Sustainable Reputations with Rating Systems*

Mehmet Ekmekci†
Northwestern University
Kellogg School of Management, MEDS.

First version: June 2007
This version: March 2010

Abstract

In a product choice game played between a long lived seller and an infinite sequence of buyers, we assume that buyers cannot observe past signals. To facilitate the analysis of applications such as online auctions (e.g. eBay), online shopping search engines (e.g. BizRate.com) and consumer reports, we assume that a central mechanism observes all past signals, and makes public announcements every period. The set of announcements and the mapping from observed signals to the set of announcements is called a rating system. We show that, absent reputation effects, information censoring cannot improve attainable payoffs. However, if there is an initial probability that the seller is a commitment type that plays a particular strategy every period, then there exists a finite rating system and an equilibrium of the resulting game such that, the expected present discounted payoff of the seller is almost his Stackelberg payoff *after every history*. This is in contrast to Cripps, Mailath and Samuelson (2004), where it is shown that reputation effects do not last forever in such games if buyers can observe all past signals. We also construct finite rating systems that increase payoffs of almost all buyers, while decreasing the seller's payoff.

Keywords: Reputations, Rating Systems, Online Reputation Mechanisms, Disappearing Reputations, Permanent Reputations.

JEL: *D82*

*This paper is based on the first chapter of my Ph.D. thesis when I was at Princeton University. I am indebted to Faruk Gul for guidance at every stage of this work. I am also grateful for suggestions and comments to Wioletta Dziuda, Aytek Erdil, Gabor Virag and Andrea Wilson. I also thank Nuh Aygun Dalkiran and Siddarth Madhav for research assistance. All remaining errors are my own.

†2001 Sheridan Road, Evanston, IL 60208. E-mail: m-ekmekci@northwestern.edu.

Introduction

In some markets trade requires trust. Shopping from an online store, buyers make the payment before they receive the product, and the store "promises" to deliver the product on time as advertised. Some goods have characteristics that reveal themselves only after they are used for some period of time, such as resistance, or endurance. At the time of the payment a user cannot observe these characteristics. Some examples of such markets are internet auction sites such as eBay.

Legal punishment systems constitute one way of achieving trust in bilateral trade. The buyer can go to the court, and claim his rights if the seller cheats. Hence, the seller can commit himself to following the requirements of the agreement because cheating is costly to him. However these punishment systems are in general too costly and in some situations it is difficult to prove that the seller has cheated because of incomplete contracts or imperfect monitoring.

One notable feature of online markets is that after every transaction the buyer and seller may leave feedback about the transaction, and this feedback becomes part of their identity which is publicly available to all users. In most cases, a seller does not encounter with the same buyer again, so a potential buyer cannot use his own past experience to evaluate the seller. However, the feedbacks are publicly observable, so the potential buyer learns past performance of the seller and decides whether to buy the product or not buy the product. If the seller cheats he receives bad feedback, and future buyers will not be willing to trade with him. So the seller's incentive to cheat is weakened or even may be eliminated by the loss of potential future profits that cheating will cause.

However, feedbacks carry only partial information about the intended action of the seller. Therefore these settings carry various potential inefficiencies. Imperfect monitoring may lead to punishments (through no trade) after unlucky bad signals although the seller didn't cheat. This in turn will cause inefficiencies that can be as severe as a total breakdown of all trade activity. Fudenberg and Levine [12], Fudenberg, Levine and Maskin [13] and Dellarocas [8] study examples of these situations and analyze possible inefficiencies.

Fudenberg and Levine [11] show that in such games even a small amount of uncertainty about the type of the seller might be enough to sustain cooperation and induce trust, at least in the early stages of a long-term relationship. The seller can achieve payoffs arbitrarily close to what he would get if he could commit himself publicly to playing a particular strategy in the stage game, even if the actions are observed imperfectly. Their finding is very strong however the implications don't last forever: i) The payoff is calculated at the beginning of the game (first period). ii) How the game is going to be played in the distant rounds of the game is not explored.

The long run features of the equilibrium play of these games have been explored only recently by Cripps, Mailath and Samuelson [6] (CMS hereafter). They show that reputation effects do not last forever in repeated games with one sided private information and imperfect monitoring: i) The payoff of the long run player will eventually be close to an equilibrium payoff of the repeated game without any uncertainty on the type of the long run player. ii) The play of the game on the equilibrium path will be eventually very similar to that of the repeated game without any uncertainty on the type of the long run player.

CMS further suggest that one should incorporate some other mechanism into the model in order to prevent reputation effects from disappearing. The purpose of this paper is to propose such a mechanism in a model of repeated moral hazard problems with incomplete information. The mechanism chooses the information transmission rule to increase efficiency. One way to alleviate the moral hazard problem would be to design transfers conditional on signals to alleviate moral hazard problem. However, we don't observe sellers being rewarded by monetary transfers for having good reviews by the buyers. In our model, sellers are punished or rewarded by the promises of the future relationship. In the equilibrium we construct, higher ratings correspond to higher equilibrium values for the seller and higher ratings are gained with more effort. So one motivation for this paper is to explain why and how information censoring may serve to facilitate (almost) efficient trade forever.

The examples in the literature that sustain non-disappearing reputation effects assume that the type of the player is governed by a stochastic process through time, rather than being determined once and for all at the beginning of the game. This is called replacement. Holmstrom [14], Cole, Dow and English [5], Mailath and Samuelson [21], and Phelan [23] maintain permanent reputations by assuming particular types of replacement in their models. Wiseman [25] and [26] also study whether reputations are sustainable in models where learning is exogenous.

In this paper we propose a new mechanism that determines the information each short run player observes about the past play of the game. In particular, a central authority that observes the full sequence of past signals *censors* the information that the short run players observe. We show that it is possible to censor the information in a way that enables the long run player to build reputation at all times.

That decreasing the quality of the observed signals may help facilitate efficient trade and alleviate moral hazard problem appears also in Kandori and Obara [16]. Kandori [15] shows that the set of pure-strategy sequential equilibrium payoff set of a repeated game where players have a common discount factor expands when the quality of the signals observed in each period improves in Blackwell's sense. The main difference between our paper and these papers is that in our model there is incomplete information. Moreover, the information censoring technology we allow for is much richer than these papers study.

In contrast to the literature, we assume that a rating system (e.g. a central computer) observes the play of the game each period (that is either all actions, or the realizations of the signals). Short run players do not have any information about the past play of the game other than what the rating system provides. Hence the rating system may censor the information observed by the short-run players.

Information censoring could be done in many ways. Showing only summary statistics about past performance data, like the average performance or time weighted average of past data, showing only the most recent data, refining the performance data into a binary form and showing the sum of the past n performances are some examples. There are various examples of information censoring in practice. Institutions that make consumer reports collect data about a product, or a firm over time, and these data are "processed" before the customer sees them. Every new piece of information is not reflected in the report in its most transparent form. Another example is "shopping search engines". These engines give scores to online stores based on the information they collect about them either by customer reviews or by shopping from the stores themselves. The scores are updated as new information arrives, and usually the customer does not have access to all pieces of information separately. Online marketplaces such as eBay also provide censored information about the seller and the buyer. Sometimes they show only feedbacks given during the most recent month. In other cases each user is given a score, an integer between 1 and 100, that is the difference between the number of good and bad feedbacks in the last 6 months.

In this paper we focus on a particular censoring method. A finite rating system describes past performance with a number from a finite rating set $S = \{0, 1, ..., n\}$ as follows. At the initial period the party to be rated is assigned a random number from the rating set. At the end of each period the rating may decrease or increase by one, or stay the same depending on the signal of the current period. If the current rating is already the lowest (highest) possible rating, then at the end of the period the rating either stays the same or increases (decreases) by one. The rule governing the transition from one rating to another after an observed performance level (e.g. signal, feedback, etc.) is called a "transition rule".

Dellarocas [8], Bakos and Dellarocas [4] study a repeated model of bilateral exchange environment with a moral hazard problem on the long run player's side. In their model the seller is the long run player and the buyers are short run players. There is no incomplete information (such as reputation effects) in their model. Short run players do not observe the outcomes of past play. Instead, an institution called a reputation mechanism observes past outcomes with some imperfection, and may disclose some or all of this information. They use the techniques developed in Fudenberg and Levine [12] to show that maximum efficiency can be attained by a two state randomization device, and that any attainable payoff vector is bounded away from the Pareto frontier of the stage game. They also show that if the monitoring imperfection is sufficiently large, then the resulting inefficiency may be as severe as the destruction of all trade possibilities. A more detailed coverage of reputation models with moral hazard problems can be found in Bar-Isaac and Tadelis [3] and in Bar-Isaac [2].

In our model there is incomplete information about the type of the long run player. One of these types is a "normal" type that has the usual payoff structure and has the moral hazard problem. The other types are commitment types that play the same strategy every period, either for morality reasons or because they are boundedly rational, or simply because their payoff structure is different.

Information disclosure by rating systems serves a dual role in our model. The first one is learning the type of the long run player and transmitting this information to the short run players. This will enable the short run players to learn the true type of the seller they are facing with a high probability. Unless the seller fully mimics a commitment type the rating system learns his true type after observing sufficiently many signals. If this information is fully disclosed to the buyers, and if it turns out that the seller is not a commitment type with high probability, then the equilibrium play is almost like that of complete information game, and inefficiency is inevitable. The rating system's second role in our model is to "forget" some of the past data, and allow the normal type of the long run player to build a reputation even in the distant future.

In theorem 0, we show that without the possibility of commitment types, the equilibrium payoff of the long run player can't be improved by the introduction of a rating system. When combined with the inefficiency results in repeated moral hazard games with imperfect monitoring (see for instance Fudenberg and Levine [12]), this theorem highlights that rating systems can facilitate indefinite efficient play only when there is incomplete information.

In theorem 1, we construct a particular finite rating system. We show that under a mild assumption on the commitment types, our rating system allows the long run player to get almost his Stackelberg payoff after every history in an equilibrium of the game. In the reputation literature, there are very few equilibrium constructions and the properties of the equilibrium play in such games is not yet well understood. We are also able to construct the equilibrium itself and along the equilibrium long-run player plays his Stackelberg strategy at almost every period. Our main contribution is to show that information censoring can provide a mechanism by which reputation gives commitment power to the agent even in the distant future.

Theorem 2 is about the payoffs of the short run players. This is a crucial point ignored in previous studies on reputation. Theorem 2 says that for sufficiently low frequencies of the commitment type(s), the seller can be made to commit to any effort level less than the effort level of the most hard working commitment type available in the type space. We also show that for each point on the Pareto frontier of the underlying game, there exists a finite rating system that implements that point in the long run.

Next we introduce our model with and without incomplete information. We present our main results on permanent reputations and welfare of buyers. We conclude with a discussion of the assumptions of our model.

