



Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
<http://economics.sas.upenn.edu/pier>

PIER Working Paper 13-009

“Reputation in the Presence of Noisy Exogenous Learning”

by

Ju Hu

<http://ssrn.com/abstract=2213830>

Reputation in the Presence of Noisy Exogenous Learning

Ju Hu*

February 6, 2013

Abstract

This paper studies the reputation effect in which a long-lived player faces a sequence of uninformed short-lived players and the uninformed players receive informative but noisy exogenous signals about the type of the long-lived player. We provide an explicit lower bound on all Nash equilibrium payoffs of the long-lived player. The lower bound shows when the exogenous signals are sufficiently noisy and the long-lived player is patient, he can be assured of a payoff strictly higher than his minmax payoff.

Keywords: Reputation, repeated games, learning, relative entropy

JEL Classification: C73; D82; D83

*Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104; Email: juhu1@sas.upenn.edu. I am grateful to George Mailath for his insightful suggestions and encouragement. All remaining errors are mine.

1 Introduction

This paper studies the reputation effect in the long-run interactions in which a long-lived player faces a sequence of uninformed short-lived players and the uninformed players receive informative but noisy exogenous signals about the type of the long-lived player. In the canonical reputation models without exogenous learning (Fudenberg and Levine (1989), Fudenberg and Levine (1992)), the long-lived player can effectively build a reputation by mimicking the behavior of a commitment type because the short-lived player will play a best response to the commitment action in all but a finite number of periods after always seeing the commitment action. The underlying reason is the fact that the short-lived player cannot be surprised too many times: every time the short-lived player expects the commitment action with small probability and yet this action is actually chosen, the posterior belief on this commitment type jumps up, but at the same time the beliefs can not exceed unity. However this “finite number of surprises” intuition does not carry over to the case with exogenous learning. It is still true that each surprise leads to a discrete jump of the posterior beliefs. But after a surprise during the periods of no surprises, the exogenous learning can drive down the posterior beliefs. After a long history without surprises, the posterior beliefs may return to the original level, resulting in another surprise. Typically, this can happen infinitely many times. Hence in the presence of exogenous learning, there is no guarantee that we have a finite number of surprises.

Wiseman (2009) first presented an infinitely repeated chain store game example with perfect monitoring and exogenous signals taking two possible values. He shows that when the long-lived player is sufficiently patient and there is sufficient noise in the signals, the long-lived player can effectively build a reputation and assure himself of a payoff strictly higher than his minmax payoff.

This paper extends Wiseman (2009) to more general reputation models with exogenous learning. We provide an *explicit* lower bound on all Nash equilibrium payoffs to the long-lived player. The lower bound is characterized by the commitment action, discount factor, prior belief and how noisy the learning process is. For fixed commitment action and discount factor, the lower bound increases in both prior probability and noise in the exogenous signals. This is intuitive as a higher prior probability on the commitment type and a noisier and slower exogenous learning process both correspond to easier reputation building. When the long-lived player become sufficiently patient, the effect of the prior probability vanishes while that of the exogenous learning remains. This is again intuitive because the prior probability represents the

cost of reputation building in the initial periods. When the long-lived player places arbitrarily high weight on future periods, the cost in the initial periods becomes negligible. In contrast, learning has a long run effect. The longer the history, the more the uninformed player can learn about the type of his opponent. Hence the effect of learning remains even if the long-lived player become sufficiently patient.

Not surprisingly, the lower bound we derive is generally lower than that if there is no exogenous learning, reflecting the negative effect of learning on reputation building. In the case that signals are completely uninformative, these two bounds coincide. Nonetheless, when the signals are sufficiently noisy, the lower bound shows that in any Nash equilibrium, the long-lived player is assured of a payoff strictly higher than his minmax value.

To derive the lower bound, we apply the relative entropy approach first introduced by Gossner (2011) to the study of reputations. Gossner (2011) uses this approach to the standard reputation game in Fudenberg and Levine (1992) and obtains an explicit lower bound on all equilibrium payoffs. He also shows when the commitment types are sufficiently rich and the long-lived player is arbitrarily patient, the lower bound is exactly the Stackelberg payoff which confirms the result in Fudenberg and Levine (1992). Ekmekci, Gossner, and Wilson (2012) applied this method to the reputation game in which the type of the long-lived player is governed by an underlying stochastic process. They calculate explicit lower bounds for all equilibrium payoffs at the beginning of the game and all continuation payoffs. In these two papers, relative entropy only serves as a measure of prediction errors. However, in this paper, in addition to a measure of prediction errors, the concept of relative entropy is also naturally adapted to the learning situation as a measure of noise in the exogenous signals. This again makes relative entropy as a more suitable tool.

The rest of the paper is organized as follows. In section 2, we describe the reputation model with exogenous learning and introduce relative entropy. Section 3 presents and discusses the main result, which is proved in Section 4.

2 Model

2.1 Reputation game with exogenous learning

We consider the canonical reputation model (Mailath and Samuelson (2006), Chapter 15) in which a fixed stage game is infinitely repeated. The stage game is a two-player simultaneous-move finite game of private monitoring. Denote by A_i the finite set of

