relevant state at a rate that is uniform across a rich class of general learning environments. We then apply this theorem to the reputation setting to show that SR players learn the state θ at a rate that is uniform across *all* equilibria.

5.1 *Formal details of the proof of Theorem 1*

We now provide details of the proof of Theorem 1. Proofs not provided in the text can be found in the Appendices. We first extend the notion of ε-entropy confirming best response of Gossner (2011) to our framework.[29] To state this, first recall the definition of the Kullback–Leibler divergence of two probability measures: Given two probability measures $P, Q \in \Delta(Y)$,

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

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 3. Let $(\kappa, \varepsilon) \in [0,1]^2$. Then $\alpha_2 \in \mathcal{A}_2$ is a (κ, ε)-confirming best response at (α_1, θ) if there exists some $\lambda \in \Delta(A_1 \times \Theta)$ such that

(i) $\alpha_2 \in B_2(\lambda)$

(ii) $D(\psi(\cdot|\alpha_1, \alpha_2, \theta)\|\psi(\cdot|\lambda, \alpha_2)) \leq \varepsilon$ (see footnote 30)[30]

(iii) $\mathbf{marg}_\Theta \lambda(\theta) \geq 1-\kappa$.

We let $\mathrm{CBR}_{\kappa,\varepsilon}(\alpha_1, \theta)$ be the set of all (κ, ε)-confirming best responses at (α_1, θ).

[29] Fudenberg and Levine (1992) provide a similar definition that uses the notion of total variational distance between probability measures instead of Kullback–Leibler divergence.

[30] We define $\psi(\cdot|\lambda, \alpha_2) := \sum_{a_1,a_2,\theta} \psi(\cdot|a_1, a_2, \theta)\lambda(a_1, \theta)\alpha_2(a_2)$.

The following lemma motivates the definition of (κ, ε)-confirming best responses and shows that for ε small, if short-run players play an $(\varepsilon, \varepsilon)$-confirming best response, then the LR player obtains payoffs close to those as if the SR player were best responding with perfect knowledge of both the LR player's action and state.

LEMMA 1. *For every* $\alpha_1 \in \mathcal{A}_1$,

$$\liminf_{\varepsilon \to 0} \inf_{\alpha_2 \in CBR_{\varepsilon,\varepsilon}(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta) \geq \inf_{\alpha_2 \in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta).$$

PROOF. By Assumption 1 and the property that $D(P|Q) = 0$ if and only if $P = Q$, we have that $\mathrm{CBR}_{0,0}(\alpha_1, \theta) = B_2(\alpha_1, \theta)$. Moreover, $\mathrm{CBR}_{\varepsilon,\varepsilon}(\alpha_1, \theta)$ is upper hemi-continuous with respect to ε, and so the inequality follows. □

Notice that a $(1, \varepsilon)$-confirming best response is essentially the extension of the idea of ε-entropy confirming best response in Gossner (2011) to the current setting. Under a $(1, \varepsilon)$-confirming best response, condition (iii) in Definition 3 is trivially satisfied and so the definition only requires that the public signal distribution associated with the belief λ required to sustain α_2 as a best response be ε-close in Kullback–Leibler divergence to the true distribution of public signals under the action profile (α_1, α_2) and state θ. When κ is small, condition (iii) additionally requires that λ indeed places large probability on the state θ. This additional requirement is important in Lemma 1, since generally, $\liminf_{\varepsilon \to 0} \inf_{\alpha_2 \in \mathrm{CBR}_{1,\varepsilon}(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta)$ may be strictly less than $\inf_{\alpha_2 \in B_2(\alpha_1,\theta)} u_1(\alpha_1, \alpha_2, \theta)$.

The following lemma constitutes the key step in the proof of the main theorem, which shows that if the LR deviates to play σ^{β_1} in any equilibrium, then the SR plays strategies consistent with $(\varepsilon, \varepsilon)$-confirming best responses in all but a finite number of periods with very large probability. Formally, define the following set of histories given an equilibrium σ and a type $\omega \in \Omega$ who plays strategies that only depend on $H^t \times \Theta$ [31]

$$\mathcal{M}^{\sigma,(\omega,\theta)}(J, \kappa, \varepsilon) := \left\{h^\infty \in H^\infty : \left|\left\{t : \sigma_2\left(h^t\right) \notin \mathrm{CBR}_{\kappa,\varepsilon}\left(\sigma_1\left(\omega, h^t, \theta\right), \theta\right)\right\}\right| < J\right\}.$$

These are the set of public histories, h^∞, where type ω and the SR players together play action profiles that are (κ, ε)-confirming best responses at state θ in all but J periods. The following lemma provides a lower bound on the probability of such histories that applies uniformly across all equilibria.

LEMMA 2. *Suppose that* $\mu(\omega^{\beta_1}) > 0$. *Then for every* $\varepsilon > 0$, *there exists some* J *such that* $\inf_{\sigma \in \mathbf{BNE}^\delta} \pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\left(\mathcal{M}^{\sigma,(\omega^{\beta_1},\theta)}(J, \varepsilon, \varepsilon)\right) \geq 1 - 2\varepsilon$.[32]

[31] In the analysis, we are concerned with these sets only for types ω^{β_1} who play strategies that only depend on $H^t \times \Theta$. Therefore, the restriction to such types is not restrictive.

[32] Notice that in this lemma, we do not necessarily require that $\beta_1 \in \mathcal{S}^{\varepsilon'}$ for some $\varepsilon' > 0$ small. Later in the proof of Theorem 1, when we use this lemma, we will use Lemma 2 for the particular case in which $\beta_1 \in \mathcal{S}^{\varepsilon'}$ for $\varepsilon' > 0$ small to ensure that by mimicking ω^{β_1}, the LR can ensure high payoffs.

There are two aspects of the lemma above that are worth emphasis. First is that the set of histories in $\mathcal{M}^{\sigma,(\omega,\theta)}(J,\varepsilon,\varepsilon)$ ensures that players play action profiles consistent with $(\varepsilon,\varepsilon)$-confirming best responses in all but J periods. One could weaken this to analyze the probability of the set of histories in $\mathcal{M}^{\sigma,(\omega,\theta)}(J,1,\varepsilon)$ that only require players to play action profiles consistent with $(1,\varepsilon)$-confirming best responses in all but J periods as in Gossner (2011). Indeed, the arguments of Fudenberg and Levine (1986) and Gossner (2011) imply a uniform lower bound on the probability of such histories across all equilibria. However, this is insufficient for our reputation theorems since as previously discussed, Lemma 1 does not apply to $(1,\varepsilon)$-confirming best responses.

The conclusion of the above lemma does not hold for any arbitrary type ω and holds *only* for types ω^{β_1}. This is again because the definition of $\mathcal{M}^{\sigma,(\omega,\theta)}(J,\varepsilon,\varepsilon)$ requires SR players to hold approximately correct beliefs on the state θ in all but J periods. In particular, if ω were a stationary commitment type, then $\pi_\infty^{\sigma,(\omega,\theta)}(\mathcal{M}^{\sigma,(\omega,\theta)}(J,\varepsilon,\varepsilon)|\omega,\theta)$ may actually be quite small for some equilibria, σ.

We prove Lemma 2 in Section 5.2. Before this, we present the proof of Theorem 1, which is now immediate.

Proof of Theorem 1. Define $\underline{u} := \min_{a\in A}\min_{\theta\in\Theta} u_1(a,\theta)$. and choose any θ. We will show that there exists some $\delta^* < 1$ such that whenever $\delta > \delta^*$, $U_1(\sigma,\theta;\delta) > u_1^*(\theta) - \rho$ for all $\sigma \in \mathbf{BNE}^\delta$. 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 1, there exists some $\varepsilon \in (0,\varepsilon^*)$ such that

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

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

By Lemma 2, there exists some J such that for every equilibrium, σ,

$$\pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\big(\mathcal{M}^{\sigma,(\omega^{\beta_1},\theta)}(J,\varepsilon,\varepsilon)\big) \geq 1-2\varepsilon.$$

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 payoff

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

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

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

5.2 *Proving Lemma 2*

We now prove our key lemma, which follows in a straightforward manner from the following two lemmas. The complete proof of Lemma 2 is provided in Appendix D. To simplify notation, given $C \subseteq \Omega \times \Theta$, define

$$\phi_t^\sigma(\cdot|h^t) = \mathbf{marg}_Y \pi_t^\sigma(\cdot|h^t),\ \phi_t^{\sigma,C}(\cdot|h^t) = \mathbf{marg}_Y \pi_t^\sigma(\cdot|h^t, C).$$

In words, $\phi_t^\sigma(\cdot|h^t)$ is the distribution over $y_t \in Y$ in period t in the equilibrium σ, conditional on the public history h^t.[33] Additionally, $\phi_t^{\sigma,C}$ is this distribution when conditioned on the event $(\omega, \theta) \in C$.

