Building Routines: Learning, Cooperation and the Dynamics of Incomplete Relational Contracts

Sylvain Chassang*, †

Princeton University

February 25, 2008

Abstract

This paper studies how agents learn to cooperate when the details of what cooperation means are not common knowledge. It considers a repeated game in which one of the players has incomplete information about when and how she should expect her partner to cooperate. Initially, monitoring is imperfect and cooperation can be sustained only at the cost of inefficient punishment. However, as the players' common history grows, the uninformed player will learn to predict when and how cooperation should take place. This learning allows players to establish cooperative routines that sustain cooperation more efficiently and reduce the partnership's sensitivity to adverse economic conditions. The paper also shows that because revealing information is costly, obtaining more information is not always desirable. This will cause efficient equilibria to be path-dependent.

KEYWORDS: cooperation, relational contracts, incomplete contracts, imperfect monitoring, learning, routines.
JEL classification codes: C72, C73, D23

*I thank Roland Benabou, Glenn Ellison, Richard Holden, Esteban Rossi-Hansberg, Lars Stole, Satoru Takahashi, participants at the MIT organizational economics lunch and seminar participants at the Chicago GSB for many helpful discussions. I am particularly indebted to Robert Gibbons for inspiration and advice. Funding from MIT's Program on Innovation in Markets and Organizations is gratefully acknowledged.

†E-mail: chassang@princeton.edu, Land-Mail: Department of Economics, Princeton University, Princeton, NJ 08544, USA.

1 Introduction

Economic agents frequently resolve conflicts of interest without resorting to formal contracts. Even when formal contracts are available and used, agents often rely on informal understandings (relational contracts) to adapt their joint behavior to circumstances that cannot be properly assessed by outside courts. Since Bull (1987), economists have analyzed such informal understandings as equilibria of repeated games: for instance MacLeod and Malcomson (1988) emphasize the role of the total surplus created by the relationship in order to overcome individuals' reneging temptations; Baker, Gibbons and Murphy (1994, 2002) analyze the interaction between formal contracts, firm boundaries and informal agreements; more recently, Levin (2003) highlights how imperfect subjective monitoring may result in inefficient rigidity and inefficient termination.

Most of these repeated-game models have been concerned with the difficulties of *maintaining* cooperation. This paper focuses on the particular hurdles involved in *building* a successful informal understanding. The starting point of the analysis is to recognize that relational contracts aren't necessarily complete. In particular the details of how and when cooperation should occur are unlikely to be common knowledge at the onset of the relationship, even between the parties themselves. For instance, a new plant manager is unlikely to have a perfect understanding of what her maintenance team can and cannot do; a firm's new CEO may have only incomplete understanding of how research teams should operate; and so on... The question is then: how do players go about building a common agreement? Or formulated in a slightly different way, how do they go about specifying the contingencies of their relational contract?

The model considers two players engaged in an infinitely repeated game with one sided-incomplete information. Every period player 1 decides whether to stay at a cost k or exit until the next period.[1] Whenever player 1 stays, player 2 obtains a profit π and has the

[1] Note that exit is not permanent. Staying vs. exiting is simply a sunk investment decision required to interact in the current period.

opportunity to reciprocate by taking some action a from a large action set $\mathcal{A}$. In a particular period t, only a subset $A(w_t) \subset \mathcal{A}$ of actions are available to player 2, but the set $A(w_t)$ is entirely characterized by a publicly observable i.i.d. state of the world w_t. Hence, even though the set of available actions varies, both players are able to verify which actions are available in a particular period. When player 2 takes an action a, it yields a random benefit $\tilde{b}(a) \in \{0, b(a)\}$ (independent of w_t) to player 1, and comes at a deterministic cost $c(a)$ to player 2. There are two types of actions: *productive* actions that are costly to player 2 and will potentially yield benefits to player 1 ($c(a) > 0$ and $b(a) > 0$), and *unproductive* actions that are costless to player 2 but will yield no benefits to player 1 ($c(a) = 0$ and $b(a) = 0$). Players are asymmetrically informed about the mapping $c(\cdot)$: player 2 knows her cost $c(a)$ and hence knows which actions will potentially lead to a benefit $b(a) > 0$; player 1 knows neither the cost $c(a)$ nor which actions potentially yield $b(a) > 0$. Every period t, player 1 can observe both the state of the world w_t and the action a_t taken by player 2. The main modeling innovation here is to distinguish between the observability of an informative state of the world (state w_t) and the ability to interpret what it implies for payoffs (the values of $b(a)$ and $c(a)$ for $a \in A(w_t)$). This distinction introduces the possibility of learning in an imperfect public monitoring framework.

The joint dynamics of cooperation and learning unfold as follows. Initially, player 1 does not know which actions are productive and hence does not know the states w_t in which she should expect cooperation (i.e. expect player 2 to take a productive action). Hence, if initially player 2 takes an action a_0 and it yields benefit $\tilde{b}(a_0) = 0$, player 1 does not know if player 2 took a costly productive action which unfortunately failed, or whether player 2 simply took a costless unproductive action. This means that at the onset of the relationship, monitoring is imperfect and player 1 may have to use inefficient punishment on the equilibrium path to induce cooperation. However, once action a_0 yields $\tilde{b}(a_0) > 0$, player 1 can identify it as a productive action and the monitoring problem disappears: player 2 can be induced to take action a_0 at no efficiency cost in future periods. Indeed, even if action

a_0 fails to yield a benefit ($\tilde{b}(a_0) = 0$) in subsequent periods, player 1 can monitor whether action a_0 was available and whether it was taken, so that there is no further suspicion of moral hazard and no punishment is required. However if player 2 takes a new action a_1 that player 1 hasn't identified yet, a failure to yield benefits may again lead to inefficient punishment on the equilibrium path. At some point the efficiency costs of identifying new productive actions may dominate the benefits of obtaining additional information. Then, it will be optimal for player 1 to stop learning and obtain benefits using only actions she has identified as productive up to now. At this point we say that the relationship has become a routine.[2]

The paper focuses on Pareto efficient equilibria and makes a number of predictions. The first is that initially, while learning occurs and the parties are specifying the contingencies of their informal understanding, relationships will be very sensitive to adverse shocks. More specifically, while learning occurs, a bad realization $\tilde{b}(a) = 0$, will be followed by punishment on the equilibrium path. However, once learning is over and the relationship becomes a routine, it becomes resilient in the sense that adverse shocks – i.e. the realization of $\tilde{b}(a)$ being 0 – will no longer trigger inefficient breakdowns.

The second prediction is that it is not necessarily optimal to reveal all the existing information: the costs of inducing further revelation can dominate the potential gains of using a more efficient routine. Because different information may get revealed depending on the particular realization of the states w_t, partnerships that are identical ex-ante may end up in different long-run routines that use different sets of actions with different degrees of efficiency. This implies that random events occurring at the onset of the relationship can have a long-term impact on the way players approach cooperation. In fact such path dependence can occur even when the players have no uncertainty about what the Pareto frontier under

[2] In this model, equilibrium behavior is called routine if learning has stopped and players use a fixed subset of actions. This use of the term is intended to be in the spirit of Nelson and Winter (1982), if not identical in details. One important similarity in the two usages is that, except for the extreme case of an equilibrium with no learning at all, routines take time (and, more importantly, shared experience) to develop.

complete information would be. In this sense the model is not about finding out whether cooperation is sustainable or not, but rather about figuring out the details of how to get to the Pareto frontier. A corollary is that existing routines will naturally remain sticky even in the presence of overwhelming evidence that they are not optimal. This mechanism would naturally lead to the kind of persistent performance differential across seemingly identical organizations that Gibbons et al. (2007) document. Appendix A pushes the model further in this direction by considering a tractable subclass of strategies and highlighting the link between the long-run distribution of routines and fundamentals of the relational contract.

The paper draws largely from the literature on imperfect public monitoring initiated by Green and Porter (1984), and developed in Abreu, Pearce and Stacchetti (1986, 1990) and Fudenberg, Levine and Maskin (1994). At the onset of the relationship, player 1's inability to understand the circumstances of player 2 generates a monitoring problem that results in inefficient punishment. The particular nature of the informational asymmetry bears much resemblance to the models of Athey and Bagwell (2001), Levin (2003) and Athey, Bagwell and Sanchirico (2004) who also consider situations in which actions are observable but the players' costs for those actions is private. The contribution of the current paper is to introduce the possibility of learning and analyze the joint dynamics of learning and cooperation. This connects the paper to the literature on repeated games with incomplete information developed by Hart (1985), Shalev (1994), Sorin (1999) and more recently by Cripps and Thomas (2003) or Peski (2007). Those papers study learning in a more general class of games but focus only on the case where discount factors go to 1. In contrast, the results highlighted in this paper occur for discount factors strictly below one, which makes the analysis delicate. This paper also shares much of the spirit of Watson (1999, 2002) who considers a partnership game in which players try to screen bad types by delaying full-fledged cooperation. The focus of Watson (1999, 2002) however is on how players figure out whether cooperation is sustainable or not, whereas the current paper is about figuring the details of how to actually implement cooperation. As a result, the papers are complementary and their predictions are mostly orthogonal.

The paper is organized as follows: Section 2 defines the setup and introduces the notion of routines; Section 3 characterizes general properties of the game and highlights the efficiency costs of information revelation; Section 4 explores in more details the case with two productive actions and shows how path dependence arises naturally; Section 5 concludes. Proofs are contained in Appendix D.

2 The Setup

The framework of the paper is defined formally in Sections 2.1 and 2.2 ; Section 2.3 discusses the model and how it might be interpreted ; Section 2.4 deals with the specific question of communication.

2.1 Actions, timing and payoffs

Consider a game with two players $i \in \{1, 2\}$, infinite horizon, and discrete time $t \in \mathbb{N}$. Players share the same discount factor δ. Each period t, player 1 decides to either *stay* or *exit* until the next period. Player 2 has a countably infinite set of actions $\mathcal{A}$.[3] Each period t, player 2 chooses an action a from a set of available actions $A(w_t) \subset \mathcal{A}$ which is different across periods. The set of available actions $A(w_t)$ is entirely characterized by an i.i.d. state of the world $w_t = \{w_t^a\}_{a\in\mathcal{A}} \in \{0, 1\}^{\mathcal{A}}$: action a is available ($a \in A(w_t)$) if and only if $w_t^a = 1$. Actions are independently available with probability p. Each period t consists of the two following stages:

Stage 1: player 1 decides to stay or exit. If player 1 exits, both players get 0 flow payoffs and the game moves on to period $t + 1$. If player 1 stays she experiences a cost $k > 0$ while player 2 gets a benefit $\pi > 0$.

[3]The action set $\mathcal{A}$ could be thought of as $\mathbb{N}$ except that $\mathbb{N}$ has a natural ordering which $\mathcal{A}$ doesn't have. Appendix B on communication associates actions to labels in $\mathbb{N}$ but insures that the labelings used by players 1 and 2 are independent so that the players cannot communicate about actions using numbers in $\mathbb{N}$.

If player 1 stays, a publicly observable i.i.d. state $w_t \in \{0,1\}^{\mathcal{A}}$ is drawn.

Stage 2: if player 1 has stayed, player 2 then gets to choose an action $a \in A(w_t)$. Taking action a has a deterministic cost $c(a)$ for player 2 and generates a random benefit $\tilde{b}(a) \in \{0, b(a)\}$ to player 1.

The benefit $\tilde{b}(a)$ takes the form

$$\tilde{b}(a) = \begin{cases} b(a) & \text{with proba } q \text{ (the action succeeds)} \\ 0 & \text{with proba } 1-q \text{ (the action fails)} \end{cases}$$

where $b(a)$ is a deterministic value.

There exists a number N of productive actions $\{a_0, \ldots, a_{N-1}\} \subset \mathcal{A}$. By extension $\mathcal{N}$ will be used to denote the set of productive actions. Productive actions are costly to player 2 but yield strictly positive expected benefits to player 1, while unproductive actions are free for player 2 but yield no benefit to player 1. More precisely, whenever $a \in \mathcal{N}$, then $c(a) = c > 0$ and $b(a) > 0$ and whenever $a \notin \mathcal{N}$ then $c(a) = 0$ and $b(a) = 0$.

