Reputations in Repeated Games*

George J. Mailath
Department of Economics
University of Pennsylvania
Philadelphia, PA 19104
gmailath@econ.upenn.edu

Larry Samuelson
Department of Economics
Yale University
New Haven, CT 06520
Larry.Samuelson@yale.edu

August 11, 2013

Abstract

This paper, prepared for the *Handbook of Game Theory*, volume 4 (Peyton Young and Shmuel Zamir, editors, Elsevier Press), surveys work on reputations in repeated games of incomplete information.

*We have benefited from our collaboration and discussions with many coauthors and colleagues over the course of many years, and we are deeply grateful to them. We thank Martin Cripps, Olivier Gossner, Yuichi Yamamoto, and especially Ju Hu and Zehao Hu for helpful comments. We thank the National Science Foundation (SES-0961540 and SES-1153893) for financial support. We thank the Stockholm, Toulouse, and UNSW Schools of Economics and the Universities of Cambridge, Chicago, and York for the opportunity to teach some of this material.

Contents

1 Introduction

1.1 Reputations

The word "reputation" appears throughout discussions of everyday interactions. Firms are said to have reputations for providing good service, professionals for working hard, people for being honest, newspapers for being unbiased, governments for being free from corruption, and so on. Reputations establish links between past behavior and expectations of future behavior—one expects good service because good service has been provided in the past, or expects fair treatment because one has been treated fairly in the past. These reputation affects are so familiar as to be taken for granted. One is instinctively skeptical of a watch offered for sale by a stranger on a subway platform, but more confident of a special deal on a watch from an established jeweler. Firms proudly advertise that they are fixtures in their communities, while few customers would be attracted by a slogan of "here today, gone tomorrow."

Repeated games allow for a clean description of both the myopic incentives that agents have to behave opportunistically and, via appropriate specifications of future behavior (and so rewards and punishments), the incentives that deter opportunistic behavior. As a consequence, strategic interactions within long-run relationships have often been studied using repeated games. For the same reason, the study of reputations has been particularly fruitful in the context of repeated games, the topic of this chapter. We do not provide a comprehensive guide to the literature, since a complete list of the relevant repeated-games papers, at the hurried rate of one paragraph per paper, would leave us no room to discuss the substantive issues. Instead, we identify the key points of entry into the literature, confident that those who are interested will easily find their way past these.[1]

1.2 The Interpretive Approach to Reputations

There are two approaches to reputations in the repeated-games literature. In the first, an equilibrium of the repeated game is selected whose actions along the equilibrium path are not Nash equilibria of the stage game. Incentives to choose these actions are created by attaching less favorable continuation paths to deviations. For perhaps the most familiar example, there is an equilibrium of the repeated prisoners' dilemma (if the players are sufficiently

[1] It should come as no surprise that we recommend Mailath and Samuelson (2006) for further reading on most topics in this chapter.

patient) in which the players cooperate in every period, with any deviation from such behavior prompting relentless mutual defection.

The players who choose the equilibrium actions in such a case are often interpreted as maintaining a reputation for doing so, with a punishment-triggering deviation interpreted as the loss of one's reputation. For example, players in the repeated prisoners' dilemma are interpreted as maintaining a reputation for being cooperative, while the first instance of defection destroys that reputation.

In this approach, the link between past behavior and expectations of future behavior is an equilibrium phenomenon, holding in some equilibria but not in others. The notion of reputation is used to interpret an equilibrium strategy profile, but otherwise involves no modification of the basic repeated game and adds nothing to the formal analysis.

1.3 The Adverse Selection Approach to Reputations

The adverse selection approach to reputations considers games of incomplete information. The motivation typically stems from a game of complete information in which the players are "normal," and the game of incomplete information is viewed as a perturbation of the complete information game. In keeping with this motivation, attention is typically focused on games of "nearly" complete information, in the sense that a player whose type is unknown is very likely (but not quite certain) to be a normal type. For example, a player in a repeated game might be almost certain to have stage-game payoffs given by the prisoners' dilemma, but may with some small possibility have no other option than to play tit-for-tat.[2] Again, consistent with the perturbation motivation, it is desirable that the set of alternative types be not unduly constrained.

The idea that a player has an incentive to build, maintain, or milk his reputation is captured by the incentive that player has to manipulate the beliefs of other players about his type. The updating of these beliefs establishes links between past behavior and expectations of future behavior. We say "reputations effects" arise if these links give rise to restrictions on equilibrium payoffs or behavior that do not arise in the underlying game of complete information.

We concentrate throughout on the adverse selection approach to reputations. The basic results identify circumstances in which reputation effects

[2] This was the case in one of the seminal reputation papers, Kreps, Milgrom, Roberts, and Wilson (1982).

necessarily arise, imposing bounds on equilibrium payoffs that are in many cases quite striking.

2 Reputations with Short-Lived Players

2.1 An Example

We begin with the example of the "product-choice" game shown in Figure 1. Think of the long-lived player 1 ("he") as a firm choosing to provide either high (H) or low (L) effort. Player 2 ("she") represents a succession of customers, with a new customer in each period, choosing between a customized (c) or standardized (s) product. The payoffs reveal that high effort is costly for player 1, since L is a strictly dominant strategy in the stage game. Player 1 would like player 2 to choose the customized product c, but 2 is willing to do so only if she is sufficiently confident that 1 will choose H.

The stage game has a unique Nash equilibrium in which the firm provides low effort and the customer buys the standardized product. In the discrete-time infinite horizon game, the firm maximizes the average discounted sum of his payoffs. In the infinite horizon game of complete information, every payoff in the interval $[1, 2]$ is a subgame perfect equilibrium payoff if the firm is sufficiently patient.

Equilibria in which the firm's payoff is close to 2 have an intuitive feel to them. In these equilibria, the firm frequently exerts high effort H, so that the customer will play her best response of purchasing the customized product c. Indeed, the firm should be able to develop a "reputation" for playing H by persistently doing so. This may be initially costly for the firm, because customers may not be immediately convinced that the firm will play H and hence customers may play s for some time, but the subsequent payoff could make this investment worthwhile for a sufficiently patient firm.

Nothing in the structure of the repeated game captures this intuition. Repeated games have a recursive structure: the continuation game following any history is identical to the original game. No matter how many times

	c	s
H	$2, 3$	$0, 2$
L	$3, 0$	$1, 1$

Figure 1: The product-choice game.

the firm has previously played H, the standard theory provides no reason for customers to believe that the firm is more likely to play H now than at the beginning of the game.

Now suppose that in addition to the *normal* type of firm, i.e., the type whose payoffs are given by Figure 1, there is a *commitment*, or *behavioral*, type $\xi(H)$. This type invariably and necessarily plays H (and hence the description "commitment" or "behavioral" type). We also refer to this type as the *Stackelberg* type, since H is the action the firm would take in the unique subgame perfect equilibrium of a sequential-move game of perfect information in which the firm chooses publicly before the customer. Even if the prior probability of this behavioral type is very small, it has a dramatic effect on the set of equilibrium payoffs when the firm is sufficiently patient. It is an implication of Proposition 1 that there is then no Nash equilibrium of the incomplete information game with a payoff to the normal firm near 1.

The flavor of the analysis is conveyed by asking what must be true in an equilibrium in which the normal type of player 1 receives a payoff close to 1, when he has discount factor close to 1. For this to occur, there must be many periods in which the customer chooses the standardized product, which in turn requires that in those periods the customer expects L with high probability. But if player 1 plays H in every period, then the customer is repeatedly surprised. And, since the H-commitment type has positive prior probability, this is impossible: each time the customer is surprised the posterior probability on the H-commitment type jumps a discrete amount (leading eventually to a posterior above the level at which customers choose the customized product).

Reputation effects do much more than eliminate unintuitive outcomes; they also rescue intuitive outcomes as equilibrium outcomes. For example, the *finitely* repeated complete information product-choice game has a unique subgame perfect equilibrium, and in this equilibrium, the static Nash profile Ls is played in every period. Nonetheless, our intuition again suggests that if the game has a sufficiently long horizon, the firm should here as well be able to develop a "reputation" for playing H by persistently doing so. It is an immediate implication of the logic underlying Proposition 1 that the firm can indeed develop such a reputation in the game with incomplete information.[3]

[3]Kreps, Milgrom, Roberts, and Wilson (1982) first made this point in the context of the finitely repeated prisoners' dilemma with the commitment type playing tit-for-tat. Kreps and Wilson (1982) and Milgrom and Roberts (1982) further explore this insight in the finite horizon chain store game.

Similarly, if customers can only imperfectly monitor the firm's effort choice, the upper bound on the firm's complete information equilibrium payoff will be less than 2. Moreover, for some monitoring distributions, the complete information game has, for *all* discount factors, a unique sequential equilibrium, and in this equilibrium, the static Nash profile Ls is played in every period.[4] Proposition 1 nonetheless implies that the firm can develop a reputation for playing H in the game with incomplete information.

2.2 The Benchmark Complete Information Game

Player i has a finite action space A_i. Player 1's actions are monitored via a public signal: the signal y, drawn from a finite set Y, is realized with probability $\rho(y|a_1)$ when the action $a_1 \in A_1$ is chosen. We make the analysis more convenient here by assuming that the public signal depends only on player 1's action, returning to this assumption in Section 2.5. Since player 2 is short-lived, we further simplify notation by assuming that the period t player 2's action choice is not observed by subsequent player 2's (though it is observed by player 1). The arguments are identical (with more notation) if player 2's actions are public.

Player i's ex post stage game payoff is a function $u_i^* : A_1 \times A_2 \times Y \to \mathbb{R}$ and i's ex ante payoff is $u_i : A_1 \times A_2 \to \mathbb{R}$ is given by

$$u_i(a) := \sum\nolimits_y u_i^*(a_1, a_2, y)\rho(y|a_1).$$

We typically begin the analysis with the ex ante payoffs, as in the product-choice game of Figure 1, and the monitoring structure ρ, leaving the ex post payoff functions u_i^* to be defined implicitly.

Player 2 observes the public signals, but does not observe player 1's actions. Player 2 might draw inferences about player 1's actions from her payoffs, but since player 2 is short-lived, these inferences are irrelevant (as long as the period t player 2 does not communicate any such inference to subsequent short-lived players). Following the literature on public monitoring repeated games, such inferences can also be precluded by assuming player 2's ex post payoff does not depend on a_1, depending only on (a_2, y).

The benchmark game includes perfect monitoring games as a special case. In the perfect monitoring product-choice game for example, $A_1 = \{H, L\}$, $A_2 = \{c, s\}$, $Y = A_1$, and $\rho(y \mid a_1) = 1$ if $y = a_1$ and 0 otherwise. An imperfect public monitoring version of the product-choice game is analyzed in Section 2.4.3 (Example 2).

[4] We return to this in Example 2.

Player 1 is long lived (and discounts flow payoffs by a discount factor δ), while player 2 is short lived (living for one period). The set of private histories for player 1 is $H_1 := \cup_{t=0}^{\infty}(A_1 \times A_2 \times Y)^t$, and a behavior strategy for player 1 is a function $\sigma_1 : H_1 \rightarrow \Delta(A_1)$. The set of histories for the short-lived players is $H_2 := \cup_{t=0}^{\infty} Y^t$, and a behavior strategy for the short-lived players is a function $\sigma_2 : H_2 \rightarrow \Delta(A_2)$. Note that the period t short-lived player does not know the action choices of past short-lived players.[5]

2.3 The Incomplete Information Game and Commitment Types

The type of player 1 is unknown to player 2. A possible type of player 1 is denoted by $\xi \in \Xi$, where Ξ is a finite or countable set of types. Player 2's prior belief about 1's type is given by the distribution μ, with support Ξ.

The set of types is partitioned into *payoff types*, Ξ_1, and *commitment types*, $\Xi_2 := \Xi \backslash \Xi_1$. Payoff types maximize the average discounted value of payoffs, which depend on their type. We accordingly expand the definition of the ex post payoff function u_1^* to incorporate types, $u_1^* : A_1 \times A_2 \times Y \times \Xi_1 \rightarrow \mathbb{R}$. The ex ante payoff function $u_1 : A_1 \times A_2 \times \Xi_1 \rightarrow \mathbb{R}$ is now given by

$$u_1(a_1, a_2, \xi) := \sum_y u_1^*(a_1, a_2, y, \xi)\rho(y|a_1).$$

It is common to identify one payoff type as the *normal* type, denoted here by ξ_0. When doing so, it is also common to drop the type argument from the payoff function. It is also common to think of the prior probability of the normal type $\mu(\xi_0)$ as being relatively large, so the games of incomplete information are a seemingly small departure from the underlying game of complete information, though there is no requirement that this be the case.

Commitment types do not have payoffs, and simply play a specified repeated game strategy. While a commitment type of player 1 can be committed to any strategy in the repeated game, much of the literature focuses on *simple commitment* or *action types*: such types play the same (pure or mixed) stage-game action in every period, regardless of history.[6] For example, one simple commitment type in the product-choice game always

[5] Under our assumption that the public signal's distribution is a function of only the long-lived players's action, the analysis to be presented proceeds unchanged if the period t short-lived player knows the actions choices of past short-lived players (i.e., the short-lived player actions are public). In more general settings (where the signal distribution depends on the complete action profile), the analysis must be adjusted in obvious ways (as we describe below).

[6] This focus is due to the observation that with short-lived uninformed players, more complicated commitment types do not lead to higher reputation bounds.

exerts high effort, while another always plays high effort with probability $\frac{2}{3}$. We denote by $\xi(\alpha_1)$ the (simple commitment) type that plays the action $\alpha_1 \in \Delta(A_1)$ in every period.

Remark 1 (Payoff or commitment types) The distinction between payoff and commitment types is not clear cut. For example, pure simple commitment types are easily modeled as payoff types. We need only represent the type $\xi(a_1)$ as receiving the stage-game payoff 1 if he plays action a_1 (regardless of what signal appears or what player 2 chooses) and zero otherwise. Note that this makes the consistent play of a_1 a strictly dominant strategy in the repeated game, and that it is not enough to simply have the action be dominant in the stage game.

♦

A behavior strategy for player 1 in the incomplete information game is given by

$$\sigma_1 : H_1 \times \Xi \to \Delta(A_1),$$

such that, for all simple commitment types $\xi(\alpha_1) \in \Xi_2$,[7]

$$\sigma_1(h_1^t, \xi(\alpha_1)) = \alpha_1, \qquad \forall h_1^t \in H_1.$$

A behavior strategy for player 2 is (as in the complete information game) a map $\sigma_2 : H_2 \to \Delta(A_2)$.

The space of outcomes is given by $\Omega := \Xi \times (A_1 \times A_2 \times Y)^\infty$, with an outcome $\omega = (\xi, a_1^0 a_2^0 y^0, a_1^1 a_2^1 y^1, a_1^2 a_2^2 y^2, \ldots) \in \Omega$, specifying the type of player 1, the actions chosen and the realized signal in each period.

A profile of strategies (σ_1, σ_2), along with the prior probability over types μ (with support Ξ), induces a probability measure on the set of outcomes Ω, denoted by $\mathbf{P} \in \Delta(\Omega)$. For a fixed commitment type $\hat{\xi} = \xi(\hat{\alpha}_1)$, the probability measure on the set of outcomes Ω conditioning on $\hat{\xi}$ (and so induced by $(\hat{\sigma}_1, \sigma_2)$, where $\hat{\sigma}_1$ is the simple strategy specifying $\hat{\alpha}_1$ in every period irrespective of history), is denoted $\widehat{\mathbf{P}} \in \Delta(\Omega)$. Denoting by $\widetilde{\mathbf{P}}$ the measure induced by (σ_1, σ_2) and conditioning on $\xi \neq \hat{\xi}$, we have

$$\mathbf{P} = \mu(\hat{\xi})\widehat{\mathbf{P}} + (1 - \mu(\hat{\xi}))\widetilde{\mathbf{P}}. \tag{1}$$

Given a strategy profile σ, $U_1(\sigma, \xi)$ denotes the type-ξ long-lived player's payoff in the repeated game,

$$U_1(\sigma, \xi) := E_{\mathbf{P}}\left[(1-\delta) \sum_{t=0}^{\infty} \delta^t u_1(a^t, y^t, \xi) \,\middle|\, \xi \right].$$

[7]For convenience, we have omitted the analogous requirement that σ_1 is similarly restricted for non-simple commitment types.

Denote by $\Gamma(\mu, \delta)$ the game of incomplete information.

As usual, a Nash equilibrium is a collection of mutual best responses:

Definition 1 *A strategy profile* (σ_1', σ_2') *is a* Nash equilibrium *of the game* $\Gamma(\mu, \delta)$ *if, for all* $\xi \in \Xi_1$, σ_1' *maximizes* $U_1(\sigma_1, \sigma_2', \xi)$ *over player* 1*'s repeated game strategies, and if for all* t *and all* $h_2^t \in \mathcal{H}_2$ *that have positive probability under* (σ_1', σ_2') *and* μ *(i.e.,* $\mathbf{P}(h_2^t) > 0$*),*

$$E_{\mathbf{P}}\left[u_2(\sigma_1'(h_1^t, \xi), \sigma_2'(h_2^t)) \mid h_2^t\right] = \max_{a_2 \in A_2} E_{\mathbf{P}}\left[u_2(\sigma_1'(h_1^t, \xi), a_2) \mid h_2^t\right].$$

2.4 Reputation Bounds

The link from simple commitment types to reputation effects arises from a basic property of updating: Suppose an action α_1' is statistically identified (i.e., there is no α_1'' giving rise to the same distribution of signals) under the signal distribution ρ, and suppose player 2 assigns positive probability to the simple type $\xi(\alpha_1')$. Then, if the normal player 1 persistently plays α_1', player 2 must eventually place high probability on that action being played (and so will best respond to that action). Intuitively, since α_1' is statistically identified, in any period in which player 2 places low probability on α_1' (and so on $\xi(\alpha_1')$), the signals will typically lead player 2 to increase the posterior probability on $\xi(\alpha_1')$, and so eventually on α_1'. Consequently, there cannot be too many periods in which player 2 places low probability on α_1'.

When the action chosen is perfectly monitored by player 2 (which requires the benchmark game have perfect monitoring and α_1' be a pure action), this intuition has a direct formalization (the route followed in the original argument of Fudenberg and Levine, 1989). However, when the action is not perfectly monitored, the path from the intuition to the proof is less clear. The original argument for this case (Fudenberg and Levine, 1992) uses sophisticated martingale techniques, subsequently simplified by Sorin (1999) (an exposition can be found in Mailath and Samuelson, 2006, §15.4.2). Here we present a recent simple unified argument, due to Gossner (2011b), based on relative entropy (see Cover and Thomas, 2006, for an introduction).

2.4.1 Relative Entropy

Let X be a finite set of outcomes. The *relative entropy* or *Kullback-Leibler distance* between probability distributions p and q over X is

$$d(p\|q) := \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}.$$

By convention, $0 \log \frac{0}{q} = 0$ for all $q \in [0, 1]$ and $p \log \frac{p}{0} = \infty$ for all $p \in (0, 1]$.[8] In our applications of relative entropy, the support of q will always contain the support of p. Since relative entropy is not symmetric, we often distinguish the roles of p and q by saying that $d(p\|q)$ is the relative entropy of q with respect to p. Relative entropy is always nonnegative and only equals 0 when $p = q$ (Cover and Thomas, 2006, Theorem 2.6.3).[9]

The relative entropy of q with respect to p measures an observer's expected error in predicting $x \in X$ using the distribution q when the true distribution is p. The probability of a sample of n draws from X identically and independently distributed according to p is $\prod_x p(x)^{n_x}$, where n_x is the number of realizations of $x \in X$ in the sample. An observer who believes the data is distributed according to q assigns to the same sample probability $\prod_x q(x)^{n_x}$. The log likelihood ratio of the sample is

$$\mathcal{L}(x_1, \ldots, x_n) = \sum\nolimits_x n_x \log \frac{p(x)}{q(x)}.$$

As the sample size n grows large, the average log likelihood $\mathcal{L}(x_1, \ldots, x_n)/n$ converges almost surely to $d(p\|q)$ (and so the log likelihood becomes arbitrarily large for any $q \neq p$, since $d(p\|q) > 0$ for any such q).

While not a metric (since it is asymmetric and does not satisfy the triangle inequality), relative entropy is usefully viewed as a notion of distance. For example, Pinsker's inequality[10] bounds the relative entropy of q with respect to p from below by a function of their L^1 distance:

$$\|p - q\| \leq \sqrt{2d(p\|q)}, \tag{2}$$

[8] These equalities, for $p, q > 0$, are justified by continuity arguments. The remaining case, $0 \log \frac{0}{0} = 0$, is made to simplify statements and eliminate nuisance cases.

The logs may have any base. Both base 2 and base e are used in information theory. We use base e.

[9] Apply Jensen's inequality to $-d(p\|q) = \sum_{x \in \operatorname{supp} p} p(x) \log[q(x)/p(x)]$.

[10] See Cesa-Bianchi and Lugosi (2006, page 371) or Cover and Thomas (2006, Lemma 11.6.1); note that Cover and Thomas (2006) define relative entropy using log base 2.

where

$$\|p-q\| := \sum\nolimits_x |p(x)-q(x)| = 2\sum\nolimits_{\{x:p(x)\geq q(x)\}} |p(x)-q(x)|\,.$$

Thus, $\|p_n - q\| \to 0$ if $d(p_n\|q) \to 0$ as $n \to \infty$. While the reverse implication does not hold in general,[11] it does for full support q.

The usefulness of relative entropy arises from a chain rule. Let P and Q be two distributions over a finite product set $X \times Y$, with marginals P_X and Q_X on X, and conditional probabilities $P_Y(\cdot|x)$ and $Q_Y(\cdot|x)$ on Y given $x \in X$. The *chain rule for relative entropy* is

$$\begin{aligned} d(P\|Q) &= d(P_X\|Q_X) + \sum\nolimits_x P_X(x) d\left(P_Y(\cdot|x)\|Q_Y(\cdot|x)\right) \\ &= d(P_X\|Q_X) + E_{P_X} d\left(P_Y(\cdot|x)\|Q_Y(\cdot|x)\right). \end{aligned} \tag{3}$$

The chain rule is a straightforward calculation (Cover and Thomas, 2006, Theorem 2.5.3). The error in predicting the pair xy can be decomposed into the error predicting x, and conditional on x, the error in predicting y.

The key to the reputation bound is bounding the error in the one-step ahead predictions of the uninformed players when the long-lived player plays identically to a commitment type. The presence of the commitment type ensures that there is a "grain of truth" in player 2's beliefs which, together with the chain rule, yields a useful bound on relative entropy. The basic technical tool is the following lemma.

Lemma 1 *Let X be a finite set of outcomes. Suppose $q = \varepsilon p + (1-\varepsilon)p'$ for some $\varepsilon > 0$ and $p, p' \in \Delta(X)$. Then,*

$$d(p\|q) \leq -\log\varepsilon.$$

Proof. Since $q(x)/p(x) \geq \varepsilon$, we have

$$-d(p\|q) = \sum\nolimits_x p(x)\log\frac{q(x)}{p(x)} \geq \sum\nolimits_x p(x)\log\varepsilon = \log\varepsilon.$$

■

[11] Suppose $X = \{0,1\}$, $p_n(1) = 1 - \frac{1}{n}$, and $q(1) = 1$. Then, $\|p_n - q\| \to 0$, while $d(p_n\|q) = \infty$ for all n.

2.4.2 Bounding the One-Step Ahead Prediction Errors

Fix a (possibly mixed) action $\hat{\alpha}_1 \in \Delta(A_1)$ and suppose the commitment type $\hat{\xi} = \xi(\hat{\alpha}_1)$ has positive prior probability.

At the history h_2^t, player 2 chooses an action $\sigma_2(h_2^t)$ that is a best response to $\alpha_1(h_2^t) := E_{\mathbf{P}}[\sigma_1(h_1^t, \xi) \mid h_2^t]$, that is, a_2 has positive probability under $\sigma_2(h_2^t)$ only if it maximizes[12]

$$\textstyle\sum_{a_1} \sum_y u_2^*(a_1, a_2, y)\rho(y|a_1)\alpha_1(a_1|h_2^t).$$

At the history h_2^t, player 2's predicted distribution of the period t signals is $p(h_2^t) := \rho(\cdot|\alpha_1(h_2^t)) = \sum_{a_1} \rho(\cdot|a_1)\alpha_1(a_1|h_2^t)$, while the true distribution when player 1 plays $\hat{\alpha}_1$ is $\hat{p} := \rho(\cdot|\hat{\alpha}_1) = \sum_{a_1} \rho(\cdot|a_1)\hat{\alpha}_1(a_1)$. Hence, if player 1 is playing $\hat{\alpha}_1$, then in general player 2 is not best responding to the true distribution of signals, and his *one-step ahead prediction error* is $d\left(\hat{p} \| p(h_2^t)\right)$. However, player 2 *is* best responding to an action profile $\alpha_1(h_2^t)$ that is $d\left(\hat{p} \| p(h_2^t)\right)$-close to $\hat{\alpha}_1$ (as measured by the relative entropy of the induced signals). To bound player 1's payoff, it suffices to bound the number of periods in which $d\left(\hat{p} \| p(h_2^t)\right)$ is large.

Since player 2's beliefs assign positive probability to the event that player 1 is always playing $\hat{\alpha}_1$, then when player 1 does so, player 2's one-step prediction error must disappear asymptotically. A bound on player 1's payoff then arises from noting that *if* player 1 relentlessly plays $\hat{\alpha}_1$, then player 2 must eventually just as persistently play a best response to $\hat{\alpha}_1$. "Eventually" may be a long time, but this delay is inconsequential to a sufficiently patient player 1.

For any period t, denote the marginal of the unconditional distribution $\mathbf{P}$ on H_2^t, the space of t period histories of public signals, by $\mathbf{P}_2^t$. Similarly, the marginal on H_2^t of $\widehat{\mathbf{P}}$ (the distribution conditional on $\hat{\xi}$) is denoted $\widehat{\mathbf{P}}_2^t$. Recalling (1) and applying Lemma 1 to these marginal distributions (which have finite supports) yields

$$d(\widehat{\mathbf{P}}_2^t \| \mathbf{P}_2^t) \leq -\log \mu(\hat{\xi}). \tag{4}$$

It is worth emphasizing that this inequality holds for all t, and across all equilibria. But notice also that the bounding term is unbounded as $\mu(\hat{\xi}) \to 0$ and our interest is in the case where $\mu(\hat{\xi})$ is small.