LEMMA 3 (Merging). *Suppose that $\gamma_0(\omega, \theta) > 0$. Then for every $\varepsilon > 0$, there exists some J_1 such that in every equilibrium σ,*

$$\pi_\infty^{\sigma,(\omega,\theta)}\left(\left\{h^\infty \in H^\infty : \left|\left\{t : D\left(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t) \| \phi_t^\sigma(\cdot|h^t)\right) > \varepsilon\right\}\right| < J_1\right\}\right) \geq 1 - \varepsilon.$$

LEMMA 4 (Uniform Learning). *Suppose $\mu_0(\omega^{\beta_1}) > 0$. Then for every $\varepsilon > 0$, there exists some J_2 such that for all $\sigma \in \mathbf{BNE}^\delta$,*

$$\pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\left(\left\{h^\infty \in H^\infty : \left|\left\{t : \nu_t^\sigma(\theta|h^t) < 1 - \varepsilon\right\}\right| < J_2\right\}\right) \geq 1 - \varepsilon.$$

As in Fudenberg and Levine (1986) and Gossner (2011), Lemma 3 strengthens the classical merging results, e.g., Blackwell and Dubins (1962) and Kalai and Lehrer (1993), by establishing a uniform upper bound across all equilibria on the probability of histories in which the SR player's prediction of today's public signal distribution, $\phi_t^\sigma(\cdot|h^t)$, diverges substantially from the "true" public signal distribution, $\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)$, in more than J_1 time periods, when LR plays $\sigma(\omega)$ in state θ. The proof follows using standard merging arguments of Gossner (2011), which we include for completeness in Appendix C.

To prove Lemma 4, we show that in any state θ, by playing $\sigma^{\beta_1}(\theta)$, the LR player can ensure that the SR players learn the state θ at a rate that is uniform across all equilibria. Indeed standard arguments immediately imply that in any equilibrium, SR players learn the true state θ whenever the LR player plays $\sigma^{\beta_1}(\theta)$. However, the additional uniformity requirement requires further analysis, which we now address in Section 5.3.

5.3 *A robust learning theorem*

Consider the following general model of learning. There is a finite signal space Y and a countable state space Ξ. A *learning environment* is some $\pi \in S(Y, \Xi)$ for which $\pi_\Xi := \mathbf{marg}_\Xi \pi$ has full support on Ξ. Recall that for any $B \subseteq \Xi$, $\pi^B \in S(Y, \Xi)$ denotes the stochastic process conditional on $\xi \in B$: $\pi^B = (\pi_t(\cdot|B))_{t=0}^\infty$. Note that this allows the stochastic process, π^ξ, for any $\xi \in \Xi$, to be very general, which may potentially contain arbitrary forms of serial correlations.

[33] In fact, $\phi_t^\sigma(\cdot|h^t)$ is the SR players' subjective belief of the period t public signal after observing h^t.

To interpret, in a learning environment, at the beginning of each period $t = 1, 2, \ldots$, an observer updates her beliefs about the true state $\xi \in \Xi$ according to Bayes' rule upon the realization of a history of signals $h^t = (y_0, \ldots, y_{t-1})$. Let $\rho_t^\pi(\cdot|h^t) \in \Delta(\Xi)$ denote the observer's beliefs after observing h^t. We now describe formally our definition of robust learning.

DEFINITION 4. Let $\xi^* \in B \subseteq \Xi$ and $S^* \subseteq S(Y, \Xi)$. Then we say that an observer S^*-*robustly learns* B at ξ^* if for every $\kappa \in (0, 1)$, there exists some K such that

$$\inf_{\pi \in S^*} \pi_\infty \left(\bigcap_{t=K}^{\infty} \left\{ h^\infty : \rho_t^\pi(B|h^t) \geq 1 - \kappa \right\} | \xi^* \right) \geq 1 - \kappa.$$

Intuitively, S^*-robust learning requires an observer's beliefs to concentrate on B forever after period K with high probability for all learning environments in S^*.

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

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

This is the moment generating function of the (random) log-likelihood ratio at time t, $\log \frac{\pi_t^B(h^t)}{\pi_t^{\xi^*}(h^t)}$, when h^t is distributed according to $\pi_t^{\xi^*}$. Toward our robust learning result, let us also define

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

Roughly speaking, $\mathcal{H}_t^\pi(B, \xi^*)$ measures the informativeness of the learning environment at time t with respect to learning the relative likelihoods of B vs. ξ^*. Notice that by Jensen's inequality, $\mathcal{H}_t^\pi(B, \xi^*) \leq 1$. Intuitively, a completely uninformative learning environment attains this maximal value of $\mathcal{H}_t^\pi(B, \xi^*) = 1$. On the other hand, if the supports of π_t^B and $\pi_t^{\xi^*}$ are disjoint so that the learning environment distinguishes B from ξ^* perfectly, then $\mathcal{H}_t^\pi(B, \xi^*) = 0$. In Appendix A, we list some additional useful properties of the Hellinger transform.[34]

In the following theorem, we show that when the Hellinger transforms converge to zero (information converges to perfect information) at a fast enough rate uniformly across all learning environments, $\pi \in S^*$, then the observer S^*-robustly learns B at ξ^*.

[34]See also Torgersen (1991) and Moscarini and Smith (2002) for more details on the Hellinger transform.

THEOREM 2. *Let $S^* \subseteq S(Y, \Xi)$ and $\xi^* \in B \subseteq \Xi$. Suppose that $\inf_{\pi \in S^*} \pi_\Xi(\xi^*) > 0$ and*

$$\lim_{K \to \infty} \sup_{\pi \in S^*} \sum_{t=K}^{\infty} \mathcal{H}_t^\pi \big(B^c, \xi^*\big) = 0.$$

Then an observer S^*-robustly learns B at ξ^*.[35]

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

COROLLARY 1. *Let $\xi^* \in B_1, \ldots, B_n \subseteq \Xi$ and $S^* \subseteq S(Y, \Xi)$. Suppose that $\inf_{\pi \in S^*} \pi_\Xi(\xi^*) > 0$ and that for all $\ell = 1, 2, \ldots, n$,*

$$\lim_{K \to \infty} \sup_{\pi \in S^*} \sum_{t=K}^{\infty} \mathcal{H}_t^\pi \big(B_\ell^c, \xi^*\big) = 0.$$

Then the observer S^*-robustly learns $B_1 \cap B_2 \cap \cdots \cap B_n$ at ξ^*.

5.3.1 *Uniform signaling of the state in reputation building* Given any equilibrium, σ, the SR players face a learning environment about the state space $\Xi = \Omega \times \Theta$ along the same lines as in Section 5.3. Of course, when we view an equilibrium, σ, as a learning environment, we can also define the appropriate Hellinger transforms. Thus, for any equilibrium σ and any event $A \subseteq \Omega \times \Theta$, we define the Hellinger transform as

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

We also accordingly define

$$\mathcal{H}_t^\sigma \big(B, (\omega, \theta)\big) = \inf_{z \in [0,1]} \mathcal{H}_t^\sigma \big(z; B, (\omega, \theta)\big).$$

Through a straightforward computation in Lemma 8 in Appendix B, we show that for any $\theta' \neq \theta$,

$$\lim_{K \to \infty} \sup_{\sigma \in \mathbf{BNE}^\delta} \sum_{t=K}^{\infty} \mathcal{H}_t^\sigma \big(\Omega \times \big\{\theta'\big\}, \big(\omega^{\beta_1}, \theta\big)\big) = 0.$$

By Corollary 1, the SR players $\mathbf{BNE}^\delta$-robustly learn $\bigcap_{\theta' \neq \theta} \Omega \times (\Theta \setminus \{\theta'\}) = \Omega \times \{\theta\}$ at $(\omega^{\beta_1}, \theta)$, which proves Lemma 4.

[35] We leave open the question of whether this condition is also necessary for S^*-robust learning for future research.

6. Upper bound on payoffs

Thus far, we have focused our analysis on a lower bound of equilibrium payoffs. This section studies the tightness of the established lower bound. For this section only, we make the following assumption that the public signal distribution only depends on the action of the LR player.[36]

Assumption 3. *For all* $\theta, \alpha_1, \alpha_2, \alpha_2', \psi(\cdot|\alpha_1, \alpha_2, \theta) = \psi(\cdot|\alpha_1, \alpha_2', \theta)$. *With a slight abuse of notation, we write* $\psi(\cdot|\alpha_1, \theta)$.

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 reputation games where $|\Theta| = 1$ (and with suitable action identification assumptions), the upper bound on payoffs is independent of initial conditions such as the probability distribution over types, as long as the LR player is sufficiently patient.[37] In contrast, we show in Appendix F that the upper bound (even for very patient players) typically depends on these initial conditions of the game if $|\Theta| \geq 2$. As a result, providing a general sharp upper bound is difficult.