Model

We study an infinitely repeated moral hazard game between one long run player (player 1) and an infinite sequence of short run players. Each short run player lives for one period and serves as player 2 in the following stage game.

The Stage Game

Let $A_1 = \{L, H\}$, $A_2 = \{B, N\}$ be the set of actions available to player 1 and player 2 in the stage game. An action $a_1 \in A_1$ describes player 1's effort level. Player 1 can either exert high effort H or low effort L. The action $a_2 \in A_2$ specifies whether player 2 buys the product or does not buy the product. Players move simultaneously in the stage game.

Let $A = A_1 \times A_2$ be the set of action profiles. For any finite set X, let $\Delta(X)$ denote the set of all probability distributions over X. In particular, $\Delta(A_k)$ is the set of all mixed strategies in the stage game for $k \in \{1, 2\}$. Let s_k denote a generic element of $\Delta(A_k)$. We will refer to $s_1(H)$ as player 1's effort level. Without risk of confusion we write $a_k \in A_k$ to denote the mixed strategy s_k such that $s_k(a_k) = 1$.

Player k receives payoff $U_k(s)$ when the stage game strategy profile is $s = (s_1, s_2)$, where $U_k(s) = \sum_{(a_1,a_2)\in A} u_k(a_1, a_2)s_1(a_1)s_2(a_2)$ and $u_k : A \to \mathbb{R}$. Let B_2 denote the best response correspondence of player 2. That is,

$$B_2(s_1) = \{s_2 \in \Delta(A_2)|U_2(s_1, s_2) \geq U_2(s_1, s_2') \text{ for all } s_2' \in \Delta(A_2)\}.$$

Let $V_1(s_1)$ denote the best commitment payoff for player 1 given s_1. That is;

$$V_1(s_1) = \max_{s_2 \in B_2(s_1)} U_1(s_1, s_2)$$

Similarly, we define follower's payoff as $V_2(s_1) = \max_{s_2\in\Delta(A_2)} U_2(s_1, s_2)$. Let V_1^s denote the Stackelberg payoff for player 1: $V_1^s = \max_{s_1\in\Delta(A_1)} V_1(s_1)$.

The following conditions on the payoffs characterize the moral hazard games we consider. We discuss why we focus on such games and require the conditions in the Discussion section.

Condition 1 *(constant effort cost)* $u_1(L, a_2) - u_1(H, a_2) = c > 0$ *for* $a_2 \in A_2$.

Condition 1 says that exerting high effort is costly and the cost is constant across all actions of player 2. This condition implies that exerting low effort is a dominant strategy for player 1.

Condition 2 $u_1(a_1, B) - u_1(a_1, N) > 0$ *for each* $a_1 \in A_1$.

Condition 2 says that holding player 1's effort level fixed, he prefers that player 2 buys the product.

Condition 3 $\exists \alpha^s \in (0, 1)$ *s.t. for* $s_1(H) > \alpha^s$, $B_2(s_1) = \{B\}$ *and for* $s_1(H) < \alpha^s$, $B_2(s_1) = \{N\}$.

This condition says that if player 1 exerts enough effort, then player 2 buys the product. Observe that conditions 1 and 3 imply that in the unique Nash Equilibrium of the moral hazard game, player 1 exerts low effort and player 2 does not buy.

Condition 4 $u_1(H, B) > u_1(L, N)$.

Condition 4 says that player 1 would prefer committing to exerting high effort to the unique Nash equilibrium payoff. Moreover Conditions 1, 3 and 4 imply that the Stackelberg strategy is unique and puts probability α^s to the action H.

Signal Structure

Short run players do not observe the action of player 1, but observe a public signal that is correlated with player 1's action. The public signal is denoted y, and is drawn from a finite set of realizations, Y. Let $\rho_{a_1} \in \Delta(Y)$ denote the probability distribution of the signal given player 1's action a_1.[1] Hence the probability of observing signal y given a_1 is $\rho_{a_1}(y)$. Note that perfect monitoring model is a special case of our model. We impose the following condition on the signal structure.

Assumption 1 *(Identification) If* $a_1 \neq a_1'$ *then* $\rho_{a_1} \neq \rho_{a_1'}$.

This assumption implies that if the long run player chooses the same action in each stage game, then an outside observer would eventually be able to learn this action.

[1] We implicitly assume that the signal distribution is independent of player 2's action. This is purely for expositional simplicity and to make the construction of the rating system relatively simpler.

Repeated Game with Rating System

The stage game described above is infinitely repeated. Player 1's discount factor is $\delta < 1$. Each short run player (player 2) can only observe the outcome in his own period.

A rating system observes the action of player 2, and the signal y every period. The rating system delivers a message $r \in R$ to the current period players before they play the stage game. The set of messages R is a countable set. We depart from the literature by assuming that a short run player cannot observe past play of the game. However, before the stage game is played, he learns the public announcement made by the rating system (r) that may provide some information about the past play. Player 1 knows his own actions, public announcements, and the actions of the short run players.

Player 1, before playing period t action has a private history, consisting of the messages, actions, and signals. Let $H_1^0 = R$, and a private history for player 1 is defined recursively as $h_1^t \in H_1^t \equiv (R \times A \times Y) \times H_1^{t-1}$ for $t \geq 1$. Player 2 at period $t \geq 0$ has a private history $h_2^t \in H_2 \equiv R$. The timing of the flow of information and actions at period t is summarized below:

1- Rating system makes a public announcement $r \in R$.

2- Player 1 and player 2 choose their actions simultaneously.

3- A signal $y \in Y$ is realized.

Signal y can be interpreted in various ways. It can be a payoff relevant variable (e.g. quality of a product, how satisfactory the service was) or a payoff irrelevant variable (e.g. feedbacks from player 2, outcomes of auditing reports). In our model, it is crucial that the signal is not observed by the population. Therefore we focus on applications in which this assumption is more easily satisfied, such as online markets or consumer reports.

A rating system maps histories that consist of past actions of player 2, signals, and its own past messages into a probability distribution over a set of messages. Among several such mechanisms, we will look for an information transmission mechanism that facilitates (almost) efficient trade. For our results it will be sufficient to focus on those that have a Markovian structure. In particular, we will use rating systems whose message at period t (r^t) evolves according to a Markov process that we call a Markov transition rule. In general, the transition probabilities will depend on the action of the short run player and the observed signal, however, for our results it will suffice to consider transition probabilities that only depend on the observed signal. Below we give the formal definition of a Markov transition rule.

Definition 1 *Let R be a countable set. A map $\tau : R \times Y \rightarrow \Delta(R)$ is called a Markov transition rule.*[2]

[2] Note that more generally a Markov transition rule is a mapping $\tau : R \times A_2 \times Y \rightarrow \Delta(R)$. Since we

Formally, a rating system consists of a set of messages called a rating set, an initial distribution on the rating set and a Markov transition rule on the rating set. At time 0, the rating is determined by the initial distribution, and every period the rating is publicly announced before players take their actions. After the actions are taken, the new rating is determined by the previous rating and the outcome of the play.

Definition 2 *A triplet $\phi = (R, \tau, p_0)$ is called a rating system where R is a countable set, τ is a Markov transition rule, and $p_0 \in \Delta(R)$.*

At time 0, a rating system announces a message r_0 drawn according to an initial distribution p_0. At the end of each period t, the random variable $r^{t+1} \in R$ is realized by the transition rule τ, the signal at period t, and $r^t \in R$. At the beginning of period $t+1$, r^{t+1} is announced to period $t+1$ players. We say that a rating system is finite if the message space R is a finite set.

The Incomplete Information Game

A type space $W = \{0, 1, ..., l\}$ is a finite set. Prior to time $t = 0$, nature chooses a type for player 1 according to a probability distribution $\eta \in \Delta(W)$. We will use η_j for $\eta(j)$, and assume without loss of generality that $\eta_j > 0$ for $j = 0, 1, ..., l$. The normal type of player 1 has the payoff structure as described in the stage game, and type 0 represents the normal type. Each type $j > 0$ represents a commitment type that plays H with a probability α^j every period. Let the index set be ordered such that $\alpha^j > \alpha^{j-1}$ for $j = 2, 3...l$. We will call the strategies of the commitment types s^j, that is $s^j \in \Delta(A_1)$ where $s^j(H) = \alpha^j$. We call $\Gamma = (W, \eta)$ a type model and assume that $B_2(s^1) = \{B\}$. Also, when we use the term player 1 without specifying a type, we mean type 0 of player 1.

Let $H = \bigcup_{t=0}^{\infty} H_1^t$ be the set of all finite histories of player 1. Player 1's strategy is a map $\sigma_1 : H \to \Delta(A_1)$. Player 2's strategy is a collection of maps $\sigma_2 = \{\sigma_2^t\}_{t=0}^{\infty}$ where $\sigma_2^t : R \to \Delta(A_2)$. The strategy spaces of players 1 and 2 are Σ_1 and Σ_2 respectively.

For a given rating system ϕ, each strategy profile (σ_1, σ_2) together with the type model Γ induces a probability distribution P over all action profiles, signals and messages. The payoff to player 1 and period-t player 2 of a strategy profile (σ_1, σ_2) are:

$$U_1(\sigma_1, \sigma_2) = E^P((1-\delta)\sum_{t=0}^{\infty} \delta^t u_1(a_1, a_2)|\text{player 1's type is 0})$$

$$U_{2t}(\sigma_1, \sigma_2|r^t) = E^P(u_2(a_1, a_2)|r^t)$$

assumed that the probability distribution function on the signals, ρ, is independent of player 2's actions, and the constant cost of exerting high effort, a transition rule that depends only on the observed signal is sufficient to both learn the type of player 1 and give the incentives to player 1 to exert high effort.

We also define the payoff of player 1 after a history $h_1^t \in H_1^t$ as:

$$U_1(\sigma_1, \sigma_2 | h_1^t) = E^P((1-\delta)\sum_{t'=t}^{\infty} \delta^{t'-t} u_1(a_1, a_2) | h_1^t, \text{player 1's type is 0})$$

We will call a game $G(\delta, W, \eta, \phi)$, that has the payoffs and strategy spaces as described above, a repeated incomplete information game with rating system ϕ. We will use $G(\delta, \Gamma, \phi)$ interchangeably with $G(\delta, W, \eta, \phi)$, where $\Gamma = (W, \eta)$.

Having described the payoffs and the strategy spaces, we can now define the equilibrium concept we will be using. For a history $h \in H$ of player 1, we denote the latest message in h by $r(h)$.