Suppose $\mu(\hat{\xi})$ is indeed small. Then, in those periods when player 2's prediction under $\mathbf{P}$, $p(h_2^t)$, has a large relative entropy with respect to $\hat{p}$, she

[12] The probability assigned to $a_1 \in A_1$ by the distribution $\alpha_1(h_2^t)$ is denoted by $\alpha_1(a_1|h_2^t)$.

will (with high probability) be surprised,[13] and so will significantly increase her posterior probability that player 1 is $\hat{\xi}$. This effectively increases the size of the grain of truth from the perspective of period $t+1$, reducing the maximum relative entropy $p(h_2^{t+1})$ can have with respect to $\hat{p}$. Intuitively, for fixed $\mu(\hat{\xi})$, there cannot be too many periods in which player 2 can be surprised with high probability.

To make this intuition precise, we consider the one-step ahead prediction errors, $d(\hat{p}\|p(h_2^\tau))$. The chain rule implies

$$d(\widehat{\mathbf{P}}_2^t\|\mathbf{P}_2^t) = \sum_{\tau=0}^{t-1} E_{\widehat{\mathbf{P}}} d(\hat{p}\|p(h_2^\tau)),$$

that is, the prediction error over t periods is the total of the t *expected* one-step ahead prediction errors. Since (4) holds for all t,

$$\sum_{\tau=0}^{\infty} E_{\widehat{\mathbf{P}}} d\left(\hat{p}\|p(h_2^\tau)\right) \leq -\log \mu(\hat{\xi}). \tag{5}$$

That is, all but a finite number of expected one-step ahead prediction errors must be small.

2.4.3 From Prediction Bounds to Payoffs

It remains to connect the bound on prediction errors (5) with a bound on player 1's payoffs.

Definition 2 *An 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*

1. *α_2 is a best response to α_1'; and*
2. *$d(\rho(\cdot|\alpha_1)\|\rho(\cdot|\alpha_1')) \leq \varepsilon$.*

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

Recall that in a Nash equilibrium, at any on-the-equilibrium-path history h_2^t, player 2's action is a $d(\hat{p}\|p(h_2^t))$-entropy confirming best response to $\hat{\alpha}_1$. Suppose player 1 always plays $\hat{\alpha}_1$. We have just seen that the expected number of periods in which $d(\hat{p}\|p(h_2^t))$ is large is bounded, independently of δ. Then for δ close to 1, player 1's equilibrium payoffs will be effectively determined by player 2's ε-entropy confirming best responses for ε small.

[13] Or, more precisely, the period $(t+1)$ version of player 2 will be surprised.

Define, for all payoff types $\xi \in \Xi_1$,

$$\underline{v}^{\xi}_{\alpha_1}(\varepsilon) := \min_{\alpha_2 \in B^d_{\varepsilon}(\alpha_1)} u_1(\alpha_1, \alpha_2, \xi),$$

and denote by $\underline{w}^{\xi}_{\alpha_1}$ the largest convex function below $\underline{v}^{\xi}_{\alpha_1}$. The function $\underline{w}^{\xi}_{\alpha_1}$ is nonincreasing in ε because $\underline{v}^{\xi}_{\alpha_1}$ is. The function $\underline{w}^{\xi}_{\alpha_1}$ allows us to translate a bound on the total discounted expected one-step ahead prediction errors into a bound on the total discounted expected payoffs of player 1.

Proposition 1 *Suppose the action type $\hat{\xi} = \xi(\hat{\alpha}_1)$ has positive prior probability, $\mu(\hat{\xi}) > 0$, for some potentially mixed action $\hat{\alpha}_1 \in \Delta(A_1)$. Then, player 1 type ξ's payoff in any Nash equilibrium of the game $\Gamma(\mu, \delta)$ is greater than or equal to $\underline{w}^{\xi}_{\hat{\alpha}_1}(\hat{\varepsilon})$, where $\hat{\varepsilon} := -(1-\delta)\log \mu(\hat{\xi})$.*

It is worth emphasizing that the only aspect of the set of types and the prior that plays a role in the proposition is the probability assigned to $\hat{\xi}$. The set of types may be very large, and other quite crazy types may receive significant probability under the prior μ.

Proof. Since in any Nash equilibrium (σ'_1, σ'_2), each payoff type ξ has the option of playing $\hat{\alpha}_1$ in every period, we have

$$\begin{aligned} U_1(\sigma', \xi) &= (1-\delta)\sum\nolimits_{t=0}^{\infty} \delta^t E_{\mathbf{P}}[u_1(\sigma'_1(h^t_1), \sigma'_2(h^t_2), \xi) \mid \xi] \\ &\geq (1-\delta)\sum\nolimits_{t=0}^{\infty} \delta^t E_{\widehat{\mathbf{P}}} u_1(\hat{\alpha}_1, \sigma'_2(h^t_2), \xi) \\ &\geq (1-\delta)\sum\nolimits_{t=0}^{\infty} \delta^t E_{\widehat{\mathbf{P}}} \underline{v}^{\xi}_{\hat{\alpha}_1}(d(\hat{p}\|p(h^t_2))) \\ &\geq (1-\delta)\sum\nolimits_{t=0}^{\infty} \delta^t E_{\widehat{\mathbf{P}}} \underline{w}^{\xi}_{\hat{\alpha}_1}(d(\hat{p}\|p(h^t_2))) \end{aligned}$$

(and so, by an application of Jensen's inequality)

$$\geq \underline{w}^{\xi}_{\hat{\alpha}_1}\left((1-\delta)\sum\nolimits_{t=0}^{\infty} \delta^t E_{\widehat{\mathbf{P}}} d(\hat{p}\|p(h^t_2))\right)$$

(and so, by an application of (5))

$$\geq \underline{w}^{\xi}_{\hat{\alpha}_1}\left(-(1-\delta)\log \mu(\hat{\xi})\right).$$

■

Corollary 1 *Suppose the action type* $\hat{\xi} = \xi(\hat{\alpha}_1)$ *has positive prior probability,* $\mu(\hat{\xi}) > 0$*, for some potentially mixed action* $\hat{\alpha}_1 \in \Delta(A_1)$*. Then, for all* $\xi \in \Xi_1$ *and* $\eta > 0$*, there exists a* $\bar{\delta} < 1$ *such that, for all* $\delta \in (\bar{\delta}, 1)$*, player* 1 *type* ξ*'s payoff in any Nash equilibrium of the game* $\Gamma(\mu, \delta)$ *is greater than or equal to*

$$\underline{v}^{\xi}_{\hat{\alpha}_1}(0) - \eta.$$

Proof. Since the distribution of signals is independent of player 2's actions, if a mixture is an ε-entropy confirming best response to $\hat{\alpha}_1$, then so is every action in its support.[14] This implies that $B^d_0(\hat{\alpha}_1) = B^d_\varepsilon(\hat{\alpha}_1)$ for ε sufficiently small, and so $\underline{v}^{\xi}_{\hat{\alpha}_1}(0) = \underline{v}^{\xi}_{\hat{\alpha}_1}(\varepsilon)$ for ε sufficiently small. Hence,

$$\underline{v}^{\xi}_{\hat{\alpha}_1}(0) = \underline{w}^{\xi}_{\hat{\alpha}_1}(0) = \lim_{\varepsilon \searrow 0} \underline{w}^{\xi}_{\hat{\alpha}_1}(\varepsilon).$$

The Corollary is now immediate from Proposition 1. ■

Example 1 To illustrate Proposition 1 and its corollary, consider first the perfect monitoring product-choice game of Figure 1. Let ξ denote the payoff type with the player 1 payoffs specified. Recall that in the perfect monitoring game, the set of signals coincides with the set of firm actions. The action c is the unique best response of a customer to any action α_1 satisfying $\alpha_1(H) > \frac{1}{2}$, while s is also a best response when $\alpha_1(H) = \frac{1}{2}$. Thus, $B^d_\varepsilon(H) = \{c\}$ for all $\varepsilon < \log 2$ (since $d(H \| \alpha_1) = -\log \alpha_1(H)$), while $B^d_\varepsilon(H) = \{c, s\}$ for all $\varepsilon \geq \log 2 \approx 0.69$. That is, c is the unique ε-entropy confirming best response to H, for ε smaller than the relative entropy of the equal randomization on H and L with respect to the pure action H. This implies

$$\underline{v}^{\xi}_{H}(\varepsilon) = \begin{cases} 2, & \text{if } \varepsilon < \log 2, \\ 0, & \text{if } \varepsilon \geq \log 2. \end{cases}$$

The functions $\underline{v}^{\xi}_{H}$ and $\underline{w}^{\xi}_{H}$ are graphed in Figure 2.

Consider now the mixed action $\hat{\alpha}_1$ which plays the action H with probability $\frac{2}{3}$ (and the action L with probability $\frac{1}{3}$). The relative entropy of the equal randomization on H and L with respect to the mixed action $\hat{\alpha}_1$ is $\frac{5}{3}\log 2 - \log 3 =: \bar{\varepsilon} \approx 0.06$, and so any action with smaller relative entropy with respect to $\hat{\alpha}_1$ has c as the unique best response. This implies

$$\underline{v}^{\xi}_{\hat{\alpha}_1}(\varepsilon) = \begin{cases} 2\frac{1}{3}, & \text{if } \varepsilon < \bar{\varepsilon}, \\ \frac{1}{3}, & \text{if } \varepsilon \geq \bar{\varepsilon}. \end{cases}$$

[14] This is not true in general, see Mailath and Samuelson (2006, fn. 10, p. 480).

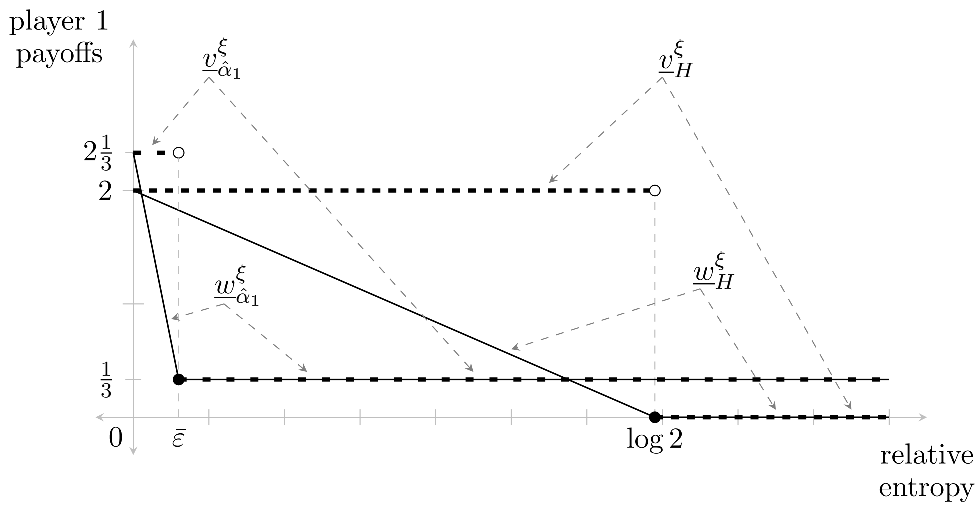


Figure 2: The functions $\underline{v}^{\xi}_{H}$, $\underline{w}^{\xi}_{H}$, $\underline{v}^{\xi}_{\hat{\alpha}_1}$, and $\underline{w}^{\xi}_{\hat{\alpha}_1}$ for the perfect monitoring version of the product-choice game in Figure 1 (the payoff type ξ has the specified player 1 payoffs). The relative entropy of $\frac{1}{2} \circ H + \frac{1}{2} \circ L$ with respect to H is $\log 2 \approx 0.69$. For higher relative entropies, s is an ε-entropy confirming best response to H. The relative entropy of $\frac{1}{2} \circ H + \frac{1}{2} \circ L$ with respect to $\hat{\alpha}_1$ is $\bar{\varepsilon} \approx 0.06$. For higher relative entropies, s is an ε-entropy confirming best response to $\hat{\alpha}_1$.

Suppose the customer puts positive probability on firm type $\xi_H := \xi(H)$. If $\mu(\xi_H)$ is large, then its log is close to zero, and we have a trivial and unsurprising reputation bound (since the customer will best respond to H from the beginning). Suppose, though, that the probability is close to zero. Then it may take many periods of imitating the action type ξ_H before the customer predicts H in every period (reflected in a large magnitude log). But since the number of periods is independent of the discount factor, for δ sufficiently close to 1, the lower bound is still close to 2: the term $(1-\delta)\log\mu(\xi)$ is close to zero.

This bound is independent of customer beliefs about the possible presence of the type $\hat{\xi} := \xi(\hat{\alpha}_1)$. A possibility of the type $\hat{\xi}$ can improve the lower bound on payoffs. However, since this type's behavior (signal distribution) is closer to the critical distribution $\frac{1}{2} \circ H + \frac{1}{2} \circ L$, the critical value of the relative entropy is significantly lower ($\bar{\varepsilon} << \log 2$; see Figure 2), and hence more periods must pass before the type ξ firm can be assured that the customer will play a best response. If the prior assigned equal probability to both $\hat{\xi}$ and ξ_H, then the bound on δ required to bound the type ξ firm's payoff by $\underline{v}^{\xi}_{\hat{\alpha}_1}(0) - \eta$ is significantly tighter than that required to bound the type ξ firm's payoff by $\underline{v}^{\xi}_{H}(0) - \eta$.

★

Example 2 One would expect reputation-based payoff bounds to be weaker in the presence of imperfect monitoring. While there is a sense in which this is true, there is another sense in which this is false.

To illustrate, we consider an imperfect monitoring version of the product-choice game. The actions of the firm are private, and the public signal is drawn from $Y := \{\underline{y}, \overline{y}\}$ according to the distribution $\rho(\overline{y} \mid a_1) = p$ if $a_1 = H$ and $q < p$ if $a_1 = L$. The ex ante payoffs for the type ξ firm and customer are again given by Figure 1.

While the best responses of the customer are unchanged from Example 1 , the ε-entropy confirming best responses are changed: The relative entropy of the mixture α_1 with respect to H is now

$$d(H\|\alpha_1) = p\log\frac{p}{\alpha_1(H)p + \alpha_1(L)q} + (1-p)\log\frac{1-p}{\alpha_1(H)(1-p) + \alpha_1(L)(1-q)}.$$

For the parameterization $p = 1 - q = \frac{2}{3}$, $d(H\|\frac{1}{2} \circ H + \frac{1}{2} \circ L) = \bar{\varepsilon} \approx 0.06$, the critical value for $\hat{\alpha}_1$ from Example 1, and so $\underline{v}^{\xi}_{H}(\varepsilon) = 2$ for $\varepsilon < \bar{\varepsilon}$, and 0 for $\varepsilon \geq \bar{\varepsilon}$. Moreover, $\underline{w}^{\xi}_{H}(0) = 2$.

Finally, the relative entropy of the critical mixture $\frac{1}{2} \circ H + \frac{1}{2} \circ L$ with respect to the mixture $\hat{\alpha}_1$ from Example 1 is approximately 0.006. Since the firm's actions are statistically identified by the signals, we also have $\underline{v}^{\xi}_{\hat{\alpha}_1}(0) = \underline{w}^{\xi}_{\hat{\alpha}_1}(0) = 2\frac{1}{3}$.

Recall that in the perfect monitoring product-choice game, the bound on the relative entropy to get a strictly positive lower bound on payoffs from $\underline{w}^{\xi}_{H}(\varepsilon)$ is $\log 2 \approx 0.69$ rather than $\bar{\varepsilon} \approx 0.06$. This implies that the required lower bound on the discount factor to get the same lower bound on payoffs is larger for the imperfect monitoring game, or equivalently, for the same discount factor, the reputation lower bound is lower under imperfect monitoring.

It is worth recalling at this point our earlier observation from Section 2.1 that reputation effects can rescue intuitive outcomes as equilibrium outcomes. Assume the actions of the customers are public; as we noted earlier, this does not affect the arguments or the reputation bounds calculated (beyond complicating notation). For the parameterization $p = 1 - q = \frac{2}{3}$, the complete information repeated game has a unique sequential equilibrium outcome, in which L is played in every period (Mailath and Samuelson, 2006, Section 7.6.2 and Proposition 10.1.1). The firm's equilibrium payoff in the complete information game is 1. In particular, while the reputation payoff bound is weaker in an absolute sense in the presence of imperfect monitoring, the relative bound (bound less the maximal payoff in any equilibrium of the complete information game) is stronger.

★

2.4.4 The Stackelberg Bound

The reputation literature has tended to focus on Stackelberg bounds. Player 1's *pure-action Stackelberg payoff* (for type ξ) is defined as

$$\bar{v}^{\xi}_1 := \sup_{a_1 \in A_1} \min_{\alpha_2 \in B(a_1)} u_1(a_1, \alpha_2, \xi),$$

where $B(a_1)$ is the set of player 2 myopic best replies to a_1. Since A_1 is finite, the supremum is attained by some action a^*_1 and any such action is an associated *Stackelberg action*,

$$a^{\xi}_1 \in \arg\max_{a_1 \in A_1} \min_{\alpha_2 \in B(a_1)} u_1(a_1, \alpha_2, \xi).$$

This is a pure action to which the type ξ player 1 would commit, if player 1 had the chance to do so (and hence the name "Stackelberg action"), given that such a commitment induces a best response from player 2.

The *mixed-action Stackelberg payoff* is defined as

$$\bar{\bar{v}}_1^\xi := \sup_{\alpha_1 \in \Delta(A_1)} \min_{\alpha_2 \in B(\alpha_1)} u_1(\alpha_1, \alpha_2, \xi).$$

Typically, the supremum is not achieved by any mixed action, and so there is no mixed-action Stackelberg type. There are, of course, mixed commitment types that, if player 2 is convinced she is facing such a type, will yield payoffs arbitrarily close to the mixed-action Stackelberg payoff.

If the signals are informative about the actions of player 1, then the set of zero-entropy confirming best replies coincides with the set of best replies and so we have the following corollary:

Corollary 2 *Suppose the actions of player* 1 *are statistically identified, i.e.,* $\rho(\cdot|a_1) \neq \rho(\cdot|a_1')$ *for all* $a_1 \neq a_1' \in A_1$*. Suppose the action type* $\xi(a_1^\xi)$ *has positive prior probability for some Stackelberg action* a_1^ξ *for the payoff type* ξ*. Then, for all* $\eta > 0$*, there exists* $\bar{\delta} \in (0,1)$ *such that for all* $\delta \in (\bar{\delta}, 1)$*, the set of player* 1 *type* ξ*'s Nash equilibrium payoffs of the game* $\Gamma(\mu, \delta)$ *is bounded below by* $\bar{v}_1^\xi - \eta$*.*

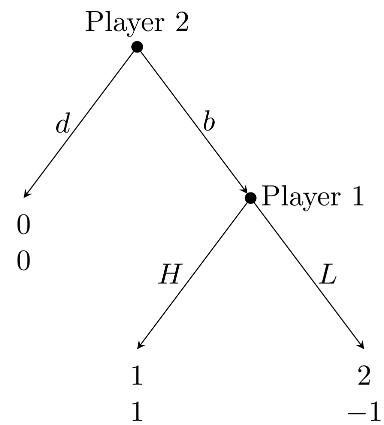
Suppose the mixed actions of player 1 *are statistically identified, i.e.,* $\rho(\cdot|\alpha_1) \neq \rho(\cdot|\alpha_1')$ *for all* $\alpha_1 \neq \alpha_1' \in \Delta(A_1)$*. Suppose the support of* μ *includes a set of action types* $\{\xi(\alpha_1) : \alpha_1 \in \Delta^*\}$*, where* Δ^* *is a countable dense subset of* $\Delta(A_1)$*. Then, for all* $\eta > 0$*, there exists* $\bar{\delta} \in (0,1)$ *such that for all* $\delta \in (\bar{\delta}, 1)$*, the set of player* 1 *type* ξ*'s Nash equilibrium payoffs of the game* $\Gamma(\mu, \delta)$ *is bounded below by* $\bar{\bar{v}}_1^\xi - \eta$*.*

2.5 More General Monitoring Structures

While the analysis in Section 2.4 is presented for the case in which the distribution of signals is a function of the actions of player 1 only, the arguments apply more generally. It is worth first noting that nothing in the argument depended on the signals being public, and so the argument applies immediately to the case of private monitoring, with signals that depend only on player 1's actions.

Suppose now that the distribution over signals depends on the actions of *both* players. The definition of ε-confirming best responses in Definition 2 is still valid, once condition 2 is adjusted to reflect the dependence of ρ on both players' actions:

$$d(\rho(\cdot|(\alpha_1, \alpha_2)) \| \rho(\cdot|(\alpha_1', \alpha_2))) < \varepsilon.$$



	d	b
H	$0, 0$	$1, 1$
L	$0, 0$	$2, -1$

Figure 3: The extensive and normal forms of the purchase game.

Proposition 1 and Corollary 1 are true in this more general setting as written. While the proof of Proposition 1 is unchanged (as a few moments of reflection will reveal), the proof of the Corollary is not. In particular, while $B^d_\varepsilon(\alpha_1)$ is still upper hemicontinuous in ε, it is not locally constant at 0 (see Gossner (2011b) for details).

The obtained reputation bounds, however, may be significantly weakened. One obvious way for signals to depend on the actions of both players is that the stage game has a nontrivial extensive form, with the public signal in each period consisting of the terminal node reached in that period. In this case, the Stackelberg action of player 1 may not be statistically identified, limiting the effects of reputations. Consider the purchase game illustrated in Figure 3. The short-lived customer first decides between "buy" (b) and "don't buy" (d), and then after b, the long-lived firm decides on the level of effort, high (H) or low (L). The extensive and normal forms are given in Figure 3. The game has three public signals, corresponding to the three terminal nodes. The distribution over the public signal is affected by the behavior of both players, and as a consequence the zero-entropy confirming best replies to H (the Stackelberg action) consist of b (the best response) and d (which is not a best response to H). There is no useful reputation bound in this case: Even if the short-lived customers assign positive probability to the possibility that the firm is the Stackelberg type, there is a sequential equilibrium in which the firm's payoff is 0 (Fudenberg and Levine, 1989, Example 4).

2.6 Temporary Reputations Under Imperfect Monitoring

As the analysis of Section 2.4 reveals, the lack of perfect monitoring of actions does not pose a difficulty for the formation of reputations. It does however pose a difficulty for its maintenance. Cripps, Mailath, and Samuelson (2004, 2007) show that under imperfect monitoring, reputations in repeated games are temporary. We present a new simpler entropy-based proof due to Gossner (2011a).

Earlier indications that reputations are temporary can be found in Benabou and Laroque (1992), Kalai and Lehrer (1995), and Mailath and Samuelson (2001). Benabou and Laroque (1992) show that the long-lived player eventually reveals her type in any Markov perfect equilibrium of a particular repeated game of strategic information transmission. Kalai and Lehrer (1995) use merging arguments to show that, under weaker conditions than we impose here, play in repeated games of incomplete information must converge to a subjective correlated equilibrium of the complete information continuation game.[15] We describe Mailath and Samuelson (2001) in Section 4.3.3.

For simplicity, we restrict attention to the case in which there are only two types of player 1, the normal type ξ_0 and the action type $\hat{\xi} = \xi(\hat{\alpha}_1)$.[16] We assume the public signals have full support (Assumption 1) under $\hat{\alpha}_1$.[17] Reputations are temporary under private monitoring as well, with an identical proof (though the notation becomes a little messier). We also assume that with sufficiently many observations, either player can correctly identify, from the frequencies of the signals, any fixed stage-game action of their opponent (Assumptions 2 and 3). We now allow the signal distribution to depend upon both players' actions (since doing so does not result in any complications to the arguments, while clarifying their nature).

Assumption 1 (Full Support) *$\rho(y|\hat{\alpha}_1, a_2) > 0$ for all $a_2 \in A_2$ and $y \in Y$.*

Assumption 2 (Identification of 1) *For all $\alpha_2 \in \Delta(A_2)$, the $|A_1|$ columns in the matrix $[\rho(y|a_1\alpha_2)]_{y\in Y, a_1\in A_1}$ are linearly independent.*

[15] This result is immediate in our context, since we examine a Nash equilibrium of the incomplete information game.

[16] The case of countably many simple action types is also covered, using the modifications described in Cripps, Mailath, and Samuelson (2004, Section 6.1).

[17] This is stronger than necessary. For example, we can easily accommodate public player 2 actions by setting $Y = Y_1 \times A_2$, and assuming the signal in Y_1 has full support.

Assumption 3 (Identification of 2) *For all $a_1 \in A_1$, the $|A_2|$ columns in the matrix $[\rho(y|a_1a_2)]_{y\in Y, a_2\in A_2}$ are linearly independent.*

We assume the actions of the short-lived player are not observed by player 1 (but see footnote 17), and so player 1's period t private history consists of the public signals and his own past actions, denoted by $h_1^t \in H_1^t := (A_1 \times Y)^t$. We continue to assume that the short-lived players do not observe past short-lived player actions, and so the private history for player 2 is denoted $h_2^t \in H_2^t := Y^t$.[18]

Given a strategy profile (σ_1, σ_2) of the incomplete information game, the short-lived player's belief in period t that player 1 is type $\hat{\xi}$ is

$$\mu^t(h_2^t) := \mathbf{P}(\hat{\xi}|h_2^t),$$

and so μ^0 is the period 0, or prior, probability assigned to $\hat{\xi}$.

Proposition 2 *Suppose Assumptions 1–3 are satisfied. Suppose player* 2 *has a unique best response $\hat{a}_2$ to $\hat{\alpha}_1$ and that $(\hat{\alpha}_1, \hat{a}_2)$ is not a Nash equilibrium of the stage game. If (σ_1, σ_2) is a Nash equilibrium of the game $\Gamma(\mu, \delta)$, then*

$$\mu^t \to 0, \qquad \widetilde{\mathbf{P}}\text{-}a.s.$$

Hence, conditional on player 1 being the normal type, player 2's posterior belief that 1 is normal almost-surely converges to one.

