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

Sylvain Chassang*, †

Princeton University

February 6, 2009

Abstract

This paper studies how agents with conflicting interests learn to cooperate when the details of cooperation are not common knowledge. It considers a repeated game in which one player has incomplete information about when and how her partner can provide benefits. Initially, monitoring is imperfect and cooperation requires inefficient punishment. As the players' common history grows, the uninformed player can learn to monitor her partner's actions, which allows players to establish more efficient cooperative routines. Because revealing information is costly it may be optimal not to reveal all the existing information, and efficient equilibria can be path-dependent.

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

*I am particularly indebted to Robert Gibbons for inspiration and advice. The paper benefited from detailed feedback by Ricardo Alonso, Alexander Frankel and several anonymous referees. I'm grateful to Roland Benabou, Esteban Rossi-Hansberg, Heikki Rantakari, Andy Skrzypacz and Satoru Takahashi for comments and advice. I thank Abhijit Banerjee, Glenn Ellison, Richard Holden, Bengt Holmström, Kevin J. Murphy, Lars Stole, Muhamet Yildiz as well as seminar participants at the Chicago GSB, Harvard, MIT, USC, LSE and the 2008 Utah Winter Business Economics Conference for many helpful discussions. Andrew Robinson provided excellent research assistance. 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

This paper studies how agents with conflicting interests go about developing a successful cooperative relationship when the details of cooperation are not common knowledge. In circumstances where formal contracts are unavailable or inconvenient, agents frequently rely on informal cooperative agreements – or relational contracts – to resolve incentive conflicts. In fact, whenever the interested parties have better information than outside courts, relational contracts can outperform formal contracts by allowing agents to adapt their joint behavior to contingencies that cannot be assessed by third parties.[1] The starting point of this paper is to recognize that even when agents have better information than outside courts, this does not mean that the details of how and when cooperation should occur are common knowledge between the parties. For instance, a new plant manager may not have a complete understanding of what her production teams can and cannot do; a firm's CEO may have only partial 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 developed in the paper considers two players engaged in an infinitely repeated game with asymmetric information. In each period, player 1 decides to stay (and interact with player 2) or exit (and skip interaction for one period). Staying is costly for player 1 and yields a benefit for player 2. If player 1 stays, player 2 can reciprocate by taking some action from a set of available actions, randomly drawn in each period from a countable superset. Player 2 can take actions of two types: *unproductive* actions that are costless to player 2 but yield zero benefit to player 1, and *productive* actions that yield benefit to player 1 with positive probability (but will also fail with some probability). Whether an action is productive or not is fixed throughout the game. As regards information, the set of available actions, the action taken by player 2, and the benefit player 1 obtains, are all publicly observed at the end of each period. The source of asymmetric information is that player 2 knows which actions are productive while player 1 does not. In this setting, player

[1] See for instance Bull (1987) or Baker, Gibbons and Murphy (1994, 2002).

1 has access to substantial information about player 2 (which actions player 2 can take and which action she does take) but lacks the ability to interpret what that information implies for payoffs (whether there are productive actions available today, and did player 2 take one). Distinguishing the availability of information and the ability to interpret it introduces the possibility of learning in an imperfect monitoring context.

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 when 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 no benefit to player 1, player 1 does not know if player 2 took a costly productive action that 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 exit on the equilibrium path to induce cooperation. However, once action a^0 yields positive benefit to player 1, 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 in subsequent periods, player 1 can monitor whether action a^0 was available and whether it was taken, so that there is no moral hazard and no punishment is required. However, if player 2 takes a new action a^1 that player 1 has not 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 two main predictions. The first is that initially, while learning occurs and the parties are specifying the contingencies

[2]In this model, equilibrium behavior is called a routine if learning has stopped and players use a fixed subset of productive 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.

of their informal understanding, relationships will be sensitive to adverse shocks. More specifically, while learning occurs, the failure of an action taken by player 2 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 failure of a productive action taken by player 2 – will no longer trigger inefficient breakdowns.

The second prediction is that it is not necessarily optimal to reveal all existing information: the costs of inducing further revelation can dominate the potential gains of using a more efficient routine. Because the information revealed depends on what actions have been available, 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 complete information would be. In this sense the model is not about finding out whether cooperation is sustainable, but rather about figuring out the details of its implementation. A corollary is that established routines may survive, 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. (2008) document.

The paper contributes to the organizational economics literature on relational contracts. Since Bull (1987), economists have analyzed such relational contracts 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; Alonso and Matouschek (2007) explore the question of relational delegation between an uninformed principal and an informed agent. These repeated-game models are concerned with the difficulties of *maintaining* cooperation. In contrast, this paper focuses on the particular hurdles involved in *building* a successful informal understanding.

The paper draws from the literature on imperfect public monitoring initiated by Green and Porter (1984), and developed in Abreu, Pearce and Stacchetti (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 are private. The contribution of the current paper is to introduce the possibility of learning how to monitor, 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 Gossner and Vieille (2003), Cripps and Thomas (2003), Wiseman (2005), or Pęski (2008). Those papers study learning in a more general class of games, but focus 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.

The 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) is on how players determine whether cooperation is sustainable or not. In contrast, the current paper takes the feasibility of cooperation as given and focuses on how players figure out the details of implementing cooperation. Finally, the paper is also related to the work of Crawford and Haller (1990) and Blume and Franco (2007), who consider the problem of how players learn to coordinate when they do not have a common language to describe actions.[3] Rather than learning to *coordinate*, the current paper focuses on the problem of learning to *cooperate* when players have incentive conflicts, and highlights the specific forces that hinder learning in such situations.

The paper is organized as follows: Section 2 defines and discusses the framework; Section

[3]See Ellison and Holden (2008) for a related paper studying the impact of limited communication on team behavior.

3 investigates general properties of the game and highlights the efficiency costs of information revelation; Section 4 explores qualitative properties of optimal learning; Section 5 concludes. Appendix A deals with technical measurability requirements. Proofs are contained in Appendix B.

2 The Setup

This section defines the framework of the paper: Section 2.1 describes players' payoffs and the timing of the game; Section 2.2 describes the information structure; Section 2.3 describes histories and strategies; Section 2.4 provides examples of situations where the model's main assumptions are reasonable.

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 $\mathcal{A} = \mathbb{N}$. Each period t, a random i.i.d. subset of actions $A_t \subset \mathcal{A}$ is available to player 2. This set A_t will be referred to as the state of the world. Actions are available with probability p, and independently of each other, so that A_t is countably infinite with probability one.[4] 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 incurs a cost $k > 0$ while player 2 gets a benefit $\pi > 0$.

If player 1 stays, an i.i.d. state of the world A_t is drawn and observed by *both* players.

Stage 2: if player 1 has stayed, player 2 then gets to choose an action $a_t \in A_t$. For any $a \in \mathcal{A}$, taking action a has a deterministic cost $c(a)$ for player 2 and generates a random

[4] More formally, in any period t, $(\mathbf{1}_{a\in A_t})_{a\in\mathcal{A}}$ is an i.i.d. sequence of Bernoulli variables such that $prob(\mathbf{1}_{a\in A_t} = 1) = p$.

benefit $\tilde{b}(a) \in \{0, b(a)\}$ for player 1.

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

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

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

There exists a number N of productive actions $\{a^0, \ldots, a^{N-1}\} \subset \mathcal{A}$.[5] By extension $\mathcal{N}$ will be used to denote the set of productive actions. Productive actions are costly for player 2 but yield strictly positive expected benefits for player 1, while unproductive actions are costless for player 2 but yield no benefit for 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 $b(a) = 0$.[6]

The paper also allows each player $i \in \{1, 2\}$ to condition her actions on a public randomization device. To reduce the notational burden, the paper does not introduce explicit notation for public randomizations. Finally, note that utility is not transferable across players, and that inefficient behavior may be required to inflict punishment.[7]

2.2 Information structure

The paper considers an asymmetric information setting in which player 2 knows which actions are productive while player 1 is entirely uninformed.

Parameters p and q are common knowledge. Player 2 observes her own cost for actions $c(\cdot)$ and hence knows the set $\mathcal{N}$ of productive actions. Player 2 also observes both the state of the world A_t and the outcome $\tilde{b}(a_t)$.

[5]Throughout the paper a^k denotes the k^{th} productive action, while a_t denotes the action player 2 takes in period t.