actions for player i in the stage game. Actions in the stage game are imperfectly observed. At the end of each period, player i only observes a private signal z_i drawn from a finite set Z_i . If an action profile $a \in A_1 \times A_2 \equiv A$ is chosen, the signal vector $z \equiv (z_1, z_2) \in Z_1 \times Z_2 \equiv Z$ is realized according to the distribution $\pi(\cdot|a) \in \Delta(Z)$.¹ The marginal distribution of player i 's private signals over Z_i is denoted by $\pi_i(\cdot|a)$. Both $\pi(\cdot|a)$ and $\pi_i(\cdot|a)$ have obvious extensions $\pi(\cdot|\alpha)$ and $\pi_i(\cdot|\alpha)$ respectively to mixed action profiles. Player i 's ex-post stage game payoff from his action a_i and private signal z_i is given by $u_i^*(a_i, z_i)$. Player i 's ex ante stage game payoff from action profile $(a_i, a_{-i}) \in A$ is $u_i(a_i, a_{-i}) = \sum_{z_i} \pi_i(z_i|a_i, a_{-i})u_i^*(a_i, z_i)$. Notice this setting includes as special cases the perfect monitoring environment (Fudenberg and Levine (1989)) in which $Z_1 = Z_2 = A$ and $\pi(z_1, z_2|a) = 1$ if and only if $z_1 = z_2 = a$, and the public monitoring environment (Fudenberg and Levine (1992)) in which $Z_1 = Z_2$ and $\pi(z_1, z_2|a) > 0$ implies $z_1 = z_2$. Player 1 is a long-lived player with discount factor $\delta \in (0, 1)$ while player 2 is a sequence of short-lived players each of whom only lives for one period. In any period t , the long-lived player 1 observes both his own previous actions and private signals, but the current generation of the short-lived player 2 only observes previous private signals of his predecessors.

There is uncertainty about the type of player 1. Let $\Xi \equiv \{\xi_0\} \cup \hat{\Xi}$ be the set of all possible types of player 1. ξ_0 is the *normal type* of player 1. His payoff in the repeated game is the average discounted sum of stage game payoffs $(1 - \delta) \sum_{t \geq 0} \delta^t u_1(a^t)$. Each $\xi(\hat{\alpha}_1) \in \hat{\Xi}$ denotes a *simple commitment type* who plays the stage game (mixed) action $\hat{\alpha}_1 \in \Delta(A_1)$ in every period independent of histories. Assume $\hat{\Xi}$ is either finite or countable. The type of player 1 is unknown to player 2. Let $\mu \in \Delta(\Xi)$ be player 2's prior belief about player 1's type, with full support.

At period $t = -1$, nature selects a type $\xi \in \Xi$ of player 1 according to the initial distribution μ . Player 2 does not observe the type of player 1. However, we assume that the uninformed player 2 has access to an exogenous channel which gradually reveals the true type of player 1. More specifically, conditional on the type ξ , a stochastic process $\{\eta_t(\xi)\}_{t \geq 0}$ generates a signal $y^t \in Y$ after every period's play, where Y is a finite set of all possible signals. To distinguish the signals $z \in Z$ generated from each period's play and the signals $y \in Y$ generated by $\{\eta_t(\xi)\}_{t \geq 0}$, we call the former *endogenous signals* and the latter *exogenous signals*. In addition to observing previous endogenous signals, each generation of player 2 also observes all the exogenous signals from earlier periods. We assume that for each type $\xi \in \Xi$, the stochastic process $\{\eta_t(\xi)\}_{t \geq 0}$ is independent and identically distributed across t .

¹For a finite set X , $\Delta(X)$ denotes the set of all probability distributions over X .

Conditional on ξ , the distribution of the exogenous signals in every period is denoted by $\rho(\cdot|\xi) \in \Delta(Y)$. Notice this assumes that the realization of the exogenous signals are independent of the play, hence it models the exogenous learning of the uninformed player 2.

For expositional convenience, we assume player 1 does not observe the exogenous signals. This assumption is not crucial for our result. The same lower bound will apply if we assume player 1 also observes the exogenous signals. A private history of player 1 in period t consists of his previous actions and endogenous signals, denoted by $h_1^t \equiv (a_1^0, z_1^0, a_1^1, z_1^1, \dots, a_1^{t-1}, z_1^{t-1}) \in H_{1t} \equiv (A_1 \times Z_1)^t$, with the usual notation $H_{10} = \{\emptyset\}$. A behavior strategy for player 1 is a map

$$\sigma_1 : \Xi \times \bigcup_{t=0}^{\infty} H_{1t} \rightarrow \Delta(A_1),$$

with the restriction that for all $\xi(\hat{\alpha}_1) \in \hat{\Xi}$,

$$\sigma_1(\xi(\hat{\alpha}_1), h_1^t) = \hat{\alpha}_1 \quad \text{for all } h_1^t \in \bigcup_{t=0}^{\infty} H_{1t}.$$

A private history of player 2 in period t contains both previous endogenous and exogenous signals, denoted by $h_2^t \equiv (z_2^0, y_2^0, z_2^1, y_2^1, \dots, z_2^{t-1}, y_2^{t-1}) \in H_{2t} \equiv (Z_2 \times Y)^t$, with $H_{20} = \{\emptyset\}$. A behavior strategy for player 2 is a map

$$\sigma_2 : \bigcup_{t=0}^{\infty} H_{2t} \rightarrow \Delta(A_2).$$

Denote by Σ_i the strategy space of player i .

Any strategy profile $\sigma \equiv (\sigma_1, \sigma_2) \in \Sigma_1 \times \Sigma_2$, together with the prior μ and the signal distributions $\{\pi(\cdot|a)\}_{a \in A}$ and $\{\rho(\cdot|\xi)\}_{\xi \in \Xi}$, induces a probability measure P^σ over the set of states $\Omega \equiv \Xi \times (A_1 \times A_2 \times Z_1 \times Z_2 \times Y)^\infty$. The measure P^σ describes how the uninformed player 2 expects play to evolve. Let \tilde{P}^σ be the conditional probability of P^σ given the event that player 1 is the normal type. The measure \tilde{P}^σ describes how play evolves if player 1 is the normal type. We use $E^\sigma[\cdot]$ (resp., $\tilde{E}^\sigma[\cdot]$) to denote the expectation with respect to the probability measure P^σ (resp., \tilde{P}^σ).

A Nash equilibrium in this reputation game is a pair of mutual best responses.