By Assumption 1, Bayes' rule determines μ^t after all histories. At any Nash equilibrium of this game, the belief μ^t is a bounded martingale and so converges $\mathbf{P}$-almost surely (and hence $\widetilde{\mathbf{P}}$- and $\widehat{\mathbf{P}}$-almost surely) to a random variable μ^∞. The idea of the proof is:

1. En route to a contradiction, assume that there is a positive $\widetilde{\mathbf{P}}$-probability event on which μ^∞ is strictly positive.

2. On this event, player 2 believes that both types of player 1 are eventually choosing the same distribution over actions $\hat{\alpha}_1$ (because otherwise player 2 could distinguish them).

3. Consequently, on a positive $\widetilde{\mathbf{P}}$-probability set of histories, eventually, player 2 will always play a best response to $\hat{\alpha}_1$.

[18] As for the earlier analysis on reputation bounds, the disappearing reputation results also hold when short-lived players observe past short-lived player actions (a natural assumption under private monitoring).

4. Assumption 3 then implies that there is a positive $\widetilde{\mathbf{P}}$-probability set of histories on which player 1 infers that player 2 is for many periods best responding to $\hat{\alpha}_1$, irrespective of the observed signals.

5. This yields the contradiction, since player 1 has a strict incentive to play differently than $\hat{\alpha}_1$.

Since $\hat{\alpha}_1$ is not a best reply to $\hat{a}_2$, there is a $\gamma > 0$, an action $\hat{a}_1 \in A_1$ receiving positive probability under $\hat{\alpha}_1$, and an $\varepsilon_2 > 0$ such that

$$\gamma < \min_{\alpha_2(\hat{a}_2) \geq 1-\varepsilon_2} \left(\max_{a_1} u_1(a_1, \alpha_2) - u_1(\hat{a}_1, \alpha_2) \right). \tag{6}$$

Define $\hat{p}(h_2^t) := \rho(\cdot|(\hat{\alpha}_1, \sigma_2(h_2^t))$ and redefine $p(h_2^t) := \rho(\cdot|\alpha_1(h_2^t), \sigma_2(h_2^t))$. Recall that $\|\cdot\|$ denotes L^1 distance. With this notation, Assumption 2 implies that there exists $\varepsilon_3 > 0$ such that, if $\left\|p(h_2^t) - \hat{p}(h_2^t)\right\| < \varepsilon_3$ (so that player 2 assigns sufficiently high probability to $\hat{\alpha}_1$), then player 2's unique best response is $\hat{a}_2$. Assumptions 1 and 2 imply that there exists $\varepsilon_1 > 0$, with

$$\varepsilon_1 < \min\left\{\varepsilon_3, \min_{(\alpha_1,\alpha_2):\alpha_1(\hat{a}_1)=0} \|\rho(\cdot|(\alpha_1, \alpha_2)) - \rho(\cdot|(\hat{\alpha}_1, \alpha_2))\|\right\},$$

such that

$$\underline{\rho} := \min_{\alpha_1, a_2} \{\rho(y|\alpha_1, a_2) : \|\rho(\cdot|\alpha_1, a_2) - \rho(\cdot|\hat{\alpha}_1, a_2)\| \leq \varepsilon_1\} > 0. \tag{7}$$

On the event

$$X^t := \left\{\left\|p(h_2^t) - \hat{p}(h_2^t)\right\| < \varepsilon_1\right\},$$

player 2's beliefs lead to her best responding to $\hat{\alpha}_1$, i.e., $\sigma_2(h_2^t) = \hat{a}_2$. The first lemma bounds the extent of player 2's $\widetilde{\mathbf{P}}$-expected surprises (i.e., periods in which player 2 both assigns a nontrivial probability to player 1 being $\hat{\xi}$ and believes $p(h_2^t)$ is far from $\hat{p}(h_2^t)$):

Lemma 2 ***(Player*** 2 ***either learns the type is normal or doesn't believe it matters)***

$$\sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[(\mu^t)^2(1 - \mathbb{1}_{X^t})\right] \leq -\frac{2\log(1-\mu^0)}{\varepsilon_1^2}$$

where $\mathbb{1}_{X^t}$ is the indicator function for the event X^t.

Proof. Since (where $\tilde{p}(h_2^t)$ is the predicted signal distribution conditional on ξ_0)

$$p(h_2^t) = \mu^t \hat{p}(h_2^t) + (1-\mu^t)\tilde{p}(h_2^t),$$

we have

$$\mu^t(p(h_2^t) - \hat{p}(h_2^t)) = (1-\mu^t)(\tilde{p}(h_2^t) - p(h_2^t)),$$

and so

$$\mu^t \left\| p(h_2^t) - \hat{p}(h_2^t) \right\| \leq \left\| p(h_2^t) - \tilde{p}(h_2^t) \right\|.$$

Then,

$$\begin{aligned}
\frac{\varepsilon_1^2}{2} \sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[(\mu^t)^2(1-\mathbb{1}_{X^t})\right] &\leq \frac{1}{2}\sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[(\mu^t)^2 \left\| p(h_2^t) - \hat{p}(h_2^t) \right\|^2\right] \\
&\leq \frac{1}{2}\sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[\left\| p(h_2^t) - \tilde{p}(h_2^t) \right\|^2\right] \\
&\leq \sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[d(\tilde{p}(h_2^t) \| p(h_2^t))\right] \\
&\leq -\log(1-\mu^0),
\end{aligned}$$

where the penultimate inequality is Pinsker's inequality (2) and the final inequality is (5) (with the normal type in the role of $\hat{\xi}$). ■

2.6.1 The Implications of Reputations Not Disappearing

Since posterior beliefs are a bounded martingale under $\mathbf{P}$, they converge $\mathbf{P}$ (and so $\widetilde{\mathbf{P}}$) almost surely, with limit μ^∞. If reputations do not disappear almost surely under $\widetilde{\mathbf{P}}$, then

$$\widetilde{\mathbf{P}}(\mu^\infty = 0) < 1,$$

and so there exists a $\lambda > 0$ and T_0 such that

$$\widetilde{\mathbf{P}}(\mu^t \geq \lambda, \forall t \geq T_0) > 0.$$

Define

$$F := \{\mu^t \geq \lambda, \forall t \geq T_0\}.$$

We first show that if reputations do not disappear almost surely under $\widetilde{\mathbf{P}}$, then eventually, with $\widetilde{\mathbf{P}}$-positive probability, player 2 must believe ξ_0 almost plays $\hat{\alpha}_1$ in every future period.

Lemma 3 ***(On F, eventually player 2 believes ξ_0 plays $\hat{\alpha}_1$)*** *Suppose $\mu^t \not\to 0$ $\widetilde{\mathbf{P}}$-a.s. There exists T_1 such that for*

$$B := \bigcap\nolimits_{t \geq T_1} X^t,$$

we have

$$\widetilde{\mathbf{P}}(B) \geq \widetilde{\mathbf{P}}(F \cap B) > 0.$$

Proof. We begin with the following calculation:

$$\begin{aligned}\sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[(\mu^t)^2(1 - \mathbb{1}_{X^t})\right] &\geq \widetilde{\mathbf{P}}(F) \sum_{t=0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[(\mu^t)^2(1 - \mathbb{1}_{X^t}) \mid F\right] \\ &\geq \widetilde{\mathbf{P}}(F)\lambda^2 \sum_{t=T_0}^{\infty} E_{\widetilde{\mathbf{P}}}\left[1 - \mathbb{1}_{X^t} \mid F\right].\end{aligned} \tag{8}$$

Lemma 2 implies the left side of (8) is finite, and so there exists $T_1 \geq T_0$ such that

$$\sum_{t \geq T_1} E_{\widetilde{\mathbf{P}}}\left[1 - \mathbb{1}_{X^t} \mid F\right] < 1.$$

Then,

$$\begin{aligned}\widetilde{\mathbf{P}}(F \cap B) &= \widetilde{\mathbf{P}}(F) - \widetilde{\mathbf{P}}(F \setminus B) \\ &= \widetilde{\mathbf{P}}(F) - \widetilde{\mathbf{P}}(F \cap (\Omega \setminus B)) \\ &= \widetilde{\mathbf{P}}(F)\left(1 - \widetilde{\mathbf{P}}(\Omega \setminus B | F)\right) \\ &\geq \widetilde{\mathbf{P}}(F)\left(1 - \sum\nolimits_{t \geq T_1} \widetilde{\mathbf{P}}(\Omega \setminus X^t | F)\right) > 0.\end{aligned}$$

■

The next lemma effectively asserts that when player 2 is eventually (under $\widetilde{\mathbf{P}}$) always playing $\hat{a}_2$, the best response to $\hat{\alpha}_1$, then the normal player 1 sees histories that lead him to be confident that for many periods, player 2 is indeed playing $\hat{a}_2$. Note that on the set B, player 2 is playing $\hat{a}_2$ in every period after T_1. Denote the filtration describing player i's information by $(\mathcal{H}_i^t)_t$.

Lemma 4 ***(On B, eventually player 1 figures out that player 2 is best responding to $\hat{\alpha}_1$)*** *Suppose $\mu^t \not\to 0$ $\widetilde{\mathbf{P}}$-a.s. For the event B from Lemma 3, for all τ, there is a subsequence (t_n) such that as $n \to \infty$,*

$$\sum_{k=1}^{\tau} \left\{1 - E_{\widetilde{\mathbf{P}}}\left[\sigma_2(h_2^{t_n+k})(\hat{a}_2) \middle| \mathcal{H}_1^{t_n}\right]\right\} \mathbb{1}_B \to 0 \quad \widetilde{\mathbf{P}}\text{-a.s.} \tag{9}$$

The convergence in (9) *holds also when* $\widetilde{\mathbf{P}}$ *is replaced by* $\mathbf{P}$.

Proof. We prove (9); obvious modifications to the argument proves it for $\mathbf{P}$. Recall that $h_1^{t+1} = (h_1^t, a_1^t y^t)$, and, for $k \geq 1$, denote player 1's k-step ahead prediction of his action and signal by $\beta_t^k(h_1^t) \in \Delta(A_1 \times Y)$, so that $\beta_t^k(h_1^t)(a_1^{t+k-1}y^{t+k-1}) = \widetilde{\mathbf{P}}(a_1^{t+k-1}y^{t+k-1} \mid h_1^t)$. Similarly, denote player 1's k-step prediction conditional on B by $\beta_{t,B}^k(h_1^t)$. The chain rule implies that, for all t and k,

$$
\begin{aligned}
& d(\beta_{t,B}^k(h_1^t) \| \beta_t^k(h_1^t)) \\
& \leq d\left(\widetilde{\mathbf{P}}(a_1^t y^t, \ldots, a_1^{t+k-1}y^{t+k-1} \mid h_1^t, B) \middle\| \widetilde{\mathbf{P}}(a_1^t y^t, \ldots, a_1^{t+k-1}y^{t+k-1} \mid h_1^t) \right) \\
= {} & E_{\widetilde{\mathbf{P}}(\cdot \mid h_1^t, B)} \sum_{k'=1}^{k} d\left(\widetilde{\mathbf{P}}(a_1^{t+k'-1}y^{t+k'-1} \mid h_1^{t+k'-1}, B) \middle\| \widetilde{\mathbf{P}}(a_1^{t+k'-1}y^{t+k'-1} \mid h_1^{t+k'-1}) \right) \\
& \qquad = E_{\widetilde{\mathbf{P}}(\cdot \mid h_1^t, B)} \sum_{k'=1}^{k} d\left(\beta_{t+k'-1,B}^1(h_1^{t+k'-1}) \| \beta_{t+k'-1}^1(h_1^{t+k'-1}) \right).
\end{aligned}
$$

Consequently, for all $h_1^{T_1}$ satisfying $\widetilde{\mathbf{P}}(h_1^{T_1}, B) > 0$, and for all k,

$$
\begin{aligned}
& \sum_{t \geq T_1} E_{\widetilde{\mathbf{P}}} \left[d(\beta_{t,B}^k(h_1^t) \| \beta_t^k(h_1^t)) \middle| h_1^{T_1}, B \right] \\
& \qquad \leq \sum_{t \geq T_1} E_{\widetilde{\mathbf{P}}} \left[\sum_{k'=1}^{k} d\left(\beta_{t+k'-1,B}^1(h_1^{t+k'-1}) \| \beta_{t+k'-1}^1(h_1^{t+k'-1}) \right) \middle| h_1^{T_1}, B \right] \\
& \qquad \leq k \sum_{t \geq T_1} E_{\widetilde{\mathbf{P}}} \left[d\left(\beta_{t,B}^1(h_1^t) \| \beta_t^1(h_1^t) \right) \middle| h_1^{T_1}, B \right] \\
& \qquad \leq -k \log \widetilde{\mathbf{P}}(B \mid h_1^{T_1}).
\end{aligned}
$$

The last inequality follows from Lemma 1 applied to the equality

$$
\begin{aligned}
\widetilde{\mathbf{P}}\left(y^{T_1}, \ldots, y^{T_1+\ell-1} \middle| h_1^{T_1} \right) = {} & \widetilde{\mathbf{P}}(B \mid h_1^{T_1}) \widetilde{\mathbf{P}}\left(y^{T_1}, \ldots, y^{T_1+\ell-1} \middle| h_1^{T_1}, B \right) \\
& + (1 - \widetilde{\mathbf{P}}(B \mid h_1^{T_1})) \widetilde{\mathbf{P}}\left(y^{T_1}, \ldots, y^{T_1+\ell-1} \middle| h_1^{T_1}, \Omega \setminus B \right)
\end{aligned}
$$

and the chain rule (via an argument similar to that leading to (5)).

Thus, for all k,

$$
E_{\widetilde{\mathbf{P}}} \left[d(\beta_{t,B}^k(h_1^t) \| \beta_t^k(h_1^t)) \middle| h_1^{T_1}, B \right] \to 0 \text{ as } t \to \infty,
$$

and so (applying Pinsker's inequality)

$$E_{\widetilde{\mathbf{P}}}\left[\|\beta^k_{t,B}(h^t_1)-\beta^k_t(h^t_1))\|\,\Big|\,h^{T_1}_1,B\right]\to 0 \text{ as } t\to\infty,$$

yielding, for all τ,

$$\sum_{k=1}^{\tau}E_{\widetilde{\mathbf{P}}}\left[\|\beta^k_{t,B}(h^t_1)-\beta^k_t(h^t_1))\|\,\Big|\,h^{T_1}_1,B\right]\to 0 \text{ as } t\to\infty. \tag{10}$$

Since

$$\|\beta^k_{t,B}(h^t_1)-\beta^k_t(h^t_1))\|\geq\kappa\left(1-E_{\widetilde{\mathbf{P}}}[\sigma_2(h^{t+k}_2)(\hat{a}_2)|h^t_1]\right),$$

where

$$\kappa:=\min_{\substack{a_1\in A_1,\\ \alpha_2\in\Delta(A_2),\alpha_2(\hat{a}_2)=0}}\|\rho(\cdot|a_1,\hat{a}_2)-\rho(\cdot|a_1,\alpha_2)\|$$

is strictly positive by Assumption 3, (10) implies

$$\sum_{k=1}^{\tau}E_{\widetilde{\mathbf{P}}}\left[1-E_{\widetilde{\mathbf{P}}}[\sigma_2(h^{t+k}_2)(\hat{a}_2)|h^t_1]\,\Big|\,h^{T_1}_1,B\right]\to 0 \text{ as } t\to\infty.$$

This implies (9), since convergence in probability implies subsequence a.e. convergence (Chung, 1974, Theorem 4.2.3). ■

Using Lemma 4, we next argue that on $B\cap F$, player 1 believes that player 2 is eventually ignoring her history while best responding to $\hat{\alpha}_1$. In the following lemma, ε_2 is from (6). The set $A_t(\tau)$ is the set of player 2 t-period histories such that player 2 ignores the *next* τ signals (the positive probability condition only eliminates 2's actions inconsistent with σ_2, since under Assumption 1, every signal realization has positive probability under $\hat{\alpha}_1$ and so under $\mathbf{P}$).

Lemma 5 ***(Eventually player*** **1** ***figures out that player*** **2** ***is best responding to*** $\hat{\alpha}_1$***, independently of signals)*** *Suppose* $\mu^t\not\to 0$ $\widetilde{\mathbf{P}}$*-a.s. For all* τ*, there is a subsequence* (t_m) *such that as* $m\to\infty$*,*

$$\widetilde{\mathbf{P}}(A_{t_m}(\tau)|\mathcal{H}^{t_m}_1)\mathbb{1}_{B\cap F}\to\mathbb{1}_{B\cap F}\quad\widetilde{\mathbf{P}}\text{-a.s.},$$

where

$$A_t(\tau):=\{h^t_2:\sigma_2(h^{t+k}_2)(\hat{a}_2)>1-\varepsilon_2/2,\ \forall h^{t+k}_2 \text{ s.t. } \mathbf{P}(h^{t+k}_2|h^t_2)>0,\\ \forall k=1,\ldots,\tau,\}.$$

Proof. Define

$$A_t^\tau := \{h_2^{t+\tau} : \sigma_2(h_2^{t+k})(\hat{a}_2) > 1 - \varepsilon_2/2,\ k = 1, \ldots, \tau\}.$$

The set A_t^τ is the set of $(t+\tau)$-period histories for player 2 at which she essentially best responds to $\hat{\alpha}_1$ for the *last* τ periods. Note that, viewed as subsets of Ω, $A_t(\tau) \subset A_t^\tau$.[19]

We then have

$$\begin{aligned}
&\sum_{k=1}^{\tau} E_{\widetilde{\mathbf{P}}} \left\{ 1 - E_{\widetilde{\mathbf{P}}} \left[\sigma_2(h_2^{t+k})(\hat{a}_2) \middle| h_1^t \right] \middle| B \cap F \right\} \\
&\qquad = E_{\widetilde{\mathbf{P}}} \left\{ E_{\widetilde{\mathbf{P}}} \left[\textstyle\sum_{k=1}^{\tau} \left(1 - \sigma_2(h_2^{t+k})(\hat{a}_2) \right) \middle| h_1^t \right] \middle| B \cap F \right\} \\
&\qquad = E_{\widetilde{\mathbf{P}}} \left\{ E_{\widetilde{\mathbf{P}}} \left[\textstyle\sum_{k=1}^{\tau} \left(1 - \sigma_2(h_2^{t+k})(\hat{a}_2) \right) \middle| h_1^t, A_t^\tau \right] \widetilde{\mathbf{P}}(A_t^\tau | h_1^t) \middle| B \cap F \right\} \\
&+ E_{\widetilde{\mathbf{P}}} \left\{ E_{\widetilde{\mathbf{P}}} \left[\textstyle\sum_{k=1}^{\tau} \left(1 - \sigma_2(h_2^{t+k})(\hat{a}_2) \right) \middle| h_1^t, \Omega \setminus A_t^\tau \right] (1 - \widetilde{\mathbf{P}}(A_t^\tau | h_1^t)) \middle| B \cap F \right\}.
\end{aligned}$$

Dropping the first term and using the implied lower bound from $\Omega \setminus A_t^\tau$ on $\sum_{k=1}^{\tau}(1 - \sigma_2(h_2^{t+k})(\hat{a}_2))$ yields

$$\begin{aligned}
&\sum_{k=1}^{\tau} E_{\widetilde{\mathbf{P}}} \left\{ 1 - E_{\widetilde{\mathbf{P}}} \left[\sigma_2(h_2^{t+k})(\hat{a}_2) \middle| h_1^t \right] \middle| B \cap F \right\} \\
&\qquad\qquad \geq \frac{\varepsilon_2}{2} \left(1 - E_{\widetilde{\mathbf{P}}} \left\{ \widetilde{\mathbf{P}}(A_t^\tau | h_1^t) \middle| B \cap F \right\} \right).
\end{aligned}$$

Lemma 4 then implies

$$\lim_{n \to \infty} E_{\widetilde{\mathbf{P}}} \left\{ \widetilde{\mathbf{P}}(A_{t_n}^\tau | \mathcal{H}_1^{t_n}) \middle| B \cap F \right\} = 1.$$

As before, this then implies that on a subsequence (t_ℓ) of (t_n), we have, as $\ell \to \infty$,

$$\widetilde{\mathbf{P}}(A_{t_\ell}^\tau | \mathcal{H}_1^{t_\ell}) \mathbb{1}_{B \cap F} \to \mathbb{1}_{B \cap F} \quad \widetilde{\mathbf{P}}\text{-a.s.} \tag{11}$$

Thus, the normal player 1 eventually (on $B \cap F$) assigns probability 1 to $(t_\ell + \tau)$-period histories for player 2 at which player 2 essentially best responds to $\hat{\alpha}_1$ for the *last* τ periods. It remains to argue that this convergence holds when $A_{t_\ell}(\tau)$ replaces $A_{t_\ell}^\tau$.

[19] More precisely, if $h_2^t(\omega)$ is the t-period player 2 history under $\omega \in \Omega$, then

$$\{\omega : h_2^t(\omega) \in A_t(\tau)\} \subset \{\omega : h_2^{t+\tau}(\omega) \in A_t^\tau\}.$$

A similar argument to that proving (11) shows that as $m \to \infty$,

$$\mathbf{P}(A^{\tau}_{t_m}|\mathcal{H}^{t_m}_1)\mathbb{1}_{B\cap F} \to \mathbb{1}_{B\cap F} \quad \mathbf{P}\text{-a.s.}$$

(where the subsequence (t_m) can be chosen so that (11) still holds).[20]

Since $B \cap F \in \mathcal{H}^{\infty}_2$ and $\mathcal{H}^t_2 \subset \mathcal{H}^t_1$, this implies

$$\mathbf{P}(A^{\tau}_{t_m}|\mathcal{H}^{t_m}_2)\mathbb{1}_{B\cap F} \to \mathbb{1}_{B\cap F} \quad \mathbf{P}\text{-a.s.}$$

We claim that for all $\omega \in B \cap\ F$, for sufficiently large m if $h^{t_m+\tau}_2(\omega) \in A^{\tau}_{t_m}$, then $h^{t_m}_2(\omega) \in A_t(\tau)$. This then implies the desired result.

For suppose not. Then, for infinitely many t_m,

$$h^{t_m}_2(\omega) \notin A_{t_m}(\tau) \quad \text{and} \quad h^{t_m+\tau}_2(\omega) \in A^{\tau}_{t_m}.$$

At any such t_m, since there is at least one τ period continuation of the history $h^{t_m}_2(\omega)$ that is not in $A^{\tau}_{t_m}$, we have (from Assumption 1) $\widehat{\mathbf{P}}(A^{\tau}_{t_m} \mid \mathcal{H}^{t_m}_2)(\omega) \le 1 - \underline{\rho}^{\tau}$, where $\underline{\rho} > 0$ is defined in (7). Moreover, on F, $\mu^{t_m} \ge \lambda$ for $t_m \ge T_0$. But this yields a contradiction, since these two imply that $\mathbf{P}(A^{\tau}_{t_m} \mid \mathcal{H}^{t_m}_2)(\omega)$ is bounded away from 1 infinitely often:

$$\begin{aligned}\mathbf{P}(A^{\tau}_{t_m} \mid \mathcal{H}^{t_m}_2)(\omega) &\le (1-\mu^{t_m}) + \mu^{t_m}\widehat{\mathbf{P}}(A^{\tau}_{t_m} \mid \mathcal{H}^{t_m}_2)(\omega) \\ &\le 1-\mu^{t_m} + \mu^{t_m}(1-\underline{\rho}^{\tau}) = 1-\lambda\underline{\rho}^{\tau}.\end{aligned}$$

■

Indeed, player 2 "knows" that player 1 believes that player 2 is eventually ignoring her history:

Lemma 6 ***(Eventually player* 2 *figures out that player* 1 *figures out...)*** *Suppose $\mu^t \not\to 0$ $\widetilde{\mathbf{P}}$-a.s. For all τ, there is a subsequence (t_m) such that as $m \to \infty$,*

$$E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_{t_m}(\tau)|\mathcal{H}^{t_m}_1)\middle|\,\mathcal{H}^{t_m}_2\right]\mathbb{1}_{B\cap F} \to \mathbb{1}_{B\cap F}, \qquad \widetilde{\mathbf{P}}\text{-a.s.}$$

Proof. Let (t_m) be the subsequence identified in Lemma 5. Conditioning on $\mathcal{H}^{t_m}_2$, Lemma 5 implies

$$E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_{t_m}(\tau)|\mathcal{H}^{t_m}_1)\mathbb{1}_{B\cap F}\middle|\,\mathcal{H}^{t_m}_2\right] - E_{\widetilde{\mathbf{P}}}\left[\mathbb{1}_{B\cap F}|\,\mathcal{H}^{t_m}_2\right] \to 0, \qquad \widetilde{\mathbf{P}}\text{-a.s.}$$

[20] Assumption 1 is weaker than full support imperfect monitoring, requiring only that all signals have positive probability under $\widehat{\mathbf{P}}$. Under full support imperfect monitoring ($\rho(y|a) > 0$ for all a), the following argument is valid with $\widetilde{\mathbf{P}}$ replacing $\mathbf{P}$, without the need to introduce a new subsequence.

The observation that $E_{\widetilde{\mathbf{P}}}[\mathbb{1}_{B\cap F} \mid \mathcal{H}_2^{t_m}]$ converges $\widetilde{\mathbf{P}}$-almost surely to $\mathbb{1}_{B\cap F}$ (since $B \cap F$ is in $\mathcal{H}_2^{\infty}$) yields

$$E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_{t_m}(\tau)|\mathcal{H}_1^{t_m})\mathbb{1}_{B\cap F}\middle|\mathcal{H}_2^{t_m}\right] \to \mathbb{1}_{B\cap F}, \qquad \widetilde{\mathbf{P}}\text{-a.s.},$$

implying the lemma.[21] ■

2.6.2 The Contradiction and Conclusion of the Proof

For fixed δ, there is a τ such that

$$(1-\delta)\gamma > 2\delta^{\tau} \max_a |u_1(a)|,$$

that is, the loss of γ (defined in (6)) in one period exceeds any possible potential gain deferred τ periods.