In any given period t, the state of the world $w_t = \{w_t^a\}_{a \in \mathcal{A}}$ is an i.i.d. sequence of Bernoulli variables such that for all $a \in \mathcal{A}$ and $t \in \mathbb{N}$, $prob(w_t^a = 1) = p$. Finally, before they take actions, each of the players can run a finite number of public coin flips with outcomes in $\{0,1\}$, where the probability of each outcome is set by the players and publicly known. We will denote by $x_t = (x_t^1, x_t^2)$ the outcomes of coin flips by player 1 and by player 2. This allows for public randomizations.

2.2 Information structures

We will be considering two information structures: a full information setting in which both players are fully informed about which actions are productive, and an asymmetric information setting in which player 2 knows which actions are productive but player 1 is uninformed.

Full Information: Parameters p and q are common knowledge. Both player 1 and player 2 know the mappings $c(\cdot)$ and $b(\cdot)$. The state w_t, the action a_t, and the realization of $\tilde{b}(a_t)$ are publicly observed. Let Γ_{FI} denote the corresponding game.

Asymmetric Information: Parameters p and q are common knowledge. Player 2 knows the mappings $c(\cdot)$ and $b(\cdot)$ and observes both the state of the world w_t and the outcome $\tilde{b}(a_t)$.

Player 1 observes the state w_t, the action a_t taken by player 2 at time t and the realization of $\tilde{b}(a_t) \in \{0, b(a_t)\}$. Player 1 has an improper uniform prior over which actions are productive so that

$$\forall A \subset \mathcal{A},\ \forall a \in A, \quad Prob_1 \left\{ c(a) = c \mid \text{card } A \cap \mathcal{N} = k \right\} \quad = \quad \frac{k}{\text{card } A}.$$

Player 1 (correctly) believes that the number N of productive actions is drawn according to a distribution B over $\mathbb{N}$. Conditionally on N, player 1 (correctly) believes that the vector $(b(a_1), \ldots, b(a_N))$ is drawn according to some symmetric distribution A_N over $[\underline{b}, \overline{b}\,]^N$, where $\underline{b} > 0$. Let Γ_{AI} denote this asymmetric information game.

2.3 Interpreting the model

The model presented in Section 2.1 makes a number of simplifications. For instance communication is not explicitly considered (see Section 2.4 for a discussion) and players do not have the ability to transfer cash (see Appendix C for a discussion). Still the model can be interpreted as a rough approximation of a number of economic settings in which the details of what cooperation means are initially unclear. For instance it could describe the problem of a new plant manager who needs to figure out when a machine failure can be dealt with by her maintenance team and when the machine must be replaced. Here action a corresponds to a specific maintenance procedure, the state w_t corresponds to the set of maintenance procedures that are feasible given what spare parts are available, what machine seems to be failing, what time is available, and the set $\mathcal{N}$ of productive actions are the actions that will get the production line running rather than simply allow the maintenance team to look like it is working.

It is important to note that the model does not assume that player 1 does not know what she wants. Rather, she does not know what inputs from player 2 are needed to get the outcome she wants. For instance the plant manager knows that she wants to minimize production costs and delays, but she doesn't know what maintenance operations will lead to that outcome. In fact, the action a could be interpreted fairly broadly. For instance imagine the problem of a firm's CEO trying to figure what kind of topics her research teams should be allowed to work on. Should scientists always be allowed to pursue their own interests? What areas are more likely to yield potential benefits for the firm? In that setting, the action $a \in \mathbb{N}$ corresponds to the choice of a research topic by the research team, while the state w_t is the set of topics that could be explored given the team's abilities, the availability of material, research opportunities and availability of funding.

In all those settings, the players need to solve a complex learning problem to figure out the specific details of how to implement cooperation. As Sections 3 and 4 show, even the simple model of Section 2 can yield useful insights on how this learning process might unfold.

2.4 Indescribability, communication and the role of the large action space

Explicit communication is not part of the model as it is described.[4] This is not because the mechanism described in the paper is incompatible with communication. Rather, the results presented here hold independently of the particular communication mechanism as long as it satisfies the following indescribability assumption: players cannot successfully describe

[4]Note however that implicit communication may be possible via actions

actions $a \in \mathcal{A}$ until they have been *taken* by player 2.

As highlighted in Appendix B, this indescribability assumption is not incompatible with players referring to their joint history. For instance player 2 could send a message defining "the set of actions that were available at history h but not at history h'". However because the set of possible actions is very large, this particular statement will define an infinite number of actions and hence conveys no information to player 1. The only way to convey information is to make statements about actions that have been taken. This is one of the main reasons why the model considers an infinite number of actions and a large state space. As in Al-Najjar, Anderlini and Felli (2006), it provides a framework in which actions can remain indescribable although players have a rich language. To a large extent, the model could be interpreted as a model of joint language building in which actions need to be taken before players can successfully communicate about them.

It is also useful to note that although the paper is about incomplete (relational) contracts, it is not subject to the critique of Maskin and Tirole (1999). This is not because the optimal contract does not use the indescribable information as in Segal (1999) or Hart (1999), or because the contract is relational rather than formal. Rather it's because (before learning occurs) the payoff implications of state w_t and action a_t are not common knowledge among players, even ex-post. For this reason indescribability is not innocuous and inefficient punishment will occur on the equilibrium path.

2.5 Strategies and solution concept

Let $d_t \in \{S, E\}$ denote player 1's decision to stay or exit at time t. When $d_t = E$, the variables x_t, w_t, a_t and $\tilde{b}(a_t)$ are set to $\emptyset$. For any two histories h and h', let $h \sqcup h'$ denote the concatenated history composed of history h followed by history h'. We distinguish three types of histories:

(i) the set $\mathcal{H}$ of histories h_{t+1} of the form $h_{t+1} = \{d_1, x_1, w_1, a_1, \tilde{b}(a_1), \ldots, d_t, x_t, w_t, a_t, \tilde{b}(a_t)\}$, corresponding to the history of play at the end of period t;

(ii) the set $\mathcal{H}^1$ of histories h^1_{t+1} of the form $h^1_{t+1} = h_t \sqcup \{x^1_{t+1}\}$, corresponding to the history of play at player 1's decision node in period $t+1$;

(iii) the set $\mathcal{H}^2$ of histories h^2_{t+1} of the form $h^2_{t+1} = h_t \sqcup \{d_{t+1}, x_{t+1}, w_{t+1}\}$ corresponding to player 2's information set at her decision node in period $t+1$.

A strategy of player 1 is a mapping $s_1 : \mathcal{H}^1 \to \{S, E\}$ and a strategy of player 2 is a mapping $s_2 : \mathcal{H}^2 \to \mathcal{A}$. We denote by $f(d, a|h_t, s_1, s_2)$ the distribution of actions in period t given strategies s_1, s_2 and history h_t. In this setting, the appropriate solution concept is perfect Bayesian equilibrium. Let $V_1(s_1, s_2, h_t)$ and $V_2(s_1, s_2, h_t)$ denote the respective continuation values of players 1 and 2 at history h_t when they respectively use strategies s_1 and s_2.

2.6 Weak and strong routines

Unproductive actions of player 2 all have identical zero payoff consequences. Hence to compute payoffs we only need to keep track of the action pairs $(d, a) \in \mathcal{R} = \{S, E\} \times \mathcal{N}$, where $\mathcal{R}$ is the set of action pairs of player 1 and player 2 that are payoff-relevant. Among pairs of strategies, we define routine strategies as follows.

Definition 1 (routines)

(i) For any pair of strategies (s_1, s_2) and any history h_t, let us define the support supp$(s_1, s_2|h_t)$ of (s_1, s_2) at history h_t as the set of actions that have positive probability in period t conditionally on h_t.

(ii) Given a set $\mathcal{S} \subset \{S, E\} \times \mathcal{N}$, starting from history h_t, the pair of strategies

(s_1, s_2) *is said to follow a* ***weak routine*** *of support* $\mathcal{S}$ *if*

$$\bigcup_{h_s \text{ following } h_t} supp(s_1, s_2|h_s) \cap \mathcal{R} \quad \subseteq \quad \mathcal{S} \quad \subseteq \bigcup_{h_s \text{ preceding } h_t} supp(s_1, s_2|h_s) \cap \mathcal{R}.$$

(iii) Starting from history h_t*, a pair of strategies* (s_1, s_2) *is said to follow a* ***strong routine*** *with support* $\mathcal{S}$ *if at all attainable histories* h_s *following* h_t*,* $supp(s_1, s_2|h_s) \cap \mathcal{R} = \mathcal{S}$ *and* $f(d, a|h_s, s_1, s_2) = f(d, a|h_t, s_1, s_2)$ *(in other terms the distribution of actions is constant).*

In words, strategies become weak routines once they don't use any new relevant actions on the equilibrium path. Strategies become strong routines when the pattern of play itself becomes stationary.

Note that this notion of routine will not serve to restrict attention to a particular subset of strategies, but rather will be a property satisfied by efficient strategies in equilibrium. It is also useful to note that the routines considered in this paper differ from those described in Winter (1971) or Nelson and Winter (1982). Here being a routine is an equilibrium property of strategies rather than an exogenous constraint placed on agents' behavior. This allows to identify what circumstances in the agents' environment will shape the establishment of routines.

3 Some general results

This section begins with an informal description of what a typical path of play might look like. Then, as a benchmark, it analyzes the Pareto frontier of the game with full information Γ_{FI}. It then establishes general properties of the incomplete information game Γ_{AI}. It is shown that exit need not occur under complete information but that inefficient exit will be necessary to sustain cooperation under incomplete information. The section also highlights that on the path of Pareto efficient equilibria, exit happens only following adverse shocks (i.e. $\tilde{b}(a) = 0$) and serves only to incentivize information revelation. Once players reach a routine stage where no further information is revealed, exit will no longer be necessary and the relationship will be resilient to shocks.

3.1 A typical path of play

Before beginning a formal analysis, let us describe informally how play might look like on a typical equilibrium path of the asymmetric information game Γ_{AI}.

In the initial stages of the game, player 1's main problem is to determine what actions a she should require player 2 to take. Because player 1 is uninformed about what action a might be productive, she has to demand benefits at some arbitrary history h_t. Player 2 then takes some action a_0 and one of two things can happen. Either player 1 receives benefit $\tilde{b}(a_0) > 0$ and learns that a_0 is a productive action, or player 1 obtains benefit $\tilde{b}(a_0) = 0$. In that last case, player 1 does not know whether player 2 took a productive costly action which unfortunately failed, or whether player 2 simply took a costless unproductive action. To provide appropriate ex-ante incentives, inefficient punishment (in the form of exit) will need to occur on the equilibrium path. This corresponds to the fact that initially, monitoring is imperfect and inefficient equilibrium punishment à la Green and Porter (1984) is required.

Imagine now that $\tilde{b}(a_0) > 0$ so that player 1 identifies action a_0 as productive. In later periods where a_0 is available, player 1 can again demand player 2 to take action a_0. In this case however, there is no suspicion of moral hazard anymore: player 1 knows that a_0 is a costly productive action and observes whether it was taken or not. Hence future play need not be conditioned on the outcome of a_0. This means that even if $\tilde{b}(a_0) = 0$ in some subsequent period, no inefficient punishment is required on the equilibrium path. However, if player 2 takes another action a_1 that player 1 has never seen before, then the outcome $\tilde{b}(a_1) = 0$ would again be polluted with moral hazard and punishment would be required on the equilibrium path. This has two implications. The first one is that the partnership will be sensitive to bad shocks while learning is occurring: when a new action fails, it will be associated with potentially inefficient punishment. The second is that player 2 is more likely to be forgiven when she fails using a standard protocol (action a_0) than when she fails using a new method (action a_1) whose consequences are not common-knowledge.

As player 1 identifies more of the productive actions, the benefits of learning new productive actions will diminish, while the cost of information revelation (in the form of inefficient punishment when a new action fails to yield benefits) may remain high. Hence, it may be optimal not to learn all the available information. When learning stops, the players enter a routine mode in which they only use actions whose consequences are common-knowledge. At that point the partnership becomes resilient to adverse shocks in the sense that bad realizations of the form $\tilde{b}(a) = 0$ will no longer cause exit on the equilibrium path. Note that because the states w_t experienced by different partnerships are different, different partnerships may reveal different actions and hence end up using different routines. In fact Section 4 shows that once player 1 learns a productive action a_0, it might become optimal never to search for a new productive action a_1 even though both players know that such an action exists and (if successful) yields benefit $b(a_1) > b(a_0)$. This happens because in this model, implementing the details of cooperation is not a trivial task and requires costly information revelation. For this reason, routines can be sticky and persist even when there is convincing evidence that they are not optimal.