Definition 3 *Strategies* (σ_1, σ_2) *are Perfect Bayesian equilibrium of* $G(\delta, \Gamma, \phi)$ *if* $\forall t = 0, 1, ...,$
i) $U_1(\sigma_1, \sigma_2 | h_1^t) \geq U_1(\sigma_1', \sigma_2 | h_1^t)\ \forall\ \sigma_1' \in \Sigma_1\ , \forall h_1^t \in H$.
ii) $U_{2t}(\sigma_1, \sigma_2 | r^t) \geq U_{2t}(\sigma_1, \sigma_2' | r^t)\ \forall \sigma_2' \in \sum_2,\ \forall r^t \in R$ *with* $P(r^t) > 0$.
iii) Let $h(r^t) = \{h_1^t \in H_1^t | r(h_1^t) = r^t\}$. *If* $P(r^t) = 0$ *and* $h(r^t) \neq \{\emptyset\}$, $\sigma_{2t}(r^t)$ *is a best response to a belief* κ *where* $\kappa \in \Delta(h(r^t))$, *and player 1's type is 0.*

This is a game of incomplete information therefore the proper solution concept is Perfect Bayesian equilibrium. Players are required to update their beliefs using Bayes' rule whenever possible. Player 1 knows his type, and there is no incomplete information on his side, thus he does not perform any Bayesian updating. However, short run players learn some information about the type of player 1 through the messages delivered by the rating system. They use the message to form their beliefs about player 1's type. But Bayesian updating is not possible after histories with probability zero. In our game, the rating system is not a player, so any deviation from the equilibrium is attributed to players' past behavior. We require the short run players to play a best response to some arbitrary belief on player 1's histories that are consistent with the message they observed (if possible), and a belief that player 1's type is the normal type.[3]

The equilibrium restriction (iii) might seem weak because it gives too much freedom in how we can choose beliefs of the short run players. Therefore the equilibrium set might be larger than a more strict equilibrium concept such as sequential equilibrium. However, we will design our rating systems in such a way that all histories for player 2 occur with positive probability given any strategy profile. So we will not make use of out of equilibrium beliefs to support equilibria we construct.

[3] Obviously the beliefs are not defined after messages that are not reachable by any history. Note that there are some messages that wouldn't be possible regardless of players' strategies. For example, suppose the rating system starts at rating 1 at period 0, and moves either one rating up or stays the same after any signal. Then a rating of 3 is not possible at period 1.

Complete Information Case

In this section we analyze the complete information case of our model. This section serves as a benchmark to see the effects of information censoring on the equilibrium payoff set when there is no room for reputation formation. Information censoring can only decrease the equilibrium payoff set compared to a model in which information is not censored at all under complete information.

Most of the literature on repeated games assumed that short run players could observe all of the past signals.[4] However, we drop this assumption and instead assume that there is a rating system that observes signals every period and conveys some information to the short run players every period. In fact these models are special cases of our model, where the rating system delivers all information about past signals to every short run player. These rating systems are called transparent rating systems and are denoted ϕ^*. The formal definition of transparent rating systems (that is, message space, transition rule, and initial state) is given in the appendix.

The messages of a transparent rating system include all information about past signals of the game. We allow for public randomizations and each short run player observes the outcomes of these randomizations as well. When a rating system is transparent, the informational assumptions of our model coincide with those in the standard reputation literature where all past signals and actions of player 2 are observable by the current period players.

If there is no incomplete information about player 1's type, Perfect Bayesian equilibrium puts the same restrictions as Perfect equilibrium.

Theorem 0 *Let ϕ be a rating system, and $W = \{0\}$. The payoff to the long run player in any Perfect Bayesian equilibrium of $G(\delta, W, \eta, \phi)$ is no more than the highest payoff he can get in some Perfect Bayesian equilibrium of $G(\delta, W, \eta, \phi^*)$ for some transparent rating system ϕ^*.*

Proof. See appendix. ■

This result says that rating systems can do no better than disclosing all information if there is no incomplete information about the type of player 1. The strength of this theorem is most emphasized when used with theorem 6.1 of Fudenberg, Levine [12], which says that the payoff to the long run player is generically bounded away from his most preferred commitment payoff. We use this result as a benchmark for discussing reputation effects in our model.[5]

[4] Liu [19] and Liu and Skrzypacz [20] assume a different information structure than the earlier models. Liu [19] assumes that a short run player pays a cost to learn about previous signals, and Liu and Skrzypacz [20] assume that each short run player can recall only a finite number of past signals.

[5] Although theorem 0 focuses on rating systems, the result is still valid if we allowed for more complicated

Permanent Reputations with Rating Systems

Having discussed the inability of rating systems to increase efficiency under complete information, we will now assume that the type space W has commitment types. Each of these types plays a particular stage game strategy every period, independent of the history. When there is incomplete information about the type of player 1, he will be able to build a reputation by imitating a certain type, or trying not to look like a particular type. Our result is that rating systems are able to give commitment power to the long run player forever. First a definition of a subset of type space is given.

Definition 4 *A type space W is called a Stackelberg space if $\alpha^l > \alpha^s$.*

Stackelberg spaces have a commitment type that exerts more effort than the Stackelberg strategy. We present our main result below.

Theorem 1 *For any Stackelberg space W, η, and $u < V_1^s$, there exists $\overline{\delta} < 1$ such that for $\delta \geq \overline{\delta}$ there is a finite rating system ϕ, and a Perfect Bayesian equilibrium of $G(\delta, W, \eta, \phi)$ where the payoff to the normal type of the long run player is at least u after every history.*

Proof. See appendix. ■

The result says that with a suitable choice of a finite rating system, there is a Perfect Bayesian equilibrium of the game in which the long run player can get almost his most preferred commitment payoff after every history. This does not mean that his stage game payoffs are always almost his commitment payoff after every history. On the equilibrium path, the long run player may get period payoffs that are less than his Stackelberg payoff. However these periods do not follow each other frequently enough, so his discounted payoff calculated at the beginning of every history becomes almost his Stackelberg payoff.

Although Theorem 1 is stated at the payoff level it has corresponding behavioral implications as well. On the equilibrium path, the frequency with which the long run player exerts his most preferred commitment effort level is almost 1 after every history. So reputation never ceases to give the long run player commitment power when information is censored. In reputation models it is very difficult to construct equilibria and most of the literature studies the payoffs of players in these games. Since our proof is by construction, we actually specify the equilibrium strategies of players during the play of the game. On the equilibrium path the long run player plays his Stackelberg strategy at most of the periods.

---

information transmission mechanisms, in particular all public institutions. A public institution is one that provides precisely the same information about its future behavior to a period's players. Every rating system is public, because the future information transmission rule is common knowledge among player 1 and period-t player 2 at period t. If the transition rule depends on some information that is not in the history of player 2 but is in the history of player 1, then the corresponding institution fails to be a public institution.

Previous studies approximated the payoff of the long run player by a payoff he could get by mimicking a particular commitment type in the support of the prior distribution of commitment types. However in our model, we only require the existence of a commitment type that exerts more effort than the Stackelberg strategy. The rating system and the equilibrium has the property that at the lowest rating there is no trade, and at all other ratings there is trade. At the highest rating, player 1 exerts low effort, and player 2 still buys the product. Observe that existence of such a "reward" phase would be impossible if there were no commitment types.

However note that even when there are commitment types, if they are not likely, then player 2 would not buy the product if the normal type were to exert low effort. The rating system, along with the equilibrium strategies separate the normal type from the commitment types, and the probability of the commitment type being at a reward phase is much higher than the probability with which the normal type is at a reward phase. Hence even if the initial probability of a commitment type is very low, observing the reward phase player 2 will believe that her opponent is most likely the commitment type.

The rating systems we construct are finite. Therefore the rating system and the equilibrium we construct are not very complex. We consider the simplicity of our reputation equilibria as a virtue.

The reason why our payoff result holds after every history is that the finite rating system is a stochastic automaton with finite states, hence player 2's beliefs about her opponent's types stay bounded away from the boundaries of the unit simplex. To contrast with CMS's result, the posterior beliefs of short run players converge to the boundaries, therefore after some time the amount of uncertainty required to generate a reputation result for the fixed discount factor vanishes eventually. In our model however, the beliefs always stay away from the boundaries with a finite number of ratings. Obviously it is not true that every rating system that prevents beliefs from hitting the boundaries can give the correct incentives to player 1 to exert effort. In theorem 1 we construct the rating systems that achieve this.

Example and sketch of proof

There are 2 actions available for each player, $A_1 = \{H, L\}$ and $A_2 = \{B, N\}$. The actions are perfectly observed, hence the signals are the actions of player 1.

Player 1 is the row player, player 2 is the column player in the stage game with the payoff matrix given below:

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

In this game, the Stackelberg strategy of player 1 is to play H with probability 0.5, and his Stackelberg payoff is 1.5. The type space is $W = \{0, 1\}$ where type 0 is the normal type of player 1 and has the payoff structure above. Type 1 plays action H at every period of the game. The prior probability distribution of the type of player 1 is $\eta \in \Delta(W)$, where $\eta(0) = 2/3$, and $\eta(1) = 1/3$.[6]

In the repeated game with complete information, the highest payoff that player 1 can get in any equilibrium is 1 if he is sufficiently patient. Moreover in the game with incomplete information, if all past actions are observed by the current player 2, then player 1's highest equilibrium payoff is bounded above by $\delta + 2(1-\delta)$ which tends to 1 as δ tends to 1.[7] Our theorem says that for any payoff $u < 1.5$, there exists a rating system and an equilibrium of the game where player 1's payoff is at least u after every history of the game if player 1 is sufficiently patient.

Let $S = \{1, 2, ..., K\}$ be the rating set. Since the set of signals and the set of player 1's actions A_1 are the same for this example, we define the transition rule to be a map $\tau : S \times A_1 \to \Delta(S)$. The transition rule will depend on the discount factor δ, however we will first construct the transition rule for $\delta = 1$, and then we will specify how to modify it for $\delta < 1$.

step 1: Constructing τ for $\delta = 1$: In the table below, the vector in each cell corresponds to the probability of downgrade, no rating change and upgrade respectively. The term NA corresponds to "Not Available" and means that a downgrade when the rating is 1 and an upgrade when the rating is K are not defined. For example the cell with the vector $(1/8, 5/8, 1/4)$ means that $\tau(s, H)(s-1) = 1/8$, $\tau(s, H)(s) = 5/8$ and $\tau(s, H)(s+1) = 1/4$.

τ	$s = 1$	$s \in \{2, ..., K-1\}$	$s = K$
$y = H$	$(NA, 5/8, 3/8)$	$(1/8, 5/8, 1/4)$	$(1/8, 7/8, NA)$
$y = L$	$(NA, 5/8, 3/8)$	$(1/8, 7/8, 0)$[10]	$(1/8, 7/8, NA)$

Table 1: Transition Rule.

step 2: Equilibrium strategies Consider the strategies $\sigma_1 : S \to [0, 1]$ indicating the probability of H for player 1 and $\sigma_2 : S \to [0, 1]$ indicating the probability of B for player 2:

$$\sigma_1(1) = 0 \text{ and } \sigma_1(K) = 0 \text{ and } \sigma_1(s) = 0.5 \text{ for } s \in \{2, 3, ..., K-1\}.$$

[6] Note that in this game the actions are perfectly observed, and hence CMS's result is not applicable. However we chose this example for expositional simplicity.