For this value of τ, for $\widetilde{\mathbf{P}}(A_{t_m}(\tau)|h_1^{t_m})$ close enough to 1, optimal play by player 1 requires $\sigma_1(h_1^{t_m})(\hat{a}_1) = 0$. Lemma 6 then implies that, on $B \cap F$, eventually player 2 predicts $\sigma_1(h_1^{t_m})(\hat{a}_1) = 0$, which implies

$$\lim_{m\to\infty} \widetilde{\mathbf{P}}(X^{t_m}|B\cap F) = 0,$$

which is a contradiction, since $\widetilde{\mathbf{P}}(X^t|B) = 1$ for all $t \geq T_1$.

2.7 Interpretation

There is a tension between Propositions 1 and 2. By Proposition 1 and its Corollaries 1 and 2, the normal player 1's ability to masquerade as a commitment type places a lower bound on his payoff. And yet, according to Proposition 2, when the player 1's actions are imperfectly monitored, if player 1 is indeed normal, then player 2 must learn that this is the case.

[21] Define $g_t := E_{\widetilde{\mathbf{P}}}\left[\mathbb{1}_{B\cap F}|\mathcal{H}_2^t\right]$. Then, since g_t is $\mathcal{H}_2^t$-measurable,

$$\begin{aligned}
&E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_t(\tau)|\mathcal{H}_1^t)\mathbb{1}_{B\cap F}\middle|\mathcal{H}_2^t\right] - E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_t(\tau)|\mathcal{H}_1^t)\middle|\mathcal{H}_2^t\right]\mathbb{1}_{B\cap F} \\
&\quad = E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_t(\tau)|\mathcal{H}_1^t)(\mathbb{1}_{B\cap F} - g_t)\middle|\mathcal{H}_2^t\right] + E_{\widetilde{\mathbf{P}}}\left[\widetilde{\mathbf{P}}(A_t(\tau)|\mathcal{H}_1^t)\middle|\mathcal{H}_2^t\right](g_t - \mathbb{1}_{B\cap F}),
\end{aligned}$$

which converges to 0 $\widetilde{\mathbf{P}}$-almost surely, since $g_t \to \mathbb{1}_{B\cap F}$ $\widetilde{\mathbf{P}}$-almost surely.

Indeed, more can be said if there is a little more structure on the game.[22] If in addition to public imperfect monitoring of player 1's actions, the actions of player 2 are public, then not only does player 2 learn the type of player 1, but the continuation play of any Nash equilibrium converges to a Nash equilibrium of the complete information game (Cripps, Mailath, and Samuelson, 2004, Theorem 2).

The tension is only apparent, however, since the reputation bounds are only effective for high discount factors and only concern ex ante payoffs. In contrast, the disappearing reputation results fix the discount factor, and consider the asymptotic evolution of beliefs (and behavior).

Suppose player 2's actions are public and player 1 is either normal or the Stackelberg type. Fix $\eta > 0$ and a discount factor δ', and suppose player 1 is the normal type with probability $1 - \mu^0$. Suppose moreover, δ' is strictly larger than the bound $\bar{\delta}$ from Corollary 2, so that the normal player 1's expected payoff in any Nash equilibrium is at least $v_1^* - \eta$, where v_1^* is 1's Stackelberg payoff. By Proposition 2, over time the posterior will tend to fall. As illustrated in Figure 4, since the discount factor is fixed at δ', the posterior will eventually fall below μ', and so Corollary 2 no longer applies. However, since all signals have positive probability, there is positive probability that even after the belief has fallen below μ', some history of signals will again push the posterior above μ'. Since continuation play forms a Nash equilibrium of the incomplete information game, Corollary 2 again applies. The probability of such histories becomes vanishingly small as the histories become long.

As an example, recall the imperfect monitoring product-choice game (with public player 2 actions) with the pure Stackelberg type $\hat{\xi} = \xi(H)$ considered in Example 2. For the parameterization in that example, the complete information repeated game has a unique sequential equilibrium outcome, in which L is played in every period. Nonetheless, by Corollary 2, given any prior probability μ^0 on $\xi(H)$, for sufficiently high δ, player 1's ex ante equilibrium payoff is at least $1\frac{3}{4}$, higher than the complete information equilibrium payoff of 1. Fix such a δ. We then know that in *any* equilibrium,

[22] One might seek refuge in the observation that one can readily find equilibria in *perfect* monitoring games in which reputation effects persist indefinitely. The initial finite horizon reputation models of Kreps and Wilson (1982) and Milgrom and Roberts (1982) exhibited equilibria in which player 1 chooses the Stackelberg action in all but a relatively small number of terminal periods (see also Conlon, 2003). The analogous equilibria in infinitely repeated games feature the Stackelberg action in every period, enforced by the realization that any misstep reveals player 1 to be normal. However, it is somewhat discomforting to rely on properties of equilibria in perfect monitoring games that have no counterparts in games of imperfect monitoring.

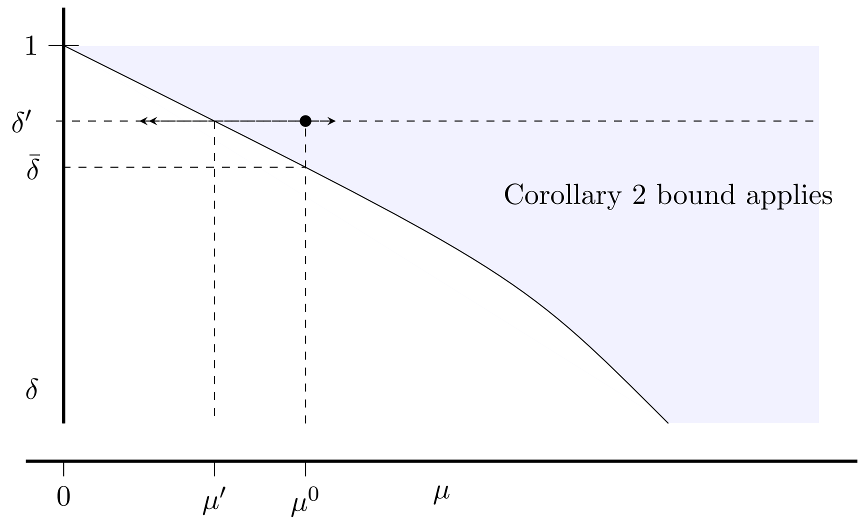


Figure 4: Illustration of reputation effects under imperfect monitoring. Suppose $\hat{\xi}$ is a Stackelberg type. For fixed $\eta > 0$, $\bar{\delta}$ is the lower bound from Corollary 2 for prior probability $\mu(\hat{\xi}) = \mu^0$. For players 1's discount factor $\delta' > \bar{\delta}$, Corollary 2 applies in period 0. The arrows indicate possible belief updates of player 2 about the likelihood of player 1 being the Stackelberg type $\hat{\xi}$. When player 1 is normal, almost surely the vertical axis is reached.

for many periods, the short-lived player must be playing c (which requires that the short-lived player in those periods assign sufficient probability to H). However, the equilibrium incentive player 1 has to cheat in each period (if only with low probability), arising from the imperfection in the monitoring, implies that (with high probability) the posterior will eventually fall sufficiently that the lack of complete information intertemporal incentives forces player 1 to play L (and player 2 to respond with s). Indeed, from Cripps, Mailath, and Samuelson (2004, Theorem 3), for high δ and for any $\varepsilon > 0$ there is an equilibrium of the incomplete information game that has ex ante player 1 payoffs over $1\frac{3}{4}$, and yet the $\widetilde{\mathbf{P}}$ probability of the event that eventually Ls is played in *every* period is at least $1 - \varepsilon$.

This discussion raises an important question: Why do we care about beliefs after histories that occur so far into the future as to be effectively discounted into irrelevance when calculating ex ante payoffs?

While the short run properties of equilibria are interesting, we believe that the long run equilibrium properties are relevant in many situations. For example, an analyst may not know the age of the relationship to which the model is to be applied. We do sometimes observe strategic interactions from a well-defined beginning, but we also often encounter on-going interactions whose beginnings are difficult to identify. Long run equilibrium properties may be an important guide to behavior in the latter cases. Alternatively, one might take the view of a social planner who is concerned with the continuation payoffs of the long-lived player and with the fate of all short-lived players, even those in the distant future. Finally, interest may center on the steady states of models with incomplete information, again directing attention to long run properties.

2.8 Exogenously Informative Signals

Section 2.6 established conditions under which reputations eventually disappear. Section 2.7 explained why this result must be interpreted carefully, with an emphasis on keeping track of which limits are taken in which order. This section, considering the impact of exogenously informative signals, provides a further illustration of the subtleties that can arise in taking limits. This section is based on Hu (2013), which provides a general analysis; Wiseman's (2009) original analysis considered the chain store example.

We consider the product-choice game of Figure 1 with the set of types $\Xi = \{\xi_0, \hat{\xi}\}$, where ξ_0 is the normal type and $\hat{\xi} = \xi(H)$ is the Stackelberg type. The stage game is a game of perfect monitoring. *Imperfect* monitoring played an important role in the disappearing-reputation result of Section 2.6. In contrast, there is no difficulty in constructing equilibria of the perfect monitoring product-choice game in which both the normal and Stackelberg types play H in every period, with customers never learning the type of player 1. It thus seems as if a game of perfect monitoring is not particularly fertile ground for studying temporary reputations.

However, suppose that at the end of every period, the players observe a public signal independent of the players' actions. In each period t, the signal z^t is an identically and independently distributed draw from $\{z_0, \hat{z}\}$. The signal is informative about player 1's type: $0 < \pi(\hat{z} \mid \xi_0) := \Pr(\hat{z} \mid \xi_0) < \Pr(\hat{z} \mid \hat{\xi}) =: \pi(\hat{z} \mid \hat{\xi}) < 1$. A sufficiently long sequence of such signals suffices for player 2 to almost surely learn player 1's type.

It now appears as if reputation arguments must lose their force. Because the public signals are unrelated to the players' actions, there is a lower bound on the rate at which player 2 learns about player 1. If player 1 is very patient, the vast bulk of his discounted expected payoff will come from periods in which player 2 is virtually certain that 1 is the normal type (assuming 1 is indeed normal). It then seems as if there is no chance to establish a reputation-based lower bound on 1's payoff. Nonetheless, for all $\varepsilon > 0$, there exists $\bar{\delta} < 1$ such that if $\delta \in (\bar{\delta}, 1)$, then player 1's equilibrium payoff is at least

$$\underline{w}_H(\ell) - \varepsilon, \tag{12}$$

where $\underline{w}_H$ is the function illustrated in Figure 2, and $\ell := d(\pi(\cdot \mid \xi_0) \| \pi(\cdot \mid \hat{\xi}))$ is the relative entropy of $\pi(\cdot \mid \hat{\xi})$ with respect to $\pi(\cdot \mid \xi_0)$.

Suppose $\pi(\hat{z} \mid \xi_0) = 1 - \pi(\hat{z} \mid \hat{\xi}) = \alpha < \frac{1}{2}$. Then, $\ell = (1 - 2\alpha) \log[(1 - \alpha)/\alpha]$. For α near zero, individual signals are very informative, and we might expect that the reputation arguments would be ineffective. This is what we find: for $\alpha < 0.22$, we have $\ell > \log 2$, and so only the conclusion that player 1's equilibrium payoff is bounded above 0, which is *less* than player 1's minmax payoff in the stage game. Hence, sufficiently precise signals can indeed preclude the construction of a reputation bound on payoffs. On the other hand, as α approaches 1/2, so that signals become arbitrarily uninformative, ℓ approaches 0 and so the bound approaches 2, the reputation bound in the game without exogenous signals. For intermediate values of α that are not too large, reputations have some force, though not as much as if the public signals were completely uninformative.

Reputations still have force with exogenous signals because the signals have full support. Suppose customers have seen a long history of exogenous signals suggesting that the firm is normal (which is likely when the firm is indeed normal). If they do not expect the normal type to exert high effort after some such history, high effort in that period results in a dramatic increase in the posterior that the firm is the Stackelberg type and hence will exert high effort in the future. While this can happen infinitely often, it can't happen too frequently (because otherwise the resulting increases in posterior overwhelm the exogenous signals, leading to a similar contradiction as for the canonical reputations argument), resulting in (12).[23]

We conclude this subsection with the proof that (12) is a lower bound on equilibrium payoffs for player 1. The space of uncertainty is now $\Omega := \{\xi_0, \hat{\xi}\} \times (A_1 \times A_2 \times Z)^\infty$. The set of endogenous signals is A_1, while the

[23] In the model with perfect monitoring with no exogenous signals, customers can only be surprised a finite number of times under the considered deviation (if not, the increases in the posterior after each surprise eventually result in a posterior precluding further surprises, since posteriors never decrease). In contrast, with exogenous signals, there is the possibility of an infinite number of surprises, since the expected posterior decreases when there are no surprises. Nonetheless, if this happens too frequently, the increasing updates from the surprises dominate the decreasing updates from the exogenous signals and again the increases in the posterior after surprises eventually preclude further surprises.

the set of exogenous signals is given by Z. Fix an equilibrium $\sigma = (\sigma_1, \sigma_2)$ of the incomplete information game. As before, σ induces the unconditional probability measure $\mathbf{P}$ on Ω, while $\hat{\xi}$ (with σ_2) induces the measure $\widehat{\mathbf{P}}$.

While it is no longer the case that there is a uniform bound on the number of large expected one-step ahead prediction errors of the form (5), there is a useful nonuniform bound.

Let $\widehat{\mathbf{Q}}$ be the measure induced on Ω by the normal type ξ_0 playing H in every period. Then, for any history $h^t \in (A_1 \times A_2 \times Z)^t$, since the exogenous signals are independent of actions,

$$\widehat{\mathbf{Q}}^t(h^t) = \widehat{\mathbf{P}}^t(h^t) \prod_{\tau=0}^{t-1} \frac{\pi(z^\tau(h^t) \mid \xi_0)}{\pi(z^\tau(h^t) \mid \hat{\xi})},$$

where, as usual, we denote the marginals on t-period histories by a superscript t. Then, since $\widehat{\mathbf{P}}^t(h^t)/\mathbf{P}^t(h^t) \leq 1/\mu(\hat{\xi})$,

$$\begin{aligned} d\left(\widehat{\mathbf{Q}}^t \middle\| \mathbf{P}^t\right) &= \sum_{h^t} \widehat{\mathbf{Q}}^t(h^t) \log \frac{\widehat{\mathbf{P}}^t(h^t)}{\mathbf{P}^t(h^t)} + \sum_{h^t} \widehat{\mathbf{Q}}^t(h^t) \sum_{\tau=0}^{t-1} \log \frac{\pi(z^\tau(h^t) \mid \xi_0)}{\pi(z^\tau(h^t) \mid \hat{\xi})} \\ &= \sum_{h^t} \widehat{\mathbf{Q}}^t(h^t) \log \frac{\widehat{\mathbf{P}}^t(h^t)}{\mathbf{P}^t(h^t)} + \sum_{\tau=0}^{t-1} \sum_{z^\tau \in Z} \pi(z^\tau \mid \xi_0) \log \frac{\pi(z^\tau \mid \xi_0)}{\pi(z^\tau \mid \hat{\xi})} \\ &\leq -\log \mu(\hat{\xi}) + t\ell. \end{aligned} \tag{13}$$

It remains to bound the total discounted number of one-step ahead prediction errors. In the notation of Section 2.4.2, $\hat{p}$ is the degenerate distribution assigning probability 1 to H, while $p(h^t)$ is the probability that 2 assigns to player 1 choosing H in period t, given the history h^t.

Then, from the chain rule (3), where the last term is the expected relative entropy of the period t z-signal predictions, we have

$$\begin{aligned} d\left(\widehat{\mathbf{Q}}^{t+1} \middle\| \mathbf{P}^{t+1}\right) &= d\left(\widehat{\mathbf{Q}}^t \middle\| \mathbf{P}^t\right) + E_{\widehat{\mathbf{Q}}^t} d(\hat{p} \| p(h^t)) + E_{\widehat{\mathbf{Q}}^t, \hat{p}} d(\widehat{\mathbf{Q}}_Z^{t+1} \| \mathbf{P}_Z^{t+1}) \\ &\geq d\left(\widehat{\mathbf{Q}}^t \middle\| \mathbf{P}^t\right) + E_{\widehat{\mathbf{Q}}^t} d(\hat{p} \| p(h^t)). \end{aligned}$$

Thus, where we normalize $d(\widehat{\mathbf{Q}}^0 \| \mathbf{P}^0) = 0$,

$$(1-\delta) \sum_{t=0}^{\infty} \delta^t E_{\widehat{\mathbf{Q}}^t} d(\hat{p} \| p(h^t)) \leq (1-\delta) \sum_{t=0}^{\infty} \delta^t \left[d\left(\widehat{\mathbf{Q}}^{t+1} \middle\| \mathbf{P}^{t+1}\right) - d\left(\widehat{\mathbf{Q}}^t \middle\| \mathbf{P}^t\right) \right]$$

$$
\begin{aligned}
&= (1-\delta)^2 \sum_{t=1}^{\infty} \delta^{t-1} d\left(\widehat{\mathbf{Q}}^t \middle\| \mathbf{P}^t\right) \\
&\leq -(1-\delta)\log \mu(\hat{\xi}) + \ell,
\end{aligned}
$$

where the second inequality follows from (13) and some algebraic manipulation. The bound now follows from an argument similar to the proof of Proposition 1.

3 Reputations with Two Long-Lived Players

Section 2 studied reputations in the most common context, that of a long-lived player facing a succession of short-lived players. This section examines the case in which player 2 is also a long-lived player.

Reputation results for the case of two long-lived players are not as strong as those for the long-lived/short-lived case, and a basic theme of the work presented in this section is the trade-off between the specificity of the model and the strength of the results. To get strong results, one must either restrict attention to seemingly quite special games, or must rely on seemingly quite special commitment types.

The standard models with two long-lived players fix a discount factor for player 2 and then examine the limit as player 1's discount factor approaches one, making player 1 arbitrarily relatively patient. There is a smaller literature that examines reputations in the case of two equally patient long-lived players, with even weaker results. As we make the players progressively more symmetric by moving from the case of a short-lived player 2, to the case of a long-lived player 2 but arbitrarily more patient player 1, to the case of two equally patient long-lived players, the results become successively weaker. This is unsurprising. Reputation results require some asymmetry. A reputation result imposes a lower bound on equilibrium payoffs, and it is typically impossible to guarantee such a payoff to both players, For example, it is typically impossible for both players to receive their Stackelberg payoffs. Some asymmetry must then lie behind a result that guarantees such a payoff to one player, and the weaker this asymmetry, the weaker the reputation result.

3.1 Types vs. Actions

Suppose we simply apply the logic of Section 2, hoping to obtain a Stackelberg reputation bound when both players are long-lived and player 1's characteristics are unknown. To keep things simple, suppose there is perfect monitoring. If the normal player 1 persistently plays the Stackelberg action and player 2 assigns positive prior probability to a type committed to that action, then player 2 must eventually attach high probability to the event that the Stackelberg action is played in the future. This argument depends only upon the properties of Bayesian belief revision, independently of whether the person holding the beliefs is long lived or short lived.

If this belief suffices for player 2 to play a best response to the Stackelberg action, as is the case when player 2 is short lived, then the remainder of the argument is straightforward. The normal player 1 must eventually receive very nearly the Stackelberg payoff in each period of the repeated game. By making player 1 sufficiently patient, we can ensure that this consideration dominates player 1's payoffs, putting a lower bound on the latter.

The key step when working with two long-lived players is thus to establish conditions under which, as player 2 becomes increasingly convinced that the Stackelberg action will appear, she must eventually play a best response to that action. This initially seems obvious. If player 2 is "very" convinced that the Stackelberg action will be played not only now but for sufficiently many periods to come, there appears to be nothing better she can do than play a stage-game best response.

This intuition misses the following possibility. Player 2 may be choosing something other than a best response to the Stackelberg action out of fear that a current best response may trigger a disastrous future punishment. This punishment would not appear if player 2 faced the Stackelberg type, but player 2 can be made confident only that she faces the Stackelberg *action*, not the Stackelberg type. The fact that the punishment lies off the equilibrium path makes it difficult to assuage player 2's fear of such punishments.

The short-lived players of Section 2 find themselves in the same situation: convinced that their long-lived opponent will play the Stackelberg action, but uncertain as to what affect their own best response to this Stackelberg action will have on future behavior. However, because they are short-lived, this uncertainty does not affect their behavior. The difference between expecting the Stackelberg action and expecting the Stackelberg type (or more generally between expecting any action and expecting the corresponding type committed to that action) is irrelevant in the case of short-lived opponents, but crucial when facing long-lived opponents.

3.2 An Example: The Failure of Reputation Effects

This section presents a simple example, adapted from Schmidt (1993), illustrating the new issues that can arise when building a reputation against

	L	C	R
T	$10, 10$	$0, 0$	$-z, 9$
B	$0, 0$	$1, 1$	$1, 0$

Figure 5: A modified coordination game, $z \in [0, 8)$.

long-lived opponents. Consider the game in Figure 5. Not only is the profile TL a Nash equilibrium of the stage game in Figure 5, it gives the player 1 his largest feasible payoff. At the same time, LC is another Nash equilibrium of the stage game, so the complete information repeated game with perfect monitoring has the infinite repetition of LC as another equilibrium outcome. It is immediate that if player 2 only assigns positive probability to the normal type (the payoff type of player 1 with payoffs shown in Figure 5) and the simple Stackelberg-action type (which always plays T), then this particular form of incomplete information again generates reputation effects.

In the example, we add another commitment type, an enforcement type, whose behavior depends on history (so this type is not simple). The idea is to use this type to induce player 2 to not always statically best respond to T. Instead, on the candidate equilibrium path, player 2 will play L in even periods, and and R in odd periods. The enforcement commitment type plays T initially, and continues with T unless player 2 stops playing L in even periods and R in odd periods, at which point the enforcement type plays B thereafter.

It is player 2's fear of triggering the out-of-equilibrium behavior of the enforcement type that will prevent player 1 from building an effective reputation. The prior distribution puts probability 0.8 on the normal type and probability 0.1 on each of the other types. At the cost of a somewhat more complicated exposition, we could replace each of these commitment types with a payoff type (as in Schmidt (1993)), and derive the corresponding behavior as part of the equilibrium.

We describe a strategy profile with an outcome path that alternates between TL and TR, beginning with TL, and then verify it is an equilibrium for patient players. An automaton representation of the profile is given in Figure 6).

> *Normal type of player* 1: Play T after any history except one in which player 1 has at least once played B, in which case play B.

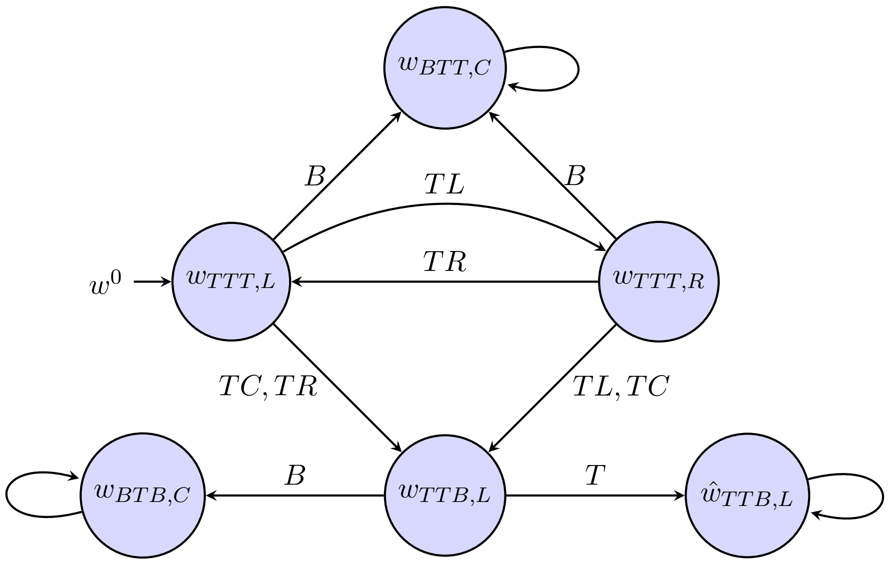


Figure 6: An automaton representation of the strategy profile illustrating the failure of reputation effects in Section 3.2. In state $w_{a_1^n a_1^s a_1^e, a_2}$ the normal type plays a_1^n, the Stackelberg type plays a_1^s (which is always equal to T), the enforcement type plays a_1^e, and player 2 plays a_2. Play begins in state $w^0 = w_{TTT,L}$. Transitions are determined by the observed actions $a_1 a_2$, and states $w_{BTB,C}$, $w_{BTT,C}$, and $\hat{w}_{TTB,L}$ are absorbing.

Player 2:

- After any history h^t featuring the play of TL in even periods and TR in odd periods, play L if t is even and R if t is odd. After any history h^t featuring the play of TL in even periods preceding $t-1$ and TR in odd periods preceding $t-1$, but in which player 1 plays B in period $t-1$, play C in period t and for every continuation history (since player 1 has revealed himself to be normal).
- After any history h^t in which h^{t-1} features the play of TL in even periods and TR in odd periods, and player 2 does not play L if $t-1$ is even or R if $t-1$ is odd, play L in period t. If 1 plays T in period t, play L in period $t+1$ and for every continuation history (interpreted as the result of attaching probability zero to the enforcement type and making no further belief revisions). If 1 plays B in period t, play C in period $t+1$ and for every continuation

history (interpreted as the result of attaching probability one to the enforcement type and making no further belief revisions).

The normal player 1's behavior depends only on his previous actions, featuring the constant play of T in equilibrium and with any switch to B triggering the subsequent constant play of B. The first item in the description of player 2's behavior describes her actions after histories in which 2 has made no deviations from equilibrium play, and the last item describes how player 2 behaves after she has deviated from equilibrium play.