The rest of the paper explore these properties in more details. Section 3.2 begins by analyzing the full-information setting.

3.2 The Pareto frontier under full-information

Proposition 1 (no exit) *Consider the full-information game Γ_{FI}. We have that either*

(i) the only equilibrium of Γ_{FI} is for player 1 to exit every period; or

(ii) whenever (s_1, s_2) is a Pareto efficient equilibrium of Γ_{FI}, then player 1 never

exits on the equilibrium path.

Proposition 1 will serve as a benchmark in the rest of the paper where inefficient exit will sometimes be required. It shows that under complete information, provided that some cooperation is sustainable, exit is unnecessary even when the action taken by player 2 fails. This is because under complete information there is no suspicion of moral hazard: player 1 knows whether or not player 2 took the appropriate action given current conditions, so there is no need to condition continuation play on the random outcome of these actions. The following proposition characterizes the Pareto frontier under complete information.

Proposition 2 (spanning the Pareto frontier under full-information) *Order productive actions $\mathcal{N} = (a_0, \cdots, a_{N-1})$ so that the corresponding benefits form an increasing sequence $b_0 \leq \ldots \leq b_{N-1}$. For any family of parameters $(r^1_0, \ldots, r^1_{N-1}) \in [0,1]^N$ and $(r^2_0, \ldots, r^2_{N-1}) \in [0,1]^N$ consider the strategies s_1 and s_2 such that on the equilibrium path,*

(i) while player 2 has never taken a productive action, if the most productive available action is a_i, then player 2 takes action a_i with probability r^1_i and takes a costless action with probability $1 - r^1_i$;

(ii) once player 2 has taken some (any) productive action, if the most productive available action is a_i, then player 2 takes action a_i with probability r^2_i and takes a costless action with probability $1 - r^2_i$;

(iii) Player 1 stays while player 2 behaves as prescribed above but exits permanently otherwise.

The Pareto frontier of Γ_{FI} is spanned by equilibria characterized by parameters $(r^1_0, \ldots, r^1_{N-1})$ and $(r^2_0, \ldots, r^2_{N-1})$ such that for all $k \in \{1,2\}$, $r^k_i > 0 \Rightarrow r^k_{i+1} = 1$.

Proposition 2 implies that the Pareto frontier can be spanned by strategies that become strong routines once player 2 has cooperated. Those strong routines take a very specific form:

the efficient way for player 2 to cooperate is to use costly actions in order of productivity so that player 2 uses an action only when she takes all actions that are more productive at the maximum feasible rate.

Note that in any equilibrium where player 1 stays in the first period, player 2 must get value greater than π. Furthermore, player 1 must get value greater than 0. This yields the following corollary.

Corollary 1 *Define $\underline{r} \equiv k/pq\bar{b}$ and $\overline{r} \equiv \delta\pi/pc$. For any Pareto efficient pair of values (V_1, V_2), there exists $r \in [\underline{r}, \overline{r}]$ such that*

$$V_1 \leq \frac{1}{1-\delta}(-k + qpr\bar{b}) \quad \textit{and} \quad V_2 \leq \frac{1}{1-\delta}(\pi - prc).$$

This establishes simple but useful bounds for Pareto efficient values. We now turn to the asymmetric information game in which player 1 has to figure out the contingencies in which to expect cooperation.

3.3 The case of asymmetric information

This section considers the asymmetric information game Γ_{AI}. Initially, when player 1 does not obtain any benefits she does not know whether player 2 took a costly productive action but it failed, or whether player 2 simply took a costless unproductive action. This means that initially monitoring is imperfect and inefficient punishment on the equilibrium path will be necessary to induce cooperation from player 2. As player 1 obtains benefits however, productive actions will be revealed, thereby allowing the players to solve the moral hazard problem. This section investigates patterns of punishment on the path of Pareto efficient equilibria. It begins with a few definitions.

Definition 2 (revelation and confirmation stage) *Consider an equilibrium (s_1, s_2). A history $h^2_{t+1} \in \mathcal{H}^2$ is called a **revelation stage** if there is non-zero probability that a pro-*

ductive action which has not been taken yet will be taken. The action a that player 2 does take is called the revealed action.

A history $h_{t+1} \in \mathcal{H}$ is called a ***confirmation stage*** *for action a if and only if $a_t = a$ and it is the first time that action a yields benefit $\tilde{b}(a) = b(a) > 0$ rather than 0.*

The next lemma highlights that from the perspective of player 1, at any history h, the only relevant actions are those that have been taken by player 2 so far.

Lemma 1 (relevant actions) *For any history h_{t+1}, consider an action $a \in \mathcal{A} \setminus \{a_s\}_{s=0,\ldots,t}$. We have that $Prob_1\{a \in \mathcal{N} | h_{t+1}\} = 0$.*

In words, at a history h_{t+1}, from the perspective of player 1, the only actions which have non-zero probability of being productive are those that have been taken by player 2 so far. This highlights that although player 1 can potentially verify the availability of a large number of actions, only the observability of actions that player 2 has taken before matters.

The next proposition establishes conditions under which inefficient punishment will happen with positive probability following the (unconfirmed) revelation of a new productive action.

Proposition 3 (costly revelation) *Consider a situation where $N' \leq N$ productive actions have been revealed and confirmed. Define*

$$\underline{V}_2^{N'} \equiv \frac{1}{1 - \delta(1-p)^{N'}} \pi \quad \textit{and} \quad \overline{V}_2 = \frac{1}{1-\delta}(\pi - pr\underline{c}).$$

Whenever

$$\delta(\overline{V}_2 - \underline{V}_2^{N'}) < c, \tag{1}$$

the revelation of a productive action $a \in \mathcal{N} \setminus \mathcal{N}'$ will necessarily require exit on the equilibrium path.

The quantity $\underline{V}_2^{N'}$ is a lower bound for the utility player 2 obtains when player 1 uses strategies that do not involve inefficient exit on the equilibrium path. It is simply the utility player 2 would get if she deviated to taking only costless actions in the future. This deviation can be detected for sure only when a confirmed costly action is available and player 2 does not take it, which happens with probability $1 - (1 - p)^{N'}$. Two important corollaries follow from Proposition 3. First, if $N' = 0$, then condition (1) is necessarily satisfied, which means that initially, inefficient exit will always happen with strictly positive probability. Second, as player 1 learns new productive actions and N' increases, condition (1) becomes harder to satisfy, which indicates that revelation becomes less costly. This happens because there are more circumstances in which player 1 can demand cooperation. This allows player 1 to detect deviations faster and reduces the need for inefficient exit.

Proposition 3 provided conditions under which exit will happen with strictly positive probability following the revelation of a new productive action. Proposition 4 now shows that such inefficient exit need only happen while learning is occurring.

Proposition 4 (no unnecessary exit) *Assume there exists an equilibrium in which player 1 can be induced to stay and consider any Pareto efficient pair of strategies* (s_1, s_2). *Player 1 never exits at an equilibrium history where there are no revealed actions that are unconfirmed.*

Proposition 3 shows that while learning is occurring, the partnership will be sensitive to adverse shocks – a revelation stage at which the action taken by player 2 fails will lead to inefficient punishment. Inversely, Proposition 4 shows that when there are no unconfirmed revealed actions, exit is unnecessary. This implies that once players stop learning (and hence follow a weak routine), the partnership becomes resilient to the adverse shocks ($\tilde{b}(a) = 0$) which necessarily occur on the equilibrium path. This contrasts with the usual imperfect public monitoring framework of Green and Porter (1984) in which some inefficient punishment is always required following adverse shocks.

4 Qualitative properties of efficient strategies in a setting with two productive actions

Proposition 3 established conditions under which the revelation of information will come at an efficiency cost. Information however may also be valuable since identifying an action that yields a high profit for player 1 at the same cost c for player 2 could improve the welfare of both. The trade-off is clear when no information has been revealed yet, since some revelation is required to induce player 1 to stay. This need not be the case anymore when one or more actions have been revealed. This section investigates qualitative aspects of optimal revelation.

For this purpose, the section focuses on a setting where it is common knowledge that there are two productive actions a_0 and a_1 that yield benefits b_0 and b_1 with $b_1 > b_0$. For greater simplicity, the following assumption is maintained throughout.

Assumption 1 *Define* $\overline{r} \equiv \delta\pi/pc$ *and* $\underline{r} = k/pqb_1$. *Parameters* b_1, π,k,c, p,q *and* δ *are such that*

(i) $\overline{r} < 1$;

(ii) $\delta(\overline{V}_2 - \underline{V}_2^D) < c$ *where* $\overline{V}_2 \equiv \frac{1}{1-\delta}(\pi - p\underline{r}c)$ *and* $\underline{V}_2^D \equiv \frac{1}{1-\delta(1-p)}\pi$;

(iii) $\delta q\overline{V}_2 > c$.

Point (i) insures that the maximum rate $\overline{r}$ at which costly actions can be taken by player 2 is strictly less than 1. This facilitates analysis and implies that under complete information, efficient equilibria use only action a_1. Point (ii) of Assumption 1 restricts attention to parameters such that revelation necessarily requires some exit on the equilibrium path. Point (iii) of Assumption 1 insures that there exists an equilibrium of Γ_{AI} in which player 2 reveals productive actions and player 1 stays in the first period.

Lemma 7 given in Appendix D insures that the region of the parameter space defined by Assumption 1 is not empty. This lemma shows in particular that the cost of productive actions c can be chosen to be sufficiently large that player 2 will never cooperate with probability one every time action a_1 (resp. a_0) is available, but sufficiently low that there is an equilibrium for which player 1 can be induced to stay.

4.1 Optimal revelation

This section considers the situation in which player 1 has already confirmed action a_0 and explores under what conditions - if any - it may be optimal that no further information be revealed in the continuation game.

Let us consider a possible revelation stage h_t^2 for action a_1 and assume that action a_1 is actually available to player 2. Since player 1 does not know which actions are productive, player 2 is tempted to take a costless unproductive action rather than a_1. Let us denote player 2 by 2T (for truthful) when she takes action a_1 and by 2D (for deviant) when she takes a different costless action. As before, lower and upper bounds to equilibrium rates of cooperation by player 2 are $\underline{r} = k/pqb_1$ and $\overline{r} = \delta\pi/pc$. Finally, we define

$$\underline{V}_2^D \equiv \frac{1}{1-\delta(1-p)}\pi \quad \textit{and} \quad \overline{V}_2 = \frac{1}{1-\delta}(\pi - p\underline{r}c).$$

By Proposition 3 we know that whenever $c > \delta(\overline{V}_2 - \underline{V}_2^D)$ player 1 will need to use inefficient exit to induce revelation. We can actually make this result slightly stronger.

Lemma 2 (losses are bounded away from 0) *Whenever $c > \delta(\overline{V}_2 - \underline{V}_2^D)$, there exists $K \in \mathbb{N}$ and $\eta > 0$ independent of b_0 such that in any equilibrium, if h_t^2 is a revelation stage, there is probability greater than η that player 1 exits at least once in the next K periods.*

Proposition 5 leverages Lemma 2 and establishes that the efficiency of further revelation will depend on the productivity differential between a_0 and a_1.

Proposition 5 (optimal learning) *Whenever Assumption 1 is satisfied, there exist* $\overline{\Delta} > \underline{\Delta} > 0$ *such that:*

(i) whenever $b_1 - b_0 > \overline{\Delta}$ *there exists* $\nu > 0$ *and* $\tau \in \mathbb{N}$ *such that on the path of all Pareto efficient equilibria, at any history* h_t *where player 1 stays, there is probability greater than* ν *that player 1 will learn action* a_1 *in the next* τ *periods;*

(ii) whenever $b_1 - b_0 < \underline{\Delta}$ *all Pareto efficient equilibria are such that once one action has been confirmed, there is no more information revelation on the equilibrium path.*

Point (i) of the proposition highlights that when the difference between b_1 and b_0 is large then on any Pareto efficient equilibrium, player 1 will keep trying to reveal action a_1. Hence, in the long run, all partnerships which are not in a state of permanent exit will have learned action a_1.