[6]The assumption that actions which benefit player 1 all have the same cost c for player 2 simplifies the analysis, but is not essential to the argument. Productive actions with different costs, or actions that benefit both players could be introduced.

[7]Note that this is a departure from the relational contracts literature, which typically assumes transferable utility. Ruling out transferable utility implies that inefficient punishment is sometimes needed to provide incentives. This property may survive under transferable utility if, for instance, player 2's cost for productive actions is private information.

In every period t in which player 1 chooses to stay, player 1 observes the state A_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 holds an improper uniform prior over which actions are productive. This implies in particular that

$$\forall A \subset \mathcal{A},\ \forall a \in A, \quad Prob_1 \{a \in \mathcal{N} \mid \text{card } A \cap \mathcal{N} = n\} \;=\; \frac{n}{\text{card } A}.$$

Both players know there is a number $N \geq 1$ of productive actions. Given a vector of productive actions $(a^0, \cdots, a^{N-1})$, the productivity vector $(b(a^0), \cdots, b(a^{N-1}))$ is drawn from a distribution B over $(\underline{b}, \overline{b})^N$, where $\overline{b}$ and $\underline{b}$ are strictly positive bounds to returns. Conditional on a set $\mathcal{N}$ of productive actions, both players hold the same prior B over the vector of benefits $(b(a^0), \cdots, b(a^{N-1}))$.[8] Let us denote by Γ_{AI} this incomplete information game.

2.3 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 A_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 between three types of histories:

(i) the set $\mathcal{H}^1$ of histories h_t^1 of the form $h_t^1 = \{d_1, A_1, a_1, \tilde{b}(a_1), \ldots, d_{t-1}, A_{t-1}, a_{t-1}, \tilde{b}(a_{t-1})\}$, corresponding to player 1's information at her decision node in period t;

(ii) the set $\mathcal{H}^{2|1}$ of histories $h_t^{2|1}$ of the form $h_t^{2|1} = h_t^1 \sqcup \{d_t, A_t\}$, corresponding to player 1's information at player 2's decision node in period t;

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

A pure strategy for player 1 is a mapping $s_1 : \mathcal{H}^1 \to \{S, E\}$. A pure strategy of player 2 is a mapping $s_2 : \mathcal{H}^2 \to \mathcal{A} = \mathbb{N}$, such that for all histories $h_t^2 \in \mathcal{H}^2$, $s_2(h_t^2) \in A_t$.

[8] In particular, player 2 knows which actions are productive (because they are costly), but has no additional information about the specific benefits generated by productive actions.

The solution concept used in the paper is perfect Bayesian equilibrium in pure strategies.[9] Because player 1 holds an improper prior, insuring that players' strategies induce well-defined beliefs over future utility requires some care. Throughout the paper, the players' behavior is required to be invariant at histories that are identical up to a relabeling of actions. This is essentially a measurability requirement. See Appendix A for a formal discussion.

The paper focuses on Pareto efficient perfect Bayesian equilibria. Pareto efficient equilibria form a focal class of equilibria and a natural benchmark in the analysis of Γ_{AI}. Furthermore, anticipating results to come, stacking the odds in favor of efficiency gives some reassurance that if long run inefficiencies occur, they are not an artifact of excessive restrictions on strategies.[10]

To ensure the existence of equilibria in which cooperation occurs and player 1 stays at least in the first period, the following assumption is maintained throughout the paper .

Assumption 1 *Parameters δ, p, q, k, c, π and $\underline{b}$ are such that*

$$\frac{k}{pq\underline{b}} < 1 \quad \textit{and} \quad q\frac{\delta}{1-\delta}\left(\pi - \frac{k}{q\underline{b}}c\right) > c.$$

Section 3 describes such a cooperative equilibrium.

2.4 Interpreting the model

As was highlighted in the introduction, a key feature of the model is that it allows for learning in an imperfect public monitoring context. Although player 1 has access to a lot of information (which actions are available and what action is taken), she cannot initially distinguish productive actions from unproductive ones. Initially, when an action fails to deliver benefits, player 1 does not know whether player 2 took an unproductive action or a productive action which failed, and punishment may be required on the equilibrium path.

[9]Note that players can condition their behavior on public randomizations. The existence of pure strategy perfect Bayesian equilibria is straightforward. Consider for instance the equilibrium in which player 1 exits every period and player 2 takes only unproductive actions.

[10]For instance, Fudenberg, Levin and Maskin (1994) show that focusing on trigger strategies severely underestimates the feasibility of cooperation in games with imperfect public monitoring.

Once an action yields positive benefits, however, player 1 learns that it is productive and can recognize it in future periods. As player 1 identifies the productivity of more actions, she can monitor more efficiently the behavior of player 2, and as long as player 2 takes actions of commonly known productivity, there is no need for inefficient punishment on the equilibrium path. In contrast, learning new actions will entail efficiency costs, and it may become optimal for players to stop revealing information. While the model makes a number of simplifying assumptions, the model's main ingredients (that players have conflicting interests, and that players have only partial understanding of *when* cooperation should take place and *what* form it should take) are likely to be present in many of the settings with which the relational contracts literature is concerned.

Imagine for instance the problem of a new manager (player 1) who must figure out how her production team (player 2) should function. Actions of player 2 correspond to various ways to organize and run production – for instance using different technologies that the team may be more or less familiar with, or assigning tasks to team members in different ways. In such a setting, the production team is likely to have private information about what are efficient ways to organize production. Moreover, the production team's preferences about how to organize production need not be aligned with those of the manager. Finally, the feasibility of actions may depend on circumstances, i.e. on what technologies and supplies are available, on the specific production task at hand, what deadlines need to be met, or on which team members are available.

Similar difficulties may arise in many of the internal governance problems that firms encounter. For instance, the model could describe the problem of an executive trying to manage a research and development team. The executive's problem is that she does not know what research environment (e.g. should there be deadlines, monetary incentives...) and what research topics are likely to yield marketable products, rather than merely suit the private desires of the research team. Similarly, the setup could describe the problem of a product manager who does not initially know the inputs of a successful marketing campaign but suspects that her marketing team overvalues creativity over efficiency.

The trade-offs identified in the paper may also be relevant to understand the process by which cooperation develops across organizations. Consider for instance the problem of a central authority, or an international organization, providing resources to a local government, and expecting in return that the government will implement sound policies. In all likelihood, the central authority will have less information than the local government about which policies are likely to yield good results, and which policies are simply in the private interest of the government officials. Furthermore, the feasibility of policies may depend on local economic and social conditions, on which political party is in power, or on the current resources of the central authority. In this environment, the central authority may require the implementation of standardized programs, even though it is common knowledge that they are suboptimal, and that the local government knows how to improve them.

In each of these settings, the players must figure out the specific details of how to implement cooperation. Sections 3 and 4 analyze how this learning process unfolds, and why incentive conflicts can lead players to stop learning inefficiently early.

3 Equilibrium patterns of punishment

This section and the next analyze the joint dynamics of learning, cooperation, and punishment. Section 3.1 provides a rapid analysis of a complete information version the game. Section 3.2 then establishes general properties of the incomplete information game Γ_{AI}.

3.1 The case of complete information

As a benchmark, let us consider the full information game Γ_{FI} in which both players know what actions are productive and what their productivity is. The main result is straightforward and simply highlights that there should be no exit on the equilibrium path of a Pareto efficient equilibrium.

Proposition 1 (no exit) *Consider the full information game Γ_{FI}. Whenever (s_1, s_2) is a Pareto efficient equilibrium of Γ_{FI}, 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. In addition to no exit occurring on the equilibrium path, it can be shown that Pareto efficiency essentially requires player 2 to take a productive action with positive probability only if she takes all more productive actions with probability 1 whenever they are available.

One can derive simple but useful bounds for Pareto efficient values. Given that the productivity of actions is bounded above by $\bar{b}$, for any equilibrium pair of values (V_1, V_2), there exists $r \in [0,1]$ such that

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

Furthermore, since player 2 has the option to never play a cooperative action, player 2 must get value greater than π in any equilibrium where player 1 stays in the first period. In addition, since player 1 has the option to exit, she must get value greater than 0. This yields that r must satisfy $r \in [\underline{r}, \overline{r}]$ where $\underline{r} \equiv k/pq\bar{b}$ and $\overline{r} \equiv \delta\pi/pc$. Hence the highest value that player 2 can expect is $\overline{V}_2 \equiv \frac{1}{1-\delta}(\pi - p\underline{r}c)$.