Definition 1. A strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*) \in \Sigma_1 \times \Sigma_2$ is a Nash equilibrium if it satisfies:

(a) for all $\sigma_1 \in \Sigma_1$,

$$\tilde{E}^{\sigma^*} \left[(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_1(a^t) \right] \geq \tilde{E}^{(\sigma_1, \sigma_2^*)} \left[(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_1(a^t) \right],$$

(b) for all $h_2^t \in \bigcup_{\tau \geq 0} H_{2\tau}$ with positive probability under P^{σ^*} ,

$$\sigma_2^*(h_2^t) \in \arg \max_{\alpha_2 \in \Delta(A_2)} E^{\sigma^*} \left[u_2(\sigma_1^*(h_1^t, \xi), \alpha_2) \middle| h_2^t \right].$$

Condition (a) states that given σ_2^* , the normal type of player 1 maximizes his expected lifetime utility. Condition (b) requires that given σ_1^* , player 2 updates his belief via Bayes' rule along the path of play and plays a myopic best response since he is short lived.

2.2 Relative entropy

The *relative entropy* between two probability distributions P and Q over a finite set X is the expected log likelihood ratio

$$d(P||Q) \equiv E_P \log \frac{P(x)}{Q(x)} = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)},$$

with the usual convention that $0 \log \frac{0}{q} = 0$ if $q \geq 0$ and $p \log \frac{p}{0} = \infty$ if $p > 0$. Relative entropy is always nonnegative and it is zero if and only if the two distributions are identical (See Cover and Thomas (2006), Gossner (2011) and Ekmekci, Gossner, and Wilson (2012) for more details on relative entropy).

Relative entropy measures the speed of the learning process of the uninformed player 2. For each commitment type $\xi(\hat{\alpha}_1) \in \hat{\Xi}$, let $\lambda_{\xi(\hat{\alpha}_1)}$ be the relative entropy of the exogenous signal distributions when player 1 is the normal type and when he is the commitment type $\xi(\hat{\alpha}_1)$, i.e.

$$\lambda_{\xi(\hat{\alpha}_1)} \equiv d(\rho(\cdot | \xi_0) || \rho(\cdot | \xi(\hat{\alpha}_1))).$$

Relative entropy measures how different the two distributions $\rho(\cdot | \xi_0)$ and $\rho(\cdot | \xi(\hat{\alpha}_1))$ are. In terms of learning, $\lambda_{\xi(\hat{\alpha}_1)}$ measures how *fast* player 2 can learn from exogenous signals that player 1 is *not* the commitment type $\xi(\hat{\alpha}_1)$ when player 1 is indeed the normal type. The larger $\lambda_{\xi(\hat{\alpha}_1)}$ is, the faster the learning process is. This is illustrated by the two polar cases. If $\lambda_{\xi(\hat{\alpha}_1)} = 0$, then the distributions of the exogenous signals when player 1 is the normal type and when he is of type $\xi(\hat{\alpha}_1)$ are identical. In this case, from the exogenous signals, player 2 can never distinguish the normal type from the commitment type $\xi(\hat{\alpha}_1)$ when player 1 is the normal type. If $\lambda_{\xi(\hat{\alpha}_1)} = \infty$, there must be some signal $y \in Y$ which will occur when player 1 is the normal type but will not occur when player 1 is the commitment type $\xi(\hat{\alpha}_1)$. Hence in this case, player 2 will learn that player 1 is not the commitment type $\xi(\hat{\alpha}_1)$ for sure in finite time

when player 1 is the normal type. For other intermediate values $0 < \lambda_{\xi(\hat{\alpha}_1)} < \infty$, conditional on the normal type, player 2 will eventually know that player 1 is not the commitment type $\xi(\hat{\alpha}_1)$.

The following assumption rules out extremely fast learning. Technically, it requires that the support of $\rho(\cdot|\xi)$ be contained in the support of $\rho(\cdot|\xi(\hat{\alpha}_1))$ for every commitment type $\xi(\hat{\alpha}_1)$.

Assumption 1. $\lambda_{\xi(\hat{\alpha}_1)} < \infty$ for all $\xi(\hat{\alpha}_1) \in \hat{\Xi}$.

Relative entropy measures the error in player 2's one step ahead prediction on the endogenous signals. Gossner (2011) first introduced the following notion of ε -entropy-confirming best response (see also Ekmekci, Gossner, and Wilson (2012)):

Definition 2. The mixed action $\alpha_2 \in \Delta(A_2)$ is an ε -entropy-confirming best response to $\alpha_1 \in \Delta(A_1)$ if there exists $\alpha'_1 \in \Delta(A_1)$ such that

- (a) α_2 is a best response to α'_1 ,
- (b) $d(\pi_2(\cdot|\alpha_1, \alpha_2) \parallel \pi_2(\cdot|\alpha'_1, \alpha_2)) \leq \varepsilon$.

The set of all ε -entropy confirming best responses to α_1 is denoted by $B_\varepsilon(\alpha_1)$.

The idea of ε -entropy-confirming best response is similar to ε -confirming best response defined in Fudenberg and Levine (1992). If player 2 plays a myopic best response α_2 to his belief that player 1 plays the action α'_1 , then player 2 believes that his endogenous signals realize according to the distribution $\pi_2(\cdot|\alpha'_1, \alpha_2)$. If the true action taken by player 1 is α_1 instead of α'_1 , then the true distribution of player 2's endogenous signals is indeed $\pi_2(\cdot|\alpha_1, \alpha_2)$. Hence player 2's one step ahead prediction error on his endogenous signals is, measured by relative entropy, $d(\pi_2(\cdot|\alpha_1, \alpha_2) \parallel \pi_2(\cdot|\alpha'_1, \alpha_2))$. The mixed action α_2 is an ε -entropy-confirming best response of α_1 if the prediction error is no greater than ε .

For any commitment type $\xi(\hat{\alpha}_1) \in \hat{\Xi}$, let

$$\underline{V}_{\xi(\hat{\alpha}_1)}(\varepsilon) \equiv \inf_{\alpha_2 \in B_\varepsilon(\hat{\alpha}_1)} u_1(\hat{\alpha}_1, \alpha_2)$$

be the lowest possible payoff to player 1 if he plays $\hat{\alpha}_1$ while player 2 plays an ε -entropy-confirming best response to $\hat{\alpha}_1$. Let $V_{\xi(\hat{\alpha}_1)}(\cdot)$ be the pointwise supremum of all convex functions below $\underline{V}_{\xi(\hat{\alpha}_1)}$. Clearly $V_{\xi(\hat{\alpha}_1)}$ is convex and nonincreasing.