We now argue that these strategies constitute an equilibrium for $z < 8$ and sufficiently large δ. First consider player 1. Along the equilibrium path, the normal type of player 1 earns a payoff that, for large δ is very close to $(10 - z)/2$ respectively. A deviation leads to a continuation payoff of 1. Hence, for any $z < 8$, there is a sufficiently large $\overline{\delta} < 1$ such that for $\delta \geq \overline{\delta}$, it is optimal for the normal player 1 to play T after every equilibrium history.

Should play ever leave the equilibrium path as a result of player 1's having chosen B, subsequent play after any continuation history constitutes a stage-game Nash equilibrium (BC) for player 2 and the normal type of player 1, and hence is optimal for player 1. Should play leave the equilibrium path because player 2 has deviated, then following the equilibrium strategies earns the player 1 a payoff of 10 in every subsequent period, and player 1 is thus playing optimally.

Now consider player 2. Along the equilibrium path, player 2 learns nothing about player 1. If player 2 deviates from the equilibrium, she has a chance to screen the types of player 1, earning a continuation payoff of 10 against the normal or Stackelberg type and a continuation payoff of at most 1 against the enforcement type. The resulting expected payoff is at most 9.1, falling short of the equilibrium payoff of almost 9.5 (for a patient player 2). This completes the argument that these strategies are an equilibrium under the conditions on discount factors that appear in the reputation result, namely that we fix δ_2 (allowed to be sufficiently large) and then let δ_1 approach 1.

By increasing the absolute value of z in Figure 5, while perhaps requiring more patience for player 1, we obtain an equilibrium in which the normal player 1's payoff is arbitrarily close to his pure minmax payoff of 1. It is thus apparent that reputation considerations can be quite ineffective.

In equilibrium, player 2 *is* convinced that she will face the Stackelberg action in every period. However, she dares not play a best response out of fear that doing so has adverse future consequences, a fear made real by the possibility of the enforcement type.[24]

Reputation arguments with two long-lived players thus cannot be simple extensions of long-lived/short-lived player results. Player 1 has the option of leading a long-lived player 2 to expect commitment behavior on the path of play, but this no longer suffices to ensure a best response from player 2, no matter how firm the belief. Despite a flurry of activity in the literature, the results are necessarily weaker and more specialized. In particular, the results for two long-lived players exploit some structure of the game or the setting to argue that player 2 will play a best response to a particular action, once convinced that player 1 is likely to play that action.

The remainder of this section illustrates the arguments and results that have appeared for the case of two long-lived players. As has the literature, we concentrate on the case in which there is uncertainty only about player 1's type, and in which player 2's discount factor is fixed while player 1's discount factor is allowed to become arbitrarily large. Hence, while we are making the players more symmetric in the sense of making both long-lived, we are still exploiting asymmetries between the players. There has been some work, such as Cripps, Dekel, and Pesendorfer (2005) and Atakan and Ekmekci (2012, 2013), which we will not discuss here, on games with two long-lived players who are equally patient, leaving only the one-sided incomplete information as the source of asymmetry.

As in the case of short-lived player 2's, Cripps, Mailath, and Samuelson (2004, 2007) establish conditions under which a long-lived player 1 facing a long-lived opponent in a game of imperfect monitoring can only maintain a temporary reputation. If player 1 is indeed normal, then with probability 1 player 2 must eventually attach arbitrarily high probability to 1's being normal. Once again, we are reminded that the ex ante payoff calculations that dominate the reputations literature may not be useful in characterizing long run behavior.

3.3 Minmax-Action Reputations

This section presents a slight extension of Schmidt (1993), illustrating a reputation effect in games with two long-lived players.

We consider a perfect monitoring repeated game with two long-lived players, 1 and 2, with finite action sets. Player 1's type is determined by a

[24] Celentani, Fudenberg, Levine, and Pesendorfer (1996, Section 5) describe an example with similar features, but involving only a normal and Stackelberg type of player 1, using the future play of the normal player 1 to punish player 2 for choosing a best response to the Stackelberg action.

probability distribution μ with a finite or countable support. The support of μ contains the normal type of player 1, ξ_0, and a collection of commitment types.

3.3.1 Minmax-Action Types

Let $\underline{v}_1(\xi_0, \delta_1, \delta_2)$ be the infimum, over the set of Nash equilibria of the repeated game, of the normal player 1's payoffs. For the action $\alpha_1 \in \Delta(A_1)$, let

$$v_1^*(\alpha_1) := \min_{a_2 \in B(\alpha_1)} u_1(\alpha_1, a_2),$$

where $B(\alpha_1)$ is the set of player 2 stage-game best responses to α_1.

The basic result is that if there exists a simple commitment type committed to an action $\hat{\alpha}_1$ minmaxing player 2, then the normal player 1, if sufficiently patient, is assured an equilibrium payoff close to $v_1^*(\hat{\alpha}_1)$:[25]

Proposition 3 *Suppose $\mu(\xi(\hat{\alpha}_1)) > 0$ for some action $\hat{\alpha}_1$ that minmaxes player 2. For any $\eta > 0$ and $\delta_2 < 1$, there exists a $\underline{\delta}_1 \in (0,1)$ such that for all $\delta_1 \in (\underline{\delta}_1, 1)$,*

$$\underline{v}_1(\xi_0, \delta_1, \delta_2) > v_1^*(\hat{\alpha}_1) - \eta.$$

Fix an action $\hat{\alpha}_1$ satisfying the criteria of the proposition. The key to establishing Proposition 3 is to characterize the behavior of player 2, on histories likely to arise when player 1 repeatedly plays $\hat{\alpha}_1$. Let $\hat{\Omega}$ be the event that player 1 always plays $\hat{\alpha}_1$. For any history h^t that arises with positive probability given $\hat{\Omega}$ (i.e., $\mathbf{P}\{\omega \in \hat{\Omega} : h^t(\omega) = h^t\} > 0$), let $E_{\mathbf{P}}[U_2(\sigma|_{h^t}) \mid \hat{\Omega}]$ be 2's expected continuation payoff, conditional on the history h^t and $\hat{\Omega}$. Let $\underline{v}_2$ be player 2's minmax payoff.

Lemma 7 *Fix $\delta_2 \in (0,1)$ and $\varepsilon > 0$. There exists $L \in \mathbb{N}$ and $\kappa > 0$ such that, for all Nash equilibria σ, pure strategies $\tilde{\sigma}_2$ satisfying $\tilde{\sigma}_2(\bar{h}^{t'}) \in \operatorname{supp} \sigma_2(\bar{h}^{t'})$ for all $\bar{h}^{t'} \in H^{t'}$, and histories $h^t \in H^t$ with positive probability under $\hat{\Omega}$, if*

$$E_{\mathbf{P}}[U_2((\sigma_1, \tilde{\sigma}_2)|_{h^t}) \mid \hat{\Omega}] \leq \underline{v}_2 - \varepsilon, \tag{14}$$

[25] An action $\hat{\alpha}_1$ minmaxes player 2 if

$$\hat{\alpha}_1 \in \arg\min_{\alpha_1 \in \Delta A_1} \max_{a_2 \in A_2} u_2(\alpha_1, a_2).$$

Schmidt (1993) considered the case in which the action $\hat{\alpha}_1$ is a pure action, i.e., that there is a pure action that mixed-action minmaxes player 2.

then there is a period $t+\tau$, $0 \le \tau \le L$, and a continuation history $h^{t+\tau}$ that has positive probability when $\hat{\alpha}_1$ is played in periods $t, t+1, \ldots, t+\tau-1$, and 2 plays $\tilde{\sigma}_2$, such that at $h^{t+\tau}$, player 2's predicted distribution of player 1's action α_1 in period $t+\tau$ satisfies $d(\hat{\alpha}_1 \| \alpha_1) > \kappa$.

Intuitively, condition (14) indicates that player 2's equilibrium strategy gives player 2 a payoff below her minmax payoff, conditional on $\hat{\alpha}_1$ always being played. But no equilibrium strategy can ever give player 2 an expected payoff lower than her minmax payoff, and hence it must be that player 2 does *not* expect $\hat{\alpha}_1$ to always be played.

Proof. Suppose, to the contrary, that for the sequence $(L_n, \kappa_n)_n$, where $L_n = n$ and $\kappa_n = n^{-1}$, there is a sequence of equilibria σ^n violating the result. Fix n, and let h^t be a history that occurs with positive probability under $\hat{\Omega}$ for which (14) holds, and such that in each of the next $L_n + 1$ periods, if player 1 has always played $\hat{\alpha}_1$ and player 2 follows $\tilde{\sigma}_2$, then player 2's predicted distribution of player 1's action α_1 satisfies $d(\hat{\alpha}_1 \| \alpha_1) \le \kappa_n$ after every positive probability history. We will derive a contradiction for sufficiently large n.

Given such posterior beliefs for player 2, an upper bound on player 2's period t expected continuation payoff is given by (where $u_2^{t+\tau}(\kappa_n)$ is player 2's expected payoff under the strategy profile $\tilde{\sigma}^n \equiv (\sigma_1^n, \tilde{\sigma}_2^n)$ in period $t+\tau$ when player 1 plays an action α_1 with $d(\hat{\alpha}_1 \| \alpha_1) \le \kappa$ in period $t+\tau$, and M is an upper bound on the magnitude of player 2's stage-game payoff),[26]

$$\begin{aligned} E_{\mathbf{P}}[U_2(\tilde{\sigma}^n|_{h^t}) \mid h^t] \le{}& (1-\delta_2)u_2^t(\kappa_n) + (1-\delta_2)\delta_2 u_2^{t+1}(\kappa_n) + (1-\delta_2)\delta_2^2 u_2^{t+2}(\kappa_n) \\ &+ \cdots + (1-\delta_2)\delta_2^{L_n} u_2^{t+L_n}(\kappa_n) + \delta_2^{L_n+1} M \\ ={}& (1-\delta_2)\sum_{\tau=0}^{L_n} \delta_2^{\tau} u_2^{t+\tau}(\kappa_n) + \delta_2^{L_n+1} M. \end{aligned}$$

For L_n sufficiently large and κ_n sufficiently small (i.e., n sufficiently large), the upper bound is within $\varepsilon/2$ of $E_{\mathbf{P}}[U_2(\tilde{\sigma}^n|_{h^t}) \mid \hat{\Omega}]$, i.e., it is close to player 2's expected continuation payoff conditioning on the event $\hat{\Omega}$. Hence, using (14), for large n

$$E_{\mathbf{P}}[U_2(\tilde{\sigma}^n|_{h^t}) \mid h^t] \le E_{\mathbf{P}}[U_2(\tilde{\sigma}^n|_{h^t}) \mid \hat{\Omega}] + \frac{\varepsilon}{2} < \underline{v}_2.$$

[26] The history h^t appears twice in the notation $E_{\mathbf{P}}[U_2(\tilde{\sigma}|_{h^t}) \mid h^t]$ for 2's expected continuation value given h^t. The subscript h^t reflects the role of h^t in determining player 2's continuation strategy. The second h^t reflects the role of the history h^t in determining the beliefs involved in calculating the expected payoff given that history.

But then player 2's continuation value at history h^t, $E_{\mathbf{P}}[U_2(\tilde{\sigma}^n|_{h^t}) \mid h^t]$, is strictly less than her minmax payoff, a contradiction. ■

Proof of Proposition 3. Fix a Nash equilibrium with player 1 payoff v_1. Fix δ_2 and η, and denote by $B_{\varepsilon'}(\hat{\alpha}_1)$ the ε'-neighborhood of $B(\hat{\alpha}_1)$. There exists $\varepsilon' > 0$ such that if $\sigma_2(h^t) \in B_{\varepsilon'}(\hat{\alpha}_1)$, then

$$u_1(\hat{\alpha}_1, \sigma_2(h^t)) > v_1^*(\hat{\alpha}_1) - \frac{\eta}{2}.$$

Let $\varepsilon := \underline{v}_2 - \max_{\alpha_2 \notin B_{\varepsilon'}(\hat{\alpha}_1)} u_2(\hat{\alpha}_1, \alpha_2) > 0$, and let L and κ be the corresponding values from Lemma 7. Lemma 7 implies that if the outcome of the game is contained in $\hat{\Omega}$, and player 2 fails to play ε'-close to a best response to $\hat{\alpha}_1$ in some period t, then there must be a period $t+\tau$, with $0 \le \tau \le L$ and a continuation history $h^{t+\tau}$ that has positive probability when $\hat{\alpha}_1$ is played in periods $t, t+1, \ldots, t+\tau-1$, and 2 plays $\tilde{\sigma}_2$, such that at $h^{t+\tau}$, player 2's one-step ahead prediction error when 1 plays $\hat{\alpha}_1$ is at least κ, i.e.,

$$d(\hat{p} \| p^{t+\tau}(h^{t+\tau})) \ge \kappa.$$

Define $\hat{\underline{\alpha}}_1 := \min\{\hat{\alpha}_1(a_1) : \hat{\alpha}_1(a_1) > 0\}$ and let $2M$ be the difference between the largest and smallest stage-game payoffs for player 1. We can bound the amount by which player 1's equilibrium payoff v_1 falls short of $v_1^*(\hat{\alpha}_1)$ as follows:

$$\begin{aligned}
v_1^*(\hat{\alpha}_1) - v_1 &\le 2M(1-\delta_1)\sum_{t=0}^{\infty} \delta_1^t E_{\widehat{\mathbf{P}}}[\mathbb{1}_{\{\tilde{\sigma}_2(h^t) \notin B_{\varepsilon'}(\hat{\alpha}_1)\}}] + \frac{\eta}{2} \\
&\le 2M(1-\delta_1)\frac{L+1}{\hat{\underline{\alpha}}_1^{L+1}\delta_1^L}\sum_{t=0}^{\infty} \delta_1^t E_{\widehat{\mathbf{P}}}[\mathbb{1}_{\{d(\hat{p}\|p^t(h^t)) > \kappa\}}] + \frac{\eta}{2} \\
&\le 2M(1-\delta_1)\frac{L+1}{\hat{\underline{\alpha}}_1^{L+1}\delta_1^L}\sum_{t=0}^{\infty} \delta_1^t \frac{1}{\kappa} E_{\widehat{\mathbf{P}}}[d(\hat{p}\|p(h^t))] + \frac{\eta}{2} \\
&\le 2M(1-\delta_1)\frac{L+1}{\hat{\underline{\alpha}}_1^{L+1}\delta_1^L}\frac{1}{\kappa}(-\log \mu(\xi(\hat{\alpha}_1))) + \frac{\eta}{2},
\end{aligned}$$

where

1. $\tilde{\sigma}_2$ in the first line is some pure strategy for 2 in the support of 2's equilibrium strategy,

2. the second inequality comes from the observation that player 2's failure to best respond to $\hat{\alpha}_1$ in different periods may be due to a belief that

1 will not play $\hat{\alpha}_1$ in the same future period after the same positive probability history (there can be at most $L+1$ such periods for each belief, and the probability of such a history is at least $\underline{\hat{\alpha}}_1^{L+1}$), and

3. the last inequality follows from (5).

Since the last term in the chain of inequalities can be made less than η for δ_1 sufficiently close to 1, the proposition is proved. ■

If there are multiple actions that minmax player 2, the relevant payoff bound corresponds to the maximum value of $v_1^*(\hat{\alpha}_1)$ over the set of such actions whose corresponding simple commitment types are assigned positive probability by player 2's prior.

3.3.2 Conflicting Interests

The strength of the bound on player 1's equilibrium payoffs depends on the nature of player 1's actions that minmax player 2. The highest reputation bound is obtained when there exists an action α_1^* that mixed action minmaxes player 2 and is also close to player 1's Stackelberg action, since the reputation bound can then be arbitrarily close to player 1's Stackelberg payoff. Restricting attention to pure Stackelberg actions and payoffs, Schmidt (1993) refers to such games as games of conflicting interests:

Definition 3 *The stage game has* conflicting interests *if a pure Stackelberg action a_1^* mixed-action minmaxes player* 2.

The payoff bound derived in Proposition 3 is not helpful in the product-choice game (Figure 1). The pure Stackelberg action H prompts a best response of c that earns player 2 a payoff of 3, above her minmax payoff of 1. The mixtures that allow player 1 to approach his mixed Stackelberg payoff of 5/2, involving a probability of H exceeding but close to 1/2, similarly prompt player 2 to choose c and earn a payoff larger than 1. The normal player 1 and player 2 thus both fare better when 1 chooses either the pure Stackelberg action or an analogous mixed action (and 2 best responds) than in the stage-game Nash equilibrium (which minmaxes player 2). This coincidence of interests precludes using the reasoning behind Proposition 3.

The prisoners' dilemma (Figure 7) is a game of conflicting interests. Player 1's pure Stackelberg action is D. Player 2's unique best response of D yields a payoff of $(0, 0)$, giving player 2 her minmax level. Proposition 3 then establishes conditions, including the presence of a player 1 type committed to D and sufficient patience on the part of player 1, under which the

	C	D
C	$2,2$	$-1,3$
D	$3,-1$	$0,0$

Figure 7: Prisoners' dilemma.

	In	Out
A	$2,2$	$5,0$
F	$-1,-1$	$5,0$

Figure 8: A simultaneous-move version of the chain-store game.

normal player 1 must earn nearly his Stackelberg payoff. In the prisoners' dilemma, however, this bound is not significant, being no improvement on the observation that player 1's payoff must be weakly individually rational.

In the normal form version of the chain store game (Figure 8), player 1 achieves his mixed Stackelberg payoff of 5 by playing the (pure) action F, prompting player 2 to choose Out and hence to receive her mixed minmax payoff of 0. We thus have a game of conflicting interests, in which (unlike the prisoners' dilemma) the reputation result has some impact. The lower bound on the normal player 1's payoff is then close to his Stackelberg payoff of 5, the highest player 1 payoff in the game.

Consider the game shown in Figure 9. The Nash equilibria of the stage game are TL, BC, and a mixed equilibrium $\left(\frac{1}{2} \circ T + \frac{1}{2} \circ B,\ \frac{2}{5} \circ L + \frac{3}{5} \circ C\right)$. Player 1 minmaxes player 2 by playing B, for a minmax value for player 2 of 0. The pure and mixed Stackelberg action for player 1 is T, against which player 2's best response is L, for payoffs of $(3,2)$. Proposition 3 accordingly

	L	C	R
T	$3,2$	$0,1$	$0,1$
B	$0,-1$	$2,0$	$0,-1$

Figure 9: A game without conflicting interests.

provides no reason to think it would be helpful for player 1 to commit to T. However, if the set of possible player 1 types includes a type committed to B (perhaps as well as a type committed to T), then Proposition 3 implies that (up to some $\eta > 0$) the normal player 1 must earn a payoff no less than 2. The presence of a type committed to B thus gives rise to a nontrivial reputation bound on player 1's payoff, though this bound falls short of his Stackelberg payoff.

3.4 Discussion

Our point of departure for studying reputations with two long-lived players was the observation that a long-lived player 2 might not play a best response to the Stackelberg action, even when convinced she will face the latter, for fear of triggering a punishment. If we are to achieve a reputation result for player 1, then there must be something in the particular application that assuages player 2's fear of punishment. Proposition 3 considers cases in which player 2 can do no better when facing the Stackelberg action than achieve her minmax payoff. No punishment can be worse for player 2 than being minmaxed, and hence no punishment type can coerce player 2 into choosing something other than a best response.

What if a commitment type does not minmax player 2? The following subsections sketch some of the alternative approaches to reputations with two long-lived players.

3.4.1 Weaker Payoff Bounds for More General Actions

For any action a_1', if player 2 puts positive prior probability on the commitment type $\xi(a_1')$, then when facing a steady stream of a_1', player 2 must eventually come to expect a_1'. If the action a_1' does not minmax player 2, we can no longer bound the number of periods in which player 2 is not best responding. We can, however, bound the number of times player 2 chooses an action that gives her a payoff less than her minmax value. This allows us to construct an argument analogous to that of Section 3.3, concluding that a lower bound on player 1's payoff is given by

$$v_1^{\dagger}(a_1') \equiv \min_{\alpha_2 \in \mathcal{D}(a_1')} u_1(a_1', \alpha_2),$$

where

$$\mathcal{D}(a_1') = \{\alpha_2 \in \Delta(A_2) \mid u_2(a_1', \alpha_2) \geq \underline{v}_2\}$$

is the set of player 2 actions that, in response to a_1', ensure that 2 receives at least her mixed minmax utility $\underline{v}_2$. In particular, Cripps, Schmidt, and

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

Figure 10: Battle of the sexes game.

Thomas (1996) show that if there exists $a_1' \in A_1$ with $\mu(\xi(a_1')) > 0$, then for any fixed $\delta_2 < 1$ and $\eta > 0$, then there exists a $\underline{\delta}_1 < 1$ such that for all $\delta_1 \in (\underline{\delta}_1, 1)$, $\underline{v}_1(\xi_0, \delta_1, \delta_2) \geq v_1^\dagger(a_1) - \eta$. This gives us a payoff bound that is applicable to any pure action for which there is a corresponding simple commitment type, though this bound will typically be lower than the Stackelberg payoff.

To illustrate this result, consider the battle of the sexes game in Figure 10. The minmax utility for player 2 is 3/4. Player 1's Stackelberg action is T, which does not minmax player 2, and hence this is not a game of conflicting interests.

The set of responses to B in which player 2 receives at least her minmax payoff is the set of actions that place at least probability 1/4 on L. Hence, if 2 assigns positive probability to a simple player 1 type committed to B, then we have a lower bound on 1's payoff of 1/4. This bound falls short of player 1's minmax payoff, and hence is not very informative. The set of responses to T in which player 2 receives at least her minmax payoff is the set of actions that place at least probability 3/4 on R. Hence, if 2 assigns positive probability to a simple player 1 type committed to T, then we have a lower bound on 1's payoff of 9/4. This bound falls short of player 1's Stackelberg payoff, but nonetheless gives us a higher bound that would appear in the corresponding game of complete information.

3.4.2 Imperfect Monitoring

The difficulty in Section 3.2 is that player 2 frequently plays an action that is not her best response to 1's Stackelberg action, in fear that playing the best response will push the game off the equilibrium path into a continuation phase where she is punished. The normal and Stackelberg types of player 1 would not impose such a punishment, but there is another punishment type who would. Along the equilibrium path player 2 has no opportunity to discern whether she is facing the normal type or the punishment type.

Celentani, Fudenberg, Levine, and Pesendorfer (1996) observe that in games of imperfect public monitoring, the sharp distinction between being on and off the equilibrium path disappears. Player 2 may then have ample opportunity to become well acquainted with player 1's behavior, including any punishment possibilities.

Consider a game with two long-lived players and finite action sets A_1 and A_2. Suppose the public monitoring distribution ρ has full support, i.e., there is a set of public signals Y with the property that for all $y \in Y$ and $a \in A_1 \times A_2$, there is positive probability $\rho(y \mid a) > 0$ of observing signal y when action profile a is played. Suppose also that Assumption 2 of Section 2.6 holds. Recall that this is an identification condition. Conditional on player 2's mixed action, different mixed actions on the part of player 1 generate different signal distributions. Given arbitrarily large amounts of data, player 2 could then distinguish player 1's actions.

It is important that player 2's actions be imperfectly monitored by player 1, so that a sufficiently wide range of player 1 behavior occurs in equilibrium. It is also important that player 2 be able to update her beliefs about the type of player 1, in response to the behavior she observes. Full-support public monitoring delivers the first condition, while the identification condition (Assumption 2) ensures the second.

We now allow player 1 to be committed to a strategy that is not simple. In the prisoners' dilemma, for example, player 1 may be committed to playing tit-for-tat rather than either always cooperating or always defecting. When player 2 is short-lived, the reputation bound on player 1's payoff cannot be improved by appealing to commitment types that are not simple. A short-lived player 2 cares only about the action she faces in the period she is active, whether this comes from a simple or more complicated commitment type. Any behavior one could hope to elicit from a short-lived player 2 can then be elicited by having her attach sufficient probability to the appropriate simple commitment type. The result described here is the first of two illustrations of how, when player 2 is long-lived, non-simple commitment types can increase the reputation bound on player 1's payoffs.

The presence of more complicated commitment types allows a stronger reputation result, but also complicates the argument. In particular, we can no longer define a lower bound on player 1's payoff simply in terms of the stage game. Instead, we first consider a finitely repeated game, of length N, in which player 1 does not discount. It will be apparent from the construction that this lack of player 1 discounting simplifies the calculations, but otherwise does not play a role. We fix a pure commitment type σ_1^N for player 1 in this finitely repeated game, and then calculate player 1's average payoff, assuming that player 1 plays the commitment strategy and player 2 plays a best response to this commitment type. Denote this payoff by $v_1^*(\sigma_1^N)$.

Now consider the infinitely repeated game, and suppose that the set of commitment types includes a type who plays σ_1^N in the first N periods, then acts as if the game has started anew and again plays σ_1^N in the next N periods, and then acts as if the game has started anew, and so on. Celentani, Fudenberg, Levine, and Pesendorfer (1996) establish a lower bound on player 1's payoff in the infinitely repeated game that approaches $v_1^*(\sigma_1^N)$ as player 1 becomes arbitrarily patient. Of course, one can take N to be as large as one would like, and can take the corresponding strategy σ_1^N to be any strategy in the N-length game, as long as one is willing to assume that the corresponding commitment type receives positive probability in the infinitely repeated game. By choosing this commitment type appropriately, one can create a lower bound on player 1's payoff that may well exceed the Stackelberg payoff of the stage game. In this sense, and in contrast to the perfect monitoring results of Sections 3.3–3.4.1, facing a long-lived opponent can strengthen reputation results.[27]

This argument embodies two innovations. The first is the use of imperfect monitoring to ensure that player 2 is not terrorized by "hidden punishments." The second is the admission of more complicated commitment types, with associated improved bounds on player 1's payoff. It thus becomes all the more imperative to consider the interpretation of the commitment types that appear in the model.