Instead, we first provide a general upper bound theorem when the probability of commitment types is small. We also show in Corollary 2 that this derived upper bound is indeed tight in a class of games where state revelation is desirable. The ideas presented here follow closely those of Mertens, Sorin, and Zamir (2014), Chapter V.3.

Definition 5. Let $p \in \Delta(\Theta)$. A state-contingent strategy $\beta \in \mathcal{B}_1$ is called nonrevealing at p if for all θ, θ' in the support of p, $\psi(\cdot|\beta(\theta), \theta) = \psi(\cdot|\beta(\theta'), \theta')$. Let $\mathrm{NR}(p)$ be the set of all $\beta \in \mathcal{B}_1$ that are nonrevealing at p.

In words, this means that if player 1 plays according to a nonrevealing strategy at p, then with probability 1, player 2's beliefs about Θ will not change regardless of the public signal she sees.

We can now define the value function, if $\mathrm{NR}(p) \neq \emptyset$, as

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

Notice that because we are interested in an upper bound, unlike in the definition of Stackelberg payoffs, we take the supremum rather than the infimum over $\alpha_2 \in B_2(\beta, p)$.[38] On the other hand, if $\mathrm{NR}(p) = \emptyset$, let us define $V(p) = \underline{u}$. Define $\mathbf{cav}V$ to be the smallest concave function that is weakly greater than V pointwise.

[36] We do not know whether the same results can be extended to environments in which the SR players' actions also affect the public signal distribution.

[37] See Fudenberg and Levine (1992) and Gossner (2011) for these results.

[38] There are settings in which this value coincides with an analogous value defined by taking the infimum over $\alpha_2 \in B_2(\beta, p)$. See Theorem 3.3 in Fudenberg and Levine (1992) for a discussion of this issue.

THEOREM 3 (Upper Bound Theorem). *Let $\varepsilon > 0$. Then there exists some $\kappa^* > 0$ and $\delta^* < 1$ such that whenever $\mu(\Omega^c) < \kappa^*$ and $\delta > \delta^*$,*

$$\sup_{\sigma \in \mathbf{BNE}^\delta} U_1(\sigma; \delta) \leq \mathbf{cav}V(\nu_0) + \varepsilon.$$

Note first that, unlike in Theorem 1, richness of the type space is not necessary for the theorem. Second, the theorem imposes a condition on the probability of commitment types. In Appendix F, we present an example in which the bound provided here does not apply when commitment types occur with large probability. The reason for the discrepancy is that when commitment type probabilities are large, the SR player's beliefs about Θ in an equilibrium, conditional on the strategic type, is no longer a martingale. In contrast, when the commitment type probabilities are small, these beliefs conditional on the strategic type's strategy follow a stochastic process that almost resembles a martingale, in which case $\mathbf{cav}V$ provides an approximate upper bound.

6.1 *Statewise payoff bounds and payoff uniqueness*

Finally, we apply Theorem 3 to a setting in which the type space satisfies richness. In the following corollary, we show that in games where $\mathbf{cav}V(\nu_0) = \sum_{\theta \in \Theta} \nu_0(\theta) u_1^*(\theta)$, a sufficiently patient LR player receives payoffs close to $u_1^*(\theta)$ in all states $\theta \in \Theta$ and all equilibria when commitment types are small in probability.

COROLLARY 2. *Suppose that $\mathbf{cav}V(\nu_0) = \sum_{\theta \in \Theta} \nu_0(\theta) u_1^*(\theta)$ and that (Ω, μ) satisfies richness. Let $\varepsilon > 0$. Then there exists some $\kappa^* > 0$ and $\delta^* < 1$ such that whenever $\mu(\Omega^c) < \kappa^*$ and $\delta > \delta^*$, then in any state $\theta \in \Theta$ and any equilibrium $\sigma \in \mathbf{BNE}^\delta$,*

$$u_1^*(\theta) - \varepsilon \leq U_1(\sigma, \theta; \delta) \leq u_1^*(\theta) + \varepsilon.$$

A key distinction between Theorem 3 and the above corollary is that we provide an upper bound on payoffs in each state. A key step in the proof of this statewise upper bound in the corollary relies on the richness of the type space. This assumption is important for the argument, as it first allows us to provide a lower bound on payoffs in each state using Theorem 1, which then together with the ex ante payoff upper bound of Theorem 3 allows us to prove the upper bound in each state.

7. CONCLUSION

We study reputation building by a long-run agent in environments in which there is uncertainty about how the agent's actions relate to observed outcomes. In contrast to the previous literature, reputation building generally requires the inclusion of dynamic commitment types: types that switch infinitely often between signaling actions and collection actions. Our main theorem shows that when such commitment types occur with positive probability, a sufficiently patient LR player obtains at least his Stackelberg payoffs (or arbitrarily close payoffs) in each state.

We conclude with some future directions for research. First, we conjecture that a similar reputation result follows even if the state is initially unknown to the LR player. The intuition is that because the LR player observes his own actions, with mild identification assumptions, he should be able to learn the state over time. However, as the LR player's beliefs evolve, he may want to to play different strategies contingent on what he has learned. As a result, the construction of commitment types would need substantial modification, since the strategies must depend on the realized signals in the long run. Second, Cripps, Mailath, and Samuelson (2004) show that in reputation models with imperfect public monitoring, the scope for reputation building disappears in the long run, since in any equilibrium, the SR players eventually learn the LR strategic player's type. Whether or not a similar result holds in a setting with monitoring uncertainty remains unclear due to identification problems across states.[39]

Appendix A: Proofs of robust learning

A.1 *Properties of the Hellinger transform*

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

Lemma 5. *Let $\pi \in S(Y, \Xi)$ and $\xi^* \notin B$, $B \subseteq \Xi$. Then $\mathcal{H}_t^\pi(z; B, \xi^*)$ satisfies the following properties:*

(i) For all t and all $z \in [0,1]$, $0 \leq \mathcal{H}_t^\pi(z; B, \xi^*) \leq 1$.

(ii) For all t and all $z \in (0,1)$, $\mathcal{H}_t^\pi(z; B, \xi^*) = 1$ if and only if $\pi_t^B(h^t) = \pi_t^{\xi^*}(h^t)$ for all h^t such that $\pi_t^{\xi^*}(h^t) > 0$.

(iii) For every $z \in [0,1]$, $\mathcal{H}_t^\pi(z; B, \xi^*)$ is weakly decreasing in t.

See Torgersen (1991) p. 40 for the first two properties. Property (iii) follows from the fact that the Hellinger transform is monotone in the Blackwell order; see Torgersen (1991) p. 358.

A.2 *Proving Theorem 2*

We use $\mathcal{H}_t^\pi(B, \xi^*)$ to provide a lower bound on the probability of learning after some time K.[40]

Lemma 6. *Let $\pi \in S(Y, \Xi)$, $\xi^* \in B \subseteq \Xi$. Suppose that $\sum_{t=0}^\infty \mathcal{H}_t^\pi(B^c, \xi^*) < +\infty$ and that $\pi_\Xi(\xi^*) > 0$. Then for all K and any $\kappa > 0$,*

$$\pi_\infty^{\xi^*}\left(\bigcap_{t=K}^\infty \left\{h^\infty : \rho_t^\pi\left(B|h^t\right) \geq 1-\kappa\right\}\right) \geq 1 - \max\left\{1, (1-\kappa)/\left(\kappa \pi_\Xi\left(\xi^*\right)\right)\right\} \sum_{t=K}^\infty \mathcal{H}_t^\pi\left(A^c, \xi^*\right).$$

[39] We thank an anonymous referee for suggesting these questions.

[40] Moscarini and Smith (2002) and Mu et al. (2021), respectively, use the Hellinger transform and the Renyi divergence (a monotone transformation of the Hellinger transform) to compare the informational value of experiments when the experiments are repeated sufficiently many times in an i.i.d. manner. In contrast, our learning environments allow for arbitrary serial correlation in public signals.