3.2 The case of asymmetric information

Under asymmetric information, player 1 must learn which actions are productive. Initially, she does not know what behavior to expect from player 2, and monitoring is imperfect. As the players' common history grows, player 1 can learn to interpret player 2's actions, and the game transitions to perfect monitoring. This section investigates patterns of punishment that are required to incentivize information revelation. It begins with a few definitions.

Definition 1 (revelation and confirmation stages) *Consider an equilibrium (s_1, s_2). A history $h_t^{2|1} \in \mathcal{H}^{2|1}$ is called a* ***revelation stage*** *if there is non-zero probability that a productive action which has not been taken yet will be taken. The action a that player 2 does take is called the revealed action.*[11]

[11] According to this definition, unproductive actions played at a revelation stage are revealed. Note that actions can be played without being revealed, provided they are not played at a revelation stage.

*A history $h^1_{t+1} \in \mathcal{H}^1$ 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.*

History $h^{2|1}_t$ corresponds to player 1's information at player 2's decision node in period t. It is a revelation stage if player 1 believes that player 2 will choose a new productive action with positive probability. A confirmation stage for some action a is a history h^1_{t+1} such that action a has been revealed in the past and yields benefits for the first time. At a confirmation stage there is no more doubt about whether player 2 took a productive action or not. Routines are defined as follows.

Definition 2 (routines) *Consider a pair of strategies (s_1, s_2), and a history h^1_{t+1}. We say that players follow a routine starting from history h^1_{t+1} if player 2 takes only confirmed or unproductive actions in the continuation game.*

In words, strategies become routines when players only use actions whose payoff consequences are common knowledge. The routines considered in this paper differ from those described in Winter (1971) or Nelson and Winter (1982). Here, being a routine is a property of equilibrium strategies rather than an exogenous constraint placed on agents' behavior. This allows one to identify what economic forces may shape the establishment of routines.

Before studying information revelation in Γ_{AI}, it is important to note that there exist equilibria of Γ_{AI} such that player 1 stays in the first period and revelation stages occur in equilibrium. This is a consequence of Assumption 1. Consider the pair of strategies (s^0_1, s^0_2) defined as follows: in the first period, player 1 stays and player 2 takes a productive action whenever one is available. If a productive action a is confirmed in period 1, continuation strategies prescribe that on the equilibrium path, player 1 stays every period and player 2 takes action a with probability $k/pq\underline{b}$ whenever it is available (using public randomizations). Off of the equilibrium path, player 1 always exits and player 2 takes unproductive actions. In the case where no productive action is confirmed in period 1, player 1 exits permanently. Let us show that under Assumption 1, (s^0_1, s^0_2) is an equilibrium. Indeed, if player 2 takes an unproductive action in period 1, her expected utility is 0. If instead a productive action is available and player 2 takes it, her expected utility is $-c + q\frac{\delta}{1-\delta}\left(\pi - \frac{k}{q\underline{b}}c\right)$, which is greater than 0 by Assumption 1. Furthermore, player 1's expected payoff when staying in period 1 is at least $-k + pq\underline{b}$, which is greater than 0 by Assumption 1. Hence, incentive compatibility holds in the first period. An identical argument shows that incentive compatibility holds in later periods as well, and (s_1^0, s_2^0) is indeed an equilibrium.

The next proposition provides sufficient conditions for inefficient punishment to happen with positive probability following the (unconfirmed) revelation of a new productive action.

Proposition 2 (costly revelation) *Consider a history $h_t^{2|1}$ where $N' < N$ actions have been revealed and denote by $\mathcal{N}'$ the corresponding set of revealed actions. Define $\underline{V}_2^{N'} \equiv \frac{1}{1-\delta(1-p)^{N'}}\pi$ and recall that $\overline{V}_2 = \frac{1}{1-\delta}(\pi - p\underline{r}c)$. Whenever*

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

then, if history $h_t^{2|1}$ is a revelation stage (for some action $a \in \mathcal{N} \setminus \mathcal{N}'$), exit must occur with strictly positive probability on the continuation path.

Value $\underline{V}_2^{N'}$ is a lower bound for the value player 2 can guarantee herself when player 1 uses strategies that do not involve 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 costly revealed action is available and player 2 does not take it, which happens with probability less than $1 - (1-p)^{N'}$. When $N' = 0$, condition (1) is necessarily satisfied, and Proposition 2 implies that initially, inefficient exit will always happen with strictly positive probability.

Proposition 2 provides conditions under which exit will happen with strictly positive probability following the revelation of a new productive action. Proposition 3 now shows that such inefficient exit need only happen following unconfirmed revelation stages.

Proposition 3 (no unnecessary exit) *In any Pareto efficient equilibrium (s_1, s_2), player 1 stays at any equilibrium history h_t^1 such that all revealed actions are confirmed.*

Note that for any equilibrium (s_1, s_2), histories h_t^1 where all revealed actions are confirmed occur with strictly positive probability since any revelation stage can lead to immediate confirmation. However, if there are unconfirmed revelation stages, exit may be required even when player 2 takes a confirmed action. Indeed, inefficient punishment may be needed to ensure that player 2 has appropriate ex ante incentives for revelation. However, it should be noted that even if some revealed actions are unconfirmed and exit may be required, strategies need not depend on the success or failure of confirmed actions. This is because the success or failure of confirmed actions does not reveal information about past play. Hence, conditioning strategies on the outcome of confirmed actions can be replaced by conditioning strategies on public randomizations.

Propositions 2 and 3 delineate the joint mechanics of cooperation and learning. While learning is occurring and new actions are being revealed, the partnership will be sensitive to adverse shocks. In particular, revelation stages at which the revealed action fails may be followed by inefficient exit. Once learning stops and the relationship routinizes, the relationship will become resilient to such shocks. First, whenever player 2 takes a confirmed action, continuation strategies need not depend on whether that action fails or not. Second, no exit should occur at a history where all revealed actions are confirmed, even if the actions taken by player 2 fail to deliver benefits. In the initial stages of the game, players interact in an imperfect public monitoring environment à la Green and Porter (1984) and some inefficient punishment is required following adverse shocks (i.e. the failure of a productive action). As the players' common history grows and the consequences of actions become common knowledge, the game becomes one of perfect monitoring and punishment becomes unnecessary in equilibrium. A corollary of this is that player 2 is more likely to be punished if she fails after using a new action whose consequences are not common knowledge, than if she fails after using a standardized protocol whose consequences are commonly known.

4 Optimal learning

Proposition 2 established conditions under which the revelation of new productive actions will come at an efficiency cost. However, learning new productive actions may 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 unambiguous 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 and confirmed. This section investigates qualitative aspects of optimal revelation.

The main result, given in Section 4.1, is a possibility result. It considers a setting with two productive actions and shows that it can be efficient to stop the revelation of actions, even though it is common knowledge that a more productive action exists and player 2 knows what action should be taken. This causes efficient partnerships to be path dependent. Section 4.2 relates the model to bandit problems and discusses how path dependence extends when there is a large but finite set of possible actions.

4.1 Properties of optimal revelation with two productive actions

For the purpose of establishing a possibility result, this section focuses on a setting where it is common knowledge that there are two productive actions, a^0 and a^1, that respectively yield benefits b^0 and b^1, with $b^1 > b^0$. Let us define

$$\overline{r} \equiv \frac{\delta\pi}{pc}\ ;\ \ \underline{V}_2^D \equiv \frac{1}{1-\delta(1-p)}\pi \ \text{ and } \ \forall l \in \{0,1\},\ \underline{r}_l \equiv \frac{k}{pqb^l}\ ;\ \ \overline{V}_2^l \equiv \frac{1}{1-\delta}(\pi - p\underline{r}_l c).$$

Value $\overline{V}_2^0$ (resp. $\overline{V}_2^1$) is the maximum value that player 2 can obtain while keeping player 1 indifferent between staying and exiting in the complete information game where a^0 (resp. a^1) is the only productive action. In addition to Assumption 1, the following assumption is maintained throughout Section 4.1.