3 Main result

For any $\delta \in (0, 1)$, let $\underline{U}_1(\delta)$ denote the infimum of all Nash equilibrium payoffs to the normal type of player 1 if the discount factor is δ . Our main result is the following:

Proposition 1. *Under Assumption 1, for all $\delta \in (0, 1)$,*

$$\underline{U}_1(\delta) \geq \sup_{\xi(\hat{\alpha}_1) \in \hat{\Xi}} V_{\xi(\hat{\alpha}_1)} \left(- (1 - \delta) \log \mu(\xi(\hat{\alpha}_1)) + \lambda_{\xi(\hat{\alpha}_1)} \right).$$

To understand the equilibrium lower bound in Proposition 1, it suffices to consider the reputation building on each $\xi(\hat{\alpha}_1) \in \hat{\Xi}$ since the overall lower bound is obtained by considering all possible commitment types. Fix a commitment type $\xi(\hat{\alpha}_1) \in \hat{\Xi}$. Proposition 1 states that in any Nash equilibrium, the normal type of player 1 is assured of a payoff no less than $V_{\xi(\hat{\alpha}_1)} \left(- (1 - \delta) \log \mu(\xi(\hat{\alpha}_1)) + \lambda_{\xi(\hat{\alpha}_1)} \right)$. Recall that $V_{\xi(\hat{\alpha}_1)}$ is a nonincreasing function. For fixed δ , this lower bound increases with $\mu(\xi(\hat{\alpha}_1))$ while decreases with $\lambda_{\xi(\hat{\alpha}_1)}$. The intuition is straightforward. A larger prior probability on the commitment type $\xi(\hat{\alpha}_1)$ makes it easier for the normal type of player 1 to build a reputation on this commitment type. In another word, the cost of reputation building in the initial periods is smaller in this case which leads to a higher lower bound. However the learning process goes against reputation building because player 2 eventually learns that player 1 is not the commitment type $\xi(\hat{\alpha}_1)$. It is then intuitive that the speed of learning matters. If the exogenous signals are sufficiently noisy, then $\lambda_{\xi(\hat{\alpha}_1)}$ is small and it is hard for player 2 to distinguish the normal type and the commitment type. This results in a rather slow learning process and hence a high lower bound. If the learning process is completely uninformative, $\lambda_{\xi(\hat{\alpha}_1)} = 0$, then the lower bound is given by $V_{\xi(\hat{\alpha}_1)} \left(- (1 - \delta) \log \mu(\xi(\hat{\alpha}_1)) \right)$ which is exactly the same lower bound derived in Gossner (2011) without exogenous learning. In general, when $\lambda_{\xi(\hat{\alpha}_1)} > 0$, the lower bound is lower than that in Gossner (2011) due to the learning effect.

Another parameter in the lower bound is player 1's discount factor δ . An interesting feature in the lower bound is that δ only appears as a coefficient for the term $\log \mu(\xi(\hat{\alpha}_1))$, not for $\lambda_{\xi(\hat{\alpha}_1)}$. This is because $-\log \mu(\xi(\hat{\alpha}_1))$ captures the cost of reputation building in the initial periods while $\lambda_{\xi(\hat{\alpha}_1)}$ is the learning effect which remains active as the game evolves. As a result, when player 1 becomes arbitrarily patient, $\delta \rightarrow 1$, the cost of reputation building in the initial periods becomes negligible since player 1 places higher and higher weight on the payoff obtained in later periods, whereas the learning effect remains unchanged. In this case, the lower bound becomes $V_{\xi(\hat{\alpha}_1)}(\lambda_{\xi(\hat{\alpha}_1)})$.² Moreover, in the presence of multiple commitment types, which commitment type is the most favorable is now ambiguous. Intuitively, this is because the effectiveness of reputation building does not only depend on the

²Since $V_{\xi(\hat{\alpha}_1)}(\varepsilon)$ is convex, it is continuous at every $\varepsilon > 0$.

stage game payoff from the commitment type but also on the learning process. Even if player 2 assigns positive probability on the Stackelberg action, committing to the Stackelberg action may not help player 1 effectively build a reputation because the exogenous signals may reveal quickly to player 2 that player 1 is not the Stackelberg commitment type. This is in a sharp contrast with the result in standard models without exogenous learning.

We use the following example which is first considered in Wiseman (2009) to illustrate the lower bound obtained in Proposition 1.

Example. There is a long-lived incumbent, player 1, facing a sequence of short-lived entrants, player 2. In every period, the entrant chooses between entering (E) and staying out (S) while the incumbent decides whether to fight (F) or accommodate (A). The stage game payoff is given in Figure 1, where $a > 1$ and $b > 0$.

	E	S
F	-1, -1	$a, 0$
A	0, b	$a, 0$

Figure 1: Chain store stage game.