There are two common views of commitment types. One is to work with commitment types that are especially natural in the setting in question. The initial appeal to commitment types, by Kreps and Wilson (1982) and Milgrom and Roberts (1982), in the context of the chain store paradox, was motivated by the presumption that entrants might be especially concerned with the possibility that the incumbent is pathologically committed to fighting. Alternatively, player 2 may have no particular idea as to what commitment types are likely, but may attach positive probability to a wide range of types, sufficiently rich that some are quite close to the Stackelberg type. Both motivations may be less obvious once one moves beyond simple commitment types. It may be more difficult to think of quite complicated commitment types as natural, and the set of such types is sufficiently large

[27] Aoyagi (1996) presents a similar analysis, with trembles instead of imperfect monitoring blurring the distinction between play on and off the equilibrium path, and with player 1 infinitely patient while player 2 discounts.

that it may be less obvious to assume that player 2's prior distribution over this set is sufficiently dispersed as to include types arbitrarily close to any particular commitment type. If so, the results built on the presence of more complicated commitment types must be interpreted with care.

3.4.3 Punishing Commitment Types

This section illustrates more starkly the effects of appropriately-chosen commitment types. Fix an action $a_1' \in A_1$, and suppose that player 2's stage-game best response a_2' satisfies $u_2(a_1', a_2') > \underline{v}_2^p := \min_{a_1} \max_{a_2} u_2(a_1, a_2)$, so that 2's best response to a_1' gives her more than her pure action minmax payoff. Let $\hat{a}_1^2$ be the action for player 1 that (pure-action) minmaxes player 2.

Consider a commitment type for player 1 who plays as follows. Play begins in phase 0. In general, phase k consists of k periods of $\hat{a}_1^2$, followed by the play of a_1'. The initial k periods of $\hat{a}_1^2$ are played regardless of any actions that player 2 takes during these periods. The length of time that phase k plays a_1', and hence the length of phase k itself, depends on player 2's actions, and this phase may never end. The rule for terminating a phase is that if player 2 plays anything other than a_2' in periods $k+1, \ldots$ of phase k, then the strategy switches to phase $k+1$.

We can interpret phase k of player 1's strategy as beginning with a punishment, in the form of k periods of minmaxing player 2, after which play switches to (a_1', a_2'). The beginning of a new phase, and hence of a new punishment, is prompted by player 2's *not* playing a_2' when called upon to do so. The commitment type thus punishes player 2, in strings of ever-longer punishments, for not playing a_2'.

Let $\hat{\sigma}_1(a_1')$ denote this strategy, with the a_1' identifying the action played by the commitment type when not punishing player 2. Evans and Thomas (1997) show that for any $\eta > 0$, if player 2 attaches positive probability to the commitment type $\hat{\sigma}_1(a_1')$, for some action profile a' with $u_2(a') > \underline{v}_2^p$, then there exists a $\underline{\delta}_2 \leq 1$ such that for all $\delta_2 \in (\underline{\delta}_2, 1)$, there in turn exists a $\underline{\delta}_1$ such that for all $\delta_1 \in (\underline{\delta}_1, 1)$, player 1's equilibrium payoff is at least $u_1(a') - \eta$. If there exists such a commitment type for player 1's Stackelberg action, then this result gives us player 1's Stackelberg payoff as an approximate lower bound on his equilibrium payoff in the game of incomplete information, as long as this Stackelberg payoff is consistent with player 2 earning more than her pure-strategy minmax.

This technique leads to reputation bounds that can be considerably higher than player 1's Stackelberg payoff. Somewhat more complicated commitment types can be constructed in which the commitment type $\hat{\sigma}_1$ plays a sequence of actions during its nonpunishment periods, rather than simply playing a fixed action a_1, and punishes player 2 for not playing an appropriate sequence in response. Using such constructions, Evans and Thomas (1997, Theorem 1) establish that the limiting lower bound on player 1's payoff, in the limit as the players become arbitrarily patient, with player 1 arbitrarily patient relative to player 2, is given by $\max\{v_1 : v \in \mathcal{F}^{\dagger p}\}$ (where $\mathcal{F}^{\dagger p}$ is that portion of the convex hull of the pure stage-game payoff profiles in which player 2 receives at least her pure minmax payoff). Hence, player 1 can be assured of the largest feasible payoff consistent with player 2's individual rationality.[28]

Like the other reputation results we have presented, this requires the uncertainty about player 1 contain appropriate commitment types. As in section 3.4.2, however, the types in this case are more complicated than the commitment types that appear in many reputation models, particularly the simple types that suffice with short-lived player 2s.[29] In this case, the commitment type not only repeatedly plays the action that brings player 1 the desired payoff, but also consistently punishes player 2 for not fulfilling her role in producing that payoff. The commitment involves behavior both on the path of a proposed outcome and on paths following deviations. Once again, work on reputations would be well served by a better-developed model of which commitment types are likely.

4 Persistent Reputations

In this section, we consider recent work that modifies the basic reputation model to obtain convenient characterizations of nontrivial long run behav-

[28] The idea behind the argument is to construct a commitment type consisting of a phase in which payoffs at least $(\max\{v_1 : v \in \mathcal{F}^{\dagger p}\} - \varepsilon, \bar{v}_2^p + \varepsilon)$ are received, as long as player 2 behaves appropriately, with inappropriate behavior triggering ever-longer punishment phases. Conditional on seeing behavior consistent with the commitment type, a sufficiently patient player 2 must then eventually find it optimal to play the appropriate response to the commitment type. For fixed, sufficiently patient player 2, making player 1 arbitrarily patient then gives the result.

[29] As Evans and Thomas (1997) note, a commitment type with punishments of arbitrary length cannot be implemented by a finite automaton. Hence, one cannot ensure that such commitment types appear in the support of player 2's prior distribution simply by assuming that this support includes the set of all strategies implementable by finite automata. Evans and Thomas (2001), working with two infinitely patient long-lived players, argue that commitment strategies capable of imposing arbitrarily severe punishments are necessary if reputation arguments are to be effective in restricting attention to efficient payoff profiles.

ior (including cycles of reputation building and destruction) and persistent reputations. We consider three classes of models. To organize our ideas, note that the fundamental force behind the disappearing reputations result of Section 2.6 is that a Bayesian with access to an unlimited number of informative signals will eventually learn a fixed parameter. The work examined in this section relaxes each element in this statement (though not all of this work was motivated by the disappearing reputations result). Section 4.1 examines models in which the short-lived players have access to only a limited amount of data. Section 4.2 considers the case in which the short-lived players are not perfect Bayesians, working instead with a misspecified model of the long-lived player's behavior. Section 4.3 considers models in which the short-lived players must learn a moving target.

In each case, reputation considerations persist indefinitely. Notice, however, that the focus in this literature, especially the work discussed in Section Section 4.3, shifts from identifying bounds on the payoffs of all equilibria to characterizing particular (persistent reputation) equilibria.

4.1 Limited Observability

We begin with a model in which player 2 cannot observe the entire history of play. Our presentation is based on a particularly simple special case of a model due to Liu (2011). Our point of departure is the product-choice game of Figure 1. Player 1 is either a normal type, whose payoffs are those of the product-choice game, or is a commitment type referred to as the "good" type, who invariably plays H. The innovation is that when constructing the associated repeated game, we assume that each successive player 2 can observe only the action taken by player 1 in the previous period. The observation of the previous-period action is perfect.

For $\delta > \frac{1}{2}$, we construct an equilibrium characterized by two states, ω^L and ω^H, determined by the action played by player 1 in the previous period. The play of the normal player 1 and player 2 is described by the automaton in Figure 11. Intuitively, we think of player 2 as being exploited in state ω^H, while player 1 builds the reputation that allows such exploitation in state ω^L.

It remains to confirm that these strategies constitute an equilibrium. A player 2 who observes L in the preceding period knows that she faces the normal player 1, who will mix equally between H and L in the next period. Player 2 is then indifferent between c and s, and hence 2's specified mixture is optimal. A player 2 who observes either nothing (in the case of period 0) or H in the previous period attaches a probability to player 1 being good

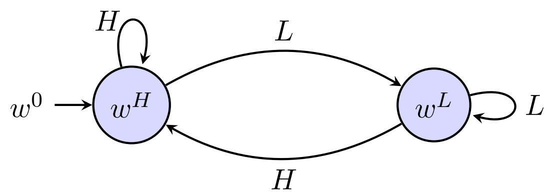


Figure 11: The play of the normal player 1 and player 2. In state ω^H, the normal player 1 plays L and player 2 plays c. In state ω^L, players randomize, with the normal player 1 playing $\frac{1}{2} \circ L + \frac{1}{2} \circ H$ and player 2 playing $p \circ c + (1-p) \circ s$, where $p := (2\delta - 1)/(2\delta)$. The transitions between states are determined by the action a_1 of player 1 in the previous period, with a_1 leading to ω^{a_1}. Play begins in ω^H.

that is at least as high (but no higher in the case of period 0) as the prior probability that player 1 is good. Player 2's actions are thus optimal if and only if the prior probability of a good type is at least 1/2. We will assume this is the case, though note that this takes us away from the common convention in reputation models that the probabilities of the commitment types can be taken to be small. Liu's (2011) more general analysis does not require such a bound on the prior.

We now turn to the normal player 1's incentives. Let $V^L(a_1)$ be player 1's payoff when the current state is ω^L and player 1 plays a_1, with $V^H(a_1)$ being the analogous payoff at the current state ω^H. Since player 1 is supposed to mix between H and L at state ω^L, he must be indifferent between the two actions, and so we must have $V^L(H) = V^L(L)$.

The values satisfy

$$\begin{aligned} V^L(H) &= (1-\delta)2p + \delta V^H(L), \\ V^L(L) &= (1-\delta)(2p+1) + \delta V^L(L), \\ V^H(L) &= (1-\delta)3 + \delta V^L(L), \\ \text{and} \qquad V^H(H) &= (1-\delta)2 + \delta V^H(L). \end{aligned}$$

Solving the second equation yields

$$V^L(L) = 2p + 1,$$

and so

$$V^H(L) = (1-\delta)3 + \delta(2p+1)$$

$$\text{and} \qquad V^L(H) = (1-\delta)2p + \delta[(1-\delta)3 + \delta(2p+1)].$$

We now choose p so that player 1 is indeed willing to mix at state ω^L, i.e., so that $V^L(H) = V^L(L)$. Solving for p gives

$$p := \frac{2\delta - 1}{2\delta}.$$

Finally, it remains to verify that it is optimal for player 1 to play L at state ω^H. It is a straightforward calculation that the normal player 1 is indifferent between L and H at ω^H, confirming that the proposed strategies are an equilibrium for $\delta \in (\frac{1}{2}, 1)$.

In this equilibrium, player 1 continually builds a reputation, only to spend this reputation once it appears. Player 2 understands that the normal player 1 behaves in this way, with player 2 falling prey to the periodic exploitation because the prior probability of the good type is sufficiently high, and player 2 receives too little information to learn player 1's actual type.

Liu and Skrzypacz (2011) consider the following variation. Once again a long-lived player faces a succession of short-lived players. Each player has a continuum of actions, consisting of the unit interval, but the game gives rise to incentives reminiscent of the product-choice game. In particular, it is a dominant strategy in the stage game for player 1 to choose action 0. Player 2's best response is increasing in player 1's action. If player 2 plays a best response to player 1's action, then 1's payoff is increasing in his action. To emphasize the similarities, we interpret player 1's action as a level of quality to produce, and player 2's action as a level of customization in the product she purchases. Player 1 may be rational, or may be a (single) commitment type who always chooses some fixed quality $q \in (0, 1]$. The short-lived players can observe only actions taken by the long-lived player, and can only observe such actions in the last K periods, for some finite K.

Liu and Skrzypacz (2011) examine equilibria in which the normal long-lived player invariably chooses either quality q (mimicking the commitment type) or quality 0 (the "most opportunistic" quality level). After any history in which the short lived-player has received K observations of q and no observations of 0 (or has only observed q, for the first $K-1$ short-lived players), the long-lived player chooses quality 0, effectively burning his reputation. This pushes the players into a reputation-building stage, characterized by the property that the short-lived players have observed at least one quality level 0 in the last K periods. During this phase the long-lived player mixes between 0 and q, until achieving a string of K straight q observations. His reputation has then been restored, only to be promptly burned. Liu and Skrzypacz (2011) establish that as long as the record length K exceeds a finite lower bound, then the limiting payoff as $\delta \to 1$ is given by the Stackelberg payoff. Moreover, they show that this bound holds after every history, giving rise to reputations that fluctuate, but are long lasting.

Ekmekci (2011) establishes a different persistent reputation result in a version of the repeated product-choice game. As usual, player 1 is a long-lived player, facing a succession of short-lived player 2's. The innovation in the model is in the monitoring structure. Ekmekci (2011) begins with the repeated product-choice game with imperfect public monitoring, and then assumes that the public signals are observed only by a mediator, described as a ratings agency. On the basis of these signals, the ratings agency announces one of a finite number of ratings to the players. The short-lived players see only the most recently announced rating, thus barring access to the information they would need to identify the long-lived player's type.

If the game is one of complete information, so that player 1 is known to be normal, then player 1's payoff with the ratings agency can be no higher than the upper bound that applies to the ordinary repeated imperfect monitoring game. However, if player 1 might also be a Stackelberg type, then there is an equilibrium in which player 1's payoff is close (arbitrarily close, for a sufficiently patient player 1) to the Stackelberg payoff after *every* history. The fact that this payoff bound holds after every history ensures that reputation effects are permanent. If the appropriate (mixed) Stackelberg type is present, then player 1's payoff may exceed the upper bound applicable in the game of complete information. Reputation effects can thus permanently expand the upper bound on 1's payoff.

In equilibrium, high ratings serve as a signal that short-lived players should buy the custom product, low ratings as a signal that the long-lived player is being punished and short-lived players should buy the standard product. The prospect of punishment creates the incentives for the long-lived player to exert high effort, and the long-lived player exerts high effort in any non-punishment period. Punishments occur, but only rarely. Short-lived players observing a rating consistent with purchasing the custom object do not have enough information to determine whether they are facing the Stackelberg type or a normal type who is currently not being punished.

4.2 Analogical Reasoning

This section considers a model, from Jehiel and Samuelson (2012), in which the inferences drawn by the short-lived players are constrained not by a lack of information, but by the fact that they have a misspecified model of the long-lived players' behavior. As usual, we consider a long-lived player 1 who faces a sequence of short-lived player 2's. It is a standard result in repeated games that the discount factor δ can be equivalently interpreted as reflecting either patience or a continuation probability, but in this case we specifically assume that conditional upon reaching period t, there is a probability $1-\delta$ that the game stops at t and probability δ that it continues.

At the beginning of the game, the long-lived player is chosen to either be normal or to be one of K simple commitment types. Player 2 assumes the normal player 1 chooses in each period according to a mixed action α_0, and that commitment type $k \in \{1, 2, \dots, K\}$ plays mixed action α_k. Player 2 is correct about the commitment types, but in general is not correct about the normal type, and this is the sense in which player 2's model is misspecified. In each period t, player 2 observes the history of play h^t (though it would suffice for player 2 to observe only the frequencies with which player 1 has played his past actions), and then updates her belief about the type of player 1 she is facing. Player 2 then chooses a best response to the expected mixed action she faces, and so chooses a best response to

$$\sum_{k=0}^{K} \mu_k(h^t)\alpha_k,$$

where $\mu_0(h^t)$ is the posterior probability of the normal type after history h^t and $\mu_k(h^t)$ (for $k \in \{1, 2, \dots, K\}$) is the posterior probability of the k^{th} commitment type.

The normal long-lived player 1 chooses a best response to player 2's strategy. The resulting strategy profile is an equilibrium if it satisfies a consistency requirement. To formulate the latter, let σ_1 and σ_2 denote the strategies of the normal player 1 and of the short-lived players 2. We denote by $\mathbf{P}^\sigma(h)$ the resulting unconditional probability that history h is reached under $\sigma := (\sigma_1, \sigma_2)$ (taking into account the probability of breakdown after each period). We then define

$$A_0 := \frac{\sum_h \mathbf{P}^\sigma(h)\sigma_1(h)}{\sum_h \mathbf{P}^\sigma(h)}.$$

We interpret A_0 as the empirical frequency of player 1's actions.[30] The

[30]The assumption that δ reflects a termination risk plays a role in this interpretation. The denominator does not equal 1 because $\mathbf{P}^\sigma(h)$ is the probability that h appears as either a terminal history, or as the initial segment of a longer history.

consistency requirement is then that

$$A_0 = \alpha_0.$$

The resulting equilibrium notion is a *sequential analogy-based expectations equilibrium* (Jehiel (2005) and Ettinger and Jehiel (2010)).

We interpret the consistency requirement on player 2's beliefs as the steady-state result of a learning process. We think of the repeated game as itself played repeatedly, though by different players in each case. At the end of each game, a record is made of the frequency with which player 1 (and perhaps player 2 as well, but this is unnecessary) has played his various actions. This in turn is incorporated into a running record listing the frequencies of player 1's actions. A short-lived player forms expectations of equilibrium play by consulting this record. In general, each repetition of the repeated game will leave its mark on the record, leading to somewhat different frequencies for the various player 1 actions in the record. This in turn will induce different behavior in the next game. However, we suppose that the record has converged, giving a steady state in which the expected frequency of player 1 play in each subsequent game matches the recorded frequency. Hence, empirical frequencies recorded in the record will match $\alpha_0, \alpha_1, \ldots, \alpha_K$, leading to the steady state captured by the sequential analogy-based expectations equilibrium.

We illustrate this model in the context of the familiar product-choice game of Figure 1. Let p^* $(= 1/2)$ be the probability of H that makes player 2 indifferent between c and s. We assume we have at least one type playing H with probability greater than p^* and one playing H with probability less than p^*. The sequential analogy-based equilibria of the product-choice game share a common structure, which we now describe.

We begin with a characterization of the short-lived players' best responses. For history h, let $n_H(h)$ be the number of times action H has been played and let $n_L(h)$ be the number of times action L has been played. The short-lived player's posterior beliefs after history h depend only on the "state" $(n_L(h), n_H(h))$, and not on the order in which the various actions have appeared.

Whenever player 2 observes H, her beliefs about player 1's type shift (in the sense of first-order stochastic dominance) toward types that are more likely to play H, with the reverse holding for an observation of L. Hence, for any given number n_L of L actions, the probability attached by player 2 to player 1 playing H is larger, the larger the number n_H of H actions. Player 2 will then play c if and only if she has observed enough H actions. More precisely, there exists an increasing function $N_H : \{0, 1, 2, \ldots\} \to \mathbb{R}$, such that for every history h, at the resulting state $(n_L, n_H) = (n_L(h), n_H(h))$,

- player 2 plays c if $n_H > N_H(n_L)$ and
- player 2 plays s if $n_H < N_H(n_L)$.

We now describe the equilibrium behavior of player 1, which also depends only on the state (n_L, n_H). There exists a function $\underline{N}_H(n_L) \leq N_H(n_L)$ such that, for sufficiently large δ and for any history h, at the resulting state $(n_L, n_H) = (n_L(h), n_H(h))$,

- player 1 plays L if $n_H > N_H(n_L + 1)$,
- player 1 plays H if $\underline{N}_H(n_L) < n_H < N_H(n_L + 1)$,
- player 1 plays L if $n_H < \underline{N}_H(n_L)$, and
- $\lim_{\delta \to 1} \underline{N}_H(n_L) < 0$ for all n_L.

Figure 12 illustrates these strategies. Note that whenever player 1 chooses H, player 2 places higher posterior probability on types that play H. Doing so will eventually induce player 2 to choose c. Player 1 is building his reputation by playing H, and then enjoying the fruits of that reputation when player 2 plays c. However, it is costly for player 1 to play H when player 2 plays s. If the number of H plays required to build a reputation is too large, player 1 may surrender all thoughts of a reputation and settle for the continual play of Ls. The cost of building a reputation depends on how patient is player 1, and a sufficiently patient player 1 inevitably builds a reputation. This accounts for the last two items in the description of player 1's best response.

If player 1 encounters a state above $\underline{N}_H$, whether at the beginning of the game or after some nontrivial history, player 1 will choose H often enough to push player 2's beliefs to the point that c is a best response. Once player 1 has induced player 2 to choose c, 1 ensures that 2 thereafter always plays c. The state never subsequently crosses the border $N_H(n_L)$. Instead, whenever the state comes to the brink of this border, 1 drives the state away with a play of H before 2 has a chance to play s.

The functions illustrated in Figure 12 depend on the distribution of possible types for player 1 and on the payoffs of the game. Depending on the shapes of these functions, the best responses we have just described combine to create three possibilities for equilibrium behavior in the product-choice

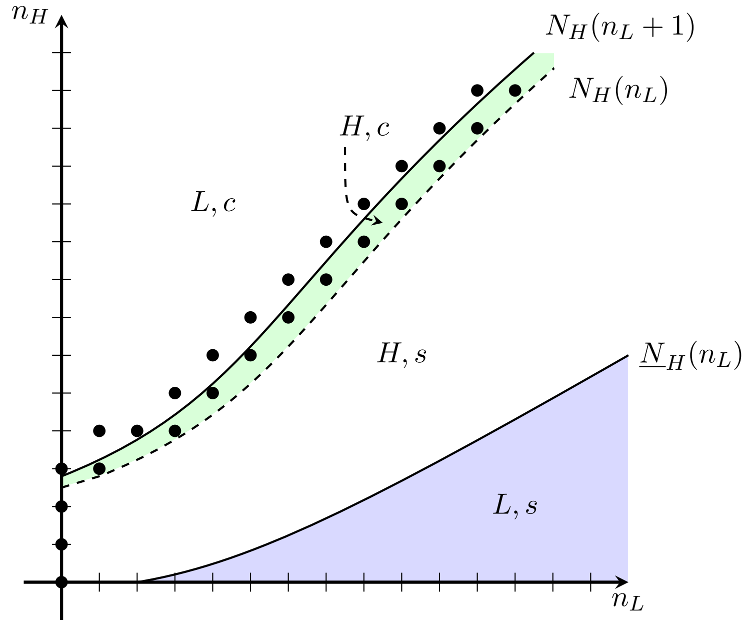


Figure 12: Behavior in the product-choice game. The state space (n_L, n_H) is divided into four regions with the players behaving as indicated in each of the four regions. The equilibrium path of play is illustrated by the succession of dots, beginning at the origin and then climbing the vertical axis in response to an initial string of H actions from player 1, with player 1 then choosing H and L so as to induce player 2 to always choose c, but to always be just at the edge of indifference between c and s.

game, that correspond to three possibilities for the intercepts of the functions $\underline{N}_H$ and N_H in Figure 12. First, it may be that $\underline{N}_H(0) > 0$. In this case, the equilibrium outcome is that the normal player 1 chooses L and player 2 chooses s in every period. Player 1 thus abandons any hope of building a reputation, settling instead for the perpetual play of the stage-game Nash equilibrium Ls. This is potentially optimal because building a reputation is costly. If player 1 is sufficiently impatient, this current cost will outweigh any future benefits of reputation building, and player 1 will indeed forego reputation building. By the same token, this will not be an equilibrium if δ is sufficiently large (but still less then one).

Second, it may be that $\underline{N}_H < 0 < N_H$, as in Figure 12. In this case, play begins with a reputation-building stage, in which player 1 chooses H

and the outcome is Hs. This continues until player 2 finds c a best response (until the state has climbed above the function N_H). Thereafter, we have a reputation-manipulation stage in which player 1 sometimes chooses L and sometimes H, selecting the latter just often enough to keep player 2 always playing c.

Alternatively, if $N_H(0) < 0$, then play begins with a reputation-spending phase in which player 1 chooses L, with outcome Lc, in the process shifting player 2's beliefs towards types that play L. This continues until player 2 is just on the verge of no longer finding c a best response (intuitively, until the state just threatens to cross the function N_H). Thereafter, we again have a reputation-manipulation stage in which player 1 sometimes chooses L and sometimes H, again selecting the latter just often enough to keep player 2 always playing c.

Which of these will be the case? For sufficiently patient players, whether one starts with a reputation-building or reputation-spending phase depends on the distribution of commitment types. If player 2's best response conditional on player 1 being a commitment type is s, then the rational player 1 must start with a reputation building phase, and we have the first of the preceding possibilities. Alternatively, If player 2's best response conditional on player 1 being a commitment type is c, then the rational player 1 must start with a reputation spending phase, and we have the second of the preceding possibilities.[31] In either case, player 2 remains perpetually uncertain as to which type of agent she faces, with her misspecified model of player 1 obviating the arguments of Cripps, Mailath, and Samuelson (2004, 2007). Player 1's equilibrium payoff approaches (as δ gets large) the mixed Stackelberg payoff, even though there may be no commitment type close to the mixed Stackelberg type.

4.3 Changing Types

Perhaps the most obvious route to a model in which reputation considerations persist is to assume that player 1's type is not fixed once-and-for-all at the beginning of the game, but is subject to persistent shocks. Intuitively, one's response upon having an uncharacteristically disappointing experience at a restaurant may be not "my previous experiences must have been atypical, and hence I should revise my posterior," but rather "perhaps something

[31] This characterization initially sounds obvious, but it is not immediate that (for example) player 2 will open the game by playing s if her best response to the commitment types is s, since 2's initial best response is an equilibrium phenomenon that also depends on the normal player 1's play. Jehiel and Samuelson (2012) fill in the details of the argument.

has changed." In this case, opponents will never be completely certain of a player's type. At first glance, it would seem that this can only reinforce the temporary-reputation arguments of Cripps, Mailath, and Samuelson (2004, 2007), making it more difficult to find persistent reputation effects. However, this very transience of reputations creates incentives to continually invest in reputation building, in order to assure opponents that one's type has not changed, opening the door for persistent reputations.

This section sketches several recent papers built around the assumption that types are uncertain but not permanently fixed. A critical assumption in all the models described in this section is that the uninformed agents do not observe whether the replacement event has occurred. These papers in turn build on a number of predecessors, such as Holmström (1982), Cole, Dow, and English (1995), and Mailath and Samuelson (2001), in which the type of the long-lived player is governed by a stochastic process.

4.3.1 Cyclic Reputations

We first consider, as a special case adapted from Phelan (2006), the product-choice game under perfect monitoring of actions, assuming that player 1 is constantly vulnerable to having his type drawn anew. The probability of such a replacement in each period is fixed, and we consider the limit in which the discount factor approaches 1.