Assumption 2 *Parameters b^0, π, k, c, p, q and $\delta < q$ are fixed and such that*

$$(i)\ \overline{r} < 1\ ;\ \ (ii)\ \underline{r}_0 > 1-p\ ;\ \ (iii)\ \delta(\overline{V}_2^0 - \underline{V}_2^D) < c\ ;\ \ (iv)\ q\frac{1}{1-\delta}\pi > \overline{V}_2^0.$$

Point (i) ensures 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, a^1 is the only productive action used in efficient equilibria. Point (ii) puts a limit on how large b^0 can be. It implies that if player 2 provides benefits to player 1 using only action a^0, then player 1 can be induced to stay only if player 2 takes action a^0 at a rate greater than $1-p$ when it is available. Point (iii) implies that when b^1 is close to b^0, so that $\overline{V}_2^1$ approaches $\overline{V}_2^0$, the revelation of a second productive action requires exit on the equilibrium path. Point (iv) implies that when b^1 becomes large, player 2 prefers obtaining $\overline{V}_2^1$ with probability q rather than obtaining $\overline{V}_2^0$ for sure. Lemma B.1 given in Appendix B ensures that the region of the parameter space defined by Assumptions 1 and 2 is not empty.[12]

Let us consider the situation in which player 1 has revealed and confirmed action a^0 and there are no unconfirmed revealed actions. This happens with strictly positive probability since a^0 can be both revealed and confirmed in the first revelation stage. The question is whether it may be optimal that no further information be revealed in the continuation game. By point (iii) of Assumption 2, we know that for b^1 close enough to b^0, it must be that $c > \delta(\overline{V}_2^1 - \underline{V}_2^D)$. By Proposition 2, this implies that inducing revelation of action a^1 will require inefficient exit. The next lemma provides a slightly stronger result.

Lemma 1 *There exists $\mu > 0$, $K \in \mathbb{N}$ and $\eta > 0$ such that for any value $b^1 \in [b^0, b^0+\mu]$, and any equilibrium of Γ_{AI}, if $h_t^{2|1}$ is a revelation stage for action a^1, then there is probability greater than η that player 1 exits at least once in the next K periods.*

A corollary of this is that when b^1 is close enough to b^0, the cost of revealing new actions is bounded away from 0. Proposition 4 leverages Lemma 1 to establish that the efficiency of

[12]It is shown that the cost c of productive actions can be large enough that player 2 will not cooperate with probability one every time action a^1 (resp. a^0) is available, but low enough that revelation of productive actions is incentive compatible.

further revelation will depend on the productivity differential between actions a^0 and a^1.

Proposition 4 (optimal learning) *Consider parameters b^0, π, k, c, p, q and δ such that Assumptions 1 and 2 are satisfied. There exist $\overline{\Delta} > \underline{\Delta} > 0$ such that for any Pareto efficient equilibrium (s_1, s_2):*

- *(i) if $b^1 - b^0 > \overline{\Delta}$, there exist $\nu > 0$ and $\tau \in \mathbb{N}$ such that at any equilibrium history h_t^1 where player 1 stays, there is probability greater than ν that action a^1 will be confirmed in the next τ periods;*

- *(ii) if $b^1 - b^0 < \underline{\Delta}$, whenever a single productive action has been revealed, and it is also confirmed, then there is no further revelation on the equilibrium path.*

Point (i) highlights that when the difference between b^1 and b^0 is large, it is efficient to keep trying to reveal action a^1. In the long run, all efficient partnerships that are not in a state of permanent exit will have confirmed action a^1.

Inversely, point (ii) shows that when the difference between b^1 and b^0 is small, 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 under Assumption 2, when the productivity difference $b^1 - b^0$ shrinks, the value of revealing information goes to 0, while the costs of revealing information do not. Note that this happens even though it is common knowledge that action a^1 exists, and there is common knowledge that player 2 knows which action it is.

A natural implication of point (ii) is that Pareto efficient equilibria may exhibit path dependence, with different partnerships ending up using different long run routines with different degrees of efficiency.

Corollary 1 (path dependence) *Pick b^1 such that $b^1 - b^0 < \underline{\Delta}$. For any Pareto efficient equilibrium (s_1, s_2), there are histories h_t^1 and $\hat{h}_t^1$, occurring with strictly positive probability, such that starting from h_t^1, (s_1, s_2) becomes a routine in which player 2 only takes productive action a^0, while starting from $\hat{h}_t^1$, (s_1, s_2) becomes a routine in which player 2 only takes productive action a^1.*

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 pieces of information are revealed, partnerships that were identical ex ante, and faced the same complete-information Pareto frontier, can end up using routines involving different productive actions and different levels of efficiency.

4.2 Discussion

Proposition 4 shows that the incentive costs of information revelation can lead partnerships to follow routines known to be suboptimal. This result relies in part on the assumption that there are infinitely many actions and that player 1 holds an improper uniform prior about which ones are optimal. With finitely many actions, it would be possible for player 1 to learn all productive actions in finite time without using costly punishment on the equilibrium path. This can be achieved by using the following "enumerative" scheme: in every period where no confirmed productive action is available, player 2 must take the unconfirmed available action that has been taken with the least frequency in the past.[13] With finitely many actions, because of the random availability of actions, player 2 will repeatedly take every possible action (i.e. enumerate the action space) and in finite time, player 1 will learn all productive actions. Because going through the action space takes time, as the number of actions grows arbitrarily large, the expected time required to learn productive actions using such an enumerative scheme will go to infinity. In this sense, Proposition 4 can be thought of as describing the limit case where the number of possible actions grows large and player 1's prior over productive actions is arbitrarily diffuse. When the set of possible actions is large but finite, productivity differentials between ex ante identical partnerships will be persistent rather than permanent.

The reason why the enumerative scheme described above reveals information without using inefficient punishment on the equilibrium path is that it does not rely on the private

[13] If there are ties, player 2 may either be allowed to choose freely between actions, or the action could be designated randomly if the players have access to public randomizations over actions. The latter scheme would prevent player 2 from delaying taking costly productive actions.

information of player 2. Player 1 requires player 2 to take specific actions independent of whether they are productive or not. In contrast, if player 1 relies on player 2 to choose which action should be taken, inefficient punishment is necessary to provide appropriate incentives. In essence, when the number of actions is finite, there are two ways for player 1 to learn which actions are productive: authoritarian learning which requires no punishment in equilibrium but is slow because relatively blind; and delegated learning which is fast because player 2 uses her private information, but is subject to moral hazard and requires inefficient punishment.[14] If there are infinitely many actions and player 1 holds a uniform improper prior, authoritarian learning becomes ineffective, and player 1 must use delegated learning (and inefficient punishment) to reveal productive actions in finite time.

The mechanism that leads players to stop learning in the model presented in the paper is closely related to the mechanism that leads a single decision maker to stop learning in a bandit problem. Here however, the cost of experimentation is endogenous and corresponds to the cost of resolving incentive conflicts between the informed and the uninformed player. This differs slightly from the usual trade-off between exploration and exploitation (see Rothschild (1974) or March (1991)) as the cost of experimentation is not the opportunity cost of forgoing current profits, but rather the anticipated cost of inducing information revelation. In particular, when incentive conflicts disappear (either because cost c goes to 0, or because the players obtain identical payoffs, as in team problems), the productivity of every productive action will be revealed in finite time.[15] To the extent that players' incentives are decision variables at the organization level, this suggests that organizational design may affect significantly the process of routinization, as well as the long run efficiency of cooperative agreements.

[14] I thank one of the referees for highlighting this point.

[15] Indeed, whenever no confirmed action is available, there is no opportunity cost of experimentation.

5 Conclusion

The model presented in this paper attempts to capture the process by which agents specify the details of a relational contract. The essence of the modeling approach is to introduce the possibility of learning how to monitor in an imperfect public monitoring context. At the onset of the relationship, although player 1 has access to a lot of information, she lacks the ability to interpret it and does not know how and when cooperation should take place. This creates moral hazard and leads to inefficient punishment on the equilibrium path. As the players' joint history grows and information is revealed in equilibrium, player 1 learns how and when to expect cooperation, and the players are able to sustain cooperation without resorting to inefficient punishment.

The model makes two main 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 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 of players 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. Because figuring out the details of a cooperative agreement requires the costly revelation of information, players may follow a routine even when there is substantial evidence that this routine is suboptimal.