[7] To see this, observe that player 1's continuation payoff at any history where he played L at least once is at most 1, and the highest stage game payoff player 1 can receive while playing H is at most 1. Therefore, until he plays L he gets at most 1 per period, and when he plays L he gets at most 2 for that period and 1 for the continuation payoff delivering the bound.

$$\sigma_2(1) = 0 \text{ and } \sigma_2(s) = 1 \text{ for } s \in \{2, 3, ..., K\}.$$

The equilibrium that we are constructing can be thought of as having 3 phases: a punishment phase, a normal phase and a reward phase.

Rating 1 serves as the punishment phase. At this rating, the transition rule doesn't depend on the observed signals, therefore player 1 plays L, and player 2 plays N.

The middle ratings serve as the ratings where player 1 is indifferent between actions H and L. In the equilibrium we construct, player 1 plays H and L with equal probabilities (Stackelberg strategy) in these ratings. This strategy, together with the transition rule implies that the probability of an upgrade (which is 1/8) is equal to the probability of a downgrade (1/8) for the normal type of player 1. The strategy of the commitment type is fixed (play H w.p. 1), therefore the probability of an upgrade (1/4) is strictly higher than the probability of a downgrade (1/8) for the commitment type in this set of ratings.

Rating K serves as the reward phase for player 1. In this rating the normal type of player 1 plays L and player 2 still plays B. This is possible since in equilibrium the frequency with which the commitment type visits the reward rating is much more than that of the normal type. In this rating the transition rule is insensitive to the signals, hence L is optimal for player 1.

step 3: Number of ratings The transition rule when coupled with the equilibrium strategies of the normal type and the commitment type induce two Markov transition matrices P^0 and P^1 on the set of ratings. In particular $P^k[i, j]$ is the probability that type k moves from rating i to rating j.

$$P^0 [i, j] = \sigma_1(i)\tau(i, H)(j) + (1 - \sigma_1(i))\tau(i, L)(j)$$

$P^1 [i, j]$ is as above where $\sigma_1(i)$ is replaced by 1. These matrices are positive recurrent and irreducible, so unique stationary distributions π^0 and π^1 exist for P^0 and P^1. In the long run, player 2 calculates the probability $p(s)$ that player 1 plays action H at rating s as below:

$$p(s) = \frac{\pi^1(s)\eta(1)}{\pi^0(s)\eta(0) + \pi^1(s)\eta(1)}1 + \frac{\pi^0(s)\eta(0)}{\pi^0(s)\eta(0) + \pi^1(s)\eta(1)}\sigma_1(s)$$

Let $pr(k, s) = \frac{\pi^k(s)\eta(k)}{\pi^0(s)\eta(0)+\pi^1(s)\eta(1)}$ be the probability a short run player who observes a rating s assigns to player 1 being a type k. The reputation of player 1 for being a commitment type is $pr(1, s)$, and is increasing in the rating s. σ_2 would be (strictly) optimal for player 2 if $pr(1, K) > 0.5$ and $pr(1, 1) < 0.5$.[11] We will choose next the number of ratings, K, in a

[11]Note that player 2 plays B only if she believes that player 1 plays H w.p. at least 1/2. Since we want

way that makes the conditions $pr(1,K)>0.5$ and $pr(1,1)<0.5$ true.

For Markov chains where there are no jumps in the transitions, the following fact will be very useful:

$$\pi^k(s)P^k[s,s+1]=\pi^k(s+1)P^k[s+1,s] \tag{1}$$

Since $P^0[s,s+1]=P^0[s+1,s]$ for $s\in\{2,3,...,K-2\}$, $\pi^0(s)=\pi^0(s')$ for $s,s'\in\{2,...,K-1\}$ using equation 1. Therefore $\pi^0(s)<1/(K-2)$ for $s\in\{2,3,...,K-1\}$. Moreover since $P^0[K-1,K]=1/8$ and $P^0[K,K-1]=1/8$, $\pi^0(K)=\pi^0(K-1)<1/(K-2)$.

For the commitment type, $P^1[s,s+1]=2P^1[s+1,s]$ for $s\in\{2,3,...,K-2\}$, $P^1[1,2]=3P^1[2,1]$ and $P^1[K-1,K]=2P^1[K,K-1]$. Note that as the number of ratings K increases $\pi^1(K)$ tends to a strictly positive number but $\pi^0(K)$ tends to zero.[12] Therefore $\lim_{K\to\infty}\frac{\pi^1(K)}{\pi^0(K)}=\infty$, and for sufficiently large K, $pr(1,K)$ becomes larger than 0.5. Moreover, $\pi^1(1)$ tends to zero faster than $\pi^0(1)$, hence $pr(0,1)$ tends to 1 as K gets very large. Hence we can choose a K such that $pr(1,K)>0.5$, and $pr(1,1)<0.5$. For the rest of the proof we fix K large enough so that $pr(1,K)>0.5$, and $pr(1,1)<0.5$.

step 4: Perturbing τ to give the correct incentive to player 1 If the beliefs of player 2 are as above, then σ_2 is a strict best response, so even if they are not exactly as above but very close, σ_2 is still a strict best response. For the normal type of player 1 however for $\delta<1$, the equilibrium strategy proposed above is not optimal. We will construct a new transition rule, τ^δ, so that the difference in the continuation values of adjacent ratings is exactly equal to $4(1-\delta)/\delta$. This is indeed the difference in the continuation values in the normal region that is necessary to make player 1 indifferent between playing H and L.[13]

Let $\tau^\delta(s,\cdot)(s-1)=\tau(s,\cdot)(s-1)+(K-s)(1-\delta)/\delta$ and $\tau^\delta(s,\cdot)(s)=\tau(s,\cdot)(s)-(K-s)(1-\delta)/\delta$ for $s>1$, and $\tau^\delta(1,\cdot)(2)=\tau(1,\cdot)(2)-(K-1)(1-\delta)/\delta$ and $\tau^\delta(1,\cdot)(1)=\tau(1,\cdot)(1)+(K-1)(1-\delta)/\delta$. First observe that the perturbed transition probabilities above are well defined for δ sufficiently close to 1.[14] Since we already fixed K, $\tau^\delta\to\tau$ when

rating K to be a reward phase where player 2 plays B, and player 1 plays L, we need $pr(1,K)>1/2$ in order to be able to sustain such a behavior in equilibrium.

[12] To see this more clearly, note that $\pi^1(s)=\pi^1(K)(1/2)^{K-s}$ for $s>1$ and $\pi^1(1)=\pi^1(K)(1/2)^{K-2}(1/3)$. Since $\sum_{s\in S}\pi^1(s)=\pi^1(K)(\sum_{s=2}^{K}(1/2)^{K-s}+(1/2)^{K-2}(1/3))=1$, $\lim_{K\to\infty}\pi^1(K)>0$.

[13] To see that this is true we need to calculate the difference between expected continuation values to player 1 from playing H and L. When player 1 plays H, the rating is upgraded w.p. 1/4 and downgraded w.p. 1/8; when he plays L the rating is upgraded w.p. 0 and downgraded w.p. 1/8. The expected difference in the final rating under two actions is $(1/4-0)=1/4$, and the expected difference in the continuation values is $(1/4)\times(4(1-\delta)/\delta)=(1-\delta)/\delta$. This is exactly the continuation value difference needed to make player 1 indifferent between incurring a cost of 1 today and not incurring any cost.

[14] That is, each transition probability is between 0 and 1, and the transition probabilities from every rating sum up to 1 for δ sufficiently high.

$\delta \to 1$, and if player 1 plays according to the equilibrium strategy, the unique stationary distribution of the transition matrices induce beliefs over actions of player 1 such that σ_2 is still optimal when δ is close to 1.

Let's show that the strategy σ_1 is optimal in the long run. The strategy profile (σ_1, σ_2) is Markovian, so the probability distribution of the stream of future payoffs depends on the current state. Let $V(s)$ denote the expected present discounted value of the strategy profile to player 1 when the state is $s \in S$. These values are determined by the following recursive equations:

$$V(s) = \left[(1-\delta)u_1(\sigma_1(s), \sigma_2(s)) + \delta E(V(s')|s))\right] \text{ where}$$

$$\begin{aligned} E(V(s')|s) \quad = & \\ & \sigma_1(s)\left[\sum_{s' \in S} \tau^{\delta}(s,H)(s')V(s')\right] + \\ & (1-\sigma_1(s))\left[\sum_{s' \in S} \tau^{\delta}(s,L)(s')V(s')\right] \end{aligned}$$

We use the method of guess and verify to find $V(s)$. Our guess is $V(s) = 1.5 - \frac{4(1-\delta)}{\delta}(K - s)$ and this can be verified by putting these values in the above system of equations.[15] Once we know the continuation values, it is straightforward to check that σ_1 is indeed optimal.

step 5: Initial periods and convergence to the long run By now we have constructed a rating system, and showed that an equilibrium strategy profile is long-run optimal, that is if the game started at $-\infty$ and has been going on for a long time and if the normal type of player 1 followed the strategy σ_1, then the belief of player 2 about the type of his opponent after observing a rating s would be the same as $pr(1,s)$. However when the game starts at period 0, player 2 believes that his opponent is a commitment type with a probability $\eta(1) = 1/3$ independent of the initial rating of player 1. Then if player 2 observes rating K during the early rounds of the game, σ_2 is not optimal given that the normal type plays action L with probability 1. We proceed as follows: Fix a large T where the first T periods serve as experimentation periods. During the experimentation periods, the normal type of player 1 plays σ_1, and when T is large enough, after period T the beliefs of player 2 are within an arbitrarily small neighborhood of $pr(1,s)$. The difficulty is to give player 1 the right incentives in these periods to play σ_1 since player 2 doesn't play

[15] To see this, one should calculate the payoffs at the 3 phases. We'll show how to calculate this for the normal region here: At rating s in the normal region, $V(s) = 2(1-\delta) + \delta(p^- V(s-1) + p^+ V(s+1) + (1 - p^- - p^+)V(s))$ where $p^- = 1/8 + (K-s)\frac{(1-\delta)}{\delta}$, $p^+ = 0$.