PROOF. If $\pi_\Xi(B^c) = 0$, then the lemma holds trivially, so let us assume that $\pi_\Xi(B^c) > 0$. First note that for any $z \in (0, 1]$,

$$\begin{aligned}
\pi_t^{\xi^*}\left(\frac{\rho_t^\pi(B^c|h^t)}{\rho_t^\pi(B|h^t)} > \frac{\kappa}{1-\kappa}\right) &\le \pi_t^{\xi^*}\left(\frac{\rho_t^\pi(B^c|h^t)}{\rho_t^\pi(\xi^*|h^t)} > \frac{\kappa}{1-\kappa}\right) \\
&= \pi_t^{\xi^*}\left(\left(\frac{\pi_\Xi(B^c)}{\pi_\Xi(\xi^*)}\right)^z \left(\frac{\pi_t^{B^c}(h^t)}{\pi_t^{\xi^*}(h^t)}\right)^z > \left(\frac{\kappa}{1-\kappa}\right)^z\right) \\
&\le \pi_t^{\xi^*}\left(\left(\frac{\pi_t^{B^c}(h^t)}{\pi_t^{\xi^*}(h^t)}\right)^z > \left(\frac{\kappa}{1-\kappa}\pi_\Xi(\xi^*)\right)^z\right) \\
&\le \left(\frac{\kappa}{1-\kappa}\pi_\Xi(\xi^*)\right)^{-z} \mathcal{H}_t^\pi(z; B^c, \xi^*) \\
&\le \max\{1, (1-\kappa)/(\kappa\pi_\Xi(\xi^*))\}\mathcal{H}_t^\pi(z; B^c, \xi^*),
\end{aligned}$$

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

$$\pi_t^{\xi^*}\left(\frac{\rho_t^\pi(B^c|h^t)}{\rho_t^\pi(B|h^t)} > \frac{\kappa}{1-\kappa}\right) \le \max\{1, (1-\kappa)/(\kappa\pi_\Xi(\xi^*)\}\mathcal{H}_t^\pi(B^c, \xi^*).$$

Then we have

$$\begin{aligned}
\pi_\infty^{\xi^*}\left(\bigcap_{t=K}^\infty \{h^\infty : \rho_t^\pi(B|h^t) \ge 1-\kappa\}\right) &= \pi_\infty^{\xi^*}\left(\bigcap_{t=K}^\infty \left\{h^\infty : \frac{\rho_t^\pi(B^c|h^t)}{\rho_t^\pi(B|h^t)} \le \frac{\kappa}{1-\kappa}\right\}\right) \\
&= 1 - \pi_\infty^{\xi^*}\left(\bigcup_{t=K}^\infty \left\{h^\infty : \frac{\rho_t^\pi(B^c|h^t)}{\rho_t^\pi(B|h^t)} > \frac{\kappa}{1-\kappa}\right\}\right) \\
&\ge 1 - \sum_{t=K}^\infty \pi_t^{\xi^*}\left(\frac{\rho_t^\pi(B^c|h^t)}{\rho_t^\pi(B|h^t)} > \frac{\kappa}{1-\kappa}\right) \\
&\ge 1 - \max\{1, (1-\kappa)/(\kappa\pi_\Xi(\xi^*)\} \sum_{t=K}^\infty \mathcal{H}_t^\pi(B^c, \xi^*).
\end{aligned}$$

□

Note that in Lemma 6, the lower bound of the probability of learning established in the above lemma depends only on four parameters: κ, K, $\pi_\Xi(\xi^*)$, and $\sum_{t=K}^\infty \mathcal{H}_t^\pi(B^c, \xi^*)$. In particular, other aspects of the learning environment do not influence this lower bound. As a result, this lemma implies Theorem 2.

PROOF OF THEOREM 2. By assumption there exists $\varepsilon > 0$ such that $\pi_\Xi(\xi^*) \ge \varepsilon$ for all $\pi \in S^*$. By Lemma 6, for every $\pi \in S^*$ and every K,

$$\pi_\infty^{\xi^*}\left(\bigcap_{t=K}^\infty \{h^\infty : \rho_t^\pi(B|h^t) \ge 1-\kappa\}\right) \ge 1 - \max\{1, (1-\kappa)/(\kappa\varepsilon)\} \sup_{\pi\in S^*} \sum_{t=K}^\infty \mathcal{H}_t^\pi(B^c, \xi^*).$$

By assumption there exists some K sufficiently large such that

$$1-\max\{1,(1-\kappa)/(\kappa\varepsilon)\}\sup_{\pi\in S^*}\sum_{t=K}^{\infty}\mathcal{H}_t^{\pi}\big(B^c,\xi^*\big)\geq 1-\kappa.$$

Therefore, $\inf_{\pi\in S^*}\pi_{\infty}^{\xi^*}(\bigcap_{t=K}^{\infty}\{h^{\infty}:\rho_t^{\pi}(B|h^t)\geq 1-\kappa\})\geq 1-\kappa$. □

A.3 *Proof of Corollary 1*

Let $\kappa>0$. By Theorem 2, there exists some K such that for all $\ell=1,2,\ldots,n$,

$$\inf_{\pi\in S^*}\pi_{\infty}^{\xi^*}\left(\bigcap_{t=K}^{\infty}\left\{h^{\infty}:\rho_t^{\pi}\big(B_{\ell}|h^t\big)\geq 1-\frac{\kappa}{n}\right\}\right)\geq 1-\frac{\kappa}{n}.$$

Therefore,

$$\begin{aligned}1-\kappa&\leq\inf_{\pi\in S^*}\pi_{\infty}^{\xi^*}\left(\bigcap_{\ell=1}^{n}\bigcap_{t=K}^{\infty}\left\{h^{\infty}:\rho_t^{\pi}\big(B_{\ell}|h^t\big)\geq 1-\frac{\kappa}{n}\right\}\right)\\&\leq\inf_{\pi\in S^*}\pi_{\infty}^{\xi^*}\left(\bigcap_{t=K}^{\infty}\big\{h^{\infty}:\rho_t^{\pi}\big(B_1\cap\cdots\cap B_n|h^t\big)\geq 1-\kappa\big\}\right).\end{aligned}$$

Appendix B: Uniform learning across all equilibria

Lemma 7. *Fix any θ. Let $\theta_j\neq\theta$. Then there exists some $\varepsilon>0$ such that for any $k\in\mathbb{N}$,*

$$\sup_{\sigma\in\mathbf{BNE}^{\delta}}\mathcal{H}_{j+n_k+1}^{\sigma}\big(\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\leq(1-\varepsilon)^k.$$

Proof. Let $z\in(0,1)$. By construction, $\psi(\cdot|\alpha_1,\alpha_2,\theta_j)\neq\psi(\cdot|\alpha_1(\theta,\theta_j),\alpha_2\theta)$ for all $\alpha_1\in\mathcal{A}_1$ and all $\alpha_2\in\mathcal{A}_2$. Then by Lemma 5, there exists some $\varepsilon>0$ such that

$$\sup_{\alpha_1\in\mathcal{A}_1,\alpha_2\in\mathcal{A}_2}\sum_{y\in Y}\pi(y|\alpha_1,\alpha_2,\theta_j)^z\pi\big(y|,\alpha_1(\theta,\theta_j),\alpha_2,\theta\big)^{1-z}\leq 1-\varepsilon.$$

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

Now consider any equilibrium σ. 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 $t=n_k+j+1$. Then by the law of iterated expectations, note that

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

Again by Lemma 5, since $\mathcal{H}_t$ is a non-increasing sequence, we have

$$\mathcal{H}_t^\sigma\big(z;\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\le(1-\varepsilon)\mathcal{H}_{t'}^\sigma\big(z;\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\le(1-\varepsilon)^k.$$

Since the above holds for fixed $z>0$, the claim also holds for the infimum over $z\in[0,1]$. Hence, for every $\sigma\in\mathbf{BNE}^\delta$,

$$\mathcal{H}_{n_k+j+1}^\sigma\big(\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\le(1-\varepsilon)^k. \qquad \square$$

LEMMA 8. *For all $\theta'\neq\theta$,*

$$\lim_{K\to\infty}\sup_{\sigma\in\mathbf{BNE}^\delta}\sum_{t=K}^{\infty}\mathcal{H}_t^\sigma\big(\Omega\times\{\theta'\},\big(\omega^{\beta_1},\theta\big)\big)=0.$$

PROOF. Let $\theta_j=\theta'$. Then by the previous lemma, there exists some $\varepsilon>0$ such that for any $k\in\mathbb{N}$,

$$\sup_{\sigma\in\mathbf{BNE}^\delta}\mathcal{H}_{j+n_k+1}^\sigma\big(\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\le(1-\varepsilon)^k.$$

Let $K\ge j+n_0+1$ and let $k(K)$ be the maximal value of k for which $j+n_k+1\le K$. For any $\sigma\in\mathbf{BNE}^\delta$, since $\mathcal{H}_t^\sigma(\Omega\times\{\theta_j\},(\omega^{\beta_1},\theta))$ is weakly decreasing in t,

$$\begin{aligned}\sum_{t=K}^{\infty}\mathcal{H}_t^\sigma\big(\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)&\le\sum_{\hat{k}=k(K)}^{\infty}\sum_{t=j+n_{\hat{k}}+1}^{j+n_{\hat{k}+1}}\mathcal{H}_t^\sigma\big(\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\\&\le\sum_{\hat{k}=k(K)}^{\infty}(m+\hat{k}+1)(1-\varepsilon)^{\hat{k}}.\end{aligned}$$

Therefore,

$$\lim_{K\to\infty}\sup_{\sigma\in\mathbf{BNE}^\delta}\sum_{t=K}^{\infty}\mathcal{H}_t^\sigma\big(\Omega\times\{\theta_j\},\big(\omega^{\beta_1},\theta\big)\big)\le\lim_{K\to\infty}\sum_{\hat{k}=k(K)}(m+\hat{k}+1)(1-\varepsilon)^{\hat{k}}=0. \qquad \square$$

APPENDIX C: MERGING AND THE PROVING LEMMA 3

The arguments in this section are results proved by Gossner (2011). We modify the arguments and notation slightly. We begin with the following key lemma of Gossner (2011).