A specific implication of the model developed in this paper is that the extent of routinization will be related to both the magnitude of incentive conflicts between players, and the efficiency cost of aligning incentives. This raises natural questions. For instance, if formal contracts determine players' stage game payoffs, can an appropriate design of formal incentives improve the revelation of useful information? Can a temporary relaxation of budget constraints favor the establishment of efficient cooperative agreements between players?

The main difficulty in answering such questions is to characterize precisely the Pareto efficient equilibria of Γ_{AI}. As Proposition 2 shows, player 1 may have to exit with positive probability following a failed revelation stage. However, since exit is inefficient, it is optimal for player 1 to try and screen whether player 2 took a productive action or not before actually exiting. This can be achieved by requiring player 2 to take the unconfirmed revealed action again. The action will yield benefits with probability q if it is indeed productive, and will fail systematically if it is unproductive. However, such statistical screening delays punishment and increases player 2's temptation to deviate. This trade-off makes solving for optimal strategies delicate. While this paper proves useful qualitative properties, obtaining more precise characterizations remains an important objective for future research.

A Improper priors, measurable strategies, and beliefs

As Section 2.3 highlights, the appropriate solution concept for game Γ_{AI} is perfect Bayesian equilibrium. However, because player 1 holds an improper prior, insuring that players' strategies induce well-defined beliefs over future utility requires some care. This is done by restricting the players' behavior to be invariant at histories that are identical up to a relabeling of actions. We begin with a few definitions.

Definition A.1 (relabeling) *Consider a bijection $\gamma : \mathcal{A} \rightarrow \mathcal{A}$, and histories h_t^i, $i \in \{1,2\}$. The relabeling $\gamma(h_t^i)$ of history h_t^i by γ is defined by*

$$\begin{aligned} \gamma(h_t^1) &= \{d_1, \gamma(A_1), \gamma(a_1), \cdots, d_{t-1}, \gamma(A_{t-1}), \gamma(a_{t-1})\} \\ \gamma(h_t^2) &= \{\gamma(\mathcal{N})\} \sqcup \gamma(h_t^1) \sqcup \{d_t, \gamma(A_t)\}. \end{aligned}$$

For any $i \in \{1,2\}$, histories h_t^i and $\hat{h}_t^i$ are equivalent up to a relabeling of actions if and only if there exists a bijection γ such that $\gamma(h_t^i) = \hat{h}_t^i$. Given an action $a \in \mathcal{A}$, $h_t^i(a)$ denotes the portion of history h_t^i that pertains to action a:

$$\begin{aligned} h_t^1(a) &= \{\mathbf{1}_{a\in A_1}, \mathbf{1}_{a=a_1}, \mathbf{1}_{a=a_1}\tilde{b}(a_1), \cdots, \mathbf{1}_{a\in A_{t-1}}, \mathbf{1}_{a=a_{t-1}}, \mathbf{1}_{a=a_{t-1}}\tilde{b}(a_{t-1})\} \\ h_t^2(a) &= \{\mathbf{1}_{a\in\mathcal{N}}\} \sqcup h_t^1(a) \sqcup \{\mathbf{1}_{a\in A_t}\}. \end{aligned}$$

Given a history h_t^i and an action $a \in \mathcal{A}$, $h_t^i(a)$ lists the public or payoff relevant features of a (e.g. has it been taken, when, did it yield benefits...). The restriction imposed on strategies is that at equivalent histories, players should take actions that have the same public or payoff relevant features. Let us define this formally.

Definition A.2 (label independent strategies) *A strategy s_1 of player 1 is label independent if and only if whenever h_t^1 and $\hat{h}_t^1$ are equivalent, then $s_1(h_t^1) = s_1(\hat{h}_t^1)$.*

A strategy s_2 of player 2 is label independent if and only if whenever h_t^2 and $\hat{h}_t^2$ are equivalent, the actions a and $\hat{a}$ prescribed by s_2 at h_t^2 and $\hat{h}_t^2$ satisfy $h_t^2(a) = \hat{h}_t^2(\hat{a})$.

Intuitively, the requirement that strategies be label independent is consistent with the assumption that productive actions are drawn according to an improper uniform distribution, and that the specific labels attached to actions are payoff irrelevant.[16]

Throughout the paper, players are constrained to use label independent strategies. Note that this is a restriction on strategies (essentially a measurability requirement) rather than a property of equilibrium. As is detailed below, this restriction ensures that players have well defined beliefs about future utility. Among other things, it rules out strategies in which player 2's decision to take or not a productive action depends on whether the index in $\mathbb{N}$ assigned to that productive action belongs to a set $\mathcal{S} \subset \mathbb{N}$ that is not measurable with respect to uniform improper priors.[17]

Let us now highlight why restricting players to use label independent strategies ensures that beliefs about future payoffs are well defined. To begin, note that if h_t^2 and $\hat{h}_t^2$ are equivalent, and $h_t^2(a) = \hat{h}_t^2(\hat{a})$, then histories $h_t^2 \sqcup \{a\}$ and $\hat{h}_t^2 \sqcup \{\hat{a}\}$ are also equivalent, and action a has the same payoff consequences conditional on h_t^2 as action $\hat{a}$ conditional on $\hat{h}_t^2$. This implies that when strategies are label independent, future utility depends only on the equivalence class of the current history. We now show that the equivalence class of a history h_t^i is characterized by a finite number of distinguishing features.

[16] Another possible interpretation is that the information available to player i, on which strategy s_i should depend, is the equivalence class of history h_t^i. Labels are then assigned to actions privately, either as actions are being taken, or to differentiate between many otherwise undistinguishable actions that could be taken.

[17] An example of such a set is $S = \cup_{k\geq 0}\{n_k, n_k+1, \cdots, m_k\}$ with $n_0 = 1$, $m_k = 2^{n_k}$ and $n_{k+1} = 2^{m_k}$.

Definition A.3 (distinguished histories) *Given any histories h_t^1 and h_t^2, we define distinguished histories by*

$$
\begin{aligned}
h_t^{1,D} &= \{h_t^1(a) |\ a \in \{a_1, \cdots, a_{t-1}\}\} \\
h_t^{2,D} &= \{h_t^2(a) |\ a \in \{a_1, \cdots, a_{t-1}\} \cup \mathcal{N}\}.
\end{aligned}
$$

The distinguished history $h_t^{i,D}$ is a finite list summarizing the portion of history h_t^i pertaining to actions that have been taken, or that player i knows to be productive. Note that $h_t^{i,D}$ records only label independent data and whenever h_t^i and $\hat{h}_t^i$ are equivalent, then $h_t^{i,D} = \hat{h}_t^{i,D}$. Conversely, the next lemma shows that the distinguished history $h_t^{i,D}$ characterizes the equivalence class of history h_t^i.

Lemma A.1 *Consider $i \in \{1, 2\}$ and histories h_t^i and $\hat{h}_t^i$ in $\mathcal{H}^i$. Whenever $h_t^{i,D} = \hat{h}_t^{i,D}$, then h_t^i and $\hat{h}_t^i$ are equivalent up to a relabeling of actions.*

Proof of Lemma A.1: Let us consider the case where player i is player 2. We denote by $\{a_s|s < t\} \cup \mathcal{N}$ actions that are either productive or have been taken at history h_t^2, and by $\{\hat{a}_s|s < t\} \cup \widehat{\mathcal{N}}$ actions that are either productive or have been taken at history $\hat{h}_t^2$.

Since $h_t^{2,D} = \hat{h}_t^{2,D}$, it must be that there is a bijection γ between $\{a_s|s < t\} \cup \mathcal{N}$ and $\{\hat{a}_s|s < t\} \cup \widehat{\mathcal{N}}$ such that for any a, $h_t^2(a) = \hat{h}_t^2(\gamma(a))$.

Now pick an action $a \notin \{a_s|s < t\} \cup \mathcal{N}$ and consider $\{a' \mid h_t^2(a') = h_t^2(a)\}$, the set of actions that are undistinguishable from a on the basis of history h_t^2. When action a varies, such sets form a partition of $\mathcal{A} \setminus \{a_s|s < t\} \cup \mathcal{N}$. The set $\{a' \mid h_t^2(a') = h_t^2(a)\}$ is countably infinite and associated to the set $\{\hat{a} \mid \hat{h}_t^2(\hat{a}) = h_t^2(a)\}$ which is also countably infinite. This implies that γ can be extended as a bijection from $\{a' \mid h_t^2(a') = h_t^2(a)\}$ to $\{\hat{a} \mid \hat{h}_t^2(\hat{a}) = h_t^2(a)\}$.

Altogether, this implies that there exists a bijection γ from $\mathcal{A}$ to $\mathcal{A}$ such that for every action a, $h_t^2(a) = \hat{h}_t^2(\gamma(a))$. This implies that $\gamma(h_t^2) = \hat{h}_t^2$. A similar proof holds for histories of player 1. ∎

Since payoffs at time t depend only on the equivalence class of histories h_t^i, Lemma A.1 implies that payoffs depend only on distinguished histories $h_t^{i,D}$. Furthermore, since strategies are label independent, the distribution of $h_t^{i,D}$ given h_{t-1}^i is well-defined and identical to that of $h_t^{i,D}$ given $h_{t-1}^{i,D}$. Together, this yields that players have well-defined beliefs over future distinguished histories, and hence well-defined beliefs about future utility.

The following lemma highlights that given a distinguished history, player 1 puts positive probability on an action being productive only if it has been revealed.

Lemma A.2 *Consider an equilibrium* (s_1, s_2) *and a distinguished history* $h_{t+1}^{1,D}$ *on the equilibrium path. Any action* $a \in \mathcal{A}$ *that has not been revealed is such that* $Prob_1\{a \in \mathcal{N} | h_{t+1}^{1,D}\} = 0$.

Proof of Lemma A.2: If action a has been taken but not at a revelation stage, then by definition of a revelation stage, there is probability 0 that action a is productive.

Let us now consider the situation where action a has never been taken. From the perspective of player 1, a is indistinguishable from any action a' such that $h_{t+1}^1(a') = h_{t+1}^1(a)$. Hence for any such action, $Prob_1(a \in \mathcal{N} | h_{t+1}^{1,D}) = Prob_1(a' \in \mathcal{N} | h_{t+1}^{1,D})$. Since there are infinitely many such actions a', this implies that $Prob_1(a \in \mathcal{N} | h_{t+1}^{1,D}) = 0$. ■

B Proofs

B.1 Results of Section 3

Proof of Proposition 1: By Assumption 1 we know there exists an equilibrium (s_1^0, s_2^0) in which player 1 stays in the first period.[18] Denote (V_1^0, V_2^0) the associated initial values. Since player 1 has the option to exit and player 2 can choose not to reciprocate, 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[19] 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, s_1 prescribes that player 1 should never stay again in equilibrium. Consider the alternative strategies $\tilde{s}_1$ and $\tilde{s}_2$ defined by:

$$\forall i \in \{1, 2\},\ \tilde{s}_i(h^i) = \left\{ \begin{array}{ll} s_i(h^i) & \text{if } \nexists h' \ s.t.\ h^i = h_t^1 \sqcup h' \\ s_i^0(\hat{h}^i) & \text{if } h^i = h_t^1 \sqcup \hat{h}^i \end{array} \right. .^{20}$$

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

Let us now consider the case in which following h_t^1 there is an attainable equilibrium history at which player 1 stays under s_1. This implies that at h_t^1 the continuation values

[18]See Section 3.2.

[19]i.e. a history that can be reached with positive probability.

[20]For conciseness, the description of strategy $\tilde{s}_2$ omits the initial element $\{\mathcal{N}\}$ from history h^2.

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:

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

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

Proof of Proposition 2: The first step of the proof is to establish that when player 1 never exits on the equilibrium path, then player 2 can guarantee herself value $\underline{V}_2^{N'}$ by deviating to the strategy in which she only takes actions that are costless.

Consider a history h_t^2 with a set of available actions A_t, such that no action in $\mathcal{N}'$ is available, i.e. $A_t \cap \mathcal{N}' = \emptyset$. Such histories happen with probability $(1-p)^{N'}$. Since this is an equilibrium history and, by assumption, player 1 never exits following histories that are attainable on the equilibrium path, there must be some non-empty set of actions $A_{Stay} \subset A_t$ such that player 1 will stay whenever player 2 takes an action $a \in A_{Stay}$. There are two cases. Either A_{Stay} is finite or infinite. If A_{Stay} is infinite, there necessarily exists a costless action $a \in A_{Stay}$. If A_{Stay} is finite, then since A_{Stay} does not include revealed actions, Lemma A.2 implies that $Prob_1(A_{Stay} \cap \mathcal{N} \neq \emptyset | h_t^{1,D}) = 0$. Since A_{Stay} is non-empty this implies that at such a history h_t^2 there always exists a costless action that player 2 can take, following which player 1 will stay. Since such histories happen with probability $(1-p)^{N'}$, this means that player 2 can guarantee herself value greater than $\frac{1}{1-\delta(1-p)^{N'}}\pi$ by only taking costless actions.

Revelation of a costly action a is incentive compatible only if player 2's continuation values satisfy

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

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 3: Consider an equilibrium history h_t^1 such that all revelation stages have been confirmed. Such a history could not be attained if player 2 had deviated at a revelation stage. It cannot be attained by a deviation of player 1 either. This implies that improving both players' equilibrium values at history h_t^1 only improves earlier incentive compatibility constraints.

From then on the proof is very similar to that of Proposition 1. In particular we know that if player 1 exited at history h_t^1, it is possible to replace the players' continuation strategies at h_t^1 by equilibrium strategies which give greater value to both players.

If player 1 stays with positive probability after h_t^1, then one may simply skip history h_t^1 and consider the modified strategies $(\tilde{s}_1, \tilde{s}_2)$,

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

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

If under the original equilibrium, player 1 exits permanently following h_t^1, then one can replace continuation strategies with any equilibrium such that player 1 stays in the first period. Such an equilibrium exists by Assumption 1, and it dominates permanent exit since in any equilibrium where player 1 stays, player 1 must get value at least 0 and player 2 must get value at least π. This concludes the proof. ∎

B.2 Results of Section 4.1

Lemma B.1 establishes that there exist parameter values such that Assumptions 1 and 2 hold together.

Lemma B.1 *Pick parameter values $k > 0$, $\pi > 0$, $\delta > 1/2$, a pair (p, q) such that $p > q > \delta$ and $\frac{1-q}{\delta} < 1 - \frac{1-pq}{pq}\frac{1-\delta}{\delta}$, and b^0 such that $\frac{1-q}{\delta} < \frac{k}{pqb^0} < 1 - \frac{1-pq}{pq}\frac{1-\delta}{\delta}$. The following hold,*

(i) $\forall c > 0, \quad \delta q \overline{V}_2^0 > \delta(\overline{V}_2^0 - \underline{V}_2^D)$.

(ii) *Let* $c_{max} = \max\{c | \delta q \overline{V}_2^0 \geq c\}$ *and* $c_{min} = \min\{c | \delta(\overline{V}_2^0 - \underline{V}_2^D) \leq c\}$. *We have that* $c_{max} > \max\{\frac{\delta\pi}{p}, c_{min}\}$.

(iii) *For any* $c \in (\max\{\frac{\delta\pi}{p}, c_{min}\}, c_{max})$, $\frac{1}{1-\delta} q\pi > \overline{V}_2^0$ *and both Assumptions 1 and 2 hold together.*

Proof of Lemma B.1: Let us begin with point (i). We have that

$$\begin{aligned}\delta q\overline{V}_2^0 &= \frac{\delta}{1-\delta}q\pi - q\frac{\delta}{1-\delta}\frac{k}{qb^0}c \\ \delta(\overline{V}_2^0 - \underline{V}_2^D) &= \frac{\delta}{1-\delta}\left(1-\frac{1-\delta}{1-\delta(1-p)}\right)\pi - \frac{\delta}{1-\delta}\frac{k}{qb^0}c.\end{aligned}$$

Note that $1-\frac{1-\delta}{1-\delta(1-p)}$ is increasing in p and that for $p=1$, it is equal to δ, which is strictly less than q. This implies that $q > 1-\frac{1-\delta}{1-\delta(1-p)}$ and hence $\delta q\overline{V}_2^0 > \delta(\overline{V}_2^0 - \underline{V}_2^D)$ for all $c > 0$. This shows point (i).

Regarding point (ii), the fact that $c_{max} > c_{min}$ simply follows from point (i). Let us now show that $c_{max} > \delta\pi/p$. We have that $c_{max} = \frac{\delta}{1-\delta}q\pi\left(\frac{\delta}{1-\delta}\frac{k}{b^0}+1\right)^{-1}$. Hence,

$$c_{max} > \delta\pi/p \iff \pi\left(\frac{\delta}{1-\delta}\frac{k}{pqb^0}pq+1\right) < \frac{1}{1-\delta}pq\pi.$$

The fact that $\frac{k}{pqb^0} < 1-\frac{1-pq}{pq}\frac{1-\delta}{\delta}$ implies this last inequality holds. This proves point (ii).

We now turn to point (iii). Let us first show that $q\frac{1}{1-\delta}\pi > \overline{V}_2^0$. We have,

$$q\frac{1}{1-\delta}\pi > \overline{V}_2^0 \iff q > 1-\frac{k}{pqb^0}\frac{pc}{\pi} \iff \frac{k}{pqb^0} > \frac{\pi}{pc}(1-q).$$

Since $c > \frac{\delta\pi}{p}$, we have that $\frac{\pi}{pc}(1-q) < (1-q)/\delta$. Since b^0 is picked such that $\frac{k}{pqb^0} > (1-q)/\delta$, we have that indeed $q\frac{1}{1-\delta}\pi > \overline{V}_2^0$. This, along with points (i) and (ii), implies that Assumptions 1 and 2 hold together. $\blacksquare$

Proof of Lemma 1: Since $\delta(\overline{V}_2^0 - \underline{V}_2^D) < c$, there exists $\mu > 0$ such that for all $b^1 \in [b^0, b^0+\mu]$, $\delta(\overline{V}_2^1 - \underline{V}_2^D) < c-\mu$. Let us now pick a value of K independent of $b^1 \in [b^0, b^0+\mu]$. For any K, define

$$\underline{V}_2^{D,K} \equiv \frac{1-\delta^{K+1}(1-p)^{K+1}}{1-\delta(1-p)}\pi.$$

As K goes to infinity, $\underline{V}_2^{D,K}$ converges to $\underline{V}_2^D$. Furthermore, since $\delta(\overline{V}_2^1 - \underline{V}_2^D) < c-\mu$, there exists K large enough, such that for all $b^1 \in [b^0, b^0+\mu]$, $\delta(\overline{V}_2^1 - \underline{V}_2^{D,K}) < c-\mu/2$.

Consider an equilibrium (s_1, s_2) and a revelation stage $h_t^{2|1}$ for action a^1. Denote by $\hat{\eta}$ the probability that player 1 exits in the next K periods. Let us consider subsequent histories h_s^2, with $t < s \leq t + K$, 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 $(1-q)^{s-t+1}(1-p)^{s-t}$ and hence following such histories, exit can only occur with probability less than

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

Out of equilibrium, if player 2 deviates by taking only costless actions, 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 at least payoff

$$\begin{aligned} \underline{V}_2^{D,K,\hat{\eta}} &\geq \sum_{s=t}^{t+K} \delta^{s-t}(1-p)^{s-t}\left(1 - \frac{\hat{\eta}}{(1-q)^{K+1}(1-p)^{K}}\right)\pi \\ &\geq \left(1 - \frac{\hat{\eta}}{(1-q)^{K+1}(1-p)^{K}}\right)\frac{1-\delta^{K+1}(1-p)^{K+1}}{1-\delta(1-p)}\pi. \end{aligned}$$

For revelation to be incentive compatible, we must have $\delta(\overline{V}_2^1 - \underline{V}_2^{D,K,\hat{\eta}}) < c$, which implies that

$$\hat{\eta} \geq \frac{(1-q)^{K+1}(1-p)^{K}}{\delta \underline{V}_2^{D,K}}\left[c - \delta(\overline{V}_2 - \underline{V}_2^{D,K})\right] \equiv \eta > 0.$$

Hence there exist $\mu > 0$, $K \in \mathbb{N}$ and $\eta > 0$ such that for all $b^1 \in [b^0, b^0 + \mu]$, at any revelation stage for action a^1, there is probability greater than η that player 1 exits in the next K periods. ■

Proof of Proposition 4: We begin with point (i). It is intuitively clear that when b^1 becomes large, revealing action a_1 creates value. However, providing incentives for revelation sometimes requires inefficient punishment, and value must be destroyed on some equilibrium paths. Hence, the delicate part of the proof is to show that after any history, including histories where inefficient punishment is required on the equilibrium path, a^1 will be confirmed with positive probability after any history where player 1 stays.

Consider a Pareto efficient equilibrium (s_1, s_2) and a history h_t^1 at which player 1 stays. We first consider the case in which action a^0 has been confirmed before h_t^1. By Assumption

2, $\frac{1}{1-\delta}q\pi > \overline{V}_2^0$. Hence, there exists $\overline{\Delta}$ high enough such that for all $b^1 > b^0 + \overline{\Delta}$, we have

$$\begin{aligned} V_2^* &\equiv q\frac{1}{1-\delta}\left(\pi - \frac{1}{q\sqrt{b^1}}c\right) > \overline{V}_2^0 \\ V_1^* &\equiv q\frac{1}{1-\delta}(-k+\sqrt{b^1}) > \frac{1}{1-\delta}(-k+b^0). \end{aligned}$$

The proof of point (i) uses the two following facts. First, by construction, there exist $\tau_1 \in \mathbb{N}$, and $\nu_1 > 0$ such that at any history h_s^2 where player 1's continuation value is greater than V_1^*, or player 2's continuation value is greater than V_2^*, there must be probability at least ν_1 that action a^1 is confirmed in the next τ_1 periods. Second, by point (ii) of Assumption 2, player 1 cannot be induced to stay if player 2 never takes action a^1, and only takes action a^0 when action a^1 is unavailable. This implies that there exist $\tau_2 \in \mathbb{N}$ and $\nu_2 > 0$ such that if player 1 stays at some history h_t^1, then player 2 must take action a^1, or take action a^0 at a history where a^1 is available, with probability at least ν_2 in the next τ_2 periods.

Consider h_s^2 with $s > t$, an equilibrium history at which player 2 takes action a^0, and action a^1 is available. Denote $V_1(h_s^2)$ and $V_2(h_s^2)$ the players' continuation values at such a history. If $V_1(h_s^2) \geq V_1^*$ or $V_2(h_s^2) \geq V_2^*$, we know that action a^1 must be taken with probability at least ν_1 in the next τ_1 periods.

Assume temporarily that at h_s^2, players' have continuation values such that $V_1(h_s^2) < V_1^*$ and $V_2(h_s^2) < V_2^*$. Let us show that if this is the case, then (s_1, s_2) cannot be efficient. Indeed, consider the modified strategies $(\hat{s}_1, \hat{s}_2)$ that coincide with (s_1, s_2) except following equilibrium history h_s^2. At history h_s^2, strategies $(\hat{s}_1, \hat{s}_2)$ prescribe that player 2 take action a^1. If a^1 is immediately confirmed, then in the continuation game, on the equilibrium path, player 1 stays every period and player 2 takes action a^1 with probability $\frac{1}{pq\sqrt{b^1}}$ whenever it is available (using public randomizations). If action a^1 fails when player 2 takes it at history h_s^2, then $(\hat{s}_1, \hat{s}_2)$ prescribe that player 1 always exits and player 2 only takes unproductive actions. Under strategies $(\hat{s}_1, \hat{s}_2)$, players obtain values V_2^* and V_1^* at history h_s^2. This increases both players' continuation values and implies that starting from h_s^2, $(\hat{s}_1, \hat{s}_2)$ is indeed an equilibrium. In particular, since taking a costly action was incentive compatible for player 2 under (s_1, s_2), it is also incentive compatible under $(\hat{s}_1, \hat{s}_2)$. Note that players obtain these higher continuation values only if actions a^0 and a^1 are both confirmed. We also know that if player 2 deviates before h_s^2, then histories at which a^0 and a^1 are both confirmed are not reachable. Hence, improving players' utility at equilibrium histories where actions a^0 and a^1 are confirmed increases continuation values on the equilibrium path but does not change player 2's payoffs upon deviation. This implies that $(\hat{s}_1, \hat{s}_2)$ is an equilibrium of the overall game. Since it dominates (s_1, s_2), which is by assumption Pareto efficient, we obtain a contradiction. This yields that $V_1(h_s^2) \geq V_1^*$ or $V_2(h_s^2) \geq V_2^*$. Altogether, this implies that whenever player 1 stays, there is probability at least $q\nu_1\nu_2$ that action a^1 will be confirmed in the next $\tau_1 + \tau_2$ periods.

We now turn to the case where a^0 is not confirmed at history h_t^1. Since player 1 stays at history h_t^1, there must be probability $\nu_2 > 0$ that player 2 takes action a^0 or a^1 in the next τ_2 periods. Consider a history h_s^2 at which player 2 takes action a^0. Since by Assumption 2, $\delta\pi/p < c$ and $q > \delta$, it follows that $(1-q)\frac{\delta}{1-\delta}\pi < c$. In words, this means that obtaining profit π forever if action a^0 fails does not cover player 2's cost of taking a productive action. Hence, there exist $\tau_3 \in \mathbb{N}$ and $\nu_3 > 0$ such that whenever player 2 takes action a^0 and a^0 is confirmed, there is probability greater than ν_3 that player 1 stays at least once in the next τ_3 periods. This puts us in the configuration discussed above. Altogether we can conclude that at h_t^1 there is probability at least $q^2\nu_1\nu_2\nu_3$ that action a^1 will be confirmed in the next $\tau_1 + \tau_2 + \tau_3$ periods. This proves point (i).

We now turn to point (ii). To begin, we consider the case where a^0 is confirmed at some history $h_{t_0}^1$ and no other action has been revealed.[21] Let us define the sets of values $\mathcal{U}_0$, $\mathcal{U}_{0,1}$ and $\mathcal{U}_{0,1}^{K,\eta}$ as follows:

(i) $\mathcal{U}_0$ is the set of Pareto efficient equilibrium values in the complete information game where only a^0 is productive, at a history $h_t^2 \in \mathcal{H}^2$ where a^0 is available.

(ii) $\mathcal{U}_{0,1}$ is the set of Pareto efficient equilibrium values in the complete information game where a^0 and a^1 are productive, at a history $h_t^2 \in \mathcal{H}^2$ where a^1 is available.

(iii) $\mathcal{U}_{0,1}^{K,\eta}$ is the set of values sustainable in the complete information game where a^0 and a^1 are productive, in equilibria such that player 1 exits with probability greater than η in the next K periods, at a history $h_t^2 \in \mathcal{H}^2$ where a^1 is available.

Consider a pair of values $(V_1, V_2) \in \mathcal{U}_{0,1}$. Since player 1 never exits in equilibrium there exists a positive number r such that $V_2 = \frac{1}{1-\delta}(\pi - prc)$. Since player 2 always has the option to take unproductive actions, we have that $V_2 \geq \pi$. This implies that $r < \frac{\delta\pi}{pc}$, which, by point (i) of Assumption 2, implies that $r < 1$. Hence (V_1, V_2) can be achieved under complete information by having player 1 never exit in equilibrium, and player 2 take action a^1 with

[21] Note that unproductive actions may have been taken at histories that are not revelation stages.

probability r when it is available. By considering the strategy in which player 1 never exits in equilibrium and player 2 takes action a^0 with probability r whenever it is available, it follows that as b^1 goes to b^0, the set of values $\mathcal{U}_{0,1}$ converges to $\mathcal{U}_0$. More formally, for all $\epsilon > 0$, there exists b^1 close enough to b^0 such that for all $(V_1, V_2) \in \mathcal{U}_{0,1}$, there exists $(\hat{V}_1, \hat{V}_2) \in \mathcal{U}_0$ such that $\hat{V}_1 \geq V_1 - \epsilon$ and $\hat{V}_2 \geq V_2 - \epsilon$.

Furthermore, for any $K \in \mathbb{N}$ and $\eta > 0$, there exists $\alpha > 0$ such that for all $(V_1', V_2') \in \mathcal{U}_{0,1}^{K,\eta}$, there exists $(V_1, V_2) \in \mathcal{U}_{0,1}$ 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 so that for all $b^1 \in (b^0, b^0 + \underline{\Delta})$, first, Lemma 1 holds, and second, whenever $(V_1', V_2') \in \mathcal{U}_{0,1}^{K,\eta}$, 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 consider a revelation stage $h_t^{2|1}$ for action a^1, such that no other revelation stages have occurred between $h_{t_0}^1$ and $h_t^{2|1}$. By Lemma 1, there exist K and $\eta > 0$ such that values (V_1^{Rev}, V_2^{Rev}) at $h_t^{2|1}$ are dominated by values in $\mathcal{U}_{0,1}^{K,\eta}$. This implies that for all $b^1 < b^0 + \underline{\Delta}$, 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 by $(V_1^{Conf,0}, V_2^{Conf,0})$ continuation values at the history $h_{t_0}^1$ 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

$$\begin{aligned}
\hat{V}_1^{Conf,0} &\equiv V_1^{Conf,0} + prob(h_t^{2|1}) \left(\hat{V}_1 - V_1^{Rev} \right) \\
\hat{V}_2^{Conf,0} &\equiv V_2^{Conf,0} + prob(h_t^{2|1}) \left(\hat{V}_2 - V_2^{Rev} \right),
\end{aligned}$$

obtained by replacing revelation values at $h_t^{2|1}$ with continuation values not involving revelation of action a^1. 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 first 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 they dominate the original values involving further revelation. The first question is whether such values correspond to a continuation equilibrium. Let us show that this is indeed the case.

Between $h_{t_0}^1$ and a consecutive revelation stage $h_t^{2|1}$ for action a^1, all revealed actions are confirmed. Proposition 3 implies that (s_1, s_2) prescribes no exit on the equilibrium path between $h_{t_0}^1$ and $h_t^{2|1}$. Hence, there exists $r > 0$ 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 point (i) of Assumption 2, we have that $c/\delta > \pi$. The fact that $\widetilde{V}_2^{Conf,0} > \pi$ implies that $r < \overline{r} < 1$. 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.

To finish the proof, we must show that incentive constraints at histories preceding $h^1_{t_0}$ still hold after changing continuation strategies at $h^1_{t_0}$. By assumption, history $h^1_{t_0}$ is such that no action is revealed and unconfirmed. This implies that $h^1_{t_0}$ is not attainable by earlier deviations from player 2. Therefore, increasing continuation values at $h^1_{t_0}$ does not increase player 2's payoffs upon deviation and increases equilibrium continuation values. As a result, increasing continuation values at $h^1_{t_0}$ can only improve earlier incentive compatibility constraints.

This concludes the proof of point (ii): for all $b^1 \in (b^0, b^0 + \underline{\Delta})$, efficient equilibria should involve no further revelation upon confirmation of a^0. An identical proof holds in the case where a^1 is confirmed and no other action has been revealed. ■

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] Alonso, Ricardo and Niko Matouschek, *Relational Delegation*, RAND Journal of Economics, Vol. 38, No. 4, pp. 1070-1089, 2007.

[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] Blume, Andreas and April Mitchell Franco, *Decentralized learning from failure*, Journal of Economic Theory, Vol. 133, No. 1, pp. 504-523, 2007.

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

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

[10] 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.

[11] Ellison, Glenn and Richard Holden, *A Theory of Rule Development*, MIT working paper, 2008.

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

[13] 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, 2008.

[14] Gossner, Olivier and Nicolas Vieille, *Strategic learning in games with symmetric information*, Games and Economic Behavior, Vol. 42, No. 1, pp. 25-47, 2003.

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

[16] Hart, Sergiu, *Nonzero-Sum Two-Person Repeated Games with Incomplete Information*, Mathematics of Operations Research, Vol. 10, No. 1, pp. 117-153, 1985.

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

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

[19] March, James, *Exploration and Exploitation in Organized Learning*, Organization Science, Vol. 2, No. 1, pp. 71-87, 1991.

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

[21] Pęski, Marcin, *Repeated Games with Incomplete Information on One Side and Discounting*, Theoretical Economics, Vol. 3, No. 1, pp. 29-84, 2008.

[22] Rothschild, Michael *A Two-Armed Bandit Theory of Market Pricing*, Journal of Economic Theory, Vol. 9, No. 2, pp. 185-202, 1974.

[23] 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.

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

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

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

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

[28] Wiseman, Thomas, *A Partial Folk Theorem for Games with Unknown Payoff Distributions*, Econometrica, Vol. 73, No. 2, pp. 629-645, 2005.