Point (ii) shows that when the difference between b_1 and b_0 is small then it will be optimal to reveal no further information once a_0 is confirmed, even though identifying action a_1 would yield unambiguous efficiency gains. This happens because Assumption 1 implies that as the difference $b_1 - b_0$ shrinks, the value of revealing information shrinks, while Lemma 2 insures that the cost of revealing information does not. It is worthwhile to notice that this happens even though it is common knowledge that action a_1 exists and there is no uncertainty about the Pareto frontier under complete information. This highlights that here, implementing the details of a relational contract requires the costly revelation of strategic information.

4.2 Path dependence

Section 4.1 highlighted that conditional on having confirmed action a_0 it may or may not be optimal to reveal action a_1. This section establishes that this will naturally lead Pareto efficient equilibria to be path dependent, with different partnerships ending up using different long run routines with different degrees of efficiency. Assumption 1 is still maintained.

Proposition 6 (path dependence) *Whenever Assumption 1 holds, there exists* $\epsilon > 0$ *such that whenever* $b_1 - b_0 < \epsilon$ *we have that*

(i) For any Pareto efficient equilibrium, there are histories h *and* h' *with strictly positive probability such that starting from* h *and* h' *the players use weak routines with disjoint supports.*

(ii) For any pair of Pareto efficient values, there exists an equilibrium (s_1, s_2) *sustaining these values and equilibrium histories* h, h'*, such that starting from* h *and* h'*, the strategies* (s_1, s_2) *follow strong routines that have disjoint supports.*

This implies that chance events occurring in the initial stages of the relationship can have a long-lasting impact on the way players approach cooperation: depending on which piece of information gets revealed, partnerships that were identical ex-ante can end up using routines involving different sets of actions and perhaps more importantly different levels of efficiency. Again it is useful to note that this happens even though the Pareto frontier under complete information is common knowledge. This suggests that the kind of routines studied here may persist even in the face of overwhelming evidence that they are not optimal.

4.3 Further comparative statics

An important implication of the model presented in this paper is that a partnership's tendency to routinize will depend on the payoffs of the underlying relational game that agents play. Section 4.1 for instance established qualitative comparative statics of optimal revelation rates with respect to the productivity differential between actions. This general approach opens up a number of questions: for instance is routinization more likely to occur when total surplus is large or when total surplus is low? or how does the tendency to routinize depend on the way surplus is split between player 1 and player 2?

To answer these questions a more explicit characterization of the Pareto efficient frontier would be useful. However, characterizing optimal continuation strategies following failed revelation is difficult. Indeed, if the action taken in a revelation stage fails to yield benefits, we know from Proposition 3 that player 1 must punish player 2 and exit at some frequency. However, efficiency would be improved if player 1 screened the two possible types 2T and 2D before exiting. Solving this optimal screening problem explicitly is delicate because it involves both selection and moral hazard. There are multiple ways for player 1 to screen types 2T and 2D, all of which delay punishment and increase the temptation to deviate (and hence increase the need for punishment). First, player 1 can demand that player 2 take the revealed action a_1 again. Type 2T will succeed in providing benefits with probability q, while type 2D will succeed with probability 0. Player 1 can then use this information to statistically distinguish the two types of player 2. Also, player 1 can exploit the fact that types 2T and 2D have different incentives to stay on the equilibrium path by demanding that player 2 take costly action a_0 whenever possible, with the idea that player 2D will prefer to signal herself rather than pay the cost c. Because solving precisely for the optimal screening equilibrium is delicate, the approach of Section 4.1 has been to focus on qualitative properties that optimal revelation schemes must satisfy.

The alternative and complementary approach explored in Appendix A is to characterize efficient equilibria among a simple class of strategies. For this subclass of equilibria, it is possible to compute explicitly how parameters of the game affect the decision to reveal further actions, and hence the long run distribution of routines. This allows to make precise comparative statics exercises. In particular it is shown that the impact of total surplus on routinization need not be monotonic and that the way total surplus is split between players can significantly affect efficient rates of revelation.

5 Conclusion

The model presented in this paper attempts to capture the process by which agents go about specifying the details of a relational contract. The essence of the modeling approach is to introduce the possibility of learning in an imperfect public monitoring context. At the onset of the relationship players do not know how and when cooperation should take place. This creates moral hazard and leads to inefficient punishment on the equilibrium path. As information gets revealed in equilibrium however, the players learn when to expect cooperation and are able to sustain cooperation without resorting to inefficient punishment.

The model has a number of predictions. The first is that relationships will be sensitive to adverse economic conditions in their initial stages, while learning is occurring. Once learning is over however, relationships routinize and become resilient to shocks. A corollary is that in this setting, failing when using a standard verifiable protocol is more likely to be forgiven than failing using a new action whose consequences are not common knowledge. This happens because taking new actions reintroduces the possibility of moral hazard. Another prediction is that because information revelation is costly, it need not be optimal for players to learn all of the available information. This implies that idiosyncratic events occurring early in the relationship might have long lasting consequences on how the parties approach cooperation. More precisely, depending on what information is revealed early on, pairs that were ex-ante identical can end up using long-run routines involving different actions and achieving different levels of efficiency. This happens even when players have no uncertainty about what the Pareto frontier would be under complete information, which highlights that the difficulties involved in building relational contracts can lead players to routinize their behavior even though there is substantial evidence that these routines are suboptimal.

An interesting aspect of this model of routines is that the cost of revealing new actions is not exogenously given but depends on environmental conditions, such as how divergent the interests of the two players are (c and k), how much surplus ($b(\cdot)$ and π) they can generate, and how much they already know about each other. This allows to link the long run distribution of routines to the fundamentals of the game agents are engaged in. Making that link more explicit is a possible avenue for future research and Appendix A pushes in that direction by characterizing Pareto efficient equilibria within a subclass of simple strategies. It shows that the tendency to routinize depends non-monotonically on potential total surplus, and that the way surplus is shared between the two players has a significant impact on optimal rates of revelation.

Appendix

Figures for Appendix A

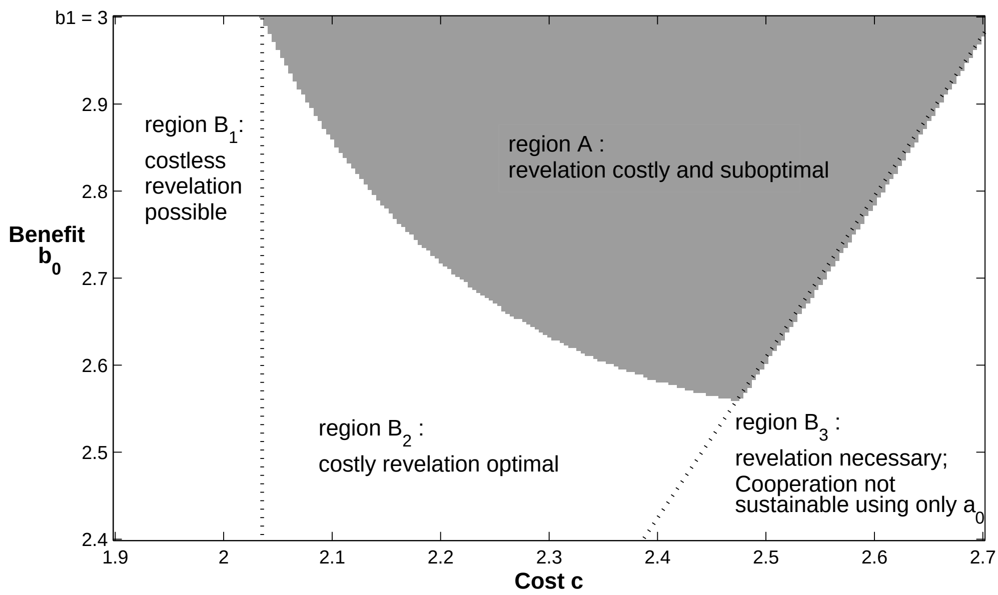

Figure 1: Optimality of revelation as a function of b_0 and c. Parameter values $q = .8$; $p = .8$; $\delta = .8$; $\pi = 2$; $k = 1$; $b_1 = 3$.

A A class of tractable strategies

This section introduces a tractable subclass of automaton strategies for which one can characterize efficient equilibria. This allows to relate the efficient degree of routinization and the economic environment players live in. Note that the strategies considered here take the form of simple automatons but will be required to be equilibria of the asymmetric information

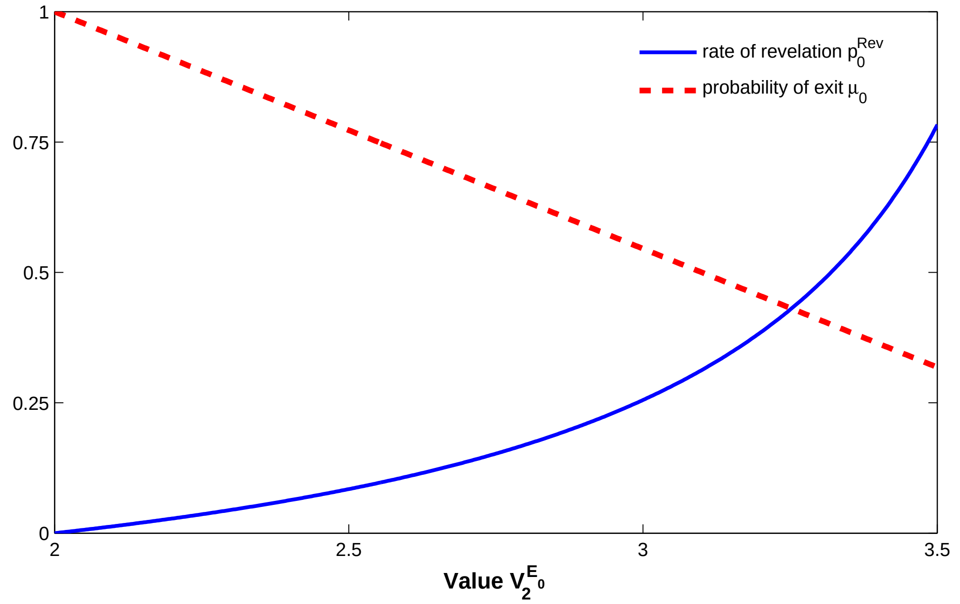

Figure 2: Exit probability and revelation rate as a function of $V_2^{E_0}$. Parameter values: $q = .8$; $p = .8$; $\delta = .8$; $c = 2.2$; $\pi = 2$; $k = 1$; $b_1 = 3$; $b_0 = 2.5$.

game Γ_{AI}. Let us first define

$$\begin{array}{lcl} \underline{V}_1^0 = 0 & ; & \overline{V}_2^0 = \frac{1}{1-\delta}\left(\pi - \frac{kc}{qb_0}\right) \\ \overline{V}_1^0 = \frac{1}{1-\delta}\left(-k + \left[\frac{\pi}{c} - \frac{1-\delta}{\delta}\right] qb_0\right) & ; & \underline{V}_2^0 = \frac{c}{\delta}. \end{array}$$

Values $\underline{V}_1^0$ and $\underline{V}_2^0$ (resp. $\overline{V}_1^0$ and $\overline{V}_2^0$) correspond to the lowest (resp. highest) equilibrium values that players 1 and 2 can obtain using strong routines involving only action a_0. Assumption 1 is maintained throughout the section. We also make the following assumption.

Assumption 2 *Parameter b_0 is high enough so that $\overline{V}_2^0 \geq c/\delta q$.*

This assumption insures that revelation can be incentive compatible even though only action a_0 will be used in the continuation game.

A.1 Automaton strategies

The class of strategies studied here are automaton strategies in which player 1 does not attempt to screen deviator types 2D upon inconclusive revelation. These automatons are defined as follows.

States. For any $C \subset \{0,1\}$, we define states

(i) E_C – this will be an exploration state in which actions $\{a_i | i \in C\}$ have already been confirmed and in which other actions might be revealed.

(ii) R_C – this will be a routine state in which actions $\{a_i | i \in C\}$ have been confirmed and no other action will be revealed. State R_C will follow inconclusive revelation occurring in state E_C.

States $E_{0,1}$ and $R_{0,1}$ are merged since no further actions can be revealed once a_0 and a_1 are confirmed.