Player 1 is initially drawn to be either a normal type or the pure Stackelberg type, who always plays H. In each period, player 1 continues to the next period with probability $\lambda > 1/2$, and is replaced by a new player 1 with probability $1 - \lambda$. In the event player 1 is replaced, player 1 is drawn to be the commitment type with probability $\hat{\mu}$, and with complementary probability is the normal type. To simplify the presentation, we assume that player 1 is initially drawn to be the commitment type with probability $(1 - \lambda)\hat{\mu}$ (see (17)).

Consider first the trigger profile: the normal player 1 always plays H if H has always been observed, plays L if ever L had been played, and player 2 plays c as long as H has been played, and plays s if ever L had been played. This profile is not an equilibrium under replacement, since after the punishment has been triggered by a play of L, the first play of H leads player 2 to believe that the normal type has been replaced by the commitment type and so 2's best response in the next period is to play c. But then the normal player 1 plays H immediately L, destroying the optimality of the punishment phase.

We construct an equilibrium when $2\delta\lambda > 1$. Let μ^t be the period t

posterior probability that player 2 attaches to the event that player 1 is the commitment type. In each period t in which μ^t is less than or equal to $\frac{1}{2}$, the normal player 1 plays H with probability

$$\alpha_1(\mu^t) := \frac{1-2\mu^t}{2(1-\mu^t)}.$$

This implies that, in such a period, player 2 faces the mixed action $\frac{1}{2} \circ H + \frac{1}{2} \circ L$, since

$$\mu^t + (1-\mu^t)\frac{1-2\mu^t}{2(1-\mu^t)} = \frac{1}{2}, \tag{15}$$

and hence is indifferent between c and s. When $\mu^t \leq \frac{1}{2}$, player 2 mixes, putting probability $\alpha_2(\mu^t)$ (to be calculated) on c, with player 2's indifference ensuring that this behavior is a best response. For values $\mu^t > \frac{1}{2}$, player 2 plays c, again a best response, and the normal player 1 chooses L. Notice that the actions of the customer and the normal player 1 depend only on the posterior probability that player 1 is the Stackelberg type, giving us a profile in Markov strategies.

It remains only to determine the mixtures chosen by player 2, which are designed so that the normal player 1 is behaving optimally. Let $\varphi(\mu \mid H)$ be the posterior probability attached to player 1 being the commitment type, given a prior $\mu \leq \frac{1}{2}$ and an observation of H. If the normal type chooses H with probability $\alpha_1(\mu)$, we have

$$\varphi(\mu \mid H) = \lambda\frac{\mu}{\mu+(1-\mu)\alpha_1(\mu)} + (1-\lambda)\hat{\mu} = 2\lambda\mu + (1-\lambda)\hat{\mu}, \tag{16}$$

using (15) for the second equality, while the corresponding calculation for L is

$$\varphi(\mu \mid L) = (1-\lambda)\hat{\mu}. \tag{17}$$

Let $\mu(0), \mu(1), \ldots, \mu(N)$ be the sequence of posterior probabilities satisfying $\mu(0) = (1-\lambda)\hat{\mu}$ and $\mu(k) = \varphi(\mu(k-1) \mid H)$ for $(k = 1, \ldots, N)$, with $\mu(N)$ being the first such probability to equal or exceed $\frac{1}{2}$.

We now attach a player 2 action to each posterior in the sequence $(\mu(k))$. Let $V(k)$ be the value to the normal player 1 of continuation play, beginning at posterior $\mu(k)$, and let $\hat{V}$ be player 1's payoff when $\mu = (1-\lambda)\hat{\mu}$. Player 2 must randomize so that the normal type of player 1 is indifferent between L and H, and hence, for $k = 0, \ldots, N-1$,

$$V(k) = (1-\delta\lambda)(2\alpha_2(\mu(k)) + 1) + \delta\lambda\hat{V} \tag{18}$$

$$= (1 - \delta\lambda)2\alpha_2(\mu(k)) + \delta\lambda V(k+1), \tag{19}$$

where the right side of the top line is the value if L is played at posterior $\mu(k)$ and the second line is the value if H is played, and where we normalize by $(1 - \delta\lambda)$. Finally,

$$V(N) = (1 - \delta\lambda)3 + \delta\lambda\hat{V},$$

since the normal player 1 chooses L with certainty for $\mu > \frac{1}{2}$.

Solving (18) for $k = 0$ gives

$$\hat{V} = 2\alpha_2(\mu(0)) + 1.$$

We then solve the equality of the right side of (18) with (19) to obtain, for $k = 1, \ldots, N$,

$$V(k) = \hat{V} + \frac{1 - \delta\lambda}{\delta\lambda}.$$

These equations tell us a great deal about the equilibrium. The value $V(0)$ is lower than the remaining values, which are all equal to another, $V(1) = V(2) = \ldots = V(N)$. This in turn implies (from (18)) that $\alpha_2(\mu(0))$ is lower than the remaining probabilities $\alpha_2(\mu(1)) = \ldots = \alpha_2(\mu(N-1)) =: \bar{\alpha}_2$. These properties reflect the special structure of the product-choice game, most notably the fact that the stage-game payoff gain to player 1 from playing L rather than H is independent of player 2's action.

It is straightforward to calculate $\alpha_2(\mu(0)) = 1 - (2\delta\lambda)^{-1}$ and $\bar{\alpha}_2 = 1$, and to confirm that a normal player 1 facing posterior $\mu(N)$ prefers to play L. All other aspects of player 1's strategy are optimal because he is indifferent between H and L at every other posterior belief.

Phelan (2006) constructs a similar equilibrium in a game that does not have the equal-gain property of the product-choice game, namely that player 1's payoff increment from playing H is independent of player 2's action. The equilibrium then has a richer dynamic structure.

The breakdown of player 1's reputation in this model is unpredictable, in the sense that the probability of a reputation-ending exertion of low effort is the same regardless of the current posterior that player 1 is good. A normal player 1 who has labored long and hard to build his reputation is just as likely to spend it as is one just starting. Once the reputation has been spent, it can be rebuilt, but only gradually, as the posterior probability of a good type gets pushed upward once more.

4.3.2 Permanent Reputations

Ekmekci, Gossner, and Wilson (2012) study a model in which persistent changes in the type of the long-lived player coexist with permanent reputation effects. The model of the preceding section is a special case of that considered here, in which the probability of a replacement is fixed. In this section, we consider the limits as the probability of a replacement approaches zero and the discount factor approaches one. As one might have expected, the order in which these limits is taken is important.

Let us once again consider the product-choice game of Figure 1. Player 1 can be one of two possible types, a normal type or a simple commitment type who invariably plays H. At the beginning of the game, player 1's type is drawn from a distribution that puts positive probability on each type. We assume that monitoring is imperfect but public; Ekmekci, Gossner, and Wilson (2012) allow private monitoring (including perfect and public monitoring as special cases).

So far, this gives us a standard imperfect monitoring reputation game. Now suppose that at the end of every period, with probability $1 - \lambda$ the long-lived player is replaced, and with probability λ the long-lived player continues until the next period. Replacement draws are independent across periods, and the type of each entering player is an independent draw from the prior distribution over types, regardless of the history that preceded that replacement.

Using entropy arguments similar to those described in Sections 2.4, 2.6, and 2.8, Ekmekci, Gossner, and Wilson (2012) show that the normal type's ex ante equilibrium payoffs in any equilibrium must be at least

$$\underline{w}_H \left(-(1-\delta)\log\mu(H) - \log\lambda\right), \tag{20}$$

and the normal type's continuation equilibrium payoffs (after any history) in any equilibrium must be at least

$$\underline{w}_H \left(-(1-\delta)\log(1-\lambda)\mu(H) - \log\lambda\right). \tag{21}$$

We consider first (20) and the limit as $\lambda \to 1$ and $\delta \to 1$. Player 1 is becoming more patient, and is also becoming increasingly likely to persist from one period to the next. The lower bound on player 1's equilibrium payoff approaches $\underline{w}_H(0)$, the pure Stackelberg payoff of 2.

This payoff bound is unsurprising. One would expect that if replacements are rare, then they will have little effect on ex ante payoffs, and hence that we can replicate the no-replacement Stackelberg payoff bound by making replacements arbitrarily rare. However, we can go beyond this bound to talk about *long-run* beliefs and payoffs. Doing so requires us to consider the lower bound in (21), which reveals the necessity of being more precise about the relative rates at which the replacement probability goes to zero and the discount factor gets large.

The bound (21) suggests that a particularly interesting limiting case arises if $(1-\delta)\ln(1-\lambda) \to 0$. For example, this is the case when $\lambda = 1-(1-\delta)^{\beta}$ for some $\beta > 0$. Under these circumstances, the replacement probability is not going to zero too fast, compared to the rate at which the discount factor goes to one. The lower bound on player 1's continuation payoff, in any Nash equilibrium and after any history, then approaches (as λ and δ get large) the pure Stackelberg payoff of 2. Notice that this reputation-based lower bound on player 1's payoff is effective after *every* history, giving a long run reputation result. Unlike the case examined by Cripps, Mailath, and Samuelson (2004, 2007), the probability that player 1 is the commitment type can never become so small as to vitiate reputation effects.

Some care is required in establishing this result, but the intuition is straightforward. The probability of a replacement imposes a lower bound on the probability of a commitment type. This lower bound might be quite small, but is large enough to ensure that reputation arguments never lose their force, as long as player 1 is sufficiently patient. Recalling our discussion from Section 2.7, the lower-bound arguments from Section 2.4 require that limits be taken in the appropriate order. For any given probability of the commitment type, there is a discount factor sufficiently large as to ensure that player 1 can build a reputation and secure a relatively large payoff. However, for a fixed discount factor, the probability of the commitment type will eventually drop so low as to be essentially irrelevant in affecting equilibrium payoffs. In light of this, consider how the limits interact in Ekmekci, Gossner, and Wilson (2012). As λ goes to one, the lower bound on the probability that player 1 is the commitment type gets very small. But if in the process δ goes to one sufficiently rapidly, then for each value of λ, and hence each (possibly tiny) probability that player 1 is the commitment type, player 1 is nonetheless sufficiently patient that reputation effects can operate. In addition, the constant (if rare) threat of replacement ensues that the lower bound on the probability of a commitment type, and hence the relevance of reputation building, applies to every history of play.

As usual, we have described the argument for the case in which there is a single commitment type, fortuitously chosen to be the Stackelberg type. A similar argument applies to the case in which there are many commitment types. The lower bound on player 1's payoff would then approach the payoff player 1 would receive if 1 were known to be the most advantageous of the possible commitment types.

4.3.3 Reputation as Separation

This section examines a stylized version of Mailath and Samuelson's (2001) model of "separating" reputations. The prospect of replacements again plays a role. However, the reputation considerations now arise not out of the desire to mimic a good commitment type, but to avoid a bad one.[32]

The underlying game is an imperfect monitoring variant of the product-choice game. In each period, the long-lived player 1 chooses either high effort (H) or low effort (L). Low effort is costless, while high effort entails a cost of c. Player 2 receives either a good outcome $\bar{y}$ or a bad outcome $\underline{y}$. The outcome received by player 2 is drawn from a distribution with probability ρ_H on a good outcome (and probability $1-\rho_H$ on a bad outcome) when the firm exerts high effort, and probability $\rho_L < \rho_H$ on a good outcome (and probability $1-\rho_L$ on a bad outcome) when the firm exerts low effort.

At the beginning of the game, player 1 is chosen to be either a normal type or an "inept" type who invariably exerts low effort. The reputation-based incentives for player 1 arise out of player 1's desire to convince player 2 that he is not inept, rather than that he is a good type. We accordingly refer to this as a model of "reputation as separation," in contrast with the more typical reputation-based incentive for a normal player 1 to pool with a commitment type.

As usual, we think of player 1 as a firm and player 2 as a customer, or perhaps as a population of customers. The customer receives utility 1 from outcome $\bar{y}$ and utility 0 from outcome $\underline{y}$. In each period, the customer purchases the product at a price equal to the customer's expected payoff: if the firm is thought to be normal with probability $\tilde{\mu}$, and if the normal firm is thought to choose high effort with probability α, then the price will be

$$p(\tilde{\mu}\alpha) := \tilde{\mu}\alpha\rho_H + (1-\tilde{\mu}\alpha)\rho_L.$$

It is straightforward to establish conditions under which there exist equilibria in which the normal firm frequently exerts high effort. Suppose the normal firm initially exerts high effort, and continues to do so as long as signal $\bar{y}$ is realized, with the period t price given by $p(\tilde{\mu}_h)$, where $\tilde{\mu}_h$ is the posterior probability of a normal firm after history h. Let signal $\underline{y}$ prompt

[32] Morris (2001) examines an alternative model of separating reputations, and the bad reputation models discussed in Section 5.1 are similarly based on a desire to separate from a bad type.

some number $K \geq 1$ periods of low effort and price ρ_L, after which play resumes with the normal firm exerting high effort. We can choose a punishment length K such that these strategies constitute an equilibrium, as long as the cost c is sufficiently small.

Given this observation, our basic question of whether there exist high-effort equilibria appears to have a straightforward, positive answer. Moreover, uncertainty about player 1's type plays no essential role in this equilibrium. The posterior probability that player 1 is normal remains unchanged in periods in which his equilibrium action is L, and tends to drift upward in periods when the normal player 1 plays H. Player 2's belief that player 1 is normal almost certainly converges to 1 (conditional on 1 being normal). Nonetheless, the equilibrium behavior persists, with strings in which player 1 chooses H interspersed with periods of punishment. Indeed, the proposed behavior remains an equilibrium even if player 1 is known to be normal from the start. We have seemingly accomplished nothing more than showing that equilibria can be constructed in repeated games in which players do not simply repeat equilibrium play of the stage game.

However, we prefer to restrict attention to Markov strategies, with the belief about player 1's type as the state variable. This eliminates the equilibrium we have just described, since under these strategies the customer's behavior depends upon the firm's previous actions as well as the customer's beliefs.

Why restrict attention to Markov equilibria? The essence of repeated games is that continuation play can be made to depend on current actions in such a way as to create intertemporal incentives. Why curtail this ability by placing restrictions on the ability to condition actions on behavior? How interesting would the repeated prisoners' dilemma be if attention were restricted to Markov equilibria?

While we describe player 2 as a customer, we have in mind cases in which the player 2 side of the game corresponds to a large (continuum) population of customers. It is most natural to think of these customers as receiving idiosyncratic noisy signals of the effort choice of the firm; by idiosyncratic we mean that the signals are independently drawn and privately observed.[33] A customer who receives a bad outcome from a service provider cannot be sure whether the firm has exerted low effort, or whether the customer has simply been unlucky (and most of the other customers have received a good

[33] There are well-known technical complications that arise with a continuum of independently distributed random variables (see, for example, Al-Najjar (1995)). Mailath and Samuelson (2006, Remark 18.1.3) describes a construction that, in the current context, avoids these difficulties.

outcome).

The imperfection in the monitoring per se is not an obstacle to creating intertemporal incentives. There is a folk theorem for games of imperfect public monitoring, even when players only receive signals that do not perfectly reveal actions (Fudenberg, Levine, and Maskin, 1994). If the idiosyncratic signals in our model were public, we could construct equilibria with a population of customers analogous to the equilibrium described above: customers coordinate their actions in punishing low effort on the part of the firm. Such coordination is possible if the customers observe each other's idiosyncratic signals (in which case the continuum of signals precisely identifies the firm's effort) or if the customers all receive the same signal (which would not identify the firm's effort, but could still be used to engineer coordinated and effective punishments).

However, we believe there are many cases in which such coordinated behavior is not possible, and so think of the idiosyncratic signals as being *private*, precluding such straightforward coordination. Nonetheless, there is now a significant literature on repeated finite games with private monitoring (beginning with Sekiguchi (1997), Piccione (2002), and Ely and Välimäki (2002)) suggesting that private signals do not preclude the provision of intertemporal incentives. While the behavioral implications of this literature are still unclear, it is clear that private signals significantly complicate the analysis. Mailath and Samuelson (2006, Section 18.1) studies the case of a continuum of customers with idiosyncratic signals.

In order to focus on the forces we are interested in without being distracted by the technical complications that arise with private signals, we work with a single customer (or, equivalently, a population of customers receiving the same signal) and rule out coordinated punishments by restricting attention to Markov equilibria.

At the end of each period, the firm continues to the next period with probability λ, but with probability $1-\lambda$, is replaced by a new firm. In the event of a replacement, the replacement's type is drawn from the prior distribution, with probability $\tilde{\mu}^0$ of a normal replacement.

A *Markov strategy* for the normal firm can be written as a mapping $\alpha : [0,1] \to [0,1]$, where $\alpha(\tilde{\mu})$ is the probability of choosing action H when the customer's posterior probability of a normal firm is $\tilde{\mu}$. The customers' beliefs after receiving a good signal are updated according to

$$\varphi(\tilde{\mu} \mid \bar{y}) = \lambda \frac{[\rho_H \alpha(\tilde{\mu}) + \rho_L(1-\alpha(\tilde{\mu}))]\,\tilde{\mu}}{[\rho_H \alpha(\tilde{\mu}) + \rho_L(1-\alpha(\tilde{\mu}))]\,\tilde{\mu} + \rho_L(1-\tilde{\mu})} + (1-\lambda)\tilde{\mu}^0,$$

with a similar expression for the case of a bad signal. The strategy α is a

Markov equilibrium if it is maximizing for the normal firm.

Proposition 4 *Suppose $\lambda \in (0,1)$ and $\tilde{\mu}^0 > 0$. There exists $\bar{c} > 0$ such that for all $0 \leq c < \bar{c}$, there exists a Markov equilibrium in which the normal firm always exerts high effort.*

Proof. Suppose the normal firm always exerts high effort. Then given a posterior probability $\tilde{\mu}$ that the firm is normal, firm revenue is given by $p(\tilde{\mu}) = \tilde{\mu}\rho_H + (1-\tilde{\mu})\rho_L$. Let $\tilde{\mu}_y := \varphi(\tilde{\mu} \mid y)$ and $\tilde{\mu}_{xy} := \varphi(\varphi(\tilde{\mu} \mid x) \mid y)$ for $x, y \in \{\underline{y}, \bar{y}\}$. Then $\tilde{\mu}_{\bar{y}\bar{y}} > \tilde{\mu}_{\bar{y}} > \tilde{\mu} > \tilde{\mu}_{\underline{y}} > \tilde{\mu}_{\underline{y}\underline{y}}$ and $\tilde{\mu}_{\bar{y}y} > \tilde{\mu}_{\underline{y}y}$ for $y \in \{\underline{y}, \bar{y}\}$. The value function of the normal firm is given by

$$V(\tilde{\mu}) = (1-\delta\lambda)(p(\tilde{\mu}) - c) + \delta\lambda\left[\rho_H V(\tilde{\mu}_{\bar{y}}) + (1-\rho_H)V(\tilde{\mu}_{\underline{y}})\right].$$

The payoff from exerting low effort and thereafter adhering to the equilibrium strategy is

$$V(\tilde{\mu}; L) := (1-\delta\lambda)p(\tilde{\mu}) + \delta\lambda\left[\rho_L V(\tilde{\mu}_{\bar{y}}) + (1-\rho_L)V(\tilde{\mu}_{\underline{y}})\right].$$

Thus, $V(\tilde{\mu}) - V(\tilde{\mu}; L)$ is given by

$$\begin{aligned} &-c(1-\delta\lambda) + \delta\lambda(1-\delta\lambda)(\rho_H - \rho_L)(p(\tilde{\mu}_{\bar{y}}) - p(\tilde{\mu}_{\underline{y}})) \\ &+ \delta^2\lambda^2(\rho_H - \rho_L)\{\rho_H[V(\tilde{\mu}_{\bar{y}\bar{y}}) - V(\tilde{\mu}_{\underline{y}\bar{y}})] + (1-\rho_H)[V(\tilde{\mu}_{\bar{y}\underline{y}}) - V(\tilde{\mu}_{\underline{y}\underline{y}})]\} \\ &\qquad \geq (1-\delta\lambda)\{-c + \delta\lambda(\rho_H - \rho_L)[p(\tilde{\mu}_{\bar{y}}) - p(\tilde{\mu}_{\underline{y}})]\}, \end{aligned} \tag{22}$$

where the inequality is established via a straightforward argument showing that V is increasing in μ.

From an application of the one shot deviation principle, it is an equilibrium for the normal firm to always exert high effort, with the implied customer beliefs, if and only if $V(\tilde{\mu}) - V(\tilde{\mu}; L) \geq 0$ for all feasible $\tilde{\mu}$. From (22), a sufficient condition for this inequality is

$$p(\tilde{\mu}_{\bar{y}}) - p(\tilde{\mu}_{\underline{y}}) \geq \frac{c}{\delta\lambda(\rho_H - \rho_L)}.$$

Because there are replacements, the left side of this inequality is bounded away from zero. There thus exists a sufficiently small c such that the inequality holds. ■

In equilibrium, the difference between the continuation value of choosing high effort and the continuation value of choosing low effort must exceed the cost of high effort. However, the value functions corresponding to high and low effort approach each other as $\tilde{\mu} \to 1$, because the values diverge only through the effect of current outcomes on future posteriors, and current outcomes have very little affect on future posteriors when customers are currently quite sure of the firm's type. The smaller the probability of an inept replacement, the closer the posterior expectation of a normal firm can approach unity. Replacements ensure that $\tilde{\mu}$ can never reach unity, and hence there is always a wedge between the high-effort and low-effort value functions. As long as the cost of the former is sufficiently small, high effort will be an equilibrium.

The possibility of replacements is important in this result. In the absence of replacements, customers eventually become so convinced the firm is normal (i.e., the posterior $\tilde{\mu}$ becomes so high), that subsequent evidence can only shake this belief very slowly. Once this happens, the incentive to choose high effort disappears (Mailath and Samuelson, 2001, Proposition 2). If replacements continually introduce the possibility that the firm has become inept, then the firm cannot be "too successful" at convincing customers it is normal, and so there is an equilibrium in which the normal firm always exerts high effort.

To confirm the importance of replacements, suppose there are no replacements, and suppose further that $\rho_H = 1 - \rho_L$. In this case, there is a unique Markov equilibrium in pure strategies, in which the normal firm exerts low effort after every history:

Under this parameter configuration, an observation of $\bar{y}$ followed by $\underline{y}$ (or $\underline{y}$ followed by $\bar{y}$) leaves the posterior at precisely the level it had attained before these observations. More generally, posterior probabilities depend only on the number of $\bar{y}$ and $\underline{y}$ observations in the history, and not on their order. We can thus think of the set of possible posterior probabilities as forming a ladder, with countably many rungs and extending infinitely in each direction, with a good signal advancing the firm one rung up and a bad signal pushing the firm down one rung. Now consider a pure Markov equilibrium, which consists simply of a prescription for the firm to exert either high effort or low effort at each possible posterior about the firm's type. If there exists a posterior at which the firm exerts low effort, then upon being reached the posterior is never subsequently revised, since normal and inept firms then behave identically, and the firm receives the lowest possible continuation payoff. This in turn implies that if the equilibrium ever calls for the firm to exert high effort, than it must also call for the firm to exert high effort at the next higher posterior. Otherwise, the next higher posterior yields the lowest possible continuation payoff, and the firm will not exert costly effort only to enhance the chances of receiving the lowest possible payoff.

We can repeat this argument to conclude that if the firm ever exerts high effort, it must do so for every higher posterior. But then for any number of periods T and $\varepsilon > 0$, we can find a posterior (close to one) at which the equilibrium calls for high effort, and with the property that no matter what effort levels the firm exerts over the next T periods, the posterior that the firm is normal will remain above $1 - \varepsilon$ over those T periods. By making T large enough that periods after T are insignificant in payoff calculations and making ε small, we can ensure that the effect of effort on the firm's revenue are overwhelmed by the cost of that effort, ensuring that the firm will exert low effort. This gives us a contradiction to the hypothesis that high effort is ever exerted.

Mailath and Samuelson (2001) show that the inability to support high effort without replacements extends beyond this simple case. These results thus combine to provide the seemingly paradoxical result that it can be good news for the firm to have customers constantly fearing that the firm might "go bad." The purpose of a reputation is to convince customers that the firm is normal and will exert high effort. As we have just seen, the problem with maintaining a reputation in the absence of replacements is that the firm essentially succeeds in convincing customers it is normal. If replacements continually introduce the possibility that the firm has turned inept, then there is an upper bound, short of unity, on the posterior $\tilde{\mu}$, and so the difference in posteriors after different signals is bounded away from zero. The incentive to exert high effort in order to convince customers that the firm is still normal then always remains.

A similar role for replacements was described by Holmström (1982) in the context of a signal-jamming model of managerial employment. The wage of the manager in his model is higher if the market posterior over the manager's type is higher, even if the manager chooses no effort. In contrast, the revenue of a firm in the current model is higher for higher posteriors only if customers also believe that the normal firm is choosing high effort. Holmström's manager always has an incentive to increase effort, in an attempt to enhance the market estimation of his talent. In contrast to the model examined here, an equilibrium then exists (without replacements) in which the manager chooses effort levels that are higher than the myopic optimum. In agreement with spirit of our analysis, however, this overexertion disappears over time, as the market's posterior concerning the manager's type approaches one.[34]

5 Discussion

Even within the context of repeated games, we have focused on a particular model of reputations. This section briefly describes some alternatives.

5.1 Outside Options and Bad Reputations

We begin with an interaction that we will interpret as involving a sequence of short-lived customers and a firm, but we allow customers an outside option that induces sufficiently pessimistic customers to abandon the firm, and so not observe any further signals.

The ability to sustain a reputation in this setting hinges crucially on whether the behavior of a firm on the brink of losing its customers makes the firm's product more or less valuable to customers. If these actions make the firm more valuable to customers, it is straightforward to identify conditions under which the firm can maintain a reputation. This section, drawing on Ely and Välimäki (2003), presents a model in which the firms' efforts to avoid a no-trade region destroy the incentives needed for a nontrivial equilibrium. We concentrate on a special case of Ely and Välimäki (2003). Ely, Fudenberg, and Levine (2008) provide a general analysis of bad reputations.