The stage game is infinitely repeated with perfect monitoring. There are two types of player 1, the normal type, denoted by ξ_0 , and a simple commitment type, denoted by $\xi(F)$ who plays the stage game Stackelberg action F in every period independent of histories. The prior probability of $\xi(F)$ is $\mu(\xi(F))$. The exogenous signals observed by player 2 only take two values: \bar{y} and \underline{y} . Assume $\rho(\bar{y}|\xi_0) = \beta$, $\rho(\bar{y}|\xi(F)) = \alpha$ and $\beta > \alpha$. Thus

$$\lambda_{\xi(F)} = \beta \log \frac{\beta}{\alpha} + (1 - \beta) \log \frac{1 - \beta}{1 - \alpha}.$$

Now we apply Proposition 1 in this setting. Because monitoring is perfect, it is easy to see $B_\varepsilon(F) = \{S\}$ when $\varepsilon < \log \frac{b+1}{b}$. Therefore, we have

$$V_{\xi(F)}(\varepsilon) = \begin{cases} a - \frac{a+1}{\log \frac{b+1}{b}} \varepsilon, & \text{if } \varepsilon < \log \frac{b+1}{b}, \\ -1 & \text{if } \varepsilon \geq \log \frac{b+1}{b}. \end{cases}$$

Proposition 1 then implies for all $\delta \in (0, 1)$

$$\underline{U}_1(\delta) \geq a - \frac{a+1}{\log \frac{b+1}{b}} \left(- (1 - \delta) \log \mu(\xi(F)) + \lambda_{\xi(F)} \right),$$

and in the limit

$$\liminf_{\delta \rightarrow 1} \underline{U}_1(\delta) \geq a - (a+1) \frac{\lambda_{\xi(F)}}{\log \frac{b+1}{b}}. \quad (1)$$

Wiseman (2009) considers symmetrically distributed signals, i.e., $\beta = 1 - \alpha > 1/2$, and derives a lower bound of $a - (a + 1) \frac{\log \frac{\beta}{1-\beta}}{\log \frac{b+1}{b}}$. Because in this symmetric case $\lambda_{\xi(F)} = (2\beta - 1) \log \frac{\beta}{1-\beta}$, this bound is lower than that in (1). As signals become less informative, i.e. $\beta \rightarrow \frac{1}{2}$, both lower bounds become arbitrarily close to player 1's Stackelberg payoff.

Although it is not surprising that exogenous learning affects reputation building, why does it take this particular form, i.e. the relative entropy of the exogenous signals? As mentioned previously, the “finite number of surprises” argument in Fudenberg and Levine (1989) does not apply because of the downward pressure on posterior beliefs due to exogenous learning. In this particular example, the uninformed entrants may enter infinitely many times even if he is always fought after any entry. Moreover, receiving the signal \bar{y} always decreases the posterior beliefs (recall $\rho(\bar{y}|\xi_0) > \rho(\bar{y}|\xi(F))$) which is the source of the downward pressure. Thus the strength of this downward pressure depends exactly on how frequently the entrants can receive the signal \bar{y} which, together with the size of surprise, in turn determines how long it takes for the posterior beliefs to return after a surprise. In other words, the size of surprise and the relative frequency of exogenous signals together determine the frequency of entries. If it takes a long time for the posterior beliefs to return, then the entrants can not enter too frequently and the incumbent can effectively build a reputation.

To see this, fix any Nash equilibrium σ . For any history h^∞ in which F is always played, let $\{\mu_t\}_{t \geq 0}$ be player 2's posterior belief on the commitment type along this history. Player 2 is willing to enter in period t only if

$$\text{Prob}(F) \equiv \mu_t + (1 - \mu_t)\sigma_1(\xi_0, h^t)(F) \leq \frac{b}{b+1}.$$

So, if player 2 enters in period t , we must have

$$\mu_t \leq \frac{b}{b+1} \tag{2}$$

and

$$\sigma_1(\xi_0, h^t)(F) \leq \frac{b}{b+1}. \tag{3}$$

We examine the odds ratio $\{\mu_t/(1 - \mu_t)\}_{t \geq 0}$ along this history. Since the entrant is always fought along this history, the odds ratio evolves as

$$\frac{\mu_{t+1}}{1 - \mu_{t+1}} = \left(\frac{\alpha}{\beta}\right)^{\mathbb{1}_{\bar{y}}(y^t)} \left(\frac{1 - \alpha}{1 - \beta}\right)^{\mathbb{1}_y(y^t)} \frac{\mu_t}{(1 - \mu_t)\sigma_1(\xi_0, h^t)(F)} \quad \forall t \geq 0,$$

where for $y \in \{\bar{y}, \underline{y}\}$, $\mathbb{1}_y$ is the indicator function, $\mathbb{1}_y(y^t) = 1$ if $y^t = y$ and 0 otherwise. Because $\sigma_1(\xi, h^t)(F)$ is always less than or equal to 1, we have

$$\frac{\mu_{t+1}}{1 - \mu_{t+1}} \geq \left(\frac{\alpha}{\beta}\right)^{\mathbb{1}_{\bar{y}}(y^t)} \left(\frac{1 - \alpha}{1 - \beta}\right)^{\mathbb{1}_y(y^t)} \frac{\mu_t}{1 - \mu_t} \tag{4}$$

if player 2 stays out in period t . Because inequality (3) holds if player 2 enters in period t , we have

$$\frac{\mu_{t+1}}{1 - \mu_{t+1}} \geq \left(\frac{\alpha}{\beta}\right)^{\mathbb{1}_{\bar{y}}(y^t)} \left(\frac{1 - \alpha}{1 - \beta}\right)^{\mathbb{1}_{\bar{y}}(y^t)} \frac{b + 1}{b} \frac{\mu_t}{1 - \mu_t} \quad (5)$$

if he enters in period t . For any $t \geq 1$, let $n_E(t)$, $n_{\bar{y}}(t)$ be the number of entries and the number of signal \bar{y} 's respectively in history h^t . Inequalities (4), (5) and simple induction imply

$$\frac{\mu_t}{1 - \mu_t} \geq \left(\frac{b + 1}{b}\right)^{n_E(t)} \left(\frac{\alpha}{\beta}\right)^{n_{\bar{y}}(t)} \left(\frac{1 - \alpha}{1 - \beta}\right)^{t - n_{\bar{y}}(t)} \frac{\mu(\xi(F))}{1 - \mu(\xi(F))} \quad \forall t \geq 1. \quad (6)$$

Moreover, if player 2 enters in period t , inequality (2) implies

$$b \geq \frac{\mu_t}{1 - \mu_t}. \quad (7)$$

Hence inequalities (6) and (7) together yield

$$b \geq \left(\frac{b + 1}{b}\right)^{n_E(t)} \left(\frac{\alpha}{\beta}\right)^{n_{\bar{y}}(t)} \left(\frac{1 - \alpha}{1 - \beta}\right)^{t - n_{\bar{y}}(t)} \frac{\mu(\xi(F))}{1 - \mu(\xi(F))} \quad (8)$$

for all t at which player 2 enters. Let $\{t_k\}_{k \geq 0}$ be the sequence of periods in which entry occurs. By taking log and dividing both sides by t_k , inequality (8) implies

$$\begin{aligned} \limsup_{k \rightarrow \infty} \frac{n_E(t_k)}{t_k} &\leq \frac{1}{\log \frac{b+1}{b}} \lim_{k \rightarrow \infty} \left[\frac{n_{\bar{y}}(t_k)}{t_k} \log \frac{\beta}{\alpha} + \left(1 - \frac{n_{\bar{y}}(t_k)}{t_k}\right) \log \frac{1 - \beta}{1 - \alpha} \right] \\ &= \frac{\lambda_{\xi(F)}}{\log \frac{b+1}{b}}, \end{aligned}$$

because $\lim_t n_{\bar{y}}(t)/t = \beta$ by law of large numbers. Because for every $t \geq 1$, there exists $k \geq 0$ such that $t_k \leq t < t_{k+1}$ and $n_E(t)/t = n_E(t_k)/t \leq n_E(t_k)/t_k$, the above inequality also holds for the whole sequence

$$\limsup_{t \rightarrow \infty} \frac{n_E(t)}{t} \leq \frac{\lambda_{\xi(F)}}{\log \frac{b+1}{b}}.$$

This inequality states exactly what we have mentioned above: the fraction of entries along a typical history is determined by the size of surprise $\frac{b+1}{b}$ and the relative frequency of the exogenous signals $\lambda_{\xi(F)}$. Lastly, because this inequality holds for all Nash equilibria, we have

$$\liminf_{\delta \rightarrow 1} \underline{U}_1(\delta) \geq \left(1 - \frac{\lambda_{\xi(F)}}{\log \frac{b+1}{b}}\right)a + \frac{\lambda_{\xi(F)}}{\log \frac{b+1}{b}}(-1) = a - (a + 1) \frac{\lambda_{\xi(F)}}{\log \frac{b+1}{b}}.$$

This is exactly the lower bound in (1).

4 Proof of Proposition 1

One important property of relative entropy is the chain rule. Let P and Q be two distributions over the product $X \times Y$ (see for example Cover and Thomas (2006) Chapter 2 and Gossner (2011)). The chain rule states that the relative entropy of P and Q can be expanded as the sum of a relative entropy and a conditional relative entropy:

$$d(P\|Q) = d(P_X\|Q_X) + E_{P_X}(d(P(\cdot|x)\|Q(\cdot|x))),$$

where P_X (resp., Q_X) is the marginal distribution of P (resp., Q) over X and $P(\cdot|x)$ (resp., $Q(\cdot|x)$) is the conditional probability of P (resp., Q) over Y given x .

Fix a commitment type $\xi(\hat{\alpha}_1) \in \hat{\Xi}$. Suppose $\sigma = (\sigma_1, \sigma_2)$ is a Nash equilibrium of the reputation game with exogenous learning. Let P^σ be the probability measure over $\Xi \times (A_1 \times A_2 \times Z_1 \times Z_2 \times Y)^\infty$ induced by σ , μ and $\{\rho(\cdot|\xi)\}_{\xi \in \Xi}$, as in section 2. Let \hat{P}^σ be the conditional probability of P^σ given the event that player 1 is the commitment type $\xi(\hat{\alpha}_1)$. The measure \hat{P}^σ determines how the play evolves if player 1 is of type $\xi(\hat{\alpha}_1)$.

Let $\sigma'_1 \in \Sigma_1$ be the strategy for player 1 in which the normal type of player 1 mimics the behavior of the commitment type $\xi(\hat{\alpha}_1)$, i.e. $\sigma'_1(\xi_0, h_1^t) = \hat{\alpha}_1$ for all $h_1^t \in \bigcup_{t \geq 0} H_{1t}$. Let $\sigma' = (\sigma'_1, \sigma_2)$. The probability measure $\tilde{P}^{\sigma'}$ (recall from section 2, $\tilde{P}^{\sigma'} = P^{\sigma'}(\cdot | \{\xi_0\} \times (A_1 \times A_2 \times Z_1 \times Z_2 \times Y)^\infty)$) describes how the normal type of player 1 expects the play to evolve if he *deviates* to the commitment strategy of $\xi(\hat{\alpha}_1)$. The only difference between $\tilde{P}^{\sigma'}$ and \hat{P}^σ is the distributions of player 2's exogenous signals. Because we assume the realizations of player 2's exogenous signals only depend on the type of player 1 and are independent of the play, for all $h^t \in (A_1 \times A_2 \times Z_1 \times Z_2 \times Y)^t$ we have

$$\tilde{P}^{\sigma'}(h^t) = \hat{P}^\sigma(h^t) \prod_{\tau=0}^{t-1} \frac{\rho(y^\tau|\xi_0)}{\rho(y^\tau|\xi(\hat{\alpha}_1))},$$

where y^0, y^1, \dots, y^{t-1} are the exogenous signals contained in the history h^t . Notice by Assumption 1, $\rho(y|\xi(\hat{\alpha}_1)) > 0$ whenever $\rho(y|\xi_0) > 0$. Hence the right hand side of the above equality is well defined.

Let P_2^σ , $\tilde{P}_2^{\sigma'}$ and \hat{P}_2^σ be the marginal distributions of P^σ , $\tilde{P}^{\sigma'}$ and \hat{P}^σ respectively on player 2's histories $(Z_2 \times Y)^\infty$, and let $\{P_{2t}^\sigma\}_{t \geq 1}$, $\{\tilde{P}_{2t}^{\sigma'}\}_{t \geq 1}$ and $\{\hat{P}_{2t}^\sigma\}_{t \geq 1}$ be the corresponding finite dimensional distributions. In period -1 before the play, player 2 believes that P_{2t}^σ is the distributions of his signals (both endogenous and exogenous) in the first t periods. However, if player 1 is the normal type and he deviates to the commitment strategy of $\xi(\hat{\alpha}_1)$, $\tilde{P}_{2t}^{\sigma'}$ is the true distribution of player 2's signals in the

first t periods. The following lemma gives an upper bound on the prediction errors in player 2's first t periods signals.

Lemma 1. *For all $t \geq 1$,*

$$d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) \leq -\log \mu(\xi(\hat{\alpha}_1)) + t\lambda_{\xi(\hat{\alpha}_1)}.$$

Proof. We show this by a simple calculation:

$$\begin{aligned} d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) &\equiv \sum_{h_2^t \in H_{2t}} \tilde{P}_{2t}^{\sigma'}(h_2^t) \log \frac{\tilde{P}_{2t}^{\sigma'}(h_2^t)}{P_{2t}^{\sigma}(h_2^t)} \\ &= \sum_{h_2^t \in H_{2t}} \tilde{P}_{2t}^{\sigma'}(h_2^t) \log \left[\frac{\hat{P}_{2t}^{\sigma}(h_2^t)}{P_{2t}^{\sigma}(h_2^t)} \prod_{\tau=0}^{t-1} \frac{\rho(y^\tau | \xi_0)}{\rho(y^\tau | \xi(\hat{\alpha}_1))} \right] \\ &= \sum_{h_2^t \in H_{2t}} \tilde{P}_{2t}^{\sigma'}(h_2^t) \log \frac{\hat{P}_{2t}^{\sigma}(h_2^t)}{P_{2t}^{\sigma}(h_2^t)} + \sum_{h_2^t \in H_{2t}} \tilde{P}_{2t}^{\sigma'}(h_2^t) \log \left(\prod_{\tau=0}^{t-1} \frac{\rho(y^\tau | \xi_0)}{\rho(y^\tau | \xi(\hat{\alpha}_1))} \right). \end{aligned}$$

Notice the second term is the relative entropy of the distributions on player 2's exogenous signals in the first t periods when player 1 is the normal type and when he is the commitment type $\xi(\hat{\alpha}_1)$. Because the exogenous signals are conditionally independent across time, the chain rule implies the second term is exactly $t\lambda_{\xi(\hat{\alpha}_1)}$. Moreover, since \hat{P}_{2t}^{σ} is obtained by conditioning P_{2t}^{σ} on the event that player 1 is the commitment type $\xi(\hat{\alpha}_1)$, we have

$$\frac{\hat{P}_{2t}^{\sigma}(h_2^t)}{P_{2t}^{\sigma}(h_2^t)} \leq \mu(\xi(\hat{\alpha}_1)) \quad \forall h_2^t \in H_{2t}.$$

Therefore the first term is no greater than $-\log \mu(\hat{\theta})$. These two observations imply the desired result. \square

For any private history $h_2^t \in \bigcup_{t \geq 0} H_{2t}$, $P_{2,t+1}^{\sigma}$ (resp., $\tilde{P}_{2,t+1}^{\sigma'}$) induces player 2's one step ahead prediction on his *endogenous* signals $z_2^t \in Z_2$, denoted by $p_{2t}^{\sigma}(\cdot | h_2^t)$ (resp., $\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t)$).³ In the equilibrium, at the information set h_2^t , player 2 believes that his endogenous signals will realize according to $p_{2t}^{\sigma}(\cdot | h_2^t)$. But if player 2 had known that player 1 was the normal type and played like the commitment type $\xi(\hat{\alpha}_1)$, then player 2 would predict his endogenous signals according to $\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t)$.

³If h_2^t has probability 0 under P^{σ} , i.e. it is not reached in the equilibrium σ , then the one step ahead prediction is not well defined. But this does not matter because we will consider the average (over h_2^t) one step prediction errors.

For any $t \geq 1$, let $\tilde{E}_{2t}^{\sigma'}[\cdot]$ denote the expectation over H_{2t} with respect to the probability measure $\tilde{P}_{2t}^{\sigma'}$. The following lemma is a direct application of the chain rule.

Lemma 2. *For all $t \geq 0$,*

$$\tilde{E}_{2t}^{\sigma'} [d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t))] \leq d(\tilde{P}_{2,t+1}^{\sigma'} \| P_{2,t+1}^{\sigma}) - d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}),$$

where $d(\tilde{P}_{2,0}^{\sigma'} \| P_{2,0}^{\sigma}) \equiv 0$.

Proof. Let $q_{2,t+1}(\cdot | h_2^t, z_2^t)$ (resp., $\tilde{q}_{2,t+1}(\cdot | h_2^t, z_2^t)$) be the one step ahead prediction on his *exogenous* signals if he had observed his past private history h_2^t and current period endogenous signal z_2^t , induced by $P_{2,t+1}^{\sigma}$ (resp., $\tilde{P}_{2,t+1}^{\sigma'}$). Because Assumption 1 and Lemma 1 implies $d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) < \infty$ for all $t \geq 1$, applying chain rule twice yields

$$\begin{aligned} & d(\tilde{P}_{2,t+1}^{\sigma'} \| P_{2,t+1}^{\sigma}) - d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) \\ &= \tilde{E}_{2t}^{\sigma'} [d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t))] + E_{2,t+1}^{\dagger} [d(\tilde{q}_{2,t+1}(\cdot | h_2^t, z_2^t) \| q_{2,t+1}(\cdot | h_2^t, z_2^t))], \end{aligned}$$

where $E_{2,t+1}^{\dagger}$ is with respect to the marginal distribution of $\tilde{P}_{2,t+1}^{\sigma'}$ over $(Z_2 \times Y)^t \times Z_2$. The desired result is obtained by noting that the last term in the above expression is nonnegative because relative entropy is always nonnegative. \square

Let $d_{\xi(\hat{\alpha}_1)}^{\delta, \sigma}$ be the expected average discounted sum of player 2's one step ahead prediction errors if player 1 is the normal type and he deviates to mimicking the commitment type $\xi(\hat{\alpha}_1)$

$$\begin{aligned} d_{\xi(\hat{\alpha}_1)}^{\delta, \sigma} &\equiv \tilde{E}^{\sigma'} \left[(1 - \delta) \sum_{t=0}^{\infty} \delta^t d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t)) \right] \\ &= (1 - \delta) \sum_{t=0}^{\infty} \delta^t \tilde{E}_{2t}^{\sigma'} [d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t))], \end{aligned}$$

where δ is player 1's discount factor.

The next lemma, combining Lemma 1 and Lemma 2, provides an upper bound for $d_{\xi(\hat{\alpha}_1)}$.

Lemma 3.

$$d_{\xi(\hat{\alpha}_1)}^{\delta, \sigma} \leq -(1 - \delta)\mu(\xi(\hat{\alpha}_1)) + \lambda_{\xi(\hat{\alpha}_1)}.$$

Proof.

$$\begin{aligned}
d_{\xi(\hat{\alpha}_1)}^{\delta, \sigma} &\leq (1 - \delta) \sum_{t=0}^{\infty} \delta^t \left(d(\tilde{P}_{2,t+1}^{\sigma'} \| P_{2,t+1}^{\sigma}) - d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) \right) \\
&= (1 - \delta) \sum_{t=0}^{\infty} \delta^t d(\tilde{P}_{2,t+1}^{\sigma'} \| P_{2,t+1}^{\sigma}) - (1 - \delta) \sum_{t=0}^{\infty} \delta^t d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) \\
&= (1 - \delta)^2 \sum_{t=1}^{\infty} \delta^{t-1} d(\tilde{P}_{2t}^{\sigma'} \| P_{2t}^{\sigma}) \\
&\leq (1 - \delta)^2 \sum_{t=1}^{\infty} \delta^{t-1} \left[-\log \mu(\xi(\hat{\theta})) + t \lambda_{\xi(\hat{\alpha}_1)} \right] \\
&= -(1 - \delta) \log \mu(\xi(\hat{\theta})) + \lambda_{\xi(\hat{\alpha}_1)},
\end{aligned}$$

where the first inequality comes from Lemma 2 and the second inequality from Lemma 1. \square

An important feature of Lemma 3 is that the upper bound on the expected prediction error is independent of P^{σ} and $\tilde{P}^{\sigma'}$, which allows us to bound player 1's payoff in *any* Nash equilibrium.

Proof of Proposition 1. In equilibrium, at any information set $h_2^t \in \bigcup_{t \geq 0} H_{2t}$ that is reached with positive probability, $\sigma_2(h_2^t)$ is a best response to $E(\sigma_1(\xi, h_1^t) | h_2^t)$ and his one step ahead prediction on his endogenous signals is $p_{2t}^{\sigma}(\cdot | h_2^t)$. If player 1 is the normal type and he deviates to mimicking $\xi(\hat{\alpha}_1)$, the one step ahead prediction is $\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t)$. Thus at any h_2^t with positive probability under $\tilde{P}^{\sigma'}$, player 2 plays a $d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t))$ -entropy confirming best response to $\hat{\alpha}_1$.⁴ Because σ is a Nash equilibrium, the deviation is not profitable. Hence in equilibrium, the payoff to the normal type is at least as high as

$$\begin{aligned}
&\tilde{E}^{\sigma'} \left[(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_1(a^t) \right] \\
&= (1 - \delta) \sum_{t=0}^{\infty} \delta^t \tilde{E}_{2t}^{\sigma'} \left[u_1(\hat{\alpha}_1, \sigma_2(h_2^t)) \right] \\
&\geq (1 - \delta) \sum_{t=0}^{\infty} \delta^t \tilde{E}_{2t}^{\sigma'} \left[V_{\xi(\hat{\alpha}_1)} \left(d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t)) \right) \right] \\
&\geq (1 - \delta) \sum_{t=0}^{\infty} \delta^t \tilde{E}_{2t}^{\sigma'} \left[V_{\xi(\hat{\alpha}_1)} \left(d(\tilde{p}_{2t}^{\sigma'}(\cdot | h_2^t) \| p_{2t}^{\sigma}(\cdot | h_2^t)) \right) \right] \\
&\geq V_{\xi(\hat{\alpha}_1)} \left(-(1 - \delta) \log \mu(\xi(\hat{\alpha}_1)) + \lambda_{\xi(\hat{\alpha}_1)} \right),
\end{aligned}$$

⁴Because $\tilde{P}^{\sigma'}$ is absolutely continuous with respect to P^{σ} .

where the second inequality comes from the definition of $V_{\xi(\hat{\alpha}_1)}$ and the last inequality from Jensen's inequality and Lemma 3. Since the Nash equilibrium σ and the commitment type $\xi(\hat{\alpha}_1)$ are arbitrary, the result follows. \square

References

- COVER, T., AND J. THOMAS (2006): *Elements of information theory*. Wiley-interscience.
- EKMEKCI, M., O. GOSSNER, AND A. WILSON (2012): "Impermanent Types and Permanent Reputations," *Journal of Economic Theory*, 147(1), 162–178.
- FUDENBERG, D., AND D. K. LEVINE (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57(4), pp. 759–778.
- (1992): "Maintaining a Reputation when Strategies are Imperfectly Observed," *The Review of Economic Studies*, 59(3), pp. 561–579.
- GOSSNER, O. (2011): "Simple Bounds on The Value of a Reputation," *Econometrica*, 79(5), pp. 1627–1641.
- MAILATH, G. J., AND L. SAMUELSON (2006): *Repeated Games and Reputations: Long-run Relationships*. Oxford University Press, USA.
- WISEMAN, T. (2009): "Reputation and exogenous private learning," *Journal of Economic Theory*, 144(3), 1352–1357.