Actions and transitions.

- For any $C \subset \{0,1\}$ with $C \neq \emptyset$, actions and transitions in state R_C are specified as follows:

  (i) Player 1 stays.

  (ii) Whenever an action a_i with $i \in C$ is available, player 2 takes the most productive such action with probability $r_i^{R_C}$.

  (iii) For all equilibrium outcomes the subsequent state is R_C.

  (iv) For out-of-equilibrium outcomes the subsequent state is $R_\emptyset$.

- In state R_C the players use routine strategies involving only productive actions in $\{a_i | i \in C\}$. By extension in $R_\emptyset$ the players use the only equilibrium strategy in which no productive action is used: player 1 exits in every period and player 2 never takes a productive action.

- For any $C \subset \{0,1\}$ with $C \neq \{0,1\}$, actions and transitions in state E_C are specified as follows:

  (i) Player 1 stays.

  (ii) Whenever action a_i with $i \in C$ is available (there is at most one such action), player 2 takes it with probability $r_i^{E_C}$.

  (iii) If no confirmed action is being taken, a revelation stage is initiated with probability ϕ_C^{Rev}. If a revelation stage is initiated, player 2 takes the most productive available action a_i with $i \in \{0,1\} \setminus C$ (and takes an unproductive action if there is no unconfirmed productive action available).

(iv) When no revelation is initiated, the next state is E_C. If revelation is initiated and a new action a is confirmed, the next state is $E_{C\cup\{a\}}$. If revelation is inconclusive the next state is $R_\emptyset$ with probability μ_C and R_C with probability $1-\mu_C$.

(v) For out-of-equilibrium outcomes the subsequent state is $R_\emptyset$.

Equilibrium play when players use automaton strategies follows a simple structure: initially the players are in an exploratory state and try to get some action revealed and confirmed. If an action is confirmed, the players stay in an exploratory state in which confirmed actions are used and new actions can be revealed. If an action is inconclusively revealed, then exploration stops and the players end up either in permanent exit (state $R_\emptyset$) or follow a routine that uses only the actions confirmed so far. For any state $S \in \{E_C, R_C | C \subset \{0,1\}\}$ let us denote (V_1^S, V_2^S) the values players 1 and 2 expect in state S. We begin by characterizing equilibrium automaton strategies.

Lemma 3 (equilibrium automaton strategies) *A pair of automaton strategies (s_1, s_2) is an equilibrium if and only if the following hold.*

(i) For any state $S \in \{E_C, R_C | C \subset \{0,1\}\}$, $V_1^S \geq 0$.

(ii) For any state $S \in \{E_C, R_C | C \subset \{0,1\}\}$, if for any $i \in C$, $r_i^S > 0$ then $V_2^S \geq c/\delta$.

(iii) For all $C \subsetneq \{0,1\}$, if $\phi_C^{Rev} > 0$, then for all $i \in \{0,1\} \setminus C$, $V_2^{E_{C\cup\{i\}}} \geq c/\delta q$ and

$$\delta q (V_2^{E_{C\cup\{i\}}} - (1-\mu_C) V_2^{R_C}) \geq c. \tag{2}$$

A.2 The Pareto frontier of automaton equilibria

We are now interested in characterizing Pareto efficient automaton equilibria. For this purpose, given any state $S \in \{E_C, R_C | C \subset \{0,1\}\}$ we define $\mathcal{V}_S$ the Pareto frontier of continuation equilibrium values sustainable by automatons starting in state S. We are interested in the properties of equilibrium automatons attaining values is $\mathcal{V}_{E_\emptyset}$.

Lemma 4 (some properties of efficient automatons) *Whenever (s_1, s_2) is Pareto efficient among the class of automaton equilibria, the following hold*

(i) For all $C \subset \{0,1\}$, continuation values $(V_1^{E_C}, V_2^{E_C})$ belong to $\mathcal{V}_{E_C}$.

(ii) $\phi_{E_\emptyset}^{Rev} > 0$ *and* $\phi_{E_1}^{Rev} = 0$

(iii) $V_2^{R_0} = \underline{V}_2^0$ *and* $1 - \mu_0 = \frac{1}{\underline{V}_2^0}(V_2^{E_{0,1}} - c/\delta q)$.

Point (i) of Lemma 4 shows that while revelation is successful, continuation play should be Pareto efficient. Point (ii) of Lemma 4 highlights that in the initial state $E_\emptyset$ some revelation must happen, otherwise player 1 would permanently exit, which is inefficient. Point (iii) states that when revelation is inconclusive in state E_0, player 2's value in state R_0 should be minimized, and that the probability of exit μ_0 is entirely pinned down by the incentive compatibility condition (2).

Upon conclusive revelation either action a_0 or a_1 gets confirmed. If action a_1 is confirmed no further revelation occurs. From Assumption 1 we know that the Pareto frontier in state E_1 and $E_{0,1}$ is identical to the Pareto frontier of the game in which only action a_1 is productive. Accordingly, the rest of this section focuses on efficient continuation strategies in state E_0. Define $p_0^{Rev} \equiv (1-pr_0^{E_0})\phi_0^{Rev}$. Given point (iii) of Lemma 4, values $V_1^{E_0}$ and $V_2^{E_0}$ must satisfy

$$\begin{aligned} V_1^{E_0} &= -k + pr_0^{E_0}qb_0 + p_0^{Rev}pqb_1 \\ &\quad +\delta p_0^{Rev}\left[pqV_1^{E_{0,1}} + (1-pq)(1-\mu_0)V_1^{R_0}\right] + \delta(1-p_0^{Rev})V_1^{E_0} \end{aligned} \tag{3}$$

$$\begin{aligned} V_2^{E_0} &= \pi - pr_0^{E_0}c - p_0^{Rev}pc \\ &\quad +\delta p_0^{Rev}\left[pqV_2^{E_{0,1}} + (1-pq)(1-\mu_0)V_2^{R_0}\right] + \delta(1-p_0^{Rev})V_1^{E_0}. \end{aligned} \tag{4}$$

Let us define $\gamma \equiv \frac{(1-\delta)(1-p_0^{Rev})}{1-\delta(1-p_0^{Rev})}$ and $\hat{r}_0 = \frac{r_0}{1-p_0^{Rev}}$. We have that $\gamma \in [0,1]$ and $\hat{r}_0 \in [0, 1/p]$. Equations (3) and (4) imply that

$$V_1^{E_0} = \gamma\frac{\left[-k + p\hat{r}_0^{E_0}qb_0\right]}{1-\delta} + (1-\gamma)\left[-k + pqb_1 + \delta pqV_1^{E_{0,1}} + \delta(1-pq)(1-\mu_0)\overline{V}_1^0\right] \tag{5}$$

$$V_2^{E_0} = \gamma\frac{\left[\pi - p\hat{r}_0^{E_0}c\right]}{1-\delta} + (1-\gamma)\left[\pi - pc + \delta pqV_2^{E_{0,1}} + \delta(1-pq)(1-\mu_0)\underline{V}_2^0\right]. \tag{6}$$

This leads us to the following definition

Definition 3 *For any* $\hat{r}_0 \in [0, 1/p]$ *and value* $V_2 \geq c/\delta q$, *we define,*

$$\begin{aligned} &V_1^0(\hat{r}_0) = \tfrac{1}{1-\delta}(-k + pq\hat{r}_0b_0) \quad ; \quad V_2^0(\hat{r}_0) = \tfrac{1}{1-\delta}(\pi - p\hat{r}_0c) \\ &V_1^1(V_2) = \tfrac{qb_1}{c}\left[\tfrac{1}{1-\delta}\pi - V_2\right] - \tfrac{1}{1-\delta}k \quad ; \quad 1-\mu_0(V_2) = (V_2 - c/\delta q)/\underline{V}_2^0 \end{aligned}$$

$$\begin{aligned} V_1^{Rev}(V_2) &= -k + pqb_1 + \delta pq V_1^1(V_2) + \delta(1-pq)(1-\mu_0)\overline{V}_1^0 \\ V_2^{Rev}(V_2) &= \pi - pc + \delta pq V_2 + \delta(1-pq)(1-\mu_0)\underline{V}_2^0 \end{aligned}$$

Equations (5) and (6) express efficient continuation values $(V_1^{E_0}, V_2^{E_0})$ as weighted averages of values with no further revelation $(V_1^0(\hat{r}_0), V_2^0(\hat{r}_0))$ and values upon immediate revelation $(V_1^{Rev}(V_2), V_2^{Rev}(V_2))$. This decomposition provides simple sufficient and necessary conditions for revelation to be efficient.

Proposition 7 (sufficient and necessary conditions for efficient revelation) *We define sets of values $\mathcal{W}_0$ and $\mathcal{W}_{Rev}$ as follows*

$$\begin{aligned} \mathcal{W}_0 &= \{(V_1^0(\hat{r}_0), V_2^0(\hat{r}_0)) | \hat{r}_0 \in [0, 1/p]\} \\ \mathcal{W}_{Rev} &= \{(V_1^{Rev}(V_2), V_2^{Rev}(V_2)) | V_2 \geq c/\delta q \text{ and } V_1^1(V_2) \geq 0\}. \end{aligned}$$

For any Pareto efficient automaton equilibrium, $p_0^{Rev} > 0$ if and only if $\mathcal{W}_{Rev}$ is not dominated by $\mathcal{W}_0$.

It is important to note that sets $\mathcal{W}_0$ and $\mathcal{W}_{Rev}$ are line segments when expressed in (V_1, V_2) coordinates. Hence checking whether $\mathcal{W}_0$ dominates $\mathcal{W}_{Rev}$ boils down to checking whether the end points of $\mathcal{W}_{Rev}$ are dominated. This allows for simple explicit computations. Figure 1 represents whether revelation is efficient or not as a function of cost c and benefit b_0. It highlights that whenever revelation is costly, then as b_0 becomes closer to b_1 revelation becomes suboptimal. Figure 1 also shows that increasing cost c has a non-monotonic effect on the optimality of revelation. In the transition from region B_1 to B_2, increasing c increases the need for inefficient exit upon inconclusive revelation. Accordingly region A corresponds to settings for which further revelation in E_0 is suboptimal. In the transition from B_2 to B_3 however, as cost c increases, greater values of b_0 are needed for revelation to be incentive compatible when the continuation strategy uses only action a_0.

The decomposition of equations (5) and (6) also allow for a simple characterization of the Pareto frontier.

Proposition 8 (characterizing the Pareto frontier) *The Pareto frontier $\mathcal{V}_{E_0}$ of automaton equilibria conditional on being in E_0 is the Pareto frontier of co hull$(\mathcal{W}_0 \cup \mathcal{W}_{Rev}) \cap \{(V_1, V_2) | (V_1, V_2) \geq (0, c/\delta q)\}$.*

This characterization of the Pareto frontier is particularly tractable since it only requires to compute the convex hull of two line segments. Moreover the following lemma highlights that for generic cases, a point on the Pareto frontier is attained by a unique automaton equilibrium.

Lemma 5 (unique implementation) *For generic values of b_0, given a value $V_2 \geq c/\delta q$, and the associated value V_1 on the frontier $\mathcal{V}_{E_0}$, there exists a unique set of values $r_0, p_0^{Rev}, V_2^{E_0,1}$ such that $(V_1^{E_0}, V_2^{E_0}) = (V_1, V_2)$.*

This lemma allows to link efficient routinization to both fundamentals of the game and the way players share surplus. Figure 2 shows how the surplus player 2 obtains affects the optimal revelation rate p_0^{Rev} and the necessary exit probability μ_0 upon inconclusive revelation. As player 2 gets a greater share of the surplus, the need for inefficient punishment decreases and it becomes efficient to reveal the more productive action a_1 at a greater rate.

B Communication

Section 2.4 briefly commented on the fact that even though the model does not allow for communication, the paper's predictions are unchanged if players can communicate, as long as their communication technology doesn't allow them to successfully communicate about actions unless those actions have already been taken by player 2. This appendix considers such a model and highlights that this indescribability assumption isn't incompatible with the players being able to communicate about their joint histories.

Player 1 labels actions in $\mathcal{A}$ with numbers in $\mathbb{N}$. Player 2 also labels action in $\mathcal{A}$ with numbers in $\mathbb{N}$ but as in Crawford and Haller (1990), the players do not share a common labeling. More precisely, if player 2 labels some action as k, then player 1 labels the same action as $\gamma(k)$ where γ is a permutation of $\mathbb{N}$. By default we will be denoting h_t and c the histories and cost function from the perspective of player 2 and $h_t \circ \gamma^{-1}$ and $c \circ \gamma^{-1}$ the same history and cost function from the perspective of player 1. Player 2 is allowed to communicate at every history h_t (communication by Player 1 isn't useful since she doesn't have private information) by sending messages belonging to a set of words $\mathcal{M}$. A message strategy of player 2 is simply a mapping $m : (c, h_t) \mapsto m(c, h_t) \in \mathcal{M}$.

The assumption that actions are indescribable by words corresponds to an assumption about the permutation γ. More specifically, while γ is fixed for the course of the game, from an ex-ante perspective it is a random variable whose realization is unknown to both players. The two players' common prior about γ is an improper uniform prior over the set

$\mathcal{G}$ of permutations[5] over $\mathbb{N}$. In particular, if γ_0 is a specific permutation and G is a finite set of permutations, the following property holds,

$$prob(\gamma = \gamma_0 \mid \gamma \in G \text{ and } \gamma_0 \in G) = \frac{1}{\operatorname{card} G}.$$

The rest of the section shows how in this framework, actions are essentially indescribable unless they have been taken by player 2. For this we need a couple of definitions.

A history h_t is a sequence $\{d_1, x_1, w_1, a_1, \tilde{b}(a_1), \ldots, d_{t-1}, x_{t-1}, w_{t-1}, a_{t-1}, \tilde{b}(a_{t-1})\}$. The subsequence $\tilde{h}_t = \{w_1, a_1, w_2, a_2, \ldots, w_t, a_{t-1}\}$ can be viewed as a sequence of subsets of $\mathbb{N}$, where for all $s \geq 0$, w_s is identified with the set of available actions $A(w_s)$ and a_s is simply identified with the singleton $\{a_s\}$. We call $\sigma(h_t)$ the σ-algebra over $\mathbb{N}$ generated by the sequence of sets $\tilde{h}_t$. Note that $\sigma(h_t)$ is finite. The partition $P(h_t)$ induced by h_t is the finest partition whose elements are measurable with respect to $\sigma(h_t)$. Each element of this partition corresponds to a set of actions that are undistinguishable based on history h_t. We say that a permutation θ is *consistent* with history h_t if for each set $P_k \in P(h_t)$, $\theta(P_k) = P_k$. Similarly, we say that a permutation θ is consistent with history $h_t \circ \gamma^{-1}$ if for each set $P_k \in P(h_t)$, $\theta \circ \gamma(P_k) = \gamma(P_k)$.The following lemma shows that a message strategy $m(\cdot, \cdot)$ can only convey information about actions up to a consistent permutation.

Lemma 6 *Consider an action $a \in \mathbb{N}$. Then for any permutation θ consistent with $h_t \circ \gamma^{-1}$ and for any message strategy $m(\cdot, \cdot)$, player 1's beliefs are such that*

$$Prob_1[a \in \gamma(\mathcal{N}) \mid m(c, h_t), h_t \circ \gamma^{-1}] = Prob_1[\theta(a) \in \gamma(\mathcal{N}) \mid m(c, h_t), h_t \circ \gamma^{-1}].$$

Proof: By the law of iterated expectation, we have that

$$\begin{aligned} &Prob_1[\theta(a) \subset \gamma(\mathcal{N}) \mid m(c, h_t), h_t \circ \gamma^{-1}] = \\ &\qquad \mathbb{E}\left[Prob_1[\theta(a) \subset \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}] \middle| m(c, h_t), h_t \circ \gamma^{-1} \right]. \end{aligned}$$

Hence it is sufficient to prove that

$$Prob_1[\theta(a) \subset \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}] = Prob_1[a \subset \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}].$$

[5]Note that $\mathcal{G}$ is in bijection with $\mathbb{R}$.

We have that for any permutation u consistent with $h_t \circ \gamma^{-1}$,

$$Prob_1[a \subset \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}] = Prob_1[a \subset u \circ \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}\}]$$

Hence,

$$\begin{aligned} Prob_1[\theta(a) \subset \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}] &= Prob_1[a \subset \theta^{-1} \circ \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}\}] \\ &= Prob_1[a \subset \gamma(\mathcal{N}) \mid \{c; h_t\}; h_t \circ \gamma^{-1}\}] \end{aligned}$$

where we used the fact that θ^{-1} is consistent with $h_t \circ \gamma^{-1}$. This concludes the proof. ■
An interesting corollary is that unless an action has been taken by player 2, no valuable information can be conveyed about it.

Corollary 2 *Consider a history h_t and consider the partition $P(h_t)$ induced by h_t. For any action a such that $a \notin h_t \circ \gamma^{-1}$ we have that*

(i) $\gamma^{-1}(a) \in P_k \in P(h_t) \Rightarrow$ *card* $P_k =$ *card* $\mathbb{N}$.

(ii) *For any message strategy $m(\cdot, \cdot)$, $Prob_1(a \in \gamma(\mathcal{N}) | m(c, h_t); h_t \circ \gamma^{-1}) = 0$.*

Point (i) holds because, when a hasn't been taken, any finite intersection of sets of $\tilde{h}_t$ that contain a is infinite. Point (ii) holds because then, the set $\{u(a) | u$ consistent with $h_t \circ \gamma^{-1}\}$ has an infinite cardinal.[6] Since for all such u, $Prob_1[a \in \gamma(\mathcal{N}) \mid m(c, h_t), h_t \circ \gamma^{-1}] = Prob_1[u(a) \in \gamma(\mathcal{N}) \mid m(c, h_t), h_t \circ \gamma^{-1}]$, this means that $Prob_1[a \in \gamma(\mathcal{N}) \mid m(c, h_t), h_t \circ \gamma^{-1}] = 0$.

Hence in this framework, communication about actions is only effective when actions have been taken. In this sense the model can be viewed as a model of communication building in which taking actions enriches the common vocabulary of the players.

C Cash and utility burning

An essential element of the model is that in order to provide incentives for revelation, player 1 must sometimes take actions against player 2 that are costly to both. This is a standard feature of games with imperfect public monitoring. In a number of environments however, it is reasonable to assume that players can either transfer cash or burn utility. This would

[6] By comparison, if a had been taken, this set would be a singleton.

tend to make incentive compatibility constraints less costly to satisfy. For instance, when revelation occurs without confirmation, player 2 could simply burn utility or even better, transfer cash.

Still it is not realistic to conclude that the availability of either cash or utility burning would dispense with the need for inefficient punishment entirely. Indeed, how player 2 values cash versus the cost of taking a cooperative action is really information that is private to player 2. This will be the case as well when player 2 takes an action that burns utility. The informational asymmetry implies that when player 1 punishes player 2 by demanding either transfers or utility burning, player 2 will still manage to save some residual utility. In that case, the only way to push player 2's value further down will be to use inefficient punishment.

Hence, there is strong presumption that the model's main predictions are robust to the introduction of transfers or utility burning although such instruments would likely make revelation less inefficient and routines less sticky.

D Proofs

D.1 Results of Section 3

Proof of Proposition 1: Assume there exists an equilibrium in which Player 1 stays at least once. Then there exists an equilibrium (s_1^0, s_2^0) in which Player 1 stays in the first period. Denote (V_1^0, V_2^0) the associated initial values. We necessarily have $V_1^0 \geq 0$ and $V_2^0 \geq \pi$. Now consider an equilibrium (s_1, s_2) on the Pareto frontier of Γ_{FI}. Assume that there is a history h_t^1 attainable on the equilibrium path at which player 1 decides to exit. We now show that (s_1, s_2) cannot be efficient.

Let us first consider the case where in the subgame starting from h_t^1, strategy s_1 prescribes that player 1 should never stay again. Consider the alternative strategies $\tilde{s}_1$ and $\tilde{s}_2$ defined by:

$$\forall i \in \{1,2\},\ \tilde{s}_i(h) = \begin{cases} s_i(h) & \text{if } \nexists h' \ s.t.\ h = h_t^1 \sqcup h' \\ s_i^0(h') & \text{if } h = h_t^1 \sqcup h' \end{cases}.$$

By construction $(\tilde{s}_1, \tilde{s}_2)$ is also an equilibrium and it strictly dominate (s_1, s_2).

Let us now consider the case in which following h_t^1 there is an attainable history at which player 1 stays under s_1. This implies that at h_t^1 the values associated with s_1 and s_2 are positive and player 2 gets strictly positive value. Consider the alternative strategies in which history h_t^1 is skipped:

$$\forall i \in \{1,2\},\ \tilde{s}_i(h) = \begin{cases} s_i(h) & \text{if } \nexists h' \ s.t.\ h = h_t^1 \sqcup h' \\ s_i(h^1 \sqcup (E, \emptyset, \emptyset, \emptyset, \emptyset) \sqcup h') & \text{if } h = h_t^1 \sqcup h' \end{cases}.$$

By construction $(\tilde{s}_1, \tilde{s}_2)$ is also an equilibrium and it strictly dominate (s_1, s_2). This concludes the proof. ■

Proof of Proposition 2: We know from Proposition 1 that on the equilibrium path of a Pareto efficient equilibrium, player 1 always decides to stay. This gives player 1 value $-\frac{1}{1-\delta}k$ and player 2 value $\frac{1}{1-\delta}\pi$. When player 2 takes productive action a, utility is transferred from player 2 to player 1 at the rate $\frac{qb(a)}{c}$. Since all productive actions have the same cost, it is efficient to have player 2 use the most productive action available.

Given $n \in \{0, \ldots, N\}$ and $r \in [0,1]$, consider strategies $(s_1, s_2)_{n,r}$ such that

(i) if $w_t(a_{N-n+1}, \ldots, a_{N-1}, a_N) \neq (0, \cdots, 0)$ player takes the highest productive action available among $(a_{N-n+1}, \ldots, a_{N-1}, a_N)$ with probability 1

(ii) if $w_t(a_{N-n+1}, \ldots, a_{N-1}, a_N) = (0, \cdots, 0)$ and $w_t(a_{N-n}) = 1$ then player 2 takes action a_{N-n-1} with probability r.

(iii) player 1 stays every period while player 2 has behaved as expected and exits otherwise.

Denote $V_1^{n,r}$ and $V_2^{n,r}$ the values players 1 and 2 obtain under these strategies.By construction $V_1^{n,r}$ and $V_2^{n,r}$ lie on the Pareto efficient *feasible* frontier. Whenever $V_1^{n,r} \geq 0$ and $\delta V_2^{n,r} \geq c$, the strategies $(s_1, s_2)_{n,r}$ also form an equilibrium. Denote n^*, r^* the values such that $V_2^{n^*,r^*} = c$. The value $V_1^{n^*,r^*}$ is the highest value player 1 can expect at a history where player 2 has taken a productive action.

In order to provide player 1 an initial value greater than $V_1^{n^*,r^*}$ we cannot increase transfers after player 2 has taken a productive action. Hence the only way to give player 1 greater initial values is to increase the likelihood that player 2 starts taking productive actions. The efficient way to transfer is again to have player 2 take the most productive action available as often as possible. This concludes the proof. ■

Proof of Lemma 1: Consider an action a that has never been taken up to history h_{t+1}. Consider the set of actions $E = \{a' \in \mathcal{A} \mid \forall s \leq t, a \in A(w_s) \Leftrightarrow a' \in A(w_s)\} \setminus \{a_s\}_{s \in \{0,\ldots,t\}}$.

Actions in E aren't distinguished by the history h_{t+1}. Hence for any action $a' \in E$, we have that $Prob_1(a \in \mathcal{N}|h_t) = Prob_1(a' \in \mathcal{N}|h_t)$. Furthermore, set E contains an infinite number of actions, hence it must be that $Prob_1(a \in \mathcal{N}|h_t) = 0$. ■

Proof of Proposition 3: Assume that player 1 never exits following histories that could be attainable on the equilibrium path. Consider some history h_t^2 with set of available actions $A(w_t)$ such that no action in $\mathcal{N}'$ is available. Such histories happen with probability $1-(1-p)^{N'}$. Since this is an equilibrium history and player 1 never exits following histories that could be attainable on the equilibrium path, there must be some non-empty set of actions $A_{Stay} \subset A(w_t)$ such that player 1 stays whenever player 2 takes an action $a \in A_{Stay}$. There are two cases. Either A_{Stay} is finite, in which case $Prob_1(A_{Stay} \cap \mathcal{N}|h_t) = 0$, or A_{Stay} is infinite, in which case there necessarily is a costless action $a_{costless} \in A_{Stay}$. Since $A_{Stay} \neq \emptyset$ this implies that at such a history h_t^2, there is always a costless action that player 2 can take, and following which player 1 will stay.

This means that player 2 can guarantee herself value greater than $\frac{1}{1-\delta(1-p)^{N'}}\pi$ by only taking costless actions. At a revelation stage, for the revelation of a costly action a to be incentive compatible it must be that

$$\delta(\mathbb{E}[V_2|a_t = a] - \mathbb{E}[V_2|a_t \neq a]) \geq c.$$

Since $\mathbb{E}[V_2|a_t = a] \leq \overline{V}_2$, the fact that $\delta(\overline{V}_2 - \underline{V}_2^{N'}) < c$ implies that $\mathbb{E}[V_2|a_t \neq a] < \underline{V}_2^{N'}$. This means that there must be inefficient exit on the equilibrium path following revelation. ■

Proof of Proposition 4: The proof is very similar to that of Proposition 1. Consider a history h such that all revelation stages have been confirmed. Strategies conditional on h do not affect player 2's temptation to deviate at any revelation stage since confirmation is unattainable if player 2 deviates. If player 1 exits at h then using the method of Proposition 1 we know that we can replace the continuation strategies at h by equilibrium strategies which give greater value to both players: if player 1 stays after h, then simply skip history h; if player 1 does not stay, then simply plug in any equilibrium where player 1 stays with some probability. ■

D.2 Results of Section 4

Lemma 7 *Given π, b_1, k and δ, pick q and p in $(0,1)$ such that*

$$q > \max\{\delta(1-k/b_1)\,;\,1-\delta(1-k/b_1)\} \;\;;\;\; p > \frac{1}{q}[1-\delta(1-k/b_1)]$$

and define

$$c_{max} \equiv \frac{\delta q \pi}{1-\delta(1-k/b_1)}.$$

There exists $\varepsilon > 0$ such that for all $c \in [c_{max}-\varepsilon, c_{max})$, Assumption 1 holds.

Proof of Lemma 7: Let us begin with point (i). For all $c < c_{max}$, we have that $q\frac{\delta}{1-\delta}(\pi - pc\underline{r}) > c$. This implies that there exists an equilibrium of Γ_{AI} in which player 1 stays in the first period. In particular, it is an equilibrium for player 1 to stay and demand that player 2 provide benefit b_1 in the first period. If this fails, then player 1 exits permanently. If this succeeds, then player 1 demand that player 2 take action a_1 at rate $\underline{r}$ whenever possible. The fact that $q\frac{1}{1-\delta}\delta(\pi - pc\underline{r}) > c$ guarantees that taking action a_1 (if possible) is incentive compatible for player 2. This shows point (i).

We now show point (ii). We simply need to prove that both properties hold as c converges to c_{max}. We have that $\overline{r} = \frac{\delta\pi}{pc}$. Hence as c converges to c_{max}, $\overline{r}$ converges to $\frac{1-\delta(1-k/b_1)}{pq}$, which is less than 1. Similarly, we have that as c converges to c_{max}, $\overline{V}_2$ converges to $c_{max}/\delta q$. Hence $\delta(\overline{V}_2 - \underline{V}_2)$ is asymptotically less than $c_{max}/q - \delta\pi$. The assumption that $q > \delta(1-k/b)$ guarantees that $c_{max}/q - \delta\pi < c_{max}$. This concludes the proof of (ii). ■

Proof of Lemma 2: Assume that for all $K \in \mathbb{N}$ and all $\eta > 0$, there exists an equilibrium (s_1, s_2) and a revelation stage such h_t^2 that following h_t^2, there is probability less than η that player 1 exits in the next K periods. The idea of the proof is to show that if this is the case, a deviating player 2 can guarantee herself payoffs arbitrarily close to $\underline{V}_2^D$ and that therefore, by the argument of Proposition 3, revelation is not incentive compatible.

Let us consider subsequent histories h_s^2 with $s \leq t+K$, such that a_1 is still unconfirmed and the confirmed action a_0 has not been available. On the equilibrium path such histories have probability at least $(1-q)^{s-t}(1-p)^{s-t}$ and hence following such histories, exit can only occur with probability less than

$$\frac{\eta}{(1-q)^{s-t}(1-p)^{s-t}} \leq \frac{\eta}{(1-q)^K(1-p)^K}.$$

Out of equilibrium, if player 2 deviates by taking only costless action, the likelihood that a_1 is still unconfirmed and the confirmed action a_0 has not been available is $(1-p)^{s-t}$. Hence using such a strategy, player 2 obtains a minimum payoff

$$\begin{aligned} V_2 &\geq \sum_{s=t}^{t+K} \delta_{s-t}(1-p)^{s-t}\left(1 - \frac{\eta}{(1-q)^K(1-p)^K}\right)\pi \\ &\geq \left(1 - \frac{\eta}{(1-q)^K(1-p)^K}\right)\frac{1-\delta_{K+1}(1-p)^{K+1}}{1-\delta(1-p)}\pi. \end{aligned}$$

For K arbitrarily large and η arbitrarily low, this lower bound is arbitrarily close to $\underline{V}_2^D$, which contradicts the incentive compatibility of revelation since $\delta(\overline{V}_2 - \underline{V}_2^D) < c$.

Hence there exists $K \in \mathbb{N}$ and $\eta > 0$ such that for any equilibrium such that h_t^2 is a revelation stage for action a_1, there is probability greater than η that player 1 exits in the next K periods. ■

Proof of Proposition 5: We begin with point (i). Given that b_1 is fixed, then for $\overline{\Delta}$ large enough, we must have that $b_0 < ck/q\pi$. This implies that there is no subgame perfect equilibrium in which player 1 stays and player 2 uses only action a_0 to transfer back utility. This implies that action a_1 must be revealed with some probability in equilibrium. Now assume that for all ν and all τ, there exists a history h at which player 1 stays and there is probability less than ν that player 1 learns action a_1 over the next τ periods. By making τ arbitrarily large and ν arbitrarily close to 0 we obtain that continuation values at the corresponding history h must be arbitrarily close to some continuation values of the game where a_0 is the only available action. Since staying is not incentive compatible in that game, we obtain a contradiction.

We now turn to point (ii). Let us first define the sets of values $\mathcal{U}_0$, $\mathcal{U}_1$ and $\mathcal{U}_1^{K,\eta}$ as follows:

(i) $\mathcal{U}_0$ is the set of Pareto efficient continuation equilibrium values, at a history $h^2 \in \mathcal{H}^2$ where a_0 is available, in the complete information game where only a_0 is productive.

(ii) $\mathcal{U}_1$ is the set of Pareto efficient continuation equilibrium values, at a history $h^2 \in \mathcal{H}^2$ where a_1 is available, in the complete information game where only a_1 is productive.

(iii) $\mathcal{U}_1^{K,\eta}$ is the set of values sustained at a history $h^2 \in \mathcal{H}^2$ where a_1 is available, in the complete information game where only a_1 is productive, by equilibria such that player 1 exits with probability greater than η in the next K period.

By Assumption 1, we have that $\delta q\overline{V}_2 > c$, which implies that there exists an equilibrium of Γ_{AI} in which player 1 stays in the first period. The fact that the inequality is strict implies that as b_0 goes to b_1, the set of values $\mathcal{U}_0$ converges to $\mathcal{U}_1$ in the sense that for all $\epsilon > 0$, there exists b_0 close enough to b_1 such that for all $(V_1, V_2) \in \mathcal{U}_1$, there exists $(\hat{V}_1, \hat{V}_2) \in \mathcal{U}_0$ such that $\hat{V}_1 \geq V_1 - \epsilon$ and $\hat{V}_2 \geq V_2 - \epsilon$.

Inversely, for any $K \in \mathbb{N}$ and $\eta > 0$, there exists $\alpha > 0$ such that for all $(V_1', V_2') \in \mathcal{U}_1^{K,\eta}$, there exists $(V_1, V_2) \in \mathcal{U}_1$ such that $V_1 \geq V_1' + \alpha$ and $V_2 \geq V_2' + \alpha$. This implies that we can pick $\underline{\Delta} > 0$ small enough that for all $b_0 > b_1 - \underline{\Delta}$, whenever $(V_1', V_2') \in \mathcal{U}_1^{K,\eta}$, there exists $(\hat{V}_1, \hat{V}_2) \in \mathcal{U}_0$ such that $\hat{V}_1 \geq V_1' + \alpha/2$ and $\hat{V}_2 \geq V_2' + \alpha/2$.

Let us now consider a revelation stage h_t^2 for action a_1. By Lemma 2, there exists K and $\eta > 0$ such that values (V_1^{Rev}, V_2^{Rev}) at h_t^2 are dominated by values in $\mathcal{U}_1^{K,\eta}$. This implies that for all $b_0 \geq b_1 - \underline{\Delta}$, there exists $(\hat{V}_1, \hat{V}_2) \in \mathcal{U}_0$ such that $\hat{V}_1 > V_1^{Rev}$ and $\hat{V}_2 > V_2^{Rev}$.

Let us denote $(V_1^{Conf,0}, V_2^{Conf,0})$ continuation values at the history h where action a_0 was confirmed. We must have $V_1^{Conf,0} \geq 0$ and $\delta V_2^{Conf,0} \geq c$. Define the pair of real numbers

$$\hat{V}_1^{Conf,0} \equiv V_1^{Conf,0} + prob(h_t^2)\left(\hat{V}_1 - V_1^{Rev}\right) \quad \text{and} \quad \hat{V}_2^{Conf,0} \equiv V_2^{Conf,0} + prob(h_t^2)\left(\hat{V}_2 - V_2^{Rev}\right)$$

obtained from replacing revelation values at h_t^2 with continuation values not involving revelation. We have that $\hat{V}_1^{Conf,0} > V_1^{Conf,0}$ and $\hat{V}_2^{Conf,0} > V_2^{Conf,0}$. Repeat the same replacement operation at all revelation stages occurring after action a_0 is confirmed. We obtain values $\widetilde{V}_1^{Conf,0} > V_1^{Conf,0}$ and $\widetilde{V}_2^{Conf,0} > V_2^{Conf,0}$. By construction these values are such that player 1 only ever gets benefit b_0 and dominate original values involving further revelation. The question is whether such values correspond to a continuation equilibrium. Let us show that this is indeed the case.

Consider $r > 0$ such that $\widetilde{V}_1^{Conf,0}$ and $\widetilde{V}_2^{Conf,0}$ can be written

$$\widetilde{V}_1^{Conf,0} = -\frac{1}{1-\delta}k + \frac{1}{1-\delta}prqb_0 \quad \text{and} \quad \widetilde{V}_2^{Conf,0} = \frac{1}{1-\delta}\pi - \frac{1}{1-\delta}prc.$$

We have that $\widetilde{V}_1^{Conf,0} > 0$ and $\delta\widetilde{V}_2^{Conf,0} > c$. By Assumption 1 this implies that $r < 1$. Hence values $\widetilde{V}_1^{Conf,0}$ and $\widetilde{V}_2^{Conf,0}$ are supported by the continuation equilibrium in which player 1 always stays on the equilibrium path and player 2 cooperates at rate r whenever action a_0 is available.

This concludes the proof: efficient equilibria should involve no further revelation upon confirmation of a_0. ■

Proof of Proposition 6: Let us first work under the assumption that there are revelation stages in which a_0 is revealed. Pick $\epsilon < \underline{\Delta}$ where $\underline{\Delta}$ was defined in Proposition 5. Point (i) follows naturally Proposition 5. Consider histories h and h' that are both revelation and confirmation stages such that at h the action being revealed is a_0 and at h' the action being revealed is a_1. It follows from Proposition 5 that following both these histories, there is no more learning. This implies that starting from these two histories, we have weak routines of respective supports $\{a_0\}$ and $\{a_1\}$.