There are two players, referred to as the firm (player 1) and the customer (player 2). We think of the customer as hiring the firm to perform a service, with the appropriate nature of the service depending upon a diagnosis that only the firm can perform. For example, the firm may be a doctor who must determine whether the patient needs to take two aspirin daily or needs a heart transplant.

The interaction is modeled as a repeated game with random states. There are two states of the world, θ_H and θ_L. In the former, the customer requires a high level of service, denoted by H, in the latter a low level denoted by L.

The stage game is an extensive form game. The state is first drawn by Nature and revealed to the firm but not the customer. The customer then

[34] Neither the market *nor the manager* knows the talent of the manager in Holmström's (1982) model. The manager's evaluation of the profitability of effort then reflects only market beliefs. In contrast, our normal firms are more optimistic about the evolution of posterior beliefs that are customers. However, the underlying mechanism generating incentives is the same.

	Hire	Not hire
H	u, u	$0, 0$
L	$-w, -w$	$0, 0$

State θ_H

	Hire	Not hire
H	$-w, -w$	$0, 0$
L	u, u	$0, 0$

State θ_L

Figure 13: Payoffs for the terminal nodes of the extensive form bad-reputation stage game, as a function of the state (θ_H or θ_L), the customer's decision of whether to hire the firm, and (if hired) the firm's choice of service. We assume $w > u > 0$.

decides whether to hire the firm. If the firm is hired, he chooses the level of service to provide.

The payoffs attached to each terminal node in the extensive form game are given in Figure 13. The firm and the customer thus have identical payoffs. Both prefer that high service be provided when necessary, and that low service be provided when appropriate. If the firm is not hired, then both players receive 0.

Given that interests are aligned, one would think there should be no difficulty in the customer and the firm achieving the obviously efficient outcome, in which the firm is always hired and the action is matched to the state. It is indeed straightforward to verify that the *stage game* presents no incentive problems. Working backwards from the observation that the only sequentially rational action for the firm is to provide action H in state θ_H and action L in state θ_L, we find a unique sequential equilibrium, supporting the efficient outcome.

Suppose now that the (extensive form) stage game is repeated. The firm is a long-lived player who discounts at rate δ. The customer is a short-lived player. Each period features the arrival of a new customer. Nature then draws the customer's state and reveals its realization to the firm (only). These draws are independent across periods, with the two states equally likely in each case. The customer decides whether to hire the firm, and the firm then chooses a level of service. At the end of the period, a public signal from the set $Y \equiv \{\emptyset, H, L\}$ is observed, indicating either that the firm was not hired ($\emptyset$) or was hired and provided either high (H) or low (L) service. Short-lived players thus learn nothing about the firm's stage-game strategy when the firm is not hired.

For large δ, the repeated game has multiple equilibria. In addition to the obvious one, in which the firm is always hired and always takes the action that is appropriate for the state, there is an equilibrium in which the firm is never hired.[35] Can we restrict attention to a subset of such equilibria by introducing incomplete information? The result of incomplete information is indeed a bound on the firm's payoff, but it is now an upper bound that consigns the firm to a surprisingly low payoff.

With probability $1 - \hat{\mu} > 0$, the firm is normal. With complementary probability $\hat{\mu} > 0$, the firm is "bad" and follows a strategy of always choosing H. A special case of Ely and Välimäki's (2003) result is then:

Proposition 5 *Assume that in any period in which the firm is believed to be normal with probability* 1*, the firm is hired. Then if the firm is sufficiently patient, there is a unique Nash equilibrium outcome in which the firm is never hired.*

Ely and Välimäki (2003) dispense with the assumption that the firm is hired whenever thought to be normal with probability 1, but this assumption is useful here in making the following intuition transparent. Fix an equilibrium strategy profile and let $\hat{\mu}^\dagger$ be the supremum of the set of posterior probabilities of a bad firm for which the firm is (in equilibrium) hired with positive probability. We note that $\hat{\mu}^\dagger$ must be less than 1 and argue that $\hat{\mu}^\dagger > 0$ is a contradiction. If the firm ever is to be hired, there must be a significant chance that the normal firm chooses L (in state θ_L), since otherwise his value to the customer is negative. Then for any posterior probability $\hat{\mu}'$ sufficiently close to $\hat{\mu}^\dagger$ at which the firm is hired, an observation of H must push the posterior of a bad firm above $\hat{\mu}^\dagger$, ensuring that the firm is never again hired. But then no sufficiently patient normal firm, facing a posterior probability $\hat{\mu}'$, would ever choose H in state θ_H. Doing so gives a payoff of $(1 - \delta)u$ (a current payoff of u, followed by a posterior above $\hat{\mu}^\dagger$ and hence a continuation payoff of 0) while choosing L reveals the firm to be normal and hence gives a higher (for large δ) payoff of $-(1-\delta)w + \delta u$. The normal firm thus cannot be induced to choose H at posterior $\hat{\mu}'$. But this now ensures that the firm will not be hired for any such posterior, giving us a contradiction to the assumption that $\hat{\mu}^\dagger$ is the supremum of the posterior probabilities for which the firm is hired.

The difficulty facing the normal firm is that an unlucky sequence of θ_H states may push the posterior probability that the firm is bad disastrously high. At this point, the normal firm will choose L in both states in a des-

[35]The structure of this equilibrium is similar to the zero-firm-payoff equilibrium of the purchase game (Section 2.5).

perate attempt to stave off a career-ending bad reputation. Unfortunately, customers will anticipate this and not hire the firm, ending his career even earlier. The normal firm might attempt to forestall this premature end by playing L (in state θ_H) somewhat earlier, but the same reasoning unravels the firm's incentives back to the initial appearance of state θ_H. We can thus never construct incentives for the firm to choose H in state θ_H, and the firm is never hired.

5.2 Investments in Reputations

It is common to speak of firms as investing in their reputations. This section presents a discrete-time version of Board and Meyer-ter-Vehn's (2012) continuous-time model, which centers around the idea that investments are needed to build reputations.

In each period, a product produced by a firm can be either high quality or low quality. As in Section 4.3.3, the customers on the other side of the market are not strategic, and simply pay a price for the good in each period equal to the probability that it is high quality.

In each period, the firm chooses not the quality of the good, but an amount $\eta \in [0, 1]$ to invest in high quality. The quality of the firm's product is initially drawn to be either low or high. At the end of each period, with probability λ, there is no change in the firm's quality. However, with probability $1 - \lambda$, a new quality draw is taken before the next period. In the event of a new quality draw at the end of period t, the probability of emerging as a high-quality firm is given by η^t, the investment made by the firm in period t.

An investment of level η costs the firm $c\eta$, with $c > 0$. The firm's payoff in each period is then the price paid by the customers in that period minus the cost of his investment. The firm has an incentive to invest because higher-quality goods receive higher prices, and a higher investment enhances the probability of a high-quality draw the next time the firm's quality level is determined.

Customers do not directly observe the quality of the firm. Instead, the customers observe either signal 0 or signal 1 in each period. If the firm is high quality in period t, then with probability ρ_H the customer receives signal 1. If the firm is low quality, then with probability ρ_L the customer receives signal 1. If $\rho_H > \rho_L$, then the arrival of a 1 is good news and pushes upward the posterior of high quality. If $\rho_H < \rho_L$, then the arrival of a 1 is bad news, and pushes downward the posterior that the good is high quality. We say that the signals give perfect good news if $\rho_L = 0$, and perfect bad news if $\rho_H = 0$.

For one extreme case, suppose $\rho_L = 0$, giving the case of perfect good news. Here, the signal 1 indicates that the firm is certainly high quality (at least until the next time its quality is redrawn). We might interpret this as a case in which the product may occasionally allow a "breakthrough" that conveys high utility, but can do so only if it is indeed high quality. For example, the product may be a drug that is always ineffective if low quality and may also often have no effect if it is high quality, but occasionally (when high quality) has dramatic effects.

Conversely, it may be that $\rho_H = 0$, giving the case of perfect bad news. Here, the product may ordinarily function normally, but if it is of low quality it may occasionally break down. The signal 1 then offers assurances that the product is low quality, at least until the next quality draw.

We focus on four types of behavior. In a full-work profile, the firm always chooses $\eta = 1$, no matter the posterior about the product's quality. Analogously, the firm chooses $\eta = 0$ for every posterior in a full-shirk profile. In a work-shirk profile, the firm sets $\eta = 1$ for all posteriors below some cutoff, and chooses $\eta = 0$ for higher posteriors. The firm thus works to increase the chance that its next draw is high quality when its posterior is relatively low, and rides on its reputation when the latter is relatively high. Turning this around, in a shirk-work profile the firm shirks for all posteriors below a cutoff and works for all higher posteriors. Here, the firm strives to maintain a high reputation but effectively surrenders to a low reputation.

In Board and Meyer-ter-Vehn's (2012) continuous-time model, equilibria for the cases of perfect good news and perfect bad news are completely characterized in terms of these four behavioral profiles. The discrete-time model is easy to describe but cumbersome to work with because it is notoriously difficult to show that the value functions are monotonic in beliefs. For the continuous-time model, Board and Meyer-ter-Vehn (2012) show the following (among other things):

- If signals are perfect good news, an equilibrium exists and this equilibrium is unique if $1 - \lambda \geq \rho_H$. Every equilibrium is either work-shirk or full-shirk. The induced process governing the firm's reputation is ergodic.

- If signals are perfect bad news again equilibrium exists, but $1 - \lambda \geq \rho_L$ does not suffice for uniqueness. Every equilibrium is either shirk-work, full-shirk or full-work. In any nontrivial shirk-work equilibrium, the reputation dynamics are not ergodic.

What lies behind the difference between perfect good news and perfect bad news? Under perfect good news, investment is rewarded by the enhanced probability of a reputation-boosting signal. This signal conveys the best information possible, namely that the firm is certainly high quality, but the benefits of this signal are temporary, as subsequent updating pushes the firm's reputation downward in recognition that its type may have changed. The firm thus goes through cycles in which its reputation is pushed to the top and the then deteriorates, until the next good signal renews the process. The firm continually invests (at all posteriors, in a full-work equilibrium, or at all sufficiently low posteriors, in a work-shirk equilibrium) in order to push its reputation upward.

In the shirk-work equilibrium under perfect bad news, bad signals destroy all hope of rebuilding a reputation, since the equilibrium hypothesis that the firm then makes no investments precludes any upward movement in beliefs. If the initial posterior as to the firm's type exceeds the shirk-work cutoff, the firm has an incentive to invest in order to ward off a devastating collapse in beliefs, but abandons all hope of a reputation once such a collapse occurs. The fact that it is optimal to not invest whenever customers expect no investment gives rise to multiple equilibria.

Some ideas reminiscent of the bad reputation model of Section 5.1 reappear here. In the case of a shirk-work equilibrium and perfect bad news, the firm can get trapped in the region of low posteriors. The customers do not literally abandon the firm, as they do in the bad-reputation model, but the absence of belief revision ensures that there is no escape. In this case, however, the firm's desire to avoid this low-posterior trap induces the firm to take high effort, which customers welcome. As a result, the unraveling of the bad reputation model does not appear here.

5.3 Continuous Time

Most reputation analyses follow the standard practice in repeated games of working with a discrete-time model, while examining the limit as the discount factor approaches one. Since we can write the discount factor as

$$\delta = e^{-r\Delta},$$

where r is the discount rate and Δ the length of a period, there are two interpretations of this limit. On the one hand, this may reflect a change in preferences, with the timing of the game remaining unchanged and the players becoming more patient ($r \to 0$). One sees shadows of this interpretation in the common statements that reputation results hold for "patient players." However, the more common view is that the increasing discount factor reflects a situation in which the players' preferences are fixed, and the game is played more frequently ($\Delta \to 0$).

If one is examining repeated games of perfect monitoring, these interpretations are interchangeable. It is less obvious that they are interchangeable under imperfect monitoring. If periods are shrinking and the game is being played more frequently, one would expect the imperfect signals to become increasingly noisy. A firm may be able to form a reasonably precise idea of how much its rivals have sold if it has a year's worth of data to examine, but may be able to learn much less from a day's worth of data. We should then seek limiting reputations results for the case in which discount factors become large and the monitoring structure is adjusted accordingly.

One convenient way to approach this problem is to work directly in the continuous-time limit. This section describes some recent work on continuous-time reputation games.

It is not obvious that continuous-time games provide a fertile ground for studying reputations. Several examples have appeared in the literature under which intertemporal incentives can lose their force as time periods shrink in complete information games, as is the case in Abreu, Milgrom, and Pearce (1991), Sannikov and Skrzypacz (2007), and Fudenberg and Levine (2007). The difficulty is that as actions become frequent, the information observed in each period provides increasingly noisy indications of actions, causing the statistical tests for cheating to yield too many false positives and trigger too many punishments, destroying the incentives.

The effectiveness of reputation considerations in continuous time hinges on order-of-limits considerations. If we fix the period length and let the discount factor approach one, then we have standard reputation results, with the underlying reasoning reproducing standard reputation arguments. Alternatively, if we fix the discount rate and let the period length go to zero, these reputation effects disappear. Here, we encounter the same informational problems that lie behind the collapse of intertemporal incentives in repeated games of complete information.

This gives us an indication of what happens in the two extreme limit orders. What happens in intermediate cases? Faingold (2005) shows that if we work with the limiting continuous-time game, then there exists a discount factor such that reputation effects persist for all higher discount factors. Hence, we can choose between discrete-time or continuous-time games to study reputations, depending on which leads to the more convenient analysis.

5.3.1 Characterizing Behavior

Reputation models have typically produced stronger characterizations of equilibrium payoffs than of equilibrium behavior. However, Faingold and Sannikov (2011) have recently exploited the convenience of continuous time to produce a complete characterization of reputation-building behavior. We confine our presentation to the description of an example, taken from their paper.

We consider a variation of the product-choice game. A long-lived firm faces a continuum of customers. At each time t, the firm chooses an effort level $a^t \in [0,1]$, while each customer chooses a level of service $b^t \in [0,3]$. The firm observes only the average service level B^t chosen by the continuum of customers, while the customers observe only the current quality level of the firm, X^t. The latter evolves according to a Brownian motion, given by $dX^t = a^t dt + dZ^t$.

The firm maximizes the discounted profits

$$\int_0^\infty e^{-rt}(B^t - a^t)dt,$$

so that the firm prefers the customers buy a higher level of service, but finds effort costly. In each period each customer chooses her level of service to maximize her instantaneous payoff, with the maximizing service level increasing in the quality of the firm and decreasing in the average service level chosen by customers. We might interpret the latter as reflecting a setting in which there is congestion in the consumption of the service.

The unique equilibrium outcome in the stage game is that the firm chooses zero effort. In the repeated game, the firm's effort levels are statistically identified—different levels of effort give rise to different processes for the evolution of the firm's quality. This opens the possibility that nontrivial incentives might be constructed in the repeated game. However, the unique equilibrium of the continuous-time repeated game features the relentless play of this stage-game Nash equilibrium.

This result is initially counterintuitive, since it is a familiar result that the long-lived player could be induced to take actions that are are not myopic best responses in a discrete-time formulation of this game. One would need only arrange the equilibrium so that "bad" signals about his actions trigger punishments. However, as we have noted, the resulting incentives can lose their force as time periods shrink, as is the case in Abreu, Milgrom, and Pearce (1991), Sannikov and Skrzypacz (2007), and Fudenberg and Levine (2007).

Now suppose that with some probability, the firm is a commitment type who always takes the maximal investment, $a = 1$. Faingold and Sannikov (2011) show that there is a unique sequential equilibrium. As the long-lived player becomes arbitrarily patient ($r \to 0$), the long-lived player's payoff converges to the Stackelberg payoff of 2.

Behind these payoffs lie rich equilibrium dynamics. The unique sequential equilibrium is a Markov equilibrium, with the posterior probability μ^t that the firm is a commitment type serving as the state variable. As μ^t approaches unity, the aggregate service level demanded by the short-lived players approaches the best response to the commitment type of 3. Smaller average service levels are demanded for smaller posteriors, though in equilibrium these levels approach 3 as the long-lived player gets more patient. The long-lived player exerts her highest effort levels at intermediate posteriors, while taking lower effort at very low or very high posteriors, with the function specifying the effort level as a function of the posterior converging to unity as the long-lived player gets more patient.

We thus have a reputation version of "number two tries harder." Firms with very high reputations rest on their laurels, finding the cost of high effort not worth the relatively small effect on customer beliefs. Firms with very low reputations abandon all hope of building their reputation. Firms in the middle labor mightily to enhance their reputations. The more patient the firm, the higher the payoff to reputation building, and hence the higher the effort profile as a function of the posterior.

5.3.2 Reputations Without Types

The standard model of reputations is centered around incomplete information concerning the long-lived player's type, with the long-lived player's reputation interpreted in terms of beliefs about his type. In contrast, Bohren (2011) shows that interesting reputation dynamics can appear without any uncertainty as to the type of the long-lived player.

Suppose that at each instant of time, a firm is characterized by a current quality, and chooses an investment. The current quality is a stock variable that is observed, while customers receive only a noisy signal of the investment. There is a continuum of customers, whose individual actions are unobservable. The model builds on Faingold and Sannikov (2011), but without incomplete information and with quality playing the role of posterior beliefs.

In each period, the current quality stock is observed, and then the firm and the customers simultaneously choose an investment level and a quantity to purchase. Higher investment levels tend to give rise to higher signals. The customer's payoff is increasing in the quality stock and the signal, and depends on the customer's quantity. The higher is the stock and the signal, the higher is the customer's payoff maximizing quantity.

The firm's payoff is increasing in the aggregate purchase decision, but is decreasing in the level of investment, which is costly. In a static version of this game, equilibrium would call for an investment level of zero from the firm. The incentive for the firm to undertake investments in the current period then arises out of the prospect of generating larger future quality stocks. In particular, the stock increases in expectation when investment exceeds the current quality level, and decreases in expectation when investment falls short of this level.

Bohren (2011) shows that there is a unique public prefect equilibrium, which is a Markov equilibrium. Once again, this equilibrium gives rise to "number two tries harder" incentives. The firm's investments are highest for an intermediate range of quality stocks, at which the firm works hard to boost its quality. The firm undertakes lower investments at lower quality levels, discouraged by the cost of building its reputation and waiting for a fortuitous quality shock to push it into more favorable territory. Similarly, the invests less for high quality levels, content to coast on its reputation. The result is a progression of product quality cycles.

References

Abreu, D., P. Milgrom, and D. Pearce (1991): "Information and Timing in Repeated Partnerships," *Econometrica*, 59(6), 1713–1733. 78, 79

Al-Najjar, N. I. (1995): "Decomposition and Characterization of Risk with a Continuum of Random Variables," *Econometrica*, 63(5), 1195–1224. 67

Aoyagi, M. (1996): "Evolution of Beliefs and the Nash Equilibrium of Normal Form Games," *Journal of Economic Theory*, 70(2), 444–469. 49

Atakan, A. E., and M. Ekmekci (2012): "Reputation in Long-Run Relationships," *Review of Economic Studies*, 79(2), 451–480. 40

——— (2013): "A Two-Sided Reputation Result with Long-Run Players," *Journal of Economic Theory*, 148(1), 376–392. 40

Benabou, R., and G. Laroque (1992): "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, 107(3), 921–958. 20

Board, S., and M. Meyer-ter-Vehn (2012): "Reputation for Quality," UCLA. 75, 76

Bohren, J. A. (2011): "Stochastic Games in Continuous Time: Persistent Actions in Long-Run Relationships," UCSD. 80, 81

Celentani, M., D. Fudenberg, D. K. Levine, and W. Pesendorfer (1996): "Maintaining a Reputation Against a Long-Lived Opponent," *Econometrica*, 64(3), 691–704. 40, 47, 49

Cesa-Bianchi, N., and G. Lugosi (2006): *Prediction, Learning, and Games*. Cambridge University Press. 9

Chung, K. L. (1974): *A Course in Probability Theory*. Academic Press, New York. 26

Cole, H. L., J. Dow, and W. B. English (1995): "Default, Settlement, and Signalling: Lending Resumption in a Reputational Model of Sovereign Debt," *International Economic Review*, 36(2), 365–385. 61

Conlon, J. (2003): "Hope Springs Eternal: Learning and the Stability of Cooperation in Short Horizon Repeated Games," *Journal of Economic Theory*, 112(1), 35–65. 30

Cover, T. M., and J. A. Thomas (2006): *Elements of Information Theory*. John Wiley & Sons, Inc., New York, second edn. 8, 9, 10

Cripps, M. W., E. Dekel, and W. Pesendorfer (2005): "Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests," *Journal of Economic Theory*, 121(2), 259–272. 40

Cripps, M. W., G. J. Mailath, and L. Samuelson (2004): "Imperfect Monitoring and Impermanent Reputations," *Econometrica*, 72(2), 407–432. 20, 30, 31, 40, 60, 61, 65

——— (2007): "Disappearing Private Reputations in Long-Run Relationships," *Journal of Economic Theory*, 134(1), 287–316. 20, 40, 60, 61, 65

Cripps, M. W., K. M. Schmidt, and J. P. Thomas (1996): "Reputation in Perturbed Repeated Games," *Journal of Economic Theory*, 69(2), 387–410. 46

Ekmekci, M. (2011): "Sustainable Reputations with Rating Systems," *Journal of Economic Theory*, 146(2), 479–503. 55

Ekmekci, M., O. Gossner, and A. Wilson (2012): "Impermanent Types and Permanent Reputations," *Journal of Economic Theory*, 147(1), 162–178. 64, 65

Ely, J., D. Fudenberg, and D. K. Levine (2008): "When is Reputation Bad?," *Games and Economic Behavior*, 63(2), 498–526. 72

Ely, J. C., and J. Välimäki (2002): "A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory*, 102(1), 84–105. 68

——— (2003): "Bad Reputation," *Quarterly Journal of Economics*, 118(3), 785–814. 72, 74

Ettinger, D., and P. Jehiel (2010): "A Theory of Deception," *American Economic Journal: Microeconomics*, 2(1), 1–20. 57

Evans, R., and J. P. Thomas (1997): "Reputation and Experimentation in Repeated Games with Two Long-Run Players," *Econometrica*, 65(5), 1153–1173. 50, 51

——— (2001): "Cooperation and Punishment," *Econometrica*, 69(4), 1061–1075. 51

Faingold, E. (2005): "Building a Reputation under Frequent Decisions," University of Pennsylvania. 78

Faingold, E., and Y. Sannikov (2011): "Reputation in Continuous-Time Games," *Econometrica*, 79(3), 773–876. 79, 80

Fudenberg, D., and D. K. Levine (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57(4), 759–778. 8, 19

——— (1992): "Maintaining a Reputation When Strategies Are Imperfectly Observed," *Review of Economic Studies*, 59(3), 561–579. 8

——— (2007): "Continuous Time Limits of Repeated Games With Imperfect Public Monitoring," *Review of Economic Dynamics*, 10(2), 173–192. 78, 79

Fudenberg, D., D. K. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62(5), 997–1039. 68

Gossner, O. (2011a): "An Entropy-Based Proof of Disappearing Reputations," Paris School of Economics. 20

——— (2011b): "Simple Bounds on the Value of a Reputation," *Econometrica*, 79(5), 1627–1641. 8, 19

Holmström, B. (1982): "Managerial Incentive Problems: A Dynamic Perspective," in *Essays in Economics and Management in Honour of Lars Wahlbeck*, pp. 209–230. Swedish School of Economics and Business Administration, Helsinki, Reprinted in *Review of Economic Studies* 66 (1), 1999, 169-182. 61, 71, 72

Hu, J. (2013): "Reputation in the Presence of Noisy Exogenous Learning," PIER Working Paper 13-009, University of Pennsylvania. 32

Jehiel, P. (2005): "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory*, 123(2), 81–104. 57

Jehiel, P., and L. Samuelson (2012): "Reputation with Analogical Reasoning," *Quarterly Journal of Economics*, 127(4), 1927–1969. 55, 60

Kalai, E., and E. Lehrer (1995): "Subjective Games and Equilibria," *Games and Economic Behavior*, 8(1), 123–163. 20

Kreps, D., P. R. Milgrom, D. J. Roberts, and R. Wilson (1982): "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 27(2), 245–252. 2, 4

Kreps, D., and R. Wilson (1982): "Reputation and Imperfect Information," *Journal of Economic Theory*, 27(2), 253–279. 4, 30, 49

Liu, Q. (2011): "Information Acquisition and Reputation Dynamics," *Review of Economic Studies*, 78(4), 1400–1425. 52, 53

Liu, Q., and A. Skrzypacz (2011): "Limited Records and Reputation," Columbia University and Graduate School of Business, Stanford. 54, 55

MAILATH, G. J., AND L. SAMUELSON (2001): "Who Wants a Good Reputation?," *Review of Economic Studies*, 68(2), 415–441. 20, 61, 66, 70, 71

——— (2006): *Repeated Games and Reputations: Long-Run Relationships.* Oxford University Press, New York, NY. 1, 8, 14, 17, 67, 68

MILGROM, P. R., AND D. J. ROBERTS (1982): "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, 27(2), 280–312. 4, 30, 49

MORRIS, S. (2001): "Political Correctness," *Journal of Political Economy*, 109(2), 231–265. 66

PHELAN, C. (2006): "Public Trust and Government Betrayal," *Journal of Economic Theory*, 130(1), 27–43. 61, 63

PICCIONE, M. (2002): "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, 102(1), 70–83. 68

SANNIKOV, Y., AND A. SKRZYPACZ (2007): "Impossibility of Collusion under Imperfect Monitoring with Flexible Production," *American Economic Review*, 97(5), 1794–1823. 78, 79

SCHMIDT, K. M. (1993): "Reputation and Equilibrium Characterization in Repeated Games of Conflicting Interests," *Econometrica*, 61(2), 325–351. 36, 37, 40, 41, 44

SEKIGUCHI, T. (1997): "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory*, 76(2), 345–361. 68

SORIN, S. (1999): "Merging, Reputation, and Repeated Games with Incomplete Information," *Games and Economic Behavior*, 29(1/2), 274–308. 8

WISEMAN, T. (2009): "Reputation and Exogenous Private Learning," *Journal of Economic Theory*, 144(3), 1352–1357. 32