LEMMA 9. *Let $\varepsilon\in(0,1)$ and $P,P'\in\Delta(X)$ for some finite set X. Suppose that $Q=\varepsilon P+(1-\varepsilon)P'$. Then*

$$D(P\|Q)\le-\log\varepsilon.$$

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

LEMMA 10. *Suppose that* $\gamma_0(\omega, \theta) > 0$. *Then for every* $\sigma \in \mathbf{BNE}^\delta$,

$$\pi_\infty^{\sigma,(\omega,\theta)}\big(\big\{h^\infty \in H^\infty : \big|\big\{t : D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big) > \varepsilon\big\}\big| \geq J\big\}\big) \leq -\frac{\log \gamma_0(\omega, \theta)}{J\varepsilon}.$$

PROOF. For every T, by the chain rule for Kullback–Leibler divergence,

$$\begin{aligned} D\big(\mathbf{marg}_{H^T}\pi_T^{\sigma,(\omega,\theta)}\|\mathbf{marg}_{H^T}\pi_T^\sigma\big) &= \mathbb{E}_{\pi_T^{\sigma,(\omega,\theta)}}\left[\sum_{t=0}^{T} D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big)\right] \\ &= \mathbb{E}_{\pi_\infty^{\sigma,(\omega,\theta)}}\left[\sum_{t=0}^{T} D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big)\right]. \end{aligned}$$

Moreover, $D(\mathbf{marg}_{H^T}\pi_T^{\sigma,(\omega,\theta)}\|\mathbf{marg}_{H^T}\pi_t^\sigma) \leq -\log\gamma_0(\omega, \theta)$ by the previous lemma. Therefore, by the monotone convergence theorem,

$$\mathbb{E}_{\pi_\infty^{\sigma,(\omega,\theta)}}\left[\sum_{t=0}^{\infty} D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big)\right] \leq -\log\gamma_0(\omega, \theta).$$

Then by Markov's inequality,

$$\begin{aligned} &\pi_\infty^{\sigma,(\omega,\theta)}\big(\big\{h^\infty \in H^\infty : \big|\big\{t : D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big) > \varepsilon\big\}\big| > J\big\}\big) \\ &\quad\leq \pi_\infty^{\sigma,(\omega,\theta)}\left(\sum_{t=0}^{\infty} D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big) > J\varepsilon\right) \\ &\quad\leq -\frac{\log\gamma_0(\omega, \theta)}{J\varepsilon}. \end{aligned}$$

□

The proof of Lemma 3 is now immediate.

PROOF OF LEMMA 3. Choose J_1 sufficiently large such that $-\frac{\log\gamma_0(\omega,\theta)}{J_1\varepsilon} < \varepsilon$. Then Lemma 3 is immediate from Lemma 10. □

APPENDIX D: PROOF OF LEMMA 2

By Lemmas 3 and 4, there exist J such that for all $\sigma \in \mathbf{BNE}^\delta$,

$$\begin{aligned} 1-\varepsilon &\leq \pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\big(\big\{h^\infty : \big|\big\{t : D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big) > \varepsilon\big\}\big| < J\big\}\big), \\ 1-\varepsilon &\leq \pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\big(\big\{h^\infty : \big|\big\{t : \nu_t^\sigma\big(\theta|h^t\big) < 1-\varepsilon\big\}\big| < J\big\}\big). \end{aligned}$$

Therefore, for all $\sigma \in \mathbf{BNE}^\delta$,

$$\begin{aligned} &\pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\big(\mathcal{M}^{\sigma,(\omega^{\beta_1},\theta)}(2J, \varepsilon, \varepsilon)\big) \\ &\quad\geq \pi_\infty^{\sigma,(\omega^{\beta_1},\theta)}\big(\big\{h^\infty : \big|\big\{t : D\big(\phi_t^{\sigma,(\omega,\theta)}(\cdot|h^t)\|\phi_t^\sigma(\cdot|h^t)\big) > \varepsilon\big\}\big|, \big|\big\{t : \nu_t^\sigma\big(\theta|h^t\big) < 1-\varepsilon\big\}\big| < J\big\}\big) \\ &\quad\geq 1-2\varepsilon. \end{aligned}$$

Appendix E: Proving Theorem 3

The proof of Theorem 3 uses ideas from Mertens, Sorin, and Zamir (2014) with some modifications. Let us begin with some notation. Given any probability vector $x \in \Delta(\Theta)$, let $\|x\|$ denote the Euclidean norm:

$$\|x\|^2 = \sum_{\theta \in \Theta} x(\theta)^2.$$

Note that if player 1 plays a strategy that induces $\lambda \in \Delta(A_1 \times \Theta)$ as the joint distribution over $A_1 \times \Theta$ and player 2 plays a_2, then player i obtains the expected utility

$$u_i(a_2, \lambda) := \mathbb{E}_\lambda\big[u_i(a_1, a_2, \theta)\big] = \sum_{a_1, \theta} u_i(a_1, a_2, \theta)\lambda(a_1, \theta).$$

We now extend the definition of a best response to ε-best response:

$$\mathrm{BR}_2^\varepsilon(\lambda) := \Big\{a_2 \in A_2 : \max_{a_2' \in A_2} u_2\big(a_2', \lambda\big) - u_2(a_2, \lambda) \le \varepsilon\Big\}.$$

Define for any $\varepsilon \ge 0$,

$$W^\varepsilon(\lambda) = \max_{a_2 \in B_2^\varepsilon(\lambda)} u_1(a_2, \lambda).$$

Finally, given $\lambda \in \Delta(A_1 \times \Theta)$, let $q(\cdot|y, \lambda)$ be the induced posterior belief about θ after observation of the signal y:

$$q(\theta|y, \lambda) = \frac{\displaystyle\sum_{a_1 \in A_1} \lambda(a_1, \theta)\psi(y|a_1, \theta)}{\displaystyle\sum_{\theta' \in \Theta}\sum_{a_1 \in A_1} \lambda\big(a_1, \theta'\big)\psi\big(y|a_1, \theta'\big)}.$$

Proposition 1. *For every $\varepsilon > 0$, there exists some $\rho > 0$ such that*

$$\mathbb{E}\big[\big\|q(\cdot|y, \lambda) - \mathbf{marg}_\Theta \lambda\big\|^2\big] < \rho \Rightarrow W^0(\lambda) \le \mathbf{cav}V(\mathbf{marg}_\Theta \lambda) + \varepsilon.$$

See Appendix H for the proof.

The following lemma provides a uniform bound (across all equilibria) on the number of times where the expected movement (in terms of $\|\cdot\|^2$ distance) in the SR players' beliefs is greater than ε.

Lemma 11. *For any $\sigma \in \mathbf{BNE}^\delta$ and any $\varepsilon > 0$,*

$$\big|\big\{t : \mathbb{E}_{\pi_\infty^\sigma}\big[\big\|\nu_{t+1}^\sigma\big(h^{t+1}\big) - \nu_t^\sigma\big(h^t\big)\big\|^2\big] \ge \varepsilon\big\}\big| \le \frac{1}{\varepsilon}.$$

Proof. Consider any time $t+1$:

$$\mathbb{E}_{\pi_\infty^\sigma}\big[\big\|\nu_{t+1}^\sigma\big(h^{t+1}\big) - \nu_0\big\|^2\big] = \mathbb{E}_{\pi_\infty^\sigma}\big[\big\|\nu_{t+1}^\sigma\big(h^{t+1}\big)\big\|^2\big] - \|\nu_0\|^2 \le 1.$$

By the martingale property of beliefs, π^σ_∞-almost surely, $\mathbb{E}_{\pi^\sigma_\infty}[\nu^\sigma_{\tau+1}(h^{\tau+1})|h^\tau] = \nu^\sigma_\tau(h^\tau)$. Therefore, it is straightforward to show that

$$1 \geq \mathbb{E}_{\pi^\sigma_\infty}\big[\big\|\nu^\sigma_{t+1}\big(h^{t+1}\big)\big\|^2\big] - \|\nu_0\|^2] = \sum_{\tau=0}^{t} \mathbb{E}_{\pi^\sigma_\infty}\big[\big\|\nu^\sigma_{\tau+1}\big(h^{\tau+1}\big) - \nu^\sigma_\tau\big(h^\tau\big)\big\|^2\big].$$

Since the above holds for every t, it implies that

$$\sum_{\tau=0}^{\infty} \mathbb{E}_{\pi^\sigma_\infty}\big[\big\|\nu^\sigma_{\tau+1}\big(h^{\tau+1}\big) - \nu^\sigma_\tau\big(h^\tau\big)\big\|^2\big] \leq 1,$$

which implies the claim. □

We can now prove Theorem 3.