Now for point (ii) consider the histories h and h' defined above. Continuation equilibria are also equilibria of complete information games in which only a_1 or a_0 are available. By Proposition 2 we know that without loss of generality, Pareto efficient equilibria of such complete information games can be taken to be strong routines. Hence the same can be done conditional on either h or h'.

Let us now return to the assumption that there are equilibrium histories such that a_0 gets revealed. Consider a revelation stage h_t^2 such that no revelation has occurred before. We have that

$$Prob_1(a_1 \in A(w_t)|h_t^2) = Prob_1(a_0 \in A(w_t)|a_1 \notin A(w_t) \text{ and } h_t^2) = p$$

Hence at a revelation stage h_t^2, we have that with probability $(1-p)p$ action a_0 is the only productive action available. If we had $b_0 = b_1$ then clearly, it is optimal for a_0 to be revealed when it is the only action available, since inconclusive revelation will necessarily result in inefficient exit. This remains true for b_0 close enough to b_1 since the inefficiency of inconclusive revelation is bounded away from 0. Hence there exists $\epsilon > 0$ such that whenever $b_0 > b_1 - \epsilon$ action a_0 gets revealed and confirmed with probability at least $(1-p)pq$. At such histories, there is no further revelation. This concludes the proof. ■

D.3 Results of Appendix A

Proof of Lemma 3: Point (i) is a simple consequence of the fact that player 1 can guarantee herself value 0 by exiting permanently. Point (ii) is an incentive compatibility condition that must hold whenever player 2 takes a costly action. Point (iii) results from the incentive compatibility condition at a revelation stage. Whenever action a_i is available, player 2 will take it if and only if

$$\delta(qV_2^{E_{C\cup\{i\}}} + (1-q)(1-\mu_C)V_2^{R_C} - (1-\mu_C)V_2^{R_C}) \geq c \iff \delta q(V_2^{E_{C\cup\{i\}}} - (1-\mu_C)V_2^{R_C}) \geq c.$$

Whenever these incentive compatibility conditions are satisfied, the automaton is an equilibrium. ■

Proof of Lemma 4: Point (i) results from the fact if player 2 has deviated in the past, the players can never be in a state E_C. Hence increasing continuation values conditional on state E_C increases continuation values on the equilibrium path, while keeping continuation values of the equilibrium path identical. This improves efficiency while maintaining incentive compatibility conditions.

Point (ii) results from the fact that because of Assumption 1, there exists an automaton equilibrium in which player 1 stays in the first period. This however requires that some revelation occur, and hence $\phi_{E_\emptyset}^{Rev} > 0$. Inversely, Assumption 1 implies that players need only to confirm action a_1 to span the entire Pareto frontier of the complete information game. Hence since additional revelation is costly, no further revelation should occur once a_1 is confirmed and $\phi_{E_1}^{Rev} = 0$.

Let us now turn to point (iii). From Lemma 3 we know that at a revelation stage, incentive compatibility requires that $\delta q(V_2^{E_{0,1}} - (1-\mu_0)V_2^{R_0}) \geq c$. Furthermore upon revelation continuation payoffs to players 1 and 2 take the form $V_1 = pqV_1^{E_{0,1}} + (1-pq)(1-\mu_0)V_1^{R_0}$ and $V_2 = pqV_2^{E_{0,1}} + (1-pq)(1-\mu_0)V_2^{R_0}$. Given $V_1^{E_{0,1}}$ and $V_2^{E_{0,1}}$, values V_2 and V_1 are both maximized when the incentive compatibility constraint binds, that is $\delta q(V_2^{E_{0,1}} - (1-\mu_0)V_2^{R_0}) = c$. This pins the value of $(1-\mu_0)V_2^{R_0}$ and the value V_2. Value V_1 is maximized when the probability of exit μ_0 is minimized. Hence $V_2^{R_0} = \underline{V}_2^0$. ■

Proof of Proposition : Let us first show that when $\mathcal{W}_0$ dominates $\mathcal{W}_{Rev}$, then $p_0^{Rev} = 0$. Indeed, assume that $\mathcal{W}_0$ dominates $\mathcal{W}_{Rev}$, and that $p_0^{Rev} > 0$. Using equations (5) and (6), we obtain that

$$V_1^{E_0} = \gamma V_1^0 + (1-\gamma)V_1^{Rev} \tag{7}$$

$$V_2^{E_0} = \gamma V_2^0 + (1-\gamma)V_2^{Rev}. \tag{8}$$

where $(V_1^0, V_2^0) \in \mathcal{W}_0$ and $(V_1^{Rev}, V_2^{Rev}) \in \mathcal{W}_{Rev}$. Since $\mathcal{W}_0$ dominate $\mathcal{W}_{Rev}$, there exist

$(\widehat{V}_1^0, \widehat{V}_2^0)$ that dominate (V_1^{Rev}, V_2^{Rev}). Replacing (V_1^{Rev}, V_2^{Rev}) in (7) and (8) by $(\widehat{V}_1^0, \widehat{V}_2^0)$ yields values $(\widehat{V}_1^{E_0}, \widehat{V}_2^{E_0})$ that dominate $(V_1^{E_0}, V_2^{E_0})$. Furthermore since $V_2^{E_0} \geq c/\delta$, then $\hat{V}_2^{E_0} \geq c/\delta$. By Assumption 1 this means that values $(\widehat{V}_1^{E_0}, \widehat{V}_2^{E_0})$ are implementable in the complete information game where only action a_0 is productive. Hence the original automaton is dominated by an automaton in which there is no revelation.

Inversely, let us now assume that $\mathcal{W}_0$ does not dominate $\mathcal{W}_{Rev}$. Pick some pair of values $\mathbf{V}^{Rev} = (V_1^{Rev}, V_2^{Rev})$ that isn't dominated by $\mathcal{W}_0$. Any weighted average $(V_1^{E_0}, V_2^{E_0})$ between $\mathbf{V}^{Rev}$ and points of $\mathcal{W}_0$ such that $V_1^{E_0} \geq 0$ and $V_2^{E_0} \geq c/\delta q$ is implementable as an automaton involving revelation. Further more $\mathcal{W}_0$ is a line segment. Hence for any pair of values $(V_1^0, V_2^0) \in \mathcal{W}_0 \cap \{V_1 \geq 0 \text{ and } V_2 \geq c/\delta q\}$ there exists a weighted average of a value in $\mathcal{W}_0$ and $\mathbf{V}^{Rev}$ that dominates it and is implementable. Hence, whenever $\mathcal{W}_0$ does not dominate $\mathcal{W}_{Rev}$ efficient automatons will exhibit some revelation. ■

Proof of Proposition 8: From equations (5) and (6) we know that values $(V_1^{E_0}, V_2^{E_0})$ are necessarily weighted averages of points in $\mathcal{W}_0$ and $\mathcal{W}_{Rev}$. A point $(V_1^{E_0}, V_2^{E_0})$ in the convex hull of $\mathcal{W}_0 \cup \mathcal{W}_{Rev}$ is implementable using an automaton if and only if $V_1^{E_0} \geq 0$ and $V_2^{E_0} \geq c/\delta q$. ■

Proof of Lemma 5: Sets $\mathcal{W}_0$ and $\mathcal{W}_{Rev}$ are line segments. For generic values of b_0, the two aren't segments of the same line. In that case any point $(V_1^{E_0}, V_2^{E_0})$ on the frontier of $\mathsf{co\ hull}\,(\mathcal{W}_0 \cup \mathcal{W}_{Rev}) \cap \{(V_1, V_2) | (V_1, V_2) \geq (0, c/\delta q)\}$ either belongs to $\mathcal{W}_0$ or $\mathcal{W}_{Rev}$ or is uniquely implemented as a weighted average between extreme points of $\mathcal{W}_0$ and $\mathcal{W}_{Rev}$. In all cases there is a unique automaton that implements $(V_1^{E_0}, V_2^{E_0})$. ■

References

[1] Abreu, Dilip, David Pearce and Ennio Stacchetti, *Toward a Theory of Discounted Repeated Games with Imperfect Monitoring*, Econometrica, Vol. 58, No. 5, pp. 1041-1063, 1990.

[2] Al-Najjar, Nabil, Luca Anderlini and Leonardo Felli, *Undescribable Events*, Review of Economic Studies, Vol. 73, No. 4, pp. 849-868, 2006.

[3] Athey, Susan and Kyle Bagwell, *Optimal Collusion with Private Information*, RAND Journal of Economics, Vol. 32, No. 3, pp. 428-465, 2001.

[4] Athey, Susan, Kyle Bagwell and Chris Sanchirico, *Collusion and Price Rigidity*, Vol. 71, No. 2, pp. 317-349, Review of Economic Studies 71 (2), 2004.

[5] Baker, George, Robert Gibbons, and Kevin Murphy, *Subjective Performance Measures in Optimal Incentive Contracts*, Quarterly Journal of Economics, Vol. 109, No. 4, pp. 1125-1156, 1994.

[6] Baker, George, Robert Gibbons, and Kevin Murphy, *Relational Contracts and the Theory of the Firm*, Quarterly Journal of Economics, Vol. 117, No. 1, pp. 39-84, 2002.

[7] Bull, Clive, *The Existence of Self-Enforcing Implicit Contracts*, Quarterly Journal of Economics, Vol. 102, No. 1, pp. 147-160, 1987.

[8] Crawford, Vincent and Hans Haller, *Learning How to Cooperate: Optimal Play in Repeated Coordination Games*, Econometrica, Vol. 58, No. 3, pp. 571-595, 1990.

[9] Cripps, Martin and Jonathan Thomas, *Some Asymptotic Results in Discounted Repeated Games of One-Sided Incomplete Information*, Mathematics of Operations Research, vol. 28, pp. 433-462, 2003.

[10] Fudenberg, Drew, David Levine, and Eric Maskin, *The Folk Theorem with Imperfect Public Monitoring*, Econometrica, Vol. 62, No. 5, pp. 997-1039, 1994.

[11] Gibbons, Robert, Rebecca Henderson, Nelson Repenning and John Sterman, *Organizational Capabilities? Suggestive Evidence and (Some) Potential Theories*, to appear in Handbook of Organizational Economics (R. Gibbons and J. Roberts, eds.), Princeton University Press.

[12] Green, Edward and Robert Porter, *Noncooperative Collusion under Imperfect Price Information*, Econometrica, Vol. 52, No. 1., pp. 87-100, 1984.

[13] Hart, Oliver, and John Moore, *Foundations of Incomplete Contracts*, Review of Economic Studies, Vol. 66, No. 1, pp. 115-138, 1999.

[14] Levin, Jonathan, *Relational Incentive Contracts*, American Economic Review, Vol. 93, No. 3, pp. 835-847, 2003.

[15] Macleod, Bentley and James Malcomson, *Reputation and Hierarchy in Dynamic Models of Employment*, Journal of Political Economy, Vol. 96, pp. 832-54, 1988.

[16] Maskin, Eric and Jean Tirole, *Unforeseen Contingencies and Incomplete Contracts*, Review of Economic Studies, Vol. 66, No. 1,pp. 83-114, 1999.

[17] Nelson, Richard, and Sidney Winter, *An Evolutionary Theory of Economic Change*, Harvard University Press, 1982.

[18] Peski, Marcin, *Repeated Games with Incomplete Information on One Side and Discounting*, Theoretical Economics, forthcoming.

[19] Segal, Ilya, *Complexity and Renegotiation: a Foundation for Incomplete Contracts*, Review of Economic Studies, Vol. 66 , No. 1, pp. 57-82, 1999.

[20] Shalev, Jonathan, *Nonzero-Sum Two-Person Repeated Games with Incomplete Information and Known own Payoffs*, Games and Economic Behavior, Vol. 7, pp. 246-259, 1994.

[21] Sorin, Sylvain, *Merging, Reputation, and Repeated Games with Incomplete Information*, Games and Economic Behavior, Vol. 29, pp. 274-308, 1999.

[22] Watson, Joel, *Starting Small and Renegotiation*, Journal of Economic Theory, Vol. 85, No. 1, pp 52-90, 1999.

[23] Watson, Joel, *Starting Small and Commitment*, Games and Economic Behavior. Vol. 38, pp. 176-199, 2002.

[24] Winter, Sidney, *Satisficing, Selection, and the Innovating Remnant*, Quarterly Journal of Economics, Vol. 85, No. 2, pp. 237-261, 1971.