COOPERATION DYNAMICS IN REPEATED GAMES OF ADVERSE SELECTION[*]

JUAN F. ESCOBAR[†] AND GASTÓN LLANES[‡]

ABSTRACT. We study dynamic games with private information. After any history, signaling reveals information that helps players coordinate their future actions, but also makes the problem of predicting the informed player's actions harder for the uninformed player. In equilibrium, the informed player may play aggressive or uncooperative actions, but his partner only tolerates a certain number of such actions. We discuss several applications of our results: We explain the cycles of cooperation and conflict observed in World War I, show that price leadership and unilateral price cuts can be part of an optimal signaling equilibrium in a repeated Bertrand game with incomplete information, show that communication between cartel members may be socially efficient in a repeated Cournot game, and study the optimal organizational design when direct communication leads to delays in decision making. Finally, we explore a model with frequent interactions and show that when types are more persistent, informational frictions are smaller.

KEYWORDS. Repeated games, private information, signaling, coordination, adaptation, collusion, communication.

Date: December 22, 2016.

[*]We thank Elton Dusha, Joe Harrington, Johannes Hörner, Romans Pancs, and several audiences for insightful comments. Escobar acknowledges financial support from Conicyt (Fondecyt No. 1130435). Llanes acknowledges financial support from Conicyt (Fondecyt No. 1150326). Escobar and Llanes acknowledge financial support from the Institute for Research in Market Imperfections and Public Policy, MIPP, ICM IS130002, Ministerio de Economía, Fomento y Turismo.

[†] Center for Applied Economics, Department of Industrial Engineering, University of Chile, E-mail: jescobar@dii.uchile.cl.

[‡] Pontificia Universidad Católica de Chile, E-mail: gaston@llanes.com.ar.

1. Introduction

Trust-based relationships often exhibit apparent deviations from cooperative behavior. For example, during World War I, frontline soldiers often refrained from attacking the enemy –provided that their restraint was reciprocated by soldiers on the other side– but unilateral aggressions did occur and triggered retaliations and mutual attacks (Ashworth, 1980). Likewise, cartel members often make unilateral price cuts, even in fully-functioning cartels (Marshall and Marx, 2013), and governments in self-enforcing trade agreements raise their import tariffs, despite the fact that such measures are detrimental for foreign partners (Bagwell and Staiger, 2005). In contrast, in most repeated game models, players never play apparently uncooperative actions on the path of play Green and Porter (1984); Rotemberg and Saloner (1986); Fudenberg and Maskin (1986); Abreu et al. (1986); Athey and Bagwell (2001).

In this paper, we shed light on this kind of phenomena by studying the scope for cooperation in a repeated game with Markovian private information. Two players make perfectly observable decisions at each round. Player 1 is privately informed about his own payoffs, which evolve according to a finite Markov chain. Importantly, since players cannot exchange cheap-talk messages, player 2 can learn about player 1's type only by observing player 1's actions.[1] We show that the combination of private information and lack of communication may result (but need not to) in apparent cooperation breaks, such as unilateral price cuts, aggressions, debt defaults, etc. These breaks substitute direct communication and may benefit the relationship by allowing the informed player to use his private information and signal the most profitable course of play. Our main theoretical results characterize a class of approximately Pareto-optimal equilibria as players become arbitrarily patient. This characterization uncovers new economic forces in repeated interactions with incomplete information and can be used in a variety of applications.

In our dynamic game, the amount of information revealed by player 1 is endogenously determined. Given any history of actions, player 1 may fully reveal his private information by separating and signaling his types. A benefit from such information revelation is that once types have been perfectly revealed by player 1's actions, the uninformed player can move on to the next round with more precise beliefs about the new type of player 1 that he will face, and this information improvement is beneficial for the relationship. A second benefit from full revelation is that player 1's payoffs do depend on his types and typically it will be in his short-run interest to choose a type-dependent action (which reveals his type). Yet, a perfectly revealing strategy need

[1] The assumption of no communication is just a simplifying one, and acknowledges the fact –articulated by Marschak and Radner (1972) and Arrow (1985) among others– that oftentimes parties encounter nontrivial communication costs. This assumption is natural in collusion applications since price discussions between competitors are generally illegal. Ashworth (1980) documents the communication problems faced by enemy troops trying to avoid confrontation during World War I. When discussing limited war, Schelling (1960) explains that "an agreement on limits is difficult to reach . . . because communication becomes difficult between adversaries in a time of war."

not be optimal: when the informed player is fully revealing his private information, it is harder for player 2 to predict player 1's current action, which hurts player 2's current payoffs. The costs of revealing information at any given history are the losses that the uninformed player has when the informed player's action is unknown.

We formally capture this tradeoff by ignoring incentive constraints and studying the problem of maximizing the average expected payoffs over all strategies. This optimization problem can be formulated as a Bellman equation in which the state variable is the public belief the uninformed player has about the current type. A solution to this equation solves the tradeoff between revealing and not revealing information, and yields an optimal equilibrium path for the repeated game with Markovian private information.

The construction of an approximately optimal equilibrium for the repeated game specifies a strategy in which the uninformed player forgives but does not forget hostile actions. To see this, consider two firms that are trying to collude in a market. Most of the time, firms are equally efficient –and therefore should fix the monopoly price and share the demand– but sometimes firm 1 is much more efficient and it is therefore desirable for the cartel to have firm 1 as the only producer. The problem is that only firm 1 knows its costs. The cartel should not allow firm 1 to freely undercut firm 2 because firm 1 would undercut even when both firms are equally efficient. We show that, more generally, the uninformed player forgives apparently hostile actions –such as price reductions–but does not forget them. Indeed, in equilibrium, the uninformed player keeps track of the number of actions played by the informed player conditional on public beliefs, and (off-path) the relationship enters a punishment phase if the path of actions seems openly mischievous.

The equilibrium strategies feature on-path dynamics that differ from previous literature. In an equilibrium with some information revelation, public beliefs determine the distribution over actions at any given history. In particular, apparently uncooperative actions (such as price cuts and price wars in a collusion application) occur on the path of play and are the optimal response of the relationship to incomplete information and lack of communication.

The assumptions of private information and no communication are natural in many long-run relationships. We illustrate our results and methods with some applications.

The first application is motivated by the live and let live system during World War I. During trench warfare, frontline soldiers often refrained from attacking the enemy. Army commanders were aware of the tendency towards non-aggression and would order raids to correct the "offensive spirit" of the troops (Ashworth, 1980; Axelrod, 1984). Battalions faced severe information asymmetries because they could not discern if aggressions were caused by opportunistic behavior or by military orders. Moreover, direct, cheap-talk communication was virtually non-existent as it was severely punished by high command. We apply our general results to explain how cooperation can arise and evolve in this type of environment. We model the relationship between soldiers as a prisoners dilemma, in which one of the sides can receive a privately observed shock that makes mutual cooperation inefficient. Our dynamic programming characterization can be used to show that aggressions can occur on the path of play. Full cooperation can be resumed after the informed side signals that army commanders left by stopping aggressions, or after a cooling-off phase in which both sides mutually attack for a fixed (but optimally chosen) number of periods. We complement our theoretical analysis with some evidence showing that soldiers actually kept an account of the number of aggressions received from the other side, suggesting that our equilibrium strategies may be a good approximation to the way soldiers actually behaved.

Our second application is to collusion with Bertrand competition. Firms trying to collude face severe informational asymmetries –local demand conditions, private technological shocks– and price discussions between competitors are illegal. We characterize the optimal collusive scheme in a Bertrand game of differentiated products in which one of the firms has private information about its demand. Consistent with case studies (Marshall and Marx, 2013), in our model *unilateral price cuts* occur on the path of play. Our repeated Bertrand game can also be interpreted as a model of *collusive price leadership* (Stigler, 1947; Markham, 1951; Scherer and Ross, 1990), in which an uninformed firm follows the informed firm's price changes.[2] We show that the dynamics of price leadership –which is the result of incomplete information and no communication– may involve significant costs for leader and follower. When demand increases, the informed firm raises its price, and experiences a short-term loss until its price raise is matched by the follower. Likewise, the follower experiences a short-term loss when the leader lowers its price after a demand reduction. These short-term losses are significant in many industries (see, for example, Clark and Houde, 2013) and our model provides a natural explanation for them.

These results extend the analysis of Bertrand games with incomplete information about marginal costs pioneered by Athey and Bagwell (2001, 2008). In Athey and Bagwell (2001), firms have iid private costs and, before choosing actions, can freely exchange messages. Athey and Bagwell (2008) extend the model to allow for Markovian private costs. In these papers, firms can be arbitrarily close to the first best collusive outcome, in which only the lowest cost firm produces and fixes the monopoly price. As Athey and Bagwell (2008) observe, communication can be dispensed with as prices can be used to signal costs at an arbitrarily low loss. But this observation crucially depends on the assumption of inelastic demand and constant returns to scale. Our results show that in more realistic Bertrand games, firms payoffs are bounded away from the perfectly collusive outcomes when the exchange of messages is costly even with arbitrarily high patience.[3]

[2] Collusive price leadership is relevant in many industries. Allen (1976), for example, documents collusive price leadership in the market of steam turbine generators in the 1960s and 1970s. In Section 5.2 we discuss additional empirical evidence.

[3] Athey et al. (2004) show conditions under which firms pool on the path of play –and therefore the cartel is bounded away from perfect collusion. But that result hinges on the restriction to strongly symmetric equilibria.

Our results reveal the constraints that lack of communication can impose in repeated interactions. In doing so, they provide the first tight characterization for the value of cheap-talk communication in repeated games. But our results can also be used to explore the value of communication in applications. We illustrate this point by studying the *social value* of communication in cartels in the context of a Cournot model with private costs. We show that communication reduces price distortions and therefore it is socially beneficial. Moreover, we show that consumers' surplus increases when cartel members communicate to coordinate production. This result confirms an informal argument made by Carlton et al. (1996) and complements Awaya and Krishna (2014) who show a strictly positive lower bound for the value of communication for the *cartel* in a repeated Bertrand game with private monitoring.

Our final application studies the decision to centralize decision making. We model the tradeoff between coordination and speed of adaptation (Roberts, 2004), by assuming that the informed agent can send a cheap-talk message communicating his type, but this message arrives to the uninformed agent with a two-period lag. Thus, centralization allows agents to coordinate better on private information, but leads to delays in decision making. Our results show that when direct communication has implicit costs, such as delays in decision making, agents may organize optimally to choose actions with limited communication.

Section 6 refines our analysis by studying a prisoners dilemma as interactions become more frequent. Following a tradition initiated by Abreu et al. (1991), we observe that as interactions become more frequent not only does the discount factor increase but also the the informed player's types become more persistent. Changing the persistence of the process of types has important effects on cooperation and equilibrium payoffs. As interactions become arbitrarily frequent, signaling becomes inexpensive compared to the benefits from more precise beliefs and, as a result, incomplete information has virtually no costs.

This paper connects to work on repeated games with Markovian private information. Athey and Bagwell (2008), Escobar and Toikka (2013), Renault et al. (2013), and Hörner et al. (2015) characterize optimal equilibria in games with communication. When players can exchange cheap-talk messages right before choosing actions, Escobar and Toikka (2013) and Hörner et al. (2015) show that the folk theorem holds. In these papers, actions have no signaling content and the dynamics of cooperation are similar to those of games with complete information and changing types if players are sufficiently patient (Rotemberg and Saloner, 1986; Dutta, 1995). We contribute to this literature by providing a new characterization for optimal equilibrium behavior in repeated games without communication. Further, our results identify new tradeoffs and inefficiencies in repeated games with incomplete information, and can be applied to a variety of economic examples.[4]

We finally observe that in games with imperfect public monitoring, players can also cycle between cooperative and uncooperative actions (Green and Porter, 1984; Abreu et al., 1986, 1990, 1991). Green and Porter (1984) and Abreu et al. (1986) study repeated games with quantity competition, and characterize equilibria with high and low price regimes. Transitions between regimes depend on the realization of an exogenous random factor affecting demand. In our adverse selection environment, in contrast, regime changes are triggered by players' actions. For example, a low-price regime (or price war) may be triggered by a price cuts, whereas a return to a high-price regime may require unilateral price rises. Abreu et al. (1991) studies a prisoners' dilemma with imperfect monitoring and shows that cooperation can be broken and never resumed in the optimal equilibrium. There is therefore room for renegotiating punishments. In our model, in contrast, virtually no value is burnt on the path of play and there is little room for renegotiation.[5]

2. Examples

In this section, we discuss two examples that illustrate some of the tradeoffs and inefficiencies arising in repeated games with Markovian private information.

2.1. A Coordination Game.

Two players, $i = 1, 2$, interact repeatedly in the coordination game in Figure 1.

	S	O
S	$1 + \alpha\theta^t, \beta$	$0, 0$
O	$0, 0$	$1 + \alpha(1 - \theta^t), \beta$

Figure 1. A repeated coordination game. $(\theta^t)_{t\geq 1}$ is a Markov chain observed only by player 1. The importance of coordination in the profile preferred by player 1 given θ^t is $\alpha > 0$. The importance of coordination for player 2 is $\beta > 0$.

[4] Other papers studying repeated games with Markovian types include Gale and Rosenthal (1994), Cole et al. (1995), and Phelan (2006). These papers focus on specific equilibria that are typically bounded away from the Pareto-frontier. Gensbittel and Renault (2015) and Pęski and Toikka (2016) characterize the value of zero-sum games with Markovian private information.

[5] Liu (2011) and Liu and Skrzypacz (2014) study games between a long-run player and a sequence of short-run players. The long-run player can be opportunistic or behavioral, and this is defined once and for all at the beginning of the game. Short-run players cannot freely access to the whole history of actions. This generates cycles of cooperation in which the long-run player builds and exploits his reputation. In those models, defaults are strategic while in our model defaults are mainly non-strategic. Acemoglu and Wolitzky (2014) study a reputation model in which players have limited and noisy observations. In all these models, memory restrictions play a key role determining cycles. The force in our model is unrelated to memory limits.

At each $t \geq 1$, θ^t is privately observed by player 1 and players simultaneously choose actions. Actions are perfectly observable. The support of θ^t is $\{0, 1\}$, $\mathrm{P}[\theta^{t+1} = \theta^t \mid \theta^t] = \lambda$, and θ^1 is drawn from the invariant distribution. We assume that $\lambda \geq 1/2$ so the Markov chain has positive persistence.

If θ^t was observed by both players at the beginning of t, then players could perfectly coordinate and play (O, O) when $\theta^t = 0$ and (S, S) when $\theta^t = 1$. This strategy profile would maximize the sum of expected total payoffs and would result in average total payoffs equal to $1 + \alpha + \beta$. Our focus is on games with incomplete information and no communication. This means that θ^t is observed only by player 1 and player 1 cannot tell the value of θ^t to player 2.

We now consider the private information case. Only for this example, we ignore incentive issues and focus on the informational value that pooling and separating strategies have.

Consider first a separating strategy profile in which player 1 fully reveals his type and player 2 mimics player 1's action in the previous period. In other words, player 1 plays S if $\theta^t = 1$ and plays O if $\theta^t = 0$. At $t + 1$, player 2 plays the action chosen by player 1 in period t. Conditional on θ^t, total payoffs in $t + 1$ equal $1 + \alpha + \beta$ with probability λ and 0 with probability $1 - \lambda$. The normalized sum of total discounted expected payoffs equals

$$(1 - \delta) \left(\frac{1 + \alpha + \beta}{2} + \sum_{t \geq 2} \delta^{t-1} \lambda \, (1 + \alpha + \beta) \right) = (1 - \delta) \, (1 + \alpha + \beta) \, \left(\frac{1}{2} + \lambda \, \frac{\delta}{1 - \delta} \right),$$

which converges to $\lambda \, (1 + \alpha + \beta)$ as $\delta \to 1$.

Alternatively, the informed player could pool his types and, for example, players could play (S, S) in each round. This means that player 2 always gets the payoff from coordination β, but player 1 receives $1 + \alpha$ when $\theta^t = 1$ and 1 when $\theta^t = 0$. The normalized sum of total discounted expected payoffs is $1 + \frac{1}{2}\alpha + \beta$.

The perfectly revealing strategy profile results in higher total payoffs than the pooling profile as players become patient iff $\lambda(1 + \alpha + \beta) > 1 + \frac{\alpha}{2} + \beta$. The revealing profile dominates when (i) λ is large (because the information generated by signaling lasts longer), or (ii) α is large (because the value of perfect coordination is high for player 1), or (iii) β is low (because otherwise player 2 values coordination and the only way to ensure such coordination occurs is by having player 1 pooling).

It is also worth noting that regardless of the strategy profile used, total expected payoffs are below the payoffs attained if information were complete: $\max\{\lambda(1+\alpha+\beta), 1+\frac{\alpha}{2}+\beta\} < 1+\alpha+\beta$. This is a general feature of our model and does not depend on the restriction on strategies used in this example. Intuitively, with incomplete information players will not be able to perfectly coordinate every round. With a separating profile, players will not coordinate a fraction $(1 - \lambda)$ of rounds (whenever the state changes), whereas with a pooling profile players will imperfectly coordinate attaining total payoffs $1 + \beta < 1 + \alpha + \beta$ half of the time. The cost of incomplete information does not vanish even as players become arbitrarily patient.

2.2. **A Prisoners Dilemma.** Two players, $i = 1, 2$, interact repeatedly in a public-good investment game. Every period, players decide whether to invest (I) or not to invest (N). Stage payoffs are equal to investment revenues minus cost. If both players invest, each player obtains a revenue of a. If only one player invests, each player obtains a revenue of b. If no player invests, both players obtain zero revenues. Let $0 < b < a$. Player 1's investment cost in period t is $\theta^t \in \{l, h\}$, where $l < h$, and player 2's investment cost is l every period. Figure 2 shows the payoff matrix.

	I	N
I	$a - \theta^t, a - l$	$b - \theta^t, b$
N	$b, b - l$	$0, 0$

Figure 2. A prisoners dilemma. Player 1's cost is privately known. Joint investment is socially desirable only when $\theta^t = l$.

Assume that $2(a - l) > 0, 2a - l - h < 0, 2b - l < 0$, and $a - l < b$. This means that playing N is a dominant action, that when the cost is low $\theta = l$ outcome (I, I) is socially desirable, whereas when the cost is high $\theta = h$ outcome (N, N) is socially desirable.

As in our previous example, at each $t \geq 1$, θ^t is privately observed by player 1 and players simultaneously choose actions. Players' actions are perfectly observable. The transitions are $\mathrm{P}[\theta^{t+1} = \theta^t \mid \theta^t] = \lambda$, and θ^1 is drawn from the invariant distribution. We assume that $\lambda \geq 1/2$ so the Markov chain has positive persistence.[6]

There are several strategies that could maximize the sum of total payoffs. Our main results imply that a revealing strategy profile σ^R in which player 1 invests iff $\theta^t = l$ and player 2 mimics player 1's previous action $a_2^t = a_1^{t-1}$ is optimal over all strategies when λ is sufficiently large and $a - b < h/2$, resulting in total average payoffs equal to $(2\lambda(a - l) - (l - 2b)(1 - \lambda))\frac{1}{2} > 0$ (details are given in Sections 4 and Section 6). The revealing strategy profile $\sigma^R = (\sigma_1^R, \sigma_2^R)$ can be formulated as

$$\sigma_1^R(\theta^t) = I \quad \text{iff} \quad \theta^t = l$$

[6] It is worth pointing out two benchmarks that are relatively easy to solve. With complete information, the type of player 1, θ^t, is publicly observed at the beginning of round t. If δ is large enough, we can construct a trigger-strategy equilibrium in which play is efficient and both players invest in t if and only if $\theta^t = l$ (Rotemberg and Saloner, 1986; Dutta, 1995). Another interesting benchmark is the case of incomplete information and communication, in which player 1 is privately informed about θ^t but can send a cheap-talk message to player 2 before actions are decided. If δ is sufficiently big, one can construct an efficient equilibrium in which player 1 truthfully reveals his type and both players invest only when $\theta^t = l$ (Escobar and Toikka, 2013).

and $\sigma_2^R(p^t) = I$ if $p^t = \lambda$ and $\sigma_2^R(p^t) = N$ if $p^t = 1 - \lambda$, where $p^t = \mathrm{P}[\theta^t = I \mid a_1^{t-1}]$ is the belief that player 2 has about θ^t after observing the action previously chosen by player 1.[7] Intuitively, the revealing strategy profile is optimal because, as in the coordination game, when the state is sufficiently persistent the relationship benefits from information revelation.

The issue of incentives is subtle. The revealing strategy profile σ^R maximizes the sum of total payoffs but whether private incentives can be aligned is non-trivial. On the one hand, player 1 should have some flexibility to choose actions and use his private information to benefit the relationship but, on the other hand, if player 1 is given full freedom to choose actions he will behave opportunistically with the purpose of maximizing his own payoffs. The problem that we face is how to balance these two forces.

Equilibrium strategies for the repeated game such that on-path play is arbitrarily close to the optimal strategy profile σ^R are constructed as follows. First, observe that ensuring player 2 behaves properly is simple as any deviation by 2 is observable and can be immediately punished by reverting to the static Nash equilibrium. Incentives for player 1 are given by noting that as play transpires, player 2 can keep checking whether player 1's behavior seems likely to have been generated from the revealing strategy σ_1^R. More precisely, note that under σ_1^R, the process of beliefs $(p^t)_{t\geq 1}$ is Markovian, with transitions that can be drawn as shown in Figure 3. By mechanically calculating probabilities using player 1's actions, the uninformed player 2 can check whether the proportions of investment and no-investment actions seem credible. For example, out of all the visits to $p^t = \lambda$, player 2 can check whether player 1 has played I in a proportion close to λ. A failure to do so would be observable and easily punished by Nash reversion.

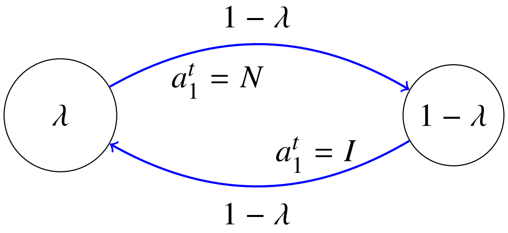


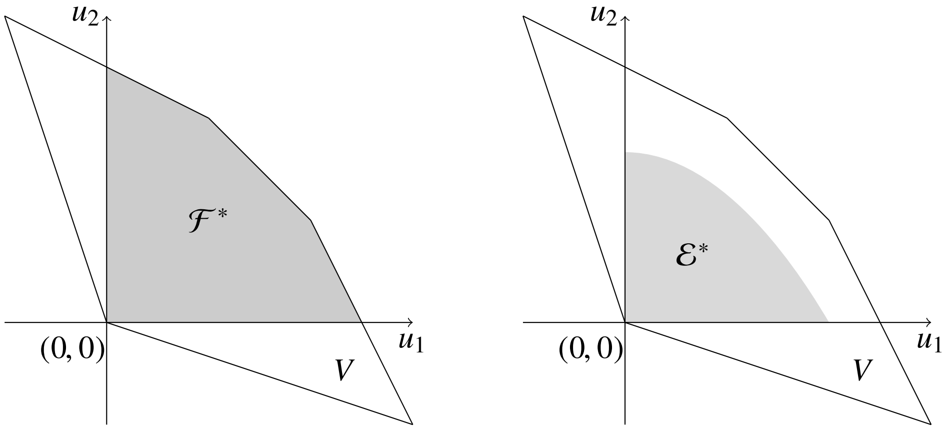
Figure 3. Dynamics of beliefs $(p^t)_{t\geq 1}$ when player 1 uses the revealing strategy σ_1^R. The support of $(p^t)_{t\geq 1}$ is the set $\{\lambda, 1 - \mu\}$.

The strategies discussed above continuously check whether player 1's actions seem credible. They are similar to strategies used in repeated games with imperfect monitoring (Radner, 1981) and in dynamic mechanism design (Jackson and Sonnenschein, 2007; Escobar and Toikka, 2013).[8] In our construction of strategies, while player 2 can tolerate some failures (i.e., periods

[7] Given the revealing strategy of player 1 σ_1^R, player 2 need not condition on the whole history of actions.

[8] As in all these papers, our strategies are derived from a test based on necessary conditions for "appropriate behavior". We then show that the necessary conditions are actually sufficient to align incentives.

in which player 2 invested but player 1 did not), he keeps track of the number of offenses, and players enter a punishment phase if that number becomes suspiciously high.[9] In other words, equilibrium strategies are so that player 2 forgives but does not forget failures.



(A) $\mathcal{F}^*$ is the set of limit equilibrium payoffs in the game with complete information or incomplete information and communication. It contains all feasible payoffs above the minmax vector $(0,0)$.

(B) $\mathcal{E}^*$ is the set of limit equilibrium payoffs in the game with incomplete information and no communication. It is strictly contained in $\mathcal{F}^*$. When signaling is too costly, $\mathcal{E}^* = \{(0,0)\}$.

Figure 4. Sets of equilibrium payoffs for games with and without communication.

Informational constraints are key to determine optimal equilibrium paths. While incentive problems disappear as players become more patient, equilibria are bounded away from first-best payoffs. Indeed, with incomplete information and communication (or with complete information), players can attain average total payoffs equal to $2(a-l)\frac{1}{2}$. Assuming the conditions under which revealing information is optimal, under incomplete information total average payoffs are $(2\lambda(a-l)-(l-2b)(1-\lambda))\frac{1}{2}$. Moreover, when the signaling costs are too high, the only equilibrium of the game is the repetition of the static Nash equilibrium even when the discount factor is arbitrarily close to 1.[10] While communication obviously expands the set of equilibria, we seem to be the first ones fully characterizing the gains from communication in a repeated game model.

[9] In this example, punishments simply consist in Nash reversion. In the general model of Section 3, punishments are more complex in order to guarantee that adhering to these punishments is incentive compatible for both players.

[10] As Hörner et al. (2015) show in their Corollary 3, the set of equilibrium payoffs in the game with communication depends on the transitions only through the invariant distribution. In contrast, in our model without communication transitions do matter to determine the equilibrium set. Another difference is that without communication the set of equilibrium payoffs is not a polytope.

3. Model

We consider a discrete-time infinitely repeated game played by 2 players. At each $t \geq 1$, player 1 is privately informed about his type $\theta^t \in \Theta$. Players simultaneously make decisions $a_i^t \in A_i$. Let $A = A_1 \times A_2$. We assume that A_1, A_2, and Θ are finite sets. Within each round t, play transpires as follows:

t.0 A randomization device χ^t is publicly realized
t.1 Player 1 is privately informed about $\theta^t \in \Theta$
t.2 Players choose actions $a_i^t \in A_i$ simultaneously
t.3 Players observe the action profile chosen $a^t \in A$

We assume players know their payoffs. The period payoff function for player 1 is $u_1(a, \theta)$, whereas player 2's payoff is $u_2(a)$. We will sometimes abuse notation and write $u_i(a, \theta)$, even when player 2's payoff does not depend on θ. Players rank flows of payoffs according to $(1 - \delta) \sum_{t \geq 1} \delta^{t-1} u_i(a^t, \theta^t)$, where $\delta < 1$ is the common discount factor. We assume that $|A_1| \geq |\Theta|$.[11]

The realizations of the randomization device are independent across time and distributed according to a uniform in $[0, 1]$. The initial type of player 1, θ^1, is drawn from a distribution $p^1 \in \Delta(\Theta)$. Player 1's private types, $(\theta^t)_{t \geq 1}$, evolve according to a Markov chain with transition matrix P on Θ. We assume that the process of types has full support: for all $\theta, \theta' \in \Theta$, $P(\theta' \mid \theta) > 0$. Let $\pi \in \Delta(\Theta)$ be the stationary distribution for P.

A (behavior) strategy for player 1 is a sequence of functions $s_1 = (s_1^t)_{t \geq 1}$ with $s_1^t \colon \Theta^t \times A^{t-1} \times [0, 1]^t \to \Delta(A_1)$, whereas a strategy for the uninformed player 2 is $s_2 = (s_2^t)_{t \geq 1}$ with $s_2^t \colon A^{t-1} \times [0, 1]^t \to \Delta(A_2)$. Any strategy profile $s = (s_1, s_2)$ induces a probability distribution over histories. We can therefore define the vector of expected payoffs given s as

$$v^\delta(s) = (1 - \delta)\mathbb{E}_s[\sum_{t \geq 1} \delta^{t-1} u(a^t, \theta^t)] \in \mathbb{R}^2.$$

Let

$$V(\delta, p^1) = \left\{ v = v^\delta(s) \in \mathbb{R}^2 \text{ for some strategy } s \right\}$$

be the set of all feasible payoffs that players can attain by employing arbitrary strategy profiles s. In passing, we note that $V(\delta, p^1) \subseteq \mathbb{R}^2$ is convex and compact.

Our definitions of strategies and set of feasible payoffs differ from those used stochastic games (Dutta, 1995; Hörner et al., 2010) and repeated games with incomplete information and communication (Escobar and Toikka, 2013; Hörner et al., 2015). The difference comes from the fact that in our model player 2 decides only based on publicly available information –the sequence of actions and public randomizations in the game.

[11] A one-sided incomplete information model is considered for expositional simplicity. We extend all our results to the two-sided incomplete information case in the Supplementary Appendix.

A strategy profile $s^* = (s_1^*, s_2^*)$ is a perfect Bayesian equilibrium if there exists a system of beliefs constructed from Bayes rule (when possible) such that s_i^* is sequentially rational (Fudenberg and Tirole, 1991). The set of perfect Bayesian equilibrium payoffs will be denoted $\mathcal{E}(\delta, p^1) \subseteq \mathbb{R}^2$. It follows that $\mathcal{E}(\delta, p^1) \subseteq V(\delta, p^1)$ for all $\delta < 1$.

4. Equilibrium Analysis

We will characterize equilibrium play in two steps. In the first step, we provide a dynamic programming formulation for efficient strategies ignoring incentive constraints. In the second step, we construct repeated game strategies that approximate the efficient benchmark. We finally show some general comparative statics results for the solutions to the dynamic programming problem.

4.1. Efficient Payoffs and Information Revelation.

This section analyzes the problem of maximizing the weighted sum of payoffs ignoring incentive constraints. This problem is formulated as a dynamic programming problem that identifies the tradeoff between revealing and not revealing information after any history.

A strategy profile s is *efficient* if for some $\alpha \in \mathbb{R}^2_{++}$, s is a solution to

$$q(\alpha) = \max\{\alpha \cdot v^\delta(s') \mid s' \text{ is a strategy profile }\}. \tag{4.1}$$

Let $s^{\alpha,\delta}$ solve (4.1). We say that $v^{\alpha,\delta} = v^\delta(s^{\alpha,\delta}) \in \mathbb{R}^2$ is an *efficient payoff vector*.[12]

Solutions to (4.1) can be found using dynamic programming tools. To see this, take the belief p^1 that player 2 has about player 1's type at the beginning of the game. The belief p^1 and the strategies used in the first round of play determine the sum of weighted payoffs in the first round. After player 1's action is observed, the strategies also determine the belief p^2 that player 2 has about the new type at the beginning of period 2. This means that the strategy profile that maximizes the weighted sum of period payoffs can be found by decomposing the discounted sum of weigthed payoffs in current and continuation payoffs using the public belief as a state variable.

To formulate the dynamic programing problem, we introduce some notation. Let $\Sigma_1 = \{\sigma_1 \colon \Theta \to A_1\}$ be a set of *controls* for player 1 and let $\Sigma = \Sigma_1 \times A_2$. An element $\sigma \in \Sigma$ is a *control profile*. Let $p \in \Delta(\Theta)$ be a belief about player 1's type given public information, and let $p(\theta)$ denote the θ-element of p. For $\sigma \in \Sigma$ and $p \in \Delta(\Theta)$, we define the vector of expected period utilities $U(\sigma, p) \in \mathbb{R}^2$ by

$$U_1(\sigma, p) = \sum_{\theta \in \Theta} u_1(\sigma_1(\theta), \sigma_2, \theta)\, p(\theta) \quad U_2(\sigma, p) = \sum_{\theta \in \Theta} u_2(\sigma_1(\theta), \sigma_2) p(\theta).$$

[12] Since any such $v^{\alpha,\delta}$ solves the problem $\max\{\alpha \cdot v \mid v \in V(\delta, p^1)\}$, the set of efficient payoff vectors v that maximize payoffs given a direction $\alpha \in \mathbb{R}^2_{++}$ is convex.

For $\alpha \in \mathbb{R}^2_{++}$, let $U^\alpha(\sigma, p) = \alpha \cdot U(\sigma, p) = \sum_{i=1}^2 \alpha_i \; U_i(\sigma, p)$ be the ex-ante weighted sum of period payoffs given $\sigma \in \Sigma$ and beliefs $p \in \Delta(\Theta)$. We also define the *Bayes operator* $B(\cdot \mid \sigma_1, p, a_1) \in \Delta(\Theta)$ as

$$B(\theta' \mid \sigma_1, p, a_1) = \sum_{\{\theta \mid \sigma_1(\theta) = a_1\}} P(\theta' \mid \theta) \, \frac{p(\theta)}{\sum_{\{\hat{\theta} \mid \sigma_1(\hat{\theta}) = a_1\}} p(\hat{\theta})} \tag{4.2}$$

whenever $\sigma_1(\hat{\theta}) = a_1$ for some $\hat{\theta}_1$ such that $p(\hat{\theta}) > 0$. In words, $B(\theta' \mid \sigma_1, p, a_1)$ is the probability that player 2 assigns to $\theta^{t+1} = \theta'$ given that at the beginning of round t his belief about θ^t was p, player 1 uses the control $\sigma_1 = \sigma_1(\theta^t)$, and player 2 observed player 1's action $a_1^t = a_1$.

For $\alpha \in \mathbb{R}^2_{++}$, consider the only solution to the Bellman equation

$$w^{\alpha,\delta}(p) = \max_{\sigma \in \Sigma} \left\{ (1-\delta) U^\alpha(\sigma, p) + \delta \sum_{a_1 \in A_1} w^{\alpha,\delta}\Big(B(\cdot \mid \sigma_1, p, a_1) \Big) \sum_{\theta \in \Theta, \sigma_1(\theta) = a_1} p(\theta) \right\} \tag{4.3}$$

for all $p \in \Delta(\Theta)$. The right hand side of this equation maximizes the weighted sum of current and continuation payoffs over all control profiles $\sigma \in \Sigma$, capturing the impact that a control has on current expected payoffs and continuation beliefs. Take $\sigma^{\alpha,\delta}(\cdot \mid p)$ as the control profile attaining the maximum in (4.3) as a function of beliefs p. A *control rule* σ is such that for all $p \in \Delta(\Theta)$, $\sigma(\cdot \mid p) \to \Sigma$. Using the control rule $\sigma^{\alpha,\delta}$, we can construct a (pure) strategy profile $s = s^{\alpha,\delta}$ from $\sigma^{\alpha,\delta}$ by setting

$$\begin{aligned} s_1^t(a^1, \ldots, a^{t-1}, \theta^1, \ldots, \theta^t, \chi^1, \ldots, \chi^t) &= \sigma_1^{\alpha,\delta}(\theta^t \mid p^t), \\ s_2^t(a^1, \ldots, a^{t-1}, \chi^1, \ldots, \chi^t) &= \sigma_2^{\alpha,\delta}(p^t), \end{aligned}$$

where p^t is the belief that player 2 has about θ^t at the beginning of t and can be recursively computed as

$$p^{t+1}(\theta) = B(\theta \mid \sigma_1^{\alpha,\delta}(\cdot \mid p^t), p^t, a_1^t) \text{ for } t \geq 1.$$

The following lemma shows that the dynamic programming formulation (4.3) provides a solution to the problem of finding efficient payoffs given weighs $\alpha \in \mathbb{R}^2_{++}$.

Lemma 1. *Let $\alpha \in \mathbb{R}^2_{++}$. Then, the value of the maximization problem (4.1) is $q(\alpha) = w^{\alpha,\delta}(p^1)$. Moreover, the strategy $s = s^{\alpha,\delta}$ constructed from $\sigma^{\alpha,\delta}$ above is a solution to (4.1).*

Like most of the literature in repeated games (Fudenberg and Maskin, 1986; Athey and Bagwell, 2008; Hörner et al., 2011), we characterize equilibrium behavior when players are sufficiently patient. It will be useful to consider efficient strategies and payoffs as $\delta \to 1$. We define the *differential discounted value* function as

$$h^{\alpha,\delta}(p) = \frac{w^{\alpha,\delta}(p)}{1-\delta} - \frac{w^{\alpha,\delta}(p^1)}{1-\delta} \tag{4.4}$$

for any $p \in \Delta(\Theta)$. Using this definition we can rewrite (4.3) as

$$h^{\alpha,\delta}(p) + w^{\alpha,\delta}(p^1) = \max_{\sigma\in\Sigma} \left\{ U^{\alpha}(\sigma,p) + \delta \sum_{a_1\in A_1} h^{\alpha,\delta}(B(\cdot \mid \sigma_1, p, a_1))\Big(\sum_{\theta\in\Theta, \sigma_1(\theta)=a} p(\theta)\Big) \right\} \quad (4.5)$$

Just to set ideas, assume that there exist subsequences $(h^{\alpha,\delta^\nu})_{\nu\geq 0}$, $(w^{\alpha,\delta^\nu})_{\nu\geq 0}$ and functions $h^\alpha : \Delta(\Theta) \to \mathbb{R}$, $w^\alpha : \Delta(\Theta) \to \mathbb{R}$ such that $h^\alpha(p) = \lim_{\nu\to\infty} h^{\alpha,\delta^\nu}(p)$ and $w^\alpha(p) = \lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p)$ for all p with $\delta^\nu \to 1$. Therefore, $\rho^\alpha = \lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p^1)$ does not depend on p^1.[13] Taking the limit in equation (4.5), we deduce that the pair $(h,\rho) = (h^\alpha, \rho^\alpha)$ solves the *average reward optimality equation* (AROE)

$$h(p) + \rho = \max_{\sigma\in\Sigma} \left\{ \alpha_1 U_1(\sigma,p) + \alpha_2 U_2(\sigma,p) + \sum_{a_1\in A_1} h(B(\cdot \mid \sigma_1, p, a_1))\Big(\sum_{\theta\in\Theta, \sigma_1(\theta)=a_1} p(\theta)\Big) \right\} \quad (4.6)$$

for all $p \in \Delta(\Theta)$. Let $\sigma^\alpha(\cdot \mid p) \in \Sigma$ be the control profile attaining the maximum in the dynamic programming problem (4.6) given $p \in \Delta(\Theta)$.

The following result establishes the key properties connecting the discounted and undiscounted dynamic programing problems.

Theorem 1 (Efficiency Theorem, Arapostathis et al. (1993))**.** *Fix $\alpha \in \mathbb{R}^2_{++}$. The following hold:*

a. *The AROE (4.6) has a solution (h^α, ρ^α) and a control rule σ^α that attains the optimum.*
b. *For any converging subsequence $h^{\alpha,\delta^\nu} \to \bar{h}$ as $\nu \to \infty$, we can take $\rho = \lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p^1)$ that does not depend on p^1, and obtain a pair $(\bar{h}, \rho)$ that solves the AROE (4.6). The function $\bar{h} \colon \Delta(\Theta) \to \mathbb{R}$ is convex.*
c. *For any strategy s, $\limsup_{\delta\to 1} \sum_{i=1}^2 \alpha_i v_i^\delta(s) \leq \lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p^1) = \rho^\alpha$.*

The first part of the Theorem ensures existence of solution. This is not obvious since (4.6) does not define a contraction map. The second part shows that such solution can be found by solving the Bellman equations as the discount factor goes to 1. The second part also establishes that $\bar{h}$ is a convex function, which means that continuation values improve when a compound lottery is resolved. The third part formally establishes that the solution $\rho \in \mathbb{R}$ to (4.6) provides a tight upper bound for the value of the discounted problem, as the discount factor goes to 1.

The AROE (4.6) is central to our analysis. The right-hand side of (4.6) captures the trade-off that an optimal control σ solves as a function of current beliefs $p \in \Delta(\Theta)$. As we show below, each of the three terms on the right-hand side of (4.6) is maximized either by a pooling or a separating rule.

A control rule σ_1 is *separating* if for any belief $p \in \Delta(\Theta)$ having positive probability in the path $(\theta^t, p^t)_{t\geq 1}$, types are separated: $\sigma_1(\theta \mid p) \neq \sigma_1(\theta' \mid p)$ for all $\theta \neq \theta'$. This means that player 1's types can be perfectly inferred after observing player 1's actions, given σ^α.

[13] To see this, note that for all $\epsilon > 0$, there exists $\bar{\nu} \in \mathbb{N}$ such that for all $\nu > \bar{\nu}$, $|w^{\alpha,\delta^\nu}(p) - w^{\alpha,\delta^\nu}(p^1) - (1-\delta^\nu)h^\alpha(p)(1-\delta)| < (1-\delta^\nu)\epsilon$. Taking the limit, it follows that $\lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p) = \lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p^1)$.

A separating control σ_1 allows player 1 to fully *reveal* his type by setting a different action for each state of the world. The problem of maximizing player 1's payoff $U_1(\sigma_1, \sigma_2, p)$ typically results in a fully revealing strategy σ_1. A second benefit of perfect information revelation is that by fully separating his types in period t, player 1 makes continuation beliefs p^{t+1} more precise and therefore player 2 faces less uncertainty about θ^{t+1} at the beginning of $t+1$. To see this, note that Theorem 1 (part b) shows that the limit differential discounted value $h(p)$ is convex in p. This means that given $p', q' \in \Delta(\Theta)$ and $\lambda \in [0,1]$, $h(\lambda p' + (1-\lambda)q') \leq \lambda h(p') + (1-\lambda)h(q')$. If player 1 uses a separating rule σ_1 in period t, he is fully resolving the uncertainty about θ^t at the end of round t and therefore maximizing $\sum_{a_1 \in A_1} h(B(\cdot \mid \tilde{\sigma}_1, p, a_1))\big(\sum_{\theta \in \Theta, \tilde{\sigma}_1(\theta)=a_1} p(\theta)\big)$ over all $\tilde{\sigma}_1 \in \Sigma_1$.

A pooling control σ_1 does not reveal information. The benefit of a pooling control is that it allows player 2 to perfectly predict player 1's current action. To see this, note that θ does not determine player 2's payoffs, and therefore the profile that maximizes player 2's expected payoff $\max_{\sigma \in \Sigma} U_2(\sigma, p)$ will typically involve a pooling rule σ_1.[14]

More generally, solutions to (4.6) will be determined by a complex mix of tradeoffs between revealing and not revealing information as time passes by.[15] The following result can be used to find those solutions in applications.

Proposition 1. *Consider a belief $p \in \Delta(\Theta)$ and a rule $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2)$ with $\bar{\sigma}_1 : \Theta \to A_1$ and $\bar{\sigma}_2 \in A_2$ such that for all $\theta \neq \theta'$, $\bar{\sigma}_1(\theta) \neq \bar{\sigma}_1(\theta')$ and*

$$\bar{\sigma} \in \arg\max_{\sigma \in \Sigma} U^{\alpha}(\sigma, p).$$

Then,

$$\bar{\sigma} \in \arg\max_{\sigma \in \Sigma} \Big\{ U^{\alpha}(\sigma, p) + \sum_{a_1 \in A_1} h(B(\cdot \mid \sigma_1, p, a_1))\big(\sum_{\theta \in \Theta, \sigma_1(\theta) = a_1} p(\theta)\big) \Big\}. \tag{4.7}$$

This proposition shows that if a rule that separates types maximizes current weighted payoffs, it also maximizes total undiscounted weighted payoffs. When current total payoffs are maximized by fully revealing, adding continuation payoffs can only reinforce the benefits from revelation.

4.2. **Equilibrium Strategies.** In this section, we investigate the conditions under which the efficient path characterized by (4.6) can be approximated by an equilibrium of the repeated game. We construct strategies in which player 1 losses credibility if his behavior does not match the efficient strategy profile. From an applied perspective, this implies that there exists an

[14] It is relatively simple to see that for an open set of full measure of payoffs for player 2, the maximization problem $\max_{\sigma \in \Sigma} U_2(\sigma, p)$ has a pooling solution σ_1. Moreover, this property also holds for all the examples presented in this paper.

[15] Problem (4.6) is similar to a bandit problem with Markovian hidden state (Keller and Rady, 1999). Separating rules maximize *exploration*. Propositions 1 and 2 show conditions under which the standard exploration vs exploitation dilemma (Bergemann and Valimaki, 2006) does not arise.

equilibrium path that is approximately equal to the path generated from the control rule σ^{α} that solves (4.6), provided players are patient enough.

A control rule σ together with the initial beliefs p^1 recursively determine a belief process $(p^t)_{t\geq 1}$ by

$$p^{t+1} = B(\cdot \mid \sigma, p^t, a_1^t) \quad \forall t \geq 1.$$

Given any control rule σ, the joint process $(\theta^t, p^t)_{t\geq 1}$ is Markovian, with p^1 and θ^1 given.

To construct equilibrium strategies for the repeated game, the main challenge is to align player 1's incentives. This is subtle because, on the one hand, we want to allow player 1 to use his private information but, on the other, allowing him to freely choose actions may open up the room for opportunistic behavior. However, player 2 can keep an account of the frequencies with which player 1 has played different actions and punish behaviors that seem, in a statistical sense, suspicious. To properly formulate how suspicious behaviors are identified, it will be useful to restrict attention to rules that generate well-behaved paths of beliefs.

Definition 1. *A control rule σ* determines a unique recurrence class *if the process $(\theta^t, p^t)_{t\geq 1}$ is a finite Markov chain having a unique recurrence class.*[16]

Note that when σ^{α} is separating, continuation beliefs come from the set $\{P(\cdot \mid \theta) \mid \theta \in \Theta\}$, the support of the process $(\theta^t, p^t)_{t\geq 1}$ is $\Theta \times (\{p^1\} \cup \{P(\cdot \mid \theta) \mid \theta \in \Theta\})$ and its unique recurrence class is $\Theta \times \{P(\cdot \mid \theta) \mid \theta \in \Theta\}$. A separating solution σ^{α} to (4.6) determines a unique recurrence class. On the other hand, when the rule pools all types along the path and the initial belief does not coincide with the stationary distribution of the transition matrix P, the path of the Markov chain $(\theta^t, p^t)_{t\geq 1}$ is countably infinite.[17] However, we will show that the restriction to efficient control rules that determine a unique recurrence class is virtually without loss.

The following result shows that relaxing the optimality requirement to allow for approximate efficiency is enough to ensure the existence of a control rule determining a unique recurrence class.

Lemma 2. *For all $\epsilon > 0$, and all $\alpha \in \mathbb{R}^2_{++}$, there exists a control rule σ, and $\bar{T} \in \mathbb{N}$ such that*

a. *σ determines a unique recurrence class; and*

b. *$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\sigma,p}[\alpha \cdot u(a^t, \theta^t)] \geq \rho^{\alpha} - \epsilon$ for al $T \geq \bar{T}$, and all p in the (finite) path of beliefs generated by σ and p_1. Moreover, when σ^{α} is a separating rule, we can take $\sigma = \sigma^{\alpha}$.*

Moreover, if σ^{α} is separating, we can take $\sigma = \sigma^{\alpha}$.

[16] In other words, a control rule determines a unique recurrence class if there exists a finite set $\mathcal{P} \subseteq \Delta(\Theta)$ such that $(\theta^t, p^t)_{t\geq 1} \subseteq \Theta \times \mathcal{P}$ and a unique subset $\mathcal{P}' \subseteq \mathcal{P}$ such that for all $(\theta, p) \in \Theta \times \mathcal{P}'$, if the Markov chain visits (θ, p), then in the next period it will stay in $\mathcal{P}'$ with probability 1, and no proper subset of $\mathcal{P}'$ has this property. See Stokey and Lucas (1989) for additional discussion.

[17] Exploring the ergodicity properties of $(\theta^t, p^t)_{t\geq 1}$ in hidden Markov models is a question dating back to Blackwell (1951). Recent developments Van Handel (2009); Tong and Van Handel (2012) do not apply to a model like ours in which the observation variable is endogenous .

When the control rule σ^α solving the AROE perfectly reveals types, this lemma is immediate and we can take $\sigma = \sigma^\alpha$. To intuitively understand this result, consider the prisoners dilemma in Section 2.2 and assume that the optimal rule is such that player 1 pools by playing I on the path of play.[18] This rule generates an infinite belief path. We can modify the rule so that after a sufficiently large number of periods, player 1's rule separates his types. This will, again, generate a new belief path that can be truncated after some time by changing the rule so that player 1's types are separated again. The modified rule determines a unique recurrence class and incurs an arbitrarily small loss in welfare.

For any control rule σ determining a unique recurrence class, the limit-average payoffs

$$v_1^\infty(\sigma) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} u_1(\sigma(\theta^t \mid p^t), \theta^t)], \quad v_2^\infty(\sigma) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} u_2(\sigma(\theta^t \mid p^t))]$$

are well defined. This follows from Proposition 8.1.1 in Puterman (2005) after noticing that the limits above are average rewards from a stationary Markov decision rule over a finite state Markov process.[19] We define $v^\infty(\sigma) = (v_i^\infty(\sigma))_{i=1,2}$.

Fix a control rule σ determining a unique recurrence class $\Theta \times \mathcal{P}$. Define $m_1^\sigma(\cdot \mid p) \in \Delta(A_1)$ as the distribution over actions given a belief $p \in \mathcal{P}$:

$$m_1^\sigma(a_1 \mid p) = \sum_{\{\theta\in\Theta \mid a_1 = \sigma_1(\theta\mid p)\}} p(\theta).$$

For $a \in A$ and $p \in \mathcal{P}$, