Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 153 (2014) 64–73

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Notes

Reputation in the presence of noisy exogenous learning ☆

Ju Hu

Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, United States

Received 6 February 2013; final version received 5 March 2014; accepted 27 May 2014

Available online 2 June 2014

Abstract

This note extends Wiseman [6] to more general reputation games with exogenous learning. Using Gossner's [4] relative entropy method, we provide an explicit lower bound on all Nash equilibrium payoffs of the long-lived player. The lower bound shows that 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.

JEL classification: C73; D82; D83

Keywords: Reputation; Repeated games; Learning; Relative entropy

1. Introduction

Wiseman [6] studies an infinitely repeated chain store reputation game in which the short-lived entrants receive noisy exogenous signals about the type of the long-lived incumbent. He shows that a sufficiently patient long-lived incumbent can effectively build a reputation and assure himself of a payoff strictly higher than his minmax payoff provided the exogenous signals are noisy enough.

☆ I am grateful to George Mailath and Yuichi Yamamoto for their insightful suggestions and encouragement, as well as an associate editor and an anonymous referee for helpful comments. I would also like to thank participants at UPenn Micro Theory Lunch Seminar for valuable discussions. All remaining errors are mine.

E-mail address: juhu1@sas.upenn.edu.

http://dx.doi.org/10.1016/j.jet.2014.05.008

This note extends Wiseman [6] to more general reputation models with exogenous learning. The analysis is built on Gossner [4] who introduces the relative entropy approach to the study of standard reputation models in Fudenberg and Levine [3] and obtains an explicit lower bound on all equilibrium payoffs. This paper shows Gossner's [4] powerful tool can also be naturally adapted to reputation models with exogenous learning.[1] We provide an explicit lower bound on all Nash equilibrium payoffs to the long-lived player which is a modification of that in Gossner [4]. The lower bound is characterized by the commitment action, discount factor, prior belief and in particular how noisy the learning process is. In general this lower bound is lower than that in Gossner [4] because of learning and these two bounds coincide if the exogenous signals are completely uninformative.

The rest of the paper is organized as follows. In Section 2, we describe the reputation model with exogenous learning. Section 3 presents the main result. Section 4 applies the obtained lower bound to the example considered in Wiseman [6] and discusses the result. All the proofs are in Appendix A.

2. Model

2.1. Reputation game with exogenous learning

We consider the canonical reputation model (Mailath and Samuelson [5], 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 $\pi(\cdot|a) \in \Delta(Z)$.[2] 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 $u_i(a_i, a_{-i}) = \sum_{z_i} \pi_i(z_i|a_i, a_{-i}) u_i^*(a_i, z_i)$. 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.

[1] Ekmekci et al. [2] also applies the relative entropy approach to the reputation game in which the type of the long-lived player is governed by an underlying stochastic process.

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

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. Conditional on ξ, the distribution of the exogenous signals in every period is denoted by $\rho(\cdot|\xi) \in \Delta(Y)$.

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

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

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

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

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

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

A strategy profile (σ_1^*, σ_2^*) is a Nash equilibrium of this reputation game if it is a pair of mutual best responses: i) given σ_2^*, the normal type of player 1 maximizes his expected lifetime utility, ii) given σ_1^*, player 2 updates his belief via Bayes' rule along the path of play and plays a myopic best response.

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.[3]

Following Gossner [4], relative entropy can be used to measure the error in player 2's one step ahead prediction of the *endogenous signals*.

[3] See Cover and Thomas [1], Gossner [4] and Ekmekci et al. [2] for more details on relative entropy.

Definition 1. 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) $d(\pi_2(\cdot|\alpha_1,\alpha_2)\|\pi_2(\cdot|\alpha_1',\alpha_2)) \le \varepsilon$.

The set of all ε-entropy confirming best responses to α_1 is denoted by $B_\varepsilon(\alpha_1)$.[4]

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.

In addition to one step ahead prediction error, relative entropy will also measure the speed of learning from *exogenous signals*. 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.

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

This value $\lambda_{\xi(\hat{\alpha}_1)}$ 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.

The following assumption rules out extremely fast learning.[5]

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

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 $\delta \in (0,1)$,*

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

To understand this lower bound, consider 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 $V_{\xi(\hat{\alpha}_1)}(-(1-\delta)\log\mu(\xi(\hat{\alpha}_1)) + \lambda_{\xi(\hat{\alpha}_1)})$. Recall that $V_{\xi(\hat{\alpha}_1)}$ is a nonincreasing function. For fixed δ,

[4] For more detailed discussion of ε-entropy-confirming best response, see Gossner [4].

[5] 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 $\xi(\hat{\alpha}_1)$.

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

Fig. 1. Chain store stage game.

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. 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 $V_{\xi(\hat{\alpha}_1)}(-(1-\delta)\log\mu(\xi(\hat{\alpha}_1)))$ which is exactly the same lower bound derived in Gossner [4] without exogenous learning. In general, when $\lambda_{\xi(\hat{\alpha}_1)} > 0$, the lower bound is lower than that in Gossner [4] due to the learning effect. If player 1 becomes arbitrarily patient, $\delta \to 1$, the lower bound becomes $V_{\xi(\hat{\alpha}_1)}(\lambda_{\xi(\hat{\alpha}_1)})$.[6] The effect of prior probability vanishes while that of learning remains unchanged. Moreover, in the presence of multiple commitment types, which commitment type is the most favorable now depends on both stage game payoffs and the speed of learning. Even if player 2 assigns positive probability on the Stackelberg type, 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.

4. An example

We use the following example which is first considered in Wiseman [6] to illustrate the lower bound obtained in Proposition 1.

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 Fig. 1, where $a > 1$ and $b > 0$.

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

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

If player 2 expects player 1 to fight with probability $\alpha(F) > \frac{b}{b+1}$, so that his prediction of player 1's play is off by $\log\frac{1}{\alpha(F)} < \log\frac{b+1}{b}$, then his unique best response is to stay out, i.e. $B_\varepsilon(F) = \{S\}$ when $\varepsilon < \log\frac{b+1}{b}$. So we have

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

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

and thus

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

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

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

and in the limit

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

Wiseman [6] 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. $\beta \to \frac{1}{2}$, both lower bounds become arbitrarily close to player 1's Stackelberg payoff.

As we have mentioned, the term $\lambda_\xi(F)$ in (1) measures the speed of exogenous learning. In particular, it measures how frequently the exogenous signals reveal to player 2 that player 1 is not a commitment type, which in turn determines how frequently player 2 enters. This point is more transparent by considering the following direct arguments.

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

$$b \geq \frac{\mu_t}{1-\mu_t}. \tag{7}$$

Hence inequalities (6) and (7) together yield

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

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

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

because $\frac{1}{t_k}\log(b\frac{1-\mu(\xi(F))}{\mu(\xi(F))}) \to 0$ as $t_k \to \infty$ and conditional on the normal type $n_{\bar{y}}(t_k)/t_k \to \beta$ 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

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

Lastly, because this inequality holds for all Nash equilibria, we have

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

This is exactly the lower bound in (1).

Appendix A

The whole proof follows the same line of arguments as in Gossner [4], modified to incorporate exogenous learning. The main idea is to estimate the payoff that player 1 would get in any equilibrium if he deviated to a commitment strategy, say $\hat{\alpha}_1$. In every period, player 2 plays a myopic best response to what he expects player 1 to play, but the actual play is always $\hat{\alpha}_1$ if the normal type deviated. If we measure this "prediction error" d_t by the relative entropy between the probability distribution over player 2's histories generated by normal type's deviation and that generated by the equilibrium play, then player 2 is actually always playing an d_t-entropy-confirming best response to $\hat{\alpha}_1$, yielding a payoff to player 1 at least as high as $\underline{V}_{\xi(\hat{\alpha}_1)}(d_t)$ in every period. The definition and convexity of $V_{\xi(\hat{\alpha}_1)}$ will then imply that ex ante player 1 can ensure himself a payoff at least as high as $V_{\xi(\hat{\alpha}_1)}((1-\delta)\sum \delta^t d_t)$ by deviation, and so this value is an lower bound for player 1's equilibrium payoff. In Gossner [4], the distribution over histories generated by normal type's deviation is exactly the same as that generated by the commitment type's play. However, in the current setting, these two distributions are in general different because the exogenous signals generated by the normal type and the commitment type are different. Nonetheless, we show here Gossner's [4] proof can also be applied to the current setting with some modifications.

Formally, fix $\sigma = (\sigma_1, \sigma_2)$ a Nash equilibrium and a commitment type $\xi(\hat{\alpha}_1) \in \hat{\Xi}$. 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 $\widehat{P}^\sigma$ be the conditional probability of P^σ given the event that player 1 is the commitment type $\xi(\hat{\alpha}_1)$. The measure $\widehat{P}^\sigma$ determines how the play evolves if player 1 is the commitment type $\xi(\hat{\alpha}_1)$.

Let σ_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 $\widetilde{P}^{\sigma'} \equiv P^{\sigma'}(\cdot|\{\xi_0\} \times (A_1 \times A_2 \times Z_1 \times Z_2 \times Y)^\infty))$ describes how the normal type of player 1 expects the play to evolve if he *deviates* to the commitment strategy of $\xi(\hat{\alpha}_1)$. As we have mentioned above, $\widetilde{P}^{\sigma'}$ and $\widehat{P}^\sigma$ differ in the distributions of player 2's exogenous signals. Because player 2's exogenous signals only depend on the type of player 1, for all $h^t \in (A_1 \times A_2 \times Z_1 \times Z_2 \times Y)^t$ we have

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

where $y^0, y^1, \cdots, y^{t-1}$ 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, $\widetilde{P}_2^{\sigma'}$ and $\widehat{P}_2^\sigma$ be the marginal distributions of P^σ, $\widetilde{P}^{\sigma'}$ and $\widehat{P}^\sigma$ respectively on player 2's histories $(Z_2 \times Y)^\infty$, and let $\{P_{2t}^\sigma\}_{t\geq 1}$, $\{\widetilde{P}_{2t}^{\sigma'}\}_{t\geq 1}$ and $\{\widehat{P}_{2t}^\sigma\}_{t\geq 1}$ be the corresponding finite dimensional distributions. The following lemma gives an upper bound on the prediction errors in player 2's first t periods signals if he expects the game to evolve as P_{2t}^σ while the normal type of player 1 deviates to $\hat{\alpha}_1$.

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

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

Proof. We show this by a simple calculation:

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

$$= \sum_{h_2^t \in H_{2t}} \widetilde{P}_{2t}^{\sigma'}\left(h_2^t\right) \log \frac{\widehat{P}_{2t}^{\sigma}(h_2^t)}{P_{2t}^{\sigma}(h_2^t)} + \sum_{h_2^t \in H_{2t}} \widetilde{P}_{2t}^{\sigma'}\left(h_2^t\right) \log\left(\prod_{\tau=0}^{t-1} \frac{\rho(y^\tau|\xi_0)}{\rho(y^\tau|\xi(\hat{\alpha}_1))}\right).$$

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 second term is simply $t\lambda_{\xi(\hat{\alpha}_1)}$. Moreover, since $\widehat{P}_{2t}^{\sigma}$ is obtained by conditioning P_{2t}^{σ} on the event that player 1 is the commitment type $\xi(\hat{\alpha}_1)$, we have

$$\frac{\widehat{P}_{2t}^{\sigma}(h_2^t)}{P_{2t}^{\sigma}(h_2^t)} \le \frac{1}{\mu(\xi(\hat{\alpha}_1))} \quad \forall h_2^t \in H_{2t}.$$

Therefore the first term is no greater than $-\log \mu(\xi(\hat{\alpha}_1))$. These two observations imply the desired result. □

For any private history $h_2^t \in \bigcup_{t\ge 0} H_{2t}$, $P_{2,t+1}^{\sigma}$ (resp., $\widetilde{P}_{2,t+1}^{\sigma'}$) 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., $\widetilde{p}_{2t}^{\sigma'}(\cdot\,|h_2^t)$).[7] 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 $\widetilde{p}_{2t}^{\sigma'}(\cdot\,|h_2^t)$.

For any $t \ge 1$, let $\widetilde{E}_{2t}^{\sigma'}[\cdot]$ denote the expectation over H_{2t} with respect to the probability measure $\widetilde{P}_{2t}^{\sigma'}$. The following lemma bounds player 2's expected one step ahead prediction error.

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

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

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

Proof. Let $q_{2,t+1}(\cdot\,|h_2^t, z_2^t)$ (resp., $\widetilde{q}_{2,t+1}(\cdot\,|h_2^t, z_2^t)$) 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., $\widetilde{P}_{2,t+1}^{\sigma'}$). Because Assumption 1 and Lemma 1 imply $d(\widetilde{P}_{2t}^{\sigma'}\|P_{2t}^{\sigma}) < \infty$ for all $t \ge 1$, applying the chain rule of relative entropy twice yields[8]

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

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

7 If 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.

8 For a formal statement of the chain rule, see Cover and Thomas [1] Section 2.5 and also Gossner [4].

Let $d^{\delta,\sigma}_{\xi(\hat{\alpha}_1)}$ 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 $\xi(\hat{\alpha}_1)$

$$d^{\delta,\sigma}_{\xi(\hat{\alpha}_1)} \equiv \widetilde{E}^{\sigma'}\left[(1-\delta)\sum_{t=0}^{\infty}\delta^t d\big(\widetilde{p}^{\sigma'}_{2t}\big(\cdot\,|h^t_2\big)\,\|\,p^{\sigma}_{2t}\big(\cdot\,|h^t_2\big)\big)\right]$$
$$= (1-\delta)\sum_{t=0}^{\infty}\delta^t \widetilde{E}^{\sigma'}_{2t}\big[d\big(\widetilde{p}^{\sigma'}_{2t}\big(\cdot\,|h^t_2\big)\,\|\,p^{\sigma}_{2t}\big(\cdot\,|h^t_2\big)\big)\big].$$

Combining Lemmas 1 and 2, a similar calculation as in Lemma 5 of Gossner [4] will imply

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

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

The final step is to show player 1 can guarantee himself a payoff at least as high as

$$V_{\xi(\hat{\alpha}_1)}\big(-(1-\delta)\log\mu\big(\xi(\hat{\alpha}_1)\big) + \lambda_{\xi(\hat{\alpha}_1)}\big)$$

by mimicking commitment type $\hat{\alpha}_1$. This directly follows from Lemma 6 in Gossner [4].

References

[1] T. Cover, J. Thomas, Elements of Information Theory, 2nd edition, Wiley–Interscience, 2006.
[2] M. Ekmekci, O. Gossner, A. Wilson, Impermanent types and permanent reputations, J. Econ. Theory 147 (1) (2012) 162–178.
[3] D. Fudenberg, D.K. Levine, Maintaining a reputation when strategies are imperfectly observed, Rev. Econ. Stud. 59 (3) (1992) 561–579.
[4] O. Gossner, Simple bounds on the value of a reputation, Econometrica 79 (5) (2011) 1627–1641.
[5] G.J. Mailath, L. Samuelson, Repeated Games and Reputations: Long-Run Relationships, Oxford University Press, USA, 2006.
[6] T. Wiseman, Reputation and exogenous private learning, J. Econ. Theory 144 (3) (2009) 1352–1357.