Proof of Theorem 3. We first provide an upper bound on

$$\mathbb{E}_{\pi^\sigma_\infty}\left[(1-\delta)\sum_{t=0}^{\infty}\delta^t u_1\big(a_1^t, a_2^t, \theta\big)\right]$$

that holds across all $\sigma \in \mathbf{BNE}^\delta$. Notice that the above payoff is *not equal to* $U_1(\sigma, \theta; \delta)$, since the expectation does not condition on ω^s. However, one can interpret the payoff above as follows. For any equilibrium $\sigma \in \mathbf{BNE}^\delta$, let $\bar{\sigma}_1$ denote the strategy, where the LR player fictitiously draws some $\omega' \in \Omega$ according to μ_0 and plays the strategy in the equilibrium, σ, associated with that type for the entirety of the repeated game.[41] Indeed $U_1(\bar{\sigma}_1, \sigma_2; \delta)$ corresponds to the payoff above.

By Proposition 1, there exists some $\rho > 0$ such that

$$\mathbb{E}_\lambda\big[\big\|q(\cdot|y, \lambda) - p\big\|^2\big] < \rho \Rightarrow W^0(\lambda) \leq \mathbf{cav}V(\mathbf{marg}_\Theta \lambda) + \varepsilon/8.$$

Choose $n \in \mathbb{N}$ such that $\frac{1}{n}(\overline{u} - \underline{u}) < \varepsilon/8$ and δ^* such that for all $\delta > \delta^*$,

$$\big(1 - \delta^{\frac{nm}{\rho}}\big)\overline{u} + \delta^{\frac{nm}{\rho}}\,\mathbf{cav}V(\nu) + \frac{\varepsilon}{4} < \mathbf{cav}V(\nu) + \frac{\varepsilon}{2}. \tag{2}$$

For any $\sigma \in \mathbf{BNE}^\delta$, let

$$\mathcal{T}^\sigma := \left\{t : \mathbb{E}_{\pi^\sigma_\infty}\big[\big\|\nu^\sigma_{t+1}\big(\cdot|h^{t+1}\big) - \nu^\sigma_t\big(\cdot|h^t\big)\big\|^2\big] \geq \frac{\rho}{n}\right\}.$$

For all $t \notin \mathcal{T}^\sigma$, by Markov's inequality, we have

$$\pi^\sigma_\infty\big(\big\|\nu^\sigma_{t+1}\big(\cdot|h^{t+1}\big) - \nu^\sigma_t\big(\cdot|h^t\big)\big\|^2 \geq \rho\big) \leq \frac{1}{n}.$$

[41] For example, if the LR player draws a commitment type ω, then the LR player plays σ^ω. If instead the LR player indeed draws ω^s, then the LR player simply plays σ_1.

	B	N
H	$1,1$	$-1,0$
L	$2,-2$	$0,0$

FIGURE 7. Quality choice.

Thus, at all $t \notin \mathcal{T}^\sigma$,

$$\mathbb{E}_{\pi^\sigma_\infty}\big[W^0\big(\nu^\sigma_t(\cdot|h^t)\big)\big] \leq \frac{1}{n}(\overline{u}-\underline{u}) + \mathbb{E}_{\pi^\sigma_\infty}\big[\mathbf{cav}V\big(\nu^\sigma_t(\cdot|h^t)\big)\big] + \varepsilon/8 \leq \mathbf{cav}V(\nu_0) + \varepsilon/4.$$

Therefore,

$$\begin{aligned} U_1(\bar{\sigma}_1, \sigma_2; \delta) &\leq (1-\delta)\sum_{t=0}^{\infty}\delta^t \mathbb{E}_{\pi^\sigma_\infty}\big[W^0(\nu^\sigma_t(\cdot|h^t)\big] \\ &= (1-\delta)\Bigg(\sum_{t\in\mathcal{T}^\sigma}\delta^t\overline{u} + \sum_{t\notin\mathcal{T}^\sigma}\delta^t\mathbb{E}_{\pi^\sigma_\infty}\big[W^0\big(\nu^\sigma_t(\cdot|h^t)\big)\big]\Bigg) \\ &\leq (1-\delta)\Bigg(\sum_{t\in\mathcal{T}^\sigma}\delta^t\overline{u} + \sum_{t\notin\mathcal{T}^\sigma}\delta^t\big(\mathbf{cav}V(\nu_0)+\varepsilon/4\big)\Bigg). \end{aligned}$$

By Lemma 11, for every $\sigma \in \mathbf{BNE}^\delta$, $|\mathcal{T}^\sigma| \leq nm/\rho$. Therefore, for all $\delta > \delta^*$ and any $\sigma \in \mathbf{BNE}^\delta$,

$$U_1(\bar{\sigma}_1, \sigma_2; \delta) \leq \big(1-\delta^{nm/\rho}\big)\overline{u} + \delta^{nm/\rho}\mathbf{cav}V(\nu_0) + \varepsilon/4 < \mathbf{cav}V(\nu_0) + \varepsilon/2.$$

Finally, note that

$$U_1(\bar{\sigma}_1, \sigma_2; \delta) \geq \big(1-\mu_0\big(\Omega^c\big)\big)U_1(\sigma_1, \sigma_2; \delta) + \mu_0\big(\Omega^c\big)\underline{u}.$$

Let $\chi^* > 0$ be such that for all $\chi < \chi^*$,

$$\frac{1}{1-\chi}\left(\mathbf{cav}V(\nu_0) + \frac{\varepsilon}{2} - \chi\underline{u}\right) < \mathbf{cav}V(\nu_0) + \varepsilon.$$

Thus, for all $\delta > \delta^*$ and $\mu_0(\Omega^c) < \chi^*$,

$$U_1(\sigma_1, \sigma_2; \delta) \leq \frac{1}{1-\mu\big(\Omega^c\big)}\left(\mathbf{cav}V(\nu_0) + \frac{\varepsilon}{2} - \mu_0\big(\Omega^c\big)\underline{u}\right) < \mathbf{cav}V(\nu_0) + \varepsilon. \qquad \square$$

APPENDIX F: EXAMPLE

The following example shows that the probability of commitment types matters for the upper bound even when δ is close to 1. Consider the quality choice game with the stage game payoffs given by Figure 7. 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 B is a best response for the SR player in the stage game if and only if $\alpha_1(C) \geq 1/2$.

$\theta=-1$	$\bar{y}$	$\underline{y}$
H	$1/6$	$5/6$
L	$4/6$	$2/6$

$\theta=1$	$\bar{y}$	$\underline{y}$
H	$5/6$	$1/6$
L	$2/6$	$4/6$

FIGURE 8. The information structure.

There are two states $\Theta=\{1,-1\}$ that only affect the signal distribution of the public signal. There are two types in the game: $\Omega=\{\omega^c,\omega^s\}$. The commitment type, ω^c, in this game is a type that always plays the mixed action, $\frac{2}{3}H\oplus\frac{1}{3}L$, regardless of the state.[42] In particular, we assume that the probability of each state is identical and the probability of the commitment type is $\mu\in(0,1)$.

The signal space is binary, $Y=\{\bar{y},\underline{y}\}$ and the information structure is given by Fig. 8.

Note that according to this information structure, $(\frac{2}{3}H\oplus\frac{1}{3}L,\theta)$ is statistically indistinguishable from $(L,-\theta)$: $\psi(\cdot|\frac{2}{3}H\oplus\frac{1}{3}L,\theta)=\psi(\cdot|L,-\theta)$. In this example, we have the following observation.

CLAIM 3. *There exists μ^* such that for all $\mu>\mu^*$ and any $\delta\in(0,1)$, there exists an equilibrium in which the strategic player obtains a payoff of 2 in both states.*

PROOF. Consider the candidate equilibrium strategy profile in which the strategic LR player always plays L. 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:

$$\lambda_t^\sigma(H|h^t)=\frac{2}{3}\mu_t^\sigma(\omega^c|h^t)=\frac{2}{3}\big(\gamma_t^\sigma(\omega^c,1|h^t)+\gamma_t^\sigma(\omega^c,-1|h^t)\big).$$

Consider the likelihood ratio

$$\frac{\gamma_t^\sigma(\omega^c,\theta|h^t)}{\gamma_t^\sigma(\omega^s,-\theta|h^t)}=\frac{\gamma_t^\sigma(\omega^c,\theta|h^0)}{\gamma_t^\sigma(\omega^s,-\theta|h^0)}=\frac{\mu}{1-\mu}.$$

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

$$\lambda_t^\sigma(H|h^t)=\frac{2}{3}\mu>\frac{1}{2}.$$

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 strategic LR player's best response is to play B at all histories. Thus, the proposed strategy profile is indeed

[42]Note that this is in reality not the mixed Stackelberg action. However, by appropriately modifying the information structure, the same conclusions hold, even if the commitment type plays some other mixed action in every period.

an equilibrium. Furthermore, according to this strategy profile, the strategic LR player's payoff is 2 in both states, concluding the proof. □

The above discussion shows that when the commitment type occurs with large probability, even an arbitrarily patient strategic LR player obtains a payoff strictly greater than the Stackelberg payoff in equilibrium. We now examine an upper bound when the commitment type probability is small.

CLAIM 4. *Let $\varepsilon > 0$. Then there exists some $\mu^* > 0$ and $\delta^* < 1$ such that for all $\mu < \mu^*$ and $\delta > \delta^*$, $U_1(\sigma, \delta) < 3/2 + \varepsilon$ for all $\sigma \in \mathbf{BNE}^\delta$.*

PROOF. Consider $V(p)$ for any $p \in \Delta(\Theta)$. Because the stage game utilities are state-independent, it is straightforward to show that

$$V(p) \leq \sup_{\alpha_1 \in \mathcal{A}_1} \max_{a_2 \in B_2(\alpha_1)} u_1(\alpha_1, a_2) = 3/2,$$

where the equality follows from a straightforward calculation. The claim then follows from Theorem 3. □

APPENDIX G: PROOF OF COROLLARY 2

The lower bound is a consequence of Theorem 1. Let us now show the upper bound. Choose some $\underline{\nu} \in (0, \min_{\theta \in \Theta} \nu_0(\theta))$.

Suppose by way of contradiction that there exists some state $\theta^* \in \Theta$ and some sequence $\delta_n \to 1$ and $\sigma^n \in \mathbf{BNE}^{\delta_n}$ such that for all n, $U_1(\sigma^n, \theta^*; \delta_n) \geq u_1^*(\theta^*) + \varepsilon$. By Theorem 3,

$$\nu_0\big(\theta^*\big)\big(u_1^*\big(\theta^*\big) + \varepsilon\big) + \limsup_{n \to \infty} \sum_{\theta \neq \theta^*} \nu_0(\theta) U_1\big(\sigma^n, \theta; \delta_n\big) < \sum_{\theta \in \Theta} \nu_0(\theta) u_1^*(\theta) + \underline{\nu}\varepsilon.$$

Together with Theorem 1, we have

$$\sum_{\theta \neq \theta^*} \nu_0(\theta) u_1^*(\theta) \leq \limsup_{n \to \infty} \sum_{\theta \neq \theta^*} \nu_0(\theta) U_1\big(\sigma^n, \theta; \delta_n\big) \leq \sum_{\theta \neq \theta^*} \nu_0(\theta) u_1^*(\theta) - \big(\nu_0\big(\theta^*\big) - \underline{\nu}\big)\varepsilon,$$

but this is a contradiction.

APPENDIX H: PROVING PROPOSITION 1

Let us first define the set

$$\hat{\mathrm{NR}}(p) := \big\{\lambda \in \Delta(A_1 \times \Theta) : \big(\lambda(\cdot|\theta)\big)_{\theta \in \Theta} \in \mathrm{NR}(p), \mathbf{marg}_\Theta \lambda = p\big\},$$
$$\hat{\mathrm{NR}} := \bigcup_{p \in \Delta(\Theta)} \hat{\mathrm{NR}}(p).$$

Notice that $V(p) = \sup_{\lambda \in \hat{\mathrm{NR}}(p)} W^0(\lambda)$. Analogously, we can define for any $\varepsilon > 0$,

$$V^\varepsilon(p) = \sup_{\lambda \in \hat{\mathrm{NR}}(p)} W^\varepsilon(\lambda).$$

Finally, define also for every $\varepsilon \geq 0$,

$$\Lambda^\varepsilon(a_2) := \left\{\lambda \in \hat{\mathrm{NR}} : a_2 \in B_2^\varepsilon(\lambda)\right\}.$$

We begin with some lemmas.

LEMMA 12. *Let $\varepsilon > 0$. Then there exists some $\rho > 0$ such that for all $\lambda \in \Delta(A_1 \times \Theta)$,*

$$\mathbb{E}\left[\left\|q(\cdot|y,\lambda) - \mathbf{marg}_\Theta \lambda\right\|^2\right] < \rho \Rightarrow \inf_{\hat{\lambda} \in \hat{\mathrm{NR}}} \|\lambda - \hat{\lambda}\| < \varepsilon.$$

See Lemma V.3.6 in Mertens, Sorin, and Zamir (2014) for the proof.

LEMMA 13. *Let $\varepsilon > 0$. Then there exists some $\rho > 0$ such that for all $\lambda, \hat{\lambda} \in \Delta(A_1 \times \Theta)$,*

$$\|\lambda - \hat{\lambda}\| < \rho \Rightarrow W^0(\lambda) \leq W^\varepsilon(\hat{\lambda}) + \varepsilon.$$

PROOF. Let $\varepsilon > 0$. First choose $\rho' > 0$ sufficiently small such that

$$\|\lambda - \hat{\lambda}\| < \rho' \Rightarrow \max_{a_2 \in A_2} \left|u_2(a_2,\lambda) - u_2(a_2,\hat{\lambda})\right| \leq \varepsilon.$$

Then there exists some $\rho \in (0,\rho')$ such that $\|\lambda - \hat{\lambda}\| \leq \rho \Longrightarrow B_2^0(\lambda) \subseteq B_2^\varepsilon(\hat{\lambda})$. Therefore, whenever $\|\lambda - \hat{\lambda}\| \leq \rho$,

$$W^0(\lambda) = \max_{a_2 \in B_2^0(\lambda)} u_1(a_2,\lambda) \leq \max_{a_2 \in B_2^\varepsilon(\hat{\lambda})} u_1(a_2,\lambda) \leq \max_{a_2 \in B_2^\varepsilon(\hat{\lambda})} u_1(a_2,\hat{\lambda}) + \varepsilon = W^\varepsilon(\hat{\lambda}) + \varepsilon. \qquad \square$$

LEMMA 14. *For every $\varepsilon > 0$, there exists some $\rho > 0$ such that for all $a_2 \in A_2$,*

$$\lambda \in \Lambda^\rho(a_2) \Rightarrow \inf_{\lambda' \in \Lambda^0(a_2)} \left\|\lambda - \lambda'\right\| < \varepsilon.$$

PROOF. Suppose otherwise. Then for some $a_2 \in A_2$ and $\varepsilon > 0$, there exists some sequence $\rho_n \to 0$ and $\lambda_n \in \Lambda^{\rho_n}(a_2)$ such that

$$\inf_{\lambda' \in \Lambda^0(a_2)} \left\|\lambda_n - \lambda'\right\| \geq \varepsilon. \tag{3}$$

By Bolzano–Weierstrass, without loss of generality, by replacing the original sequence with an appropriate subsequence, we can assume this sequence to be convergent to some limit λ. However, note that since $\lambda_n \to \lambda$ and $\lambda_n \in \Lambda^{\rho_n}(a_2)$ for all n, $\lambda \in \Lambda^0(a_2)$. This contradicts (3). $\square$

LEMMA 15. *For every $\varepsilon > 0$, there exists some $\rho^* > 0$ such that for all $\lambda \in \hat{\mathrm{NR}}$ and all $\rho < \rho^*$,*

$$W^\rho(\lambda) \leq \mathbf{cav}V(\mathbf{marg}_\Theta \lambda) + \varepsilon.$$

PROOF. First, because $\mathbf{cav}V$ and $u_1(\cdot, a_2)$ are Lipschitz continuous for all $a_2 \in A_2$, there exists some $\varepsilon' > 0$ such that whenever $\|\lambda - \lambda'\| < \varepsilon'$, then

$$\left|\mathbf{cav}V(\mathbf{marg}_\Theta \lambda) - \mathbf{cav}V\left(\mathbf{marg}_\Theta \lambda'\right)\right|, \max_{a_2 \in A_2} \left|u_1(a_2, \lambda) - u_1\left(a_2, \lambda'\right)\right| < \varepsilon/2.$$

By the previous lemma, let $\rho > 0$ be such that for all $a_2 \in A_2$,

$$\lambda \in \Lambda^\rho(a_2) \Rightarrow \inf_{\lambda' \in \Lambda^0(a_2)} \left\|\lambda - \lambda'\right\| < \varepsilon'.$$

Recall that

$$W^\rho(\lambda) = \max_{a_2 \in B_2^\rho(\lambda)} u_1(a_2, \lambda).$$

Let $a_2^\rho(\lambda) \in B_2^\rho(\lambda)$ be the solution to the above maximization problem. Thus, for every $\lambda \in \hat{\text{NR}}$, $\lambda \in \Lambda^\rho(a_2^\rho(\lambda))$. Therefore, for all $\lambda \in \hat{\text{NR}}$, there exists some $\lambda'(\lambda) \in \Lambda^0(a_2^\rho(\lambda))$ with $\|\lambda - \lambda'(\lambda)\| \leq \varepsilon'$.

Then for any $\lambda \in \hat{\text{NR}}$,

$$\begin{aligned} W^\rho(\lambda) = u_1\left(a_2^\rho(\lambda), \lambda\right) &\leq \max_{a_2 \in B_2^0(\lambda'(\lambda))} u_1(a_2, \lambda) \\ &\leq W^0\left(\lambda'(\lambda)\right) + \varepsilon/2 \\ &\leq \mathbf{cav}V\left(\mathbf{marg}_\Theta \lambda'(\lambda)\right)) + \varepsilon/2 \leq \mathbf{cav}V(\mathbf{marg}_\Theta \lambda) + \varepsilon. \end{aligned}$$ □

We can now prove Proposition 1.

PROOF OF PROPOSITION 1. By Lemma 15, there exists some $\rho^* \in (0, \varepsilon/3)$ such that for all $\hat{\lambda} \in \hat{\text{NR}}$,

$$W^{\rho^*}(\hat{\lambda}) \leq \mathbf{cav}V(\mathbf{marg}_\Theta \hat{\lambda}) + \varepsilon/3.$$

By Lemma 13 and Lipschitz continuity of $\mathbf{cav}V$, there exists some $\rho' > 0$ such that

$$\left\|\lambda - \lambda'\right\| < \rho' \Rightarrow W^0(\lambda) \leq W^{\rho^*}\left(\lambda'\right) + \rho^*, \left|\mathbf{cav}V(\mathbf{marg}_\Theta \lambda) - \mathbf{cav}V\left(\mathbf{marg}_\Theta \lambda'\right)\right| < \varepsilon/3.$$

By Lemma 12, there exists $\rho > 0$ such that for all λ for which $\mathbb{E}[\|q(\cdot|y, \lambda) - \mathbf{marg}_\Theta \lambda\|^2] < \rho$, there exists $\hat{\lambda}(\lambda) \in \hat{\text{NR}}$ such that $\|\hat{\lambda}(\lambda) - \lambda\| < \rho'$.

Thus, for any λ in which $\mathbb{E}[\|q(\cdot|y, \lambda) - \mathbf{marg}_\Theta \lambda\|^2] < \rho$, we have

$$W^0(\lambda) \leq W^{\rho^*}\left(\hat{\lambda}(\lambda)\right) + \rho^* \leq \mathbf{cav}V(\mathbf{marg}_\Theta\left(\hat{\lambda}(\lambda)\right) + 2\varepsilon/3 \leq \mathbf{cav}V(\mathbf{marg}_\Theta(\lambda) + \varepsilon.$$ □

REFERENCES

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2016), "Fragility of asymptotic agreement under Bayesian learning." *Theoretical Economics*, 11, 187–225. [0173]

Al-Najjar, Nabil I. (2009), "Decision makers as statisticians: Diversity, ambiguity, and learning." *Econometrica*, 77, 1371–1401. [0173]

Al-Najjar, Nabil I. and Mallesh M. Pai (2014), "Coarse decision making and overfitting." *Journal of Economic Theory*, 150, 467–486. [0173]

Aoyagi, Masaki (1996), "Reputation and dynamic Stackelberg leadership in infinitely repeated games." *Journal of Economic Theory*, 71, 378–393. [0172]

Atakan, Alp E. and Mehmet Ekmekci (2011), "Reputation in long-run relationships." *The Review of Economic Studies*, rdr037. [0172]

Atakan, Alp E. and Mehmet Ekmekci (2015), "Reputation in the long-run with imperfect monitoring." *Journal of Economic Theory*, 157, 553–605. [0172]

Aumann, Robert J., Michael Maschler, and Richard E. Stearns (1995), "Repeated games with incomplete information." MIT press. [0172]

Blackwell, David and Lester Dubins (1962), "Merging of opinions with increasing information." *The Annals of Mathematical Statistics*, 33, 882–886. [0171, 0190]

Celentani, Marco, Drew Fudenberg, David K. Levine, and Wolfgang Pesendorfer (1996), "Maintaining a reputation against a long-lived opponent." *Econometrica*, 64, 691–704. [0172]

Cripps, Martin W., Eddie Dekel, and Wolfgang Pesendorfer (2005), "Reputation with equal discounting in repeated games with strictly conflicting interests." *Journal of Economic Theory*, 121, 259–272. [0172]

Cripps, Martin W., George J. Mailath, and Larry Samuelson (2004), "Imperfect monitoring and impermanent reputations." *Econometrica*, 72, 407–432. [0195]

Ely, Jeffrey C., Drew Fudenberg, and David K. Levine (2008), "When is reputation bad?" *Games and Economic Behaivor*, 63, 498–526. [0179]

Ely, Jeffrey C. and Juuso Välimäki (2003), "Bad reputation." *Quarterly Journal of Economics*, 118, 785–814. [0179]

Evans, Robert and Jonathan P. Thomas (1997), "Reputation and experimentation in repeated games with two long-run players." *Econometrica*, 65, 1153–1173. [0172]

Fudenberg, Drew and David K. Levine (1986), "Limit games and limit equilibria." *Journal of Economic Theory*, 38, 261–279. [0187, 0189, 0190]

Fudenberg, Drew and David K. Levine (1989), "Reputation and equilibrium selection in games with a patient player." *Econometrica*, 57, 759–778. [0172]

Fudenberg, Drew and David K. Levine (1992), "Maintaining a reputation when strategies are imperfectly observed." *Review of Economic Studies*, 59, 561–579. [0170, 0172, 0174, 0187, 0193]

Fudenberg, Drew and David K. Levine (1993a), "Self-confirming equilibrium." *Econometrica*, 61, 523–546. [0173]

Fudenberg, Drew and David K. Levine (1993b), "Steady state learning and Nash equilibrium." *Econometrica*, 61, 547–574. [0173]

Fudenberg, Drew and Yuichi Yamamoto (2010), "Repeated games where the payoffs and monitoring structure are unknown." *Econometrica*, 78, 1673–1710. [0172]

Ghosh, Sambuddha (2014), "Multiple long-lived opponents and the limits of reputation." Report. [0172]

Gossner, Olivier (2011), "Simple bounds on the value of a reputation." *Econometrica*, 79, 1627–1641. [0171, 0172, 0187, 0188, 0189, 0190, 0193, 0198]

Hörner, Johannes and Stefano Lovo (2009), "Belief-free equilibria in games with incomplete information." *Econometrica*, 77, 453–487. [0172]

Hörner, Johannes, Stefano Lovo, and Tristan Tomala (2011), "Belief-free equilibria in games with incomplete information: Characterization and existence." *Journal of Economic Theory*, 146, 1770–1795. [0172]

Kalai, Ehud and Ehud Lehrer (1993), "Rational learning leads to Nash equilibrium." *Econometrica*, 61, 1019–1045. [0190]

Kreps, David M. and Robert J. Wilson (1982), "Reputation and imperfect information." *Journal of Economic Theory*, 27, 253–279. [0172]

Mertens, Jean-François, Sylvain Sorin, and Shmuel Zamir (2014), *Repeated Games*, volume 55. Cambridge University Press. [0172, 0193, 0200, 0205]

Milgrom, Paul R. and John Roberts (1982), "Predation, reputation and entry deterrence." *Journal of Economic Theory*, 27, 280–312. [0172]

Moscarini, Giuseppe and Lones Smith (2002), "The law of large demand for information." *Econometrica*, 70, 2351–2366. [0171, 0173, 0191, 0195]

Mu, Xiaosheng, Luciano Pomatto, Philipp Strack, and Omer Tamuz (2021), "From Blackwell dominance in large samples to Rényi divergences and back again." *Econometrica*, 89, 475–506. [0173, 0195]

Pei, Harry (2020), "Reputation effects under interdependent values." *Econometrica*, 88, 2175–2202. [0173]

Schmidt, Klaus M. (1993), "Reputation and equilibrium characterization in repeated games of conflicting interests." *Econometrica*, 61, 325–351. [0172]

Torgersen, Erik (1991), *Comparison of Statistical Experiments*, volume 36. Cambridge University Press. [0171, 0191, 0195]

Wiseman, Thomas (2005), "A partial folk theorem for games with unknown payoff distributions." *Econometrica*, 73, 629–645. [0172]

Co-editor Simon Board handled this manuscript.

Manuscript received 25 January, 2022; final version accepted 13 June, 2024; available online 25 June, 2024.