σ_2. We get around this by "banking" the payoff losses and gains of player 1 from what he would get otherwise if player 2 played σ_2. At the end of period T, the payoff gains and losses are calculated. Note that this number is bounded above by $T(1-\delta)2$, where 2 is the maximum payoff difference for player 1 in the stage game. We perturb the rating system after time T slightly so that the continuation payoff of player 1 exactly compensates for the losses and gains during the experimentation periods. [16]By using lemma 1 below about uniform convergence of Markov transition matrices, we make sure that σ_2 is still optimal in this slightly perturbed rating system. The amount of perturbation is at the order of $1-\delta$, therefore the limit result holds.

The strategy of player 2 is a pure strategy, and is a strict best response. So when the distribution of types across ratings is close to the stationary distributions, the strategy σ_2 is still a best response. Therefore, after a finite time T, σ_2 becomes optimal. We will define the transition rule in the initial periods in a way that player 1's strategy is still a best response at all times, and player 2 plays potentially a different strategy than σ_2 for the first T periods, but plays according to σ_2 after time T. We explain a precise definition of the transition rule and equilibrium strategies during these initial periods in the Appendix.

Lemma 1 *Let S be a finite set, and P be an irreducible and positive recurrent transition matrix on S with a stationary distribution $\pi \in \triangle(S)$. For every $\overline{\varepsilon} > 0$, there exists a $\varepsilon > 0$ and a T such that for any set of $T+1$ irreducible and positive recurrent transition matrices $\{P_i^*\}_{i=1,\dots,T+1}$ such that $\| P_i^* - P \|$ [17] $< \varepsilon$, and for any probability distribution $p \in \triangle(S)$, $\| p(\prod_{i=1}^{T} P_i^*)(P_{T+1}^*)^t - \pi \| < \overline{\varepsilon}$ for every positive integer t.*

Proof. Let π_{T+1} be the unique stationary distribution of P_{T+1}^*. We can choose T and ε such that $\| p(\prod_{i=1}^{T} P_i^*) - \pi_{T+1} \| < \overline{\varepsilon}/2$ for every $p \in \triangle(S)$. Also we can choose ε such that $\| \pi - \pi_{T+1} \| \leq \overline{\varepsilon}/2$. We have $\| p(\prod_{i=1}^{T} P_i^*)(P_{T+1}^*)^t - \pi \| \leq \| p(\prod_{i=1}^{T} P_i^*)(P_{T+1}^*)^t - \pi_{T+1} \|$ $+ \| \pi_{T+1} - \pi \| \leq \| p(\prod_{i=1}^{T} P_i^*) - \pi_{T+1} \| + \| \pi_{T+1} - \pi \| \leq \overline{\varepsilon}$. These inequalities follow from triangular inequality. ■

Lemma 1 says that for any irreducible and positive recurrent transition matrix P on a finite state space S, starting from any initial distribution, if we apply transition matrices very close to P for a number of periods T long enough, and then we continue by applying another transition matrix P' close to P, we will always stay close to π.

[16]Note that T is a finite number, and hence we can define the transition rule during the experimentation periods and after period T as part of a rating system with a large finite state space. Therefore our rating systems don't violate time independency.

[17]The metric is the total variation metric.

The Fate of Short run Players

So far studies concentrated on when and how reputation gives commitment power to the long run player. In many circumstances we care about the fate of the short run players, even those in the distant future. We may want to regulate the long run player to exert more effort than his most preferred commitment effort level. In our online shopping example, the rater might be interested in regulating the store to be trustable not only as much as is enough to induce the customers to buy, but more. The interests of a consumer report might be more aligned with those of the customers than those of firms.

The regulating power of rating systems depend on the type space. In general, the higher α^l, the larger is the set of implementable payoffs. We present our results when the probability of commitment types is very small.

Definition 5 *For a type space W, the set* $CP(W) = \{(V_1(s_1), V_2(s_1)) : s_1(H) < \alpha^l\}$ *is called the commitment payoff set of* W*. The set* $IRP(W)$ *is the convex hull of* $CP(W)$*, and is a subset of the set of individually rational payoffs of the stage game.*

$IRP(W)$ is the set of payoff vectors obtained when player 1 plays H with a probability $\alpha < \alpha^l$, and player 2 plays a best response to α.

Theorem 2 *For every* $\epsilon > 0$, $(U, V) \in IRP(W)$*, there exists* $\overline{\eta}_0 < 1$*, a natural number* T*,* $\overline{\delta} < 1$*, such that for* $\delta \geq \overline{\delta}$, $1 > \eta_0 > \overline{\eta}_0$ *there is a finite rating system* ϕ*, and a Perfect Bayesian equilibrium of* $G(\delta, W, \eta, \phi)$ *satisfying the following:*

i) the payoff to the normal type of the long run player is at least $U - \epsilon$ *after every history.*

ii) unconditional expected payoff of every short run player after period T *is at least* $V - \epsilon$*.*[18]

Proof. See appendix. ■

The very idea behind reputation is that a small amount of uncertainty on the type of the long run player can give him his commitment power. Our theorem says that if the uncertainty is indeed small enough, then we can regulate the long run player to exert any effort level α that is less than the effort level of the most hard working commitment type, α^l. Moreover, short run players distant enough in the future get an expected payoff of what they can get at best when the long run player's effort level is α. The short run players in the early rounds of the play do not get as much payoff as the ones in the distant future, because they are informationally inferior to them. The type of the long run player is almost revealed through the signals during these rounds, and in the distant future the effort level becomes close to α.

[18] Here unconditional refers to unconditional on the type of the long run player and on the rating at time T.

Discussion

In this paper we show that reputations can be sustained permanently if information about the past play of the game is regulated by a central mechanism. Recent results in the reputation literature pointed out that reputation effects do not last long if past outcomes of a long term play is observed by all players. We show that in a class of moral hazard games, reputations can be sustained and moral hazard can be alleviated (forever) if information is censored by a rating system.

The class of games we consider are restricted. We assume condition 1 (constant effort cost) in order to make the construction simpler. Together with the assumption that public signals don't depend on player 2's actions, this condition enables us to choose the transition rule of the rating system independent from player 2's actions. The number of actions available to buyers is restricted to two. When short run players have more than 2 actions available to them, then keeping track of their actions at every belief that can be generated in equilibrium becomes a very hard task.[19] Conditions 1,2 and 4 imply that we can have a reward phase where player 1 enjoys the play of the action profile (N, B). Condition 3 puts a monotonic relation between player 2's best response set and her beliefs about the probability with which player 1 plays H. This monotonicity allows us to more easily keep track of player 2's optimal bahavior when the rating system has a *one jump at a time* property.

References

[1] Abreu, D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041-1063.

[2] Bar-Isaac, H. (2003): "Reputation and Survival: Learning in a Dynamic Signalling Model," *Review of Economic Studies,* 70 (2), 231-251.

[3] Bar-Isaac, H., and S. Tadelis (2008): "Seller Reputation," *Foundations and Trends in Microeconomics, 4:4, 273-351.*

[4] Bakos, Y., and C. Dellarocas (2003): "Cooperation Without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms," *MIT Sloan Working Paper*, No. 4295-03.

[19] This can still be managed. In Ekmekci [9] we showed that this is possible (with a lot of additional notation and much more complicated rating system) with a certain monotonicity assumption in player 2's best response correspondence (similar to the current condition 3).

[5] 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, 365-385.

[6] Cripps, M. W., G.J. Mailath, and L. Samuelson (2004): "Imperfect Monitoring and Impermanent Reputations," *Econometrica*, 72, 407-432.

[7] Dellarocas, C. (2003): "The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49, 1407-1424.

[8] Dellarocas, C. (2004): "Efficiency and Robustness of Binary Feedback Mechanisms in Trading Environments with Moral Hazard," *MIT Sloan Working Paper*, No. 4297-03.

[9] Ekmekci, M. (2006): "Essays on Dynamic Games," Ph.D. Thesis, Princeton University.

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

[11] Fudenberg, D., and D. K. Levine (1992): "Maintaining a Reputation When Strategies are Imperfectly Observed," *Review of Economic Studies*, 59, 561-579.

[12] Fudenberg, D., and D. K. Levine (1994): "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, 62, 103-135.

[13] Fudenberg, D., and D. K. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997-1040.

[14] Holmstrom, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169-182.

[15] Kandori, M. (1992): "The Use of Information in Repeated Games with Imperfect Monitoring," *Review of Economic Studies,* 59 (3), 581-593.

[16] Kandori, M., and I. Obara (2006): "Less is More: an Observability Paradox in Repeated Games," *International Journal of Game Theory*, 34, 475-493.

[17] 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, 245-252.

[18] Kreps, D., and R. Wilson (1982): "Reputation and Imperfect Information," *Journal of Economic Theory*, 27, 253-279.

[19] Liu, Q. (2009): "Information Acquisition and Reputation Dynamics", working paper.

[20] Liu, Q., and A. Skrzypacz (2009): "Limited Records and Reputation", working paper.

[21] Mailath, G. J., and L. Samuelson (2001): "Who Wants a Good reputation?" *Review of Economic Studies*, 68, 415-441.

[22] Milgrom, P. R., and D.J. Roberts (1982): "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, 27, 280-312.

[23] Phelan, C. (2006): "Public Trust and Government Betrayal," *Journal of Economic Theory*, 127(1), 27-43

[24] Resnick, P., R. Zeckhauser, E. Friedman, and K. Kuwabara (2000): "Reputation Systems," *Communications of the ACM*, 43, 45-48.

[25] Wiseman, T. (2008): "Reputation and Impermanent Types," *Games and Economic Behavior,* 62, 190-210.

[26] Wiseman, T. (2009): "reputation and Exogenous Private Learning," *Journal of Economic Theory,* 144, 1352-1357.

APPENDIX

DEFINITIONS

Definition 6 *Let D be a countable set. A set of all finite histories is* $H_I(D) = \cup_{t=0}^{\infty}(D \times A_2 \times Y)^t \times D$.

Definition 7 *Let D be a countable set. A transition rule* $\tau : H_I(D) \times A_2 \times Y \to \Delta(H_I(D))$ *is a transparent transition rule if* $\sum_{d \in D} \tau(h, a_2, y)((h, a_2, y, d)) = 1$ *for every* $h \in H_I(D)$, $a_2 \in A_2$ *and* $y \in Y$.

Definition 8 *A rating system* $\phi = (H_I(D), \tau, p_0)$ *where τ is a transparent transition rule, and support of p_0 is D is called a transparent rating syatem, and is denoted* ϕ^*.

Proof of Theorem 0

We will show that for each equilibrium strategy profile (σ_1, σ_2) of the game $G(\delta, W, \eta, \phi)$, there exists a transparent rating system ϕ^* whose public randomization outcomes are the messages of the rating system ϕ, and there exists a strategy profile (σ_1', σ_2') that yields the same payoff to players 1 and 2, and is an equilibrium of the game $G(\delta, W, \eta, \phi^*)$.

Let (σ_1, σ_2) be a perfect equilibrium of $G(\delta, W, \eta, \phi)$. For $\phi = (R, \tau, p_0)$, let $D = R$, $\tau^* : H_I(D) \times A_2 \times Y \to \Delta(H_I(D))$ such that, $\tau^*(h, a_2, y)((h, a_2, y, r')) = \tau(r(h), a_2, y)(r')$ where $r(h) \in R$ represents the last message $r \in R$ in h.

τ^* is a transparent transition rule, and ϕ^* is a transparent rating system where $\phi^* = (H_I(D), \tau^*, p'_0)$. The support of p'_0 is the same as the support of p_0, and $p'_0(d) = p_0(d)$ for each $d \in D$. Let H_k^t and $H_k^{t\prime}$ be the set of histories of length t, H_k and H'_k be the set of all finite histories for player k in the game $G(\delta, W, \eta, \phi)$ and $G(\delta, W, \eta, \phi^*)$ respectively. The histories in the latter game include the histories in the former game, let $\varkappa : H'_1 \to H_1$ be a map where $\varkappa(h'_1)$ is the actions, signals and messages of ϕ observed by player 1 in a history in H'_1. We will construct an equilibrium in the latter game that yields the same payoffs to both players as in the former game.

Define $\sigma'_2(h'_2) = \sigma_2(r(h'_2))$. Define σ'_1 as follows:

$\sigma'_1(h_1^0) = \sigma_1(h_1^0)$ for $h_1^0 \in H_1^0 = H_1^{0\prime}$

Let $\psi_1(r) \in \Delta(\varkappa(h_1^{1\prime}))$ be the belief of player 2 at period 1 about the histories of length 1 that player 1 observes when player 2 observes the message r in the former game.

$\sigma'_1(h_1^{1\prime}) = \sum_{h_1^1 \in \varkappa(h_1^{1\prime})} \psi_1(r(h_1^{1\prime}))(h_1^1)\sigma_1(h_1^1)$

At any period $t > 1$,

$\sigma'_1(h_1^{t\prime}) = \sum_{h_1^t \in \varkappa(h_1^{t\prime})} \psi_1(r(h_1^{t\prime}))(h_1^t)\sigma_1(h_1^t)$

It is straightforward to check that σ'_1, σ'_2 is a perfect equilibrium of $G(\delta, W, \eta, \phi^*)$, and yields the same payoffs to players 1 and 2 as in the former game.

Proofs of Theorems 1-2

Proof of Theorem 1

Theorem 1:

For any Stackelberg space W, η, and $u < V_1^s$, there exists $\overline{\delta} < 1$ such that for $\delta \geq \overline{\delta}$ there is a finite rating system ϕ, and a Perfect Bayesian equilibrium of $G(\delta, W, \eta, \phi)$ where the payoff to the normal type of the long run player is at least u after every history.

Proof: We will construct the rating system in 5 steps: In steps 1,2,3 and 4 we construct a set of ratings, a transition rule and a set of strategies with the property that: if the repeated game were played from time $t = -\infty$ with the rating system and proposed strategies, then at time 0 player 1's equilibrium payoff at any rating would be more than u, and the proposed strategies would constitute an equilibrium. In step 5, we construct a new rating system that is used to prove theorem 1. This new rating system uses the rating system we construct in the first 4 steps. Since step 5 is used exactly in the same way for the proof of theorem 2, we defer this step of the proof to the end of the appendix.

Step 1: Ratings and Transition Rule

1) We will use $V_1(\alpha)$ to refer to $V_1(s_1)$ where $s_1(H) = \alpha$. Fix W, η and $u < V_1^s$. Choose $\alpha^* \in (0, 1)$ such that: i) $\alpha^* > \alpha^s$, $\nexists j : \alpha^s < \alpha^j < \alpha^*$, ii) $V_1(\alpha^*) > u$.
Our assumptions ensure that such an α^* exists. Moreover, $V_1(\alpha^*) > V_1(\alpha^j)$ for each $j > 0$. In the equilibrium we will construct, player 1 is going to play H with probability α^* most of the time.
Let $S = \{1, 2, ..., K\}$ be a set of ratings. The rating set S and the transition rule $\tau : Y \times S \to \Delta(S)$ is going to be the rating system. For the rest of the analysis, $\tau(y, s, s')$ denotes $\tau(y, s)(s')$, i.e. the probability that the rating changes from s to s' when the signal is y. $\rho(a_1, y)$ denotes $\rho_{a_1}(y)$, i.e. the probability that the signal is y if player 1 plays action a_1.
2) By identification assumption, there exists a $y^* \in Y$ such that $\rho(H, y^*) > \rho(L, y^*)$.
3) Let $m = \max\{\frac{c}{[\rho(H,y^*)-\rho(L,y^*)]/2}, 2(V_1(\alpha^*) - u_1(L, N))\}$. Recall that c is the effort cost of player 1.
4) Let $\alpha^- = [\rho(H, y^*)\alpha^* + \rho(L, y^*)(1 - \alpha^*)]\frac{c}{m[\rho(H,y^*)-\rho(L,y^*)]} < 1/2$
5) i) For $s \in \{2, 3, .., K - 1\}$:

$\tau(y, s, s - 1) = \alpha^-\ \forall y \in Y$, $\tau(y^*, s, s + 1) = \frac{c}{m[\rho(H,y^*)-\rho(L,y^*)]}$, $\tau(y, s, s + 1) = 0\ \forall y \in Y\backslash\{y^*\}$ and $\tau(y, s, s) = 1-\tau(y, s, s-1)-\tau(y, s, s+1)\ \forall y \in Y$. Note that $\tau(y^*, s, s+1) \leq 1/2$ by definition of m in item 3).

ii) For $s = 1$:

$\tau(y, 1, 2) = \frac{V_1(\alpha^*)-u_1(L,N)}{m}$ and $\tau(y, 1, 1) = 1 - \tau(y, 1, 2)\ \forall y \in Y$. Note that $\tau(y, 1, 2) \leq 1/2$ by definition of m in item 3).

iii) For $s = K$:

$\tau(y, K, K - 1) = \frac{\alpha^* c}{m}$ and $\tau(y, K, K) = 1 - \tau(y, K, K - 1)\ \forall y \in Y$.

This specification of τ is chosen to satisfy three properties:

i) Suppose we manage to make the difference in the continuation values of player 1 at adjacent ratings to be exactly $\frac{m(1-\delta)}{\delta}$ (which we will manage to do in step 4). Then at a rating $s \in S\backslash\{1, K\}$, player 1 is indifferent between actions H and L. To see this, the probability of an upgrade when player 1 plays H is $\rho(H, y^*)\tau(y^*, s, s+1)$ and when he plays L is $\rho(L, y^*)\tau(y^*, s, s + 1)$. The difference in these two probabilities is exactly $\frac{c}{m}$. Hence, the expected value of the continuation values of player 1 differ by $\frac{c}{m}\frac{m(1-\delta)}{\delta} = c\frac{(1-\delta)}{\delta}$ across playing H and L. Similarly in the initial rating and last rating playing L is optimal since the transition rule is insensitive to signals at these ratings.

ii) If the difference in the continuation values of every 2 adjacent ratings $s + 1$ and s is $\frac{m(1-\delta)}{\delta}$, then the transition rules ensure that the continuation value of being at rating s to player 1 is $V_1(\alpha^*)$.[20] To see this, let v be the equilibrium value to player 1 of being at rating

[20] Obviously in this sentence the conclusion contradicts the supposition if $\delta < 1$, since then the payoff at

$s \in S/\{1, K\}$. From i) we see that he is indifferent between playing H and L. If he plays L, he gets $u_1(L, B)$ for the period, gets downgraded w.p. $\alpha^- = [\rho(H, y^*)\alpha^* + \rho(L, y^*)(1 - \alpha^*)]\frac{c}{m[\rho(H,y^*)-\rho(L,y^*)]}$ and gets upgraded w.p. $\frac{c}{m[\rho(H,y^*)-\rho(L,y^*)]}\rho(L, y^*)$, and remains at the same rating with the remaining probability. The expected change in his rating, Δs, is

$$\frac{c}{m[\rho(H, y^*) - \rho(L, y^*)]}\rho(L, y^*) - [\rho(H, y^*)\alpha^* + \rho(L, y^*)(1 - \alpha^*)]\frac{c}{m[\rho(H, y^*) - \rho(L, y^*)]}$$

$$= \frac{c}{m[\rho(H, y^*) - \rho(L, y^*)]}\alpha^*(\rho(L, y^*) - \rho(H, y^*)) = -\alpha^*\frac{c}{m}$$

Hence, the expected change in his continuation value will be:

$$\Delta v = -\alpha^*\frac{c}{m}\frac{m(1 - \delta)}{\delta} = -\alpha^* c\frac{(1 - \delta)}{\delta}.$$

We have therefore,

$$v = (1 - \delta)u_1(L, B) + \delta(v - \alpha^* c\frac{(1 - \delta)}{\delta})$$

and consequently,

$$v = u_1(L, B) - \alpha^* c = V_1(\alpha^*)$$

The other cases, $s = 1$ and $s = K$ can be checked similarly.

iii) At a rating $s \in S\backslash\{1, K\}$, the probability of an upgrade for a normal type who plays H w.p. α^* is exactly the same as the probability of a downgrade. This follows from the choice of α^- in item 4).

Step 2: Equilibrium Strategies

For the normal type of player 1, consider the strategy $\sigma_1 : S \to [0, 1]$ that chooses the probability of playing action H at each state.

$$\sigma_1(1) = \sigma_1(K) = 0 \text{ and } \sigma_1(s) = \alpha^*, \ \forall s \in S\backslash\{1, K\}.$$

For player 2, consider the strategy $\sigma_2 : S \to [0, 1]$ that chooses the probability of playing action B at each state.

$$\sigma_2(1) = 0 \text{ and } \sigma_2(s) = 1, \ \forall s \in S\backslash\{1\}.$$

every rating would be $V(\alpha^*)$ and the continuation values of adjacent ratings would not differ by $\frac{m(1-\delta)}{\delta}$. We modify the transition rules slightly in step 4 to get the desired difference.

Step 3: Determination of Number of Ratings

In this subsection, we determine the number of ratings K in S. Let $P^k_+ = P^k[s, s+1]$ and $P^k_- = P^k[s, s-1]$ for $s \in S\backslash\{1, K\}$ denote the probability of an upgrade and downgrade over the ratings for type k that is induced by the strategies σ_1, W and τ. In particular, $P^k_+ = P^k(s, s+1) = \frac{1}{2}[p(H, y^*)\alpha^k + p(L, y^*)(1-\alpha^k)]$ for $s > 1$ and $P^k_- = P^k(s, s-1) = \alpha^-$ for $s < K$. Let $\pi^k_K(s)$ denote the joint probability that player 1's type is k and visits rating s in the long run when the rating set has K ratings.[21] The following is a known fact about Markov chains that have transitions having a one step at a time property:

$$\pi^k_K(s)P^k[s, s+1] = \pi^k_K(s+1)P^k[s+1, s], \forall s < K.$$

Hence $\pi^k_K(K-1) = \pi^k_K(K)\frac{P^k(K,K-1)}{P^k_+}$, $\pi^k_K(s-1) = \pi^k_K(s)\frac{P^k_-}{P^k_+}$ for $s \in \{2, 3, ..., K-1\}$ and $\pi^k_K(1) = \pi^k_K(2)\frac{P^k_-}{P^k(1,2)}$. Combining, we have:

$$\pi^k_K(K)[1+\frac{P^k(K,K-1)}{P^k_+}(1+(\frac{P^k_-}{P^k_+}) + (\frac{P^k_-}{P^k_+})^2 + ... + (\frac{P^k_-}{P^k_+})^{K-3} + (\frac{P^k_-}{P^k_+})^{K-3}\frac{P^k_-}{P^k(1,2)})] = \eta(k) \quad (2)$$

Since $P^k_+ > P^k_-$ for $k > 1$ and $P^0_+ = P^0_-$, $\lim_{K\to\infty}\pi^k_K(K) > 0$, $\lim_{K\to\infty}\pi^0_K(s) = 0$ for $s \leq K$ and $\lim_{K\to\infty}\frac{\pi^k_K(1)}{\pi^0_K(1)} = 0$. Therefore there exists a $\overline{K}$ such that for any $K > \overline{K}$ we have $\frac{\pi^0_K(1)}{\sum_{k=0}^{l}\pi^k_K(1)} > (1-\alpha^s)$ and $\frac{\pi^l_K(K)\alpha^l}{\sum_{k=0}^{l}\pi^k_K(K)} > \alpha^*$. We choose the number of ratings K such that the above two inequalities hold.

Step 4: ε, δ transition rules for $\delta < 1$

In this step, we change the transition rule τ constructed in step 1 slightly to make player 1's strategy long-run optimal by making the values of adjacent ratings to differ by exactly $\frac{m(1-\delta)}{\delta}$ and to give him a different but very close payoff to $V_1(\alpha^*)$. If player 2 plays according to σ_2 and the transition rules are as described as below, then the present discounted payoff to player 1 of a rating s is $V_1(\alpha^*) - \varepsilon - (K-s)\frac{(1-\delta)m}{\delta}$ for ε sufficiently small and δ sufficiently close to 1.

$\tau_{\varepsilon,\delta} : Y \times S \to \Delta(S)$ is such that:
$\tau_{\varepsilon,\delta}(y, s, s-1) = \tau(y, s, s-1) + (K-s)\frac{(1-\delta)}{\delta} + \frac{\varepsilon}{m}$ $\forall y \in Y$, $\forall s > 1$,
$\tau_{\varepsilon,\delta}(y, s, s') = \tau(y, s, s')$ $\forall y \in Y$, $\forall s > 1$, and $s' \neq s-1$
$\tau_{\varepsilon,\delta}(y, 1, 2) = \tau(y, 1, 2) - (K-1)\frac{(1-\delta)}{\delta} - \frac{\varepsilon}{m}$ $\forall y \in Y$, $s = 1$
$\tau_{\varepsilon,\delta}(y, 1, s') = \tau(y, 1, s')$ $\forall y \in Y$ and $\forall s' \neq 2$

[21] In particular, the probability that the stationary distribution of P^k places on rating s multiplied by $\eta(k)$, the probability of type k.

Note that in this definition, we perturb the transition rule τ "slightly" without changing the rating set S. The probability of upgrades or downgrades are changed by the same amount across all signals, therefore the "sensitivity" of rating changes to actions is the same as that of τ, preserving the optimality of player 1's behavior as proposed in step 2. Moreover, $\tau_{\varepsilon,\delta}$ is a well defined transition rule for ε sufficiently small and δ sufficiently close to 1.

Lemma 2 *Let $\pi^k_{\varepsilon,\delta}(s)$ be the steady state joint distribution of type k visiting rating s when the transition rule is $\tau_{\varepsilon,\delta}$ and players use strategies proposed in step 2. Then $\exists\overline{\delta} < 1$ and $\overline{\varepsilon} > 0$ such that $\forall\delta > \overline{\delta}$, $\varepsilon < \overline{\varepsilon}$, if player 2 assigned probability $p^k(s) = \frac{\pi^k_{\varepsilon,\delta}(s)}{\sum_{k=0}^{l}\pi^k_{\varepsilon,\delta}(s)}$ to his opponent being type k when she observes rating s, then*

i) σ_2 is a strict best response to σ_1

ii) Player 1's expected discounted present value at rating s is $V_1(\alpha^*) - \varepsilon - (K-s)\dfrac{(1-\delta)m}{\delta}$

iii) σ_1 is a best response to σ_2.

Proof. i) At step 3, K is chosen such that $\frac{\pi^0(1)}{\sum_{k=0}^{l}\pi^k(1)} > (1-\alpha^s)$ and $\frac{\pi^l(K)\alpha^l}{\sum_{k=0}^{l}\pi^k(K)} > \alpha^*$. By continuity of the steady state distributions with respect to the transition matrices, $\exists\overline{\delta} < 1$ and $\overline{\varepsilon} > 0$ such that $\forall\delta > \overline{\delta}$, $\varepsilon < \overline{\varepsilon}$ we have $\frac{\pi^0_{\varepsilon,\delta}(1)}{\sum_{k=0}^{l}\pi^k_{\varepsilon,\delta}(1)} > (1-\alpha^s)$ and $\frac{\pi^l_{\varepsilon,\delta}(K)\alpha^l}{\sum_{k=0}^{l}\pi^k_{\varepsilon,\delta}(K)} > \alpha^*$. Since $\sigma_1(1) = 0$, the total probability with which player 1 plays H at rating 1 is strictly less than $1-\alpha^s$, and hence $\sigma_2(1) = 0$ is a strict best response. Similarly, the total probability that player 1 plays H at rating K is at least $\frac{\pi^l_{\varepsilon,\delta}(K)\alpha^l}{\sum_{k=0}^{l}\pi^k_{\varepsilon,\delta}(K)} > \alpha^*$, hence $\sigma_2(K) = 1$ is a strict best response. Since each type of player 1 puts a probability strictly more than α^s at the ratings $s \in S\backslash\{1,K\}$, $\sigma_2(s) = 1$ is a strict best response.

ii) We'll use the method of guess and verify to argue that player 1's expected discounted present value with strategies σ_1 and σ_2 is $V(s) = V_1(\alpha^*) - \varepsilon -$ $(K-s)\frac{(1-\delta)m}{\delta}$ at a rating s with the transition rule $\tau_{\varepsilon,\delta}$.
At rating K,

$$V(K) = (1-\delta)u_1(L,B) + \delta((\frac{\alpha^*c}{m} + \frac{\varepsilon}{m})V(K-1) + (1 - \frac{\alpha^*c}{m} - \frac{\varepsilon}{m})V(K))$$

Note also that $V_1(\alpha^*) = u_1(L,B) - \alpha^*c$, combining this with our guess $V(K-1) = V_1(\alpha^*) - \varepsilon - \frac{(1-\delta)m}{\delta}$ verifies that $V(K) = V_1(\alpha^*) - \varepsilon$.
At ratings $s \in S\backslash\{1,K\}$, let $p^- = (\alpha^- + \frac{\varepsilon}{m} + (K-s)\frac{(1-\delta)}{\delta})$

$$V(s) = (1-\delta)(u_1(L,B) - \alpha^*c) + \delta(p^-V(s-1) + (\alpha^-)V(s+1) + (1-\alpha^- - p^-)V(s))$$

We use our values from the guess on the right hand side of the equation and verify that $V(s) = V_1(\alpha^*) - \varepsilon - (K - s)\frac{(1-\delta)m}{\delta}$
At rating 1, let $p^+ = (\frac{V_1(\alpha^*) - u_1(L,N)}{m} - (K-1)\frac{(1-\delta)}{\delta} - \frac{\varepsilon}{m})$

$$V(1) = (1-\delta)u_1(L,N) + \delta(p^+ V(2) + (1-p^+)V(1))$$

Again putting the guess values on the r.h.s. verifies that $V(1) = V_1(\alpha^*) - \varepsilon - \frac{(K-1)(1-\delta)m}{\delta}$.

iii) Since the ratings are insensitive to the signals at ratings 1 and K, $\sigma_1(1) = \sigma_1(K) = 0$ is straightforward. For $s \in S\backslash\{1, K\}$, adjacent ratings differ in values by $\frac{(1-\delta)}{\delta}m$, and the probability of upgrades differ by $\frac{c}{m[\rho(y^*,H) - \rho(y^*,L)]}(\rho(y^*, H) - \rho(y^*, L)) = \frac{c}{m}$, hence the impact on continuation values of action H compared to action L when discounted to present is exactly $(1-\delta)c$ which is the period cost of action H. ■

Proof of Theorem 2

Theorem 2:

For every $\epsilon > 0$, $(U, V) \in IRP(W)$, there exists $\overline{\eta}_0 < 1$, a natural number T, $\overline{\delta} < 1$, such that for $\delta \geq \overline{\delta}$, $1 > \eta_0 > \overline{\eta}_0$ there is a rating system ϕ, and a Perfect Bayesian equilibrium of $G(\delta, W, \eta, \phi)$ satisfying the following: i) the payoff to the normal type of the long run player is at least $U - \epsilon$ after every history. ii) unconditional expected payoff of every short run player after period T is at least $V - \epsilon$.

Proof

We will implement payoffs in $CP(W)$. By an initial randomization over the payoffs in $CP(W)$, we can generate any payoff in the convex hull of $IRP(W)$. Fix W, and $(U, V) \in CP(W)$. Then there is an $\alpha^* < \alpha^l$, ϵ_1, $\epsilon_2 > 0$ such that $U - \frac{\epsilon_2}{2} < V_1(\alpha^*)$, $V - \frac{\epsilon_2}{2} < U_2(\alpha^* - \epsilon_1, B)$, $\nexists \alpha^k \geq \alpha^* - \epsilon_1$. Such an α^* exists because of the structure of the games we are analyzing.

Define the set of ratings S, transition rule τ as in step 1 of the proof of theorem 1. Strategy σ_2 is the same as in theorem 1, but,

$$\sigma_1(1) = \sigma_1(K) = 0 \text{ and } \sigma_1(s) = \alpha^* - \epsilon_1, \ \forall s \in S\backslash\{1, K\}.$$

Using equation 2, let $\lim_{K\to\infty} \frac{\pi_K^k(K)}{\eta(k)} = \overline{\pi}^k$ and $\lim_{K\to\infty} \frac{\pi_K^k(1)}{\eta(k)} = \underline{\pi}^k$. Note that as ϵ_1 goes to zero, $\underline{\pi}^0$ goes to zero as well, hence we can make $\underline{\pi}^0$ arbitrarily small by choosing ϵ_1 small enough. In particular let ϵ_1 be such that $U_2((1 - \underline{\pi}^k)(\alpha^* - \epsilon_1), B) > V - \epsilon_2$.

Steps 4 and 5 remain unaltered. The number of ratings and the cutoff for the probability of the normal type, $\overline{\eta}_0$, is chosen in step 3 as below.

Step 3: Determination of Number of Ratings and choice of $\overline{\eta}_0$

Let k^* be the largest integer with $\alpha^k < \alpha^*$, then $P_+^k < P_-^k$ for $k \leq k^*$ and for $k > k^*$, $P_+^k > P_-^k$. Therefore we have $\overline{\pi}^k = 0$ and $\underline{\pi}^k > 0$ for $k \leq k^*$ and $\overline{\pi}^k > 0$ and $\underline{\pi}^k = 0$ for $k > k^*$. Therefore there exists $\epsilon_1 > 0$, $\overline{\eta}_0 < 1$ such that for $\eta(0) > \overline{\eta}_0$:

i) For $\overline{s} \in \Delta(A_1)$ with $\overline{s}(H) = \frac{\sum_{k=1}^{l} \overline{\pi}^k \eta(k)\alpha^k + \overline{\pi}^0 \eta(0)0}{\sum_{k=0}^{l} \overline{\pi}^k \eta(k)}$, $B_2(\overline{s}) = \{B\}$,

ii) for $\underline{s} \in \Delta(A_1)$ with $\underline{s}(H) = \frac{\sum_{k=1}^{l} \underline{\pi}^k \eta(k)\alpha^k + \underline{\pi}^0 \eta(0)0}{\sum_{k=0}^{l} \underline{\pi}^k \eta(k)}$, $B_2(\underline{s}) = \{N\}$,

iii) For $s(H) > (\eta(0) - \underline{\pi}^0)(\alpha^* - \epsilon_1)$, $U_2(s, B) > V - \epsilon_2$.

Moreover for every type space and the prior probability distribution function η such that $\eta(0) > \overline{\eta}_0$, there exists an integer K^* such that for $K > K^*$ we have for $\overline{s}_K(H) = \frac{\sum_{k=1}^{l} \pi_K^k(K)\alpha^k}{\sum_{k=0}^{l} \pi_K^k(K)}$, $B_2(\overline{s}_K) = \{B\}$ and $\underline{s}_K(H) = \frac{\sum_{k=1}^{l} \pi_K^k(1)\alpha^k}{\sum_{k=0}^{l} \pi_K^k(1)}$, $B_2(\underline{s}_K) = \{N\}$ and $U_2((\eta(0) - \pi_K^0(1) - \pi_K^0(K))(\alpha^* - \epsilon_1), B) > V - \epsilon_2$.

Lemma 3 *Let $\pi_{\varepsilon,\delta}^k(s)$ be the steady state joint distribution of type k visiting rating s when the transition rule is $\tau_{\varepsilon,\delta}$ and players use strategies proposed in the beginning of the proof of theorem 2. Then $\exists \overline{\delta} < 1$ and $\overline{\varepsilon} > 0$ such that $\forall \delta > \overline{\delta}$, $\varepsilon < \overline{\varepsilon}$, If player 2 assigned probability $p^k(s) = \frac{\pi_{\varepsilon,\delta}^k(s)}{\sum_{k=0}^{l} \pi_{\varepsilon,\delta}^k(s)}$ to his opponent being type k when she observes rating s, then i) σ_2 is a strict best response to σ_1 ii) Player 1's expected discounted present value at rating s is $V_1(\alpha^*) - \varepsilon -$ $(K - s)\frac{(1-\delta)m}{\delta}$ iii) σ_1 is a best response to σ_2.iv) Player 2's ex-ante expected value is at least $V - \epsilon_2$*

Proof. Proofs of ii) and iii) are exactly the same as in the proof of lemma 2. For i) note that in step 3, the choice of the number of ratings, and the choice of $\eta(0)$ ensures that at rating 1, the relative frequency of the normal type is high enough that N is a strict best response for player 2 observing rating 1. At rating K, the relative frequency of type l is high enough that B is a best response, and at all other ratings every type puts as much effort as is required to make it strictly optimal for player 2 to play B. For iv) note that the ex-ante expected probability with which player 1 plays action H is at least $(\eta(0) - \pi_K^0(1) - \pi_K^0(K))(\alpha^* - \epsilon_1)$ and hence player 2's payoff is at least $V - \epsilon_2$ from step 3. ■

Step 5 of Theorem 1 and 2: Experimentation Periods (Convergence Path)

In the early stages of the game σ_2 described above is not necessarily optimal. Therefore the continuation values we specified for player 1 in order to make him indifferent between H and L will not hold true in the early periods of the game. To deal with this we do the following trick: At any period t, if the state is (s) and player 2 plays a_2 that is possibly different than $\sigma_2(s)$, we "bank" the difference in the flow payoff to player 1 to be paid back at the end of the experimentation periods; in particular $(1-\delta)(u_1(\sigma_1(s), \sigma_2(s)) - u_1(\sigma_1(s), \sigma_2^t(s)))$

is what player 1 needs to get back. We define a string z_t which is a string of 0's and 1's such that at period $t+1$ a new string z_{t+1} is formed by attaching a 0 or 1 to z_t. Which of the numbers is attached will be determined by a probability distribution that takes into account period t flow payoff difference. The difference in the continuation values of player 1 from having a 1 attached or a 0 attached is a small but fixed number. Since the period t flow difference is at the order of $(1-\delta)$, a probability distribution that differs from the one that attaches 0 w.p. 1/2 and 1 w.p. 1/2 at an order of $(1-\delta)$ is enough to give player 1 the right incentives and expected continuation payoffs. Formally we proceed as follows: Choose T and ε in the definition of lemma 1 applied to the transition matrices induced by (τ, σ_1, W) and $\overline{\varepsilon}$ as defined in lemmas 2 and 3.

For $t \geq 1$, let $Z^t = \{0,1\}^t$ and $Z^0 = \oslash$. Let $p_0 \in \Delta^o(S)$ be the initial probability distribution over the ratings with full support. Moreover let $\sigma_2^t : S \to A_2$ be a pure strategy best response to σ_1 and the beliefs of player 2 about player 1's types, that's obtained recursively and concurrently with the definition of τ_t below:

$\tau_t : Y \times S \times Z^t \to \Delta(S \times Z^{t+1})$ is defined recursively such that:

$\tau_t(y, s, z^t, s', (z^t, 1)) = \tau_{\varepsilon,\delta}(y, s, s')[1/2 + \frac{(1-\delta)(u_1(\sigma_1(s), \sigma_2(s)) - u_1(\sigma_1(s), \sigma_2^t(s)))\delta^{T-t}}{(\varepsilon/T)}]$ and

$\tau_t(y, s, z^t, s', (z^t, 0)) = \tau_{\varepsilon,\delta}(y, s, s')[1/2 - \frac{(1-\delta)(u_1(\sigma_1(s), \sigma_2(s)) - u_1(\sigma_1(s), \sigma_2^t(s)))\delta^{T-t}}{(\varepsilon/T)}]$.

$\tau_T : Y \times S \times Z^T \to \Delta(S)$ be such that:

$\tau_T(z^T) = \tau_{(1-\frac{n(z^T)}{T})\varepsilon,\delta}$ for $z^T \in Z^T$ where $n(z^T)$ is the number of occurrences of 1 in the string z^T.

Finally we'll define the rating system $\phi_{\varepsilon,\delta}$. Let $\phi_{\varepsilon,\delta} : \cup_{t=0}^T Z^t \times S \times Y \to \Delta(\cup_{t=0}^T Z^t \times S)$ be the transition rule on the rating set $\cup_{t=0}^T Z^t \times S$ and for $z \in Z^t$ where $t < T$ define $\phi_{\varepsilon,\delta}(y, s, z) = \tau_t(y, s, z)$ and for $z \in Z^T$ let $\phi_{\varepsilon,\delta}(y, s, z, s', z) = \tau_T(y, s, z, s')$ and $\phi_{\varepsilon,\delta}(y, s, z, s', z') = 0$ for when $z \neq z'$.

Next we prove theorem 1 using lemma 2. Proof of theorem 2 is very similar, and the main difference is to use the constructions for theorem 2, and applying lemma 3 instead of lemma 2.

Theorem: *For every $\zeta > 0$ there exists $\varepsilon > 0$, $\overline{\delta} < 1$ such that for $\delta > \overline{\delta}$, there exists a perfect Bayesian Nash equilibrium of* $G(\phi_{\varepsilon,\delta}, \delta, W, \eta)$ where player 1's payoff is more than $V_1(\alpha^*) - \zeta$ after every history.

Proof. Choose $\varepsilon = \zeta/2$, and $\overline{\delta}$ large enough such that $(K-1)\frac{1-\overline{\delta}}{\overline{\delta}} m < \zeta/2$. Let P_{σ_1} be the probability distribution over $\cup_{t=0}^T Z^t \times S \times \{0, 1, ...\} \times W$ and $P_{\sigma_1}(s, z, t, k)$ denote the probability of player 1 being type k, the rating at time t being (s, z). Next define the strategy $\sigma_1^*(s, z) = \sigma_1(s)$, and $\sigma_2^t(s, z) \in B_2(\sum_{k>0} P_{\sigma_1^*}(w = k|s, z, t)\alpha^k + P_{\sigma_1^*}(w = 0|s, z, t)\sigma_1^*(s, z))$. Note that the strategies are stationary for $t \geq T$. Let the average discounted payoff of the repeated game to player 1 be $V(z^t, s)$. It is easy to check that $V(z^T, s) = V_1(\alpha^*) - (K - s)\frac{1-\delta}{\delta} m - (1 - \frac{n(z^T)}{T})\varepsilon$ (from lemma 2). Moreover, one can easily verify that $V(z^t, s) =$

$V_1(\alpha^*) - (K-s)\frac{1-\delta}{\delta}m - \delta^{T-t}(\varepsilon(1-[n(z^t)+(T-t)1/2]))$. Once we know the continuation values, we can check that there is no profitable deviation for player 1 at any history (after time $t = T$ this follows from lemma 2. For $t < T$, the transition rules on the string z^t ensures that incentives are correct). To check that strategy of player 2 is optimal, note that until T, σ_2^t is a best response (by definition above), and after T the beliefs are $\overline{\varepsilon}$ close to the steady state beliefs generated by (τ, σ_1, W), and hence applying lemma 2 delivers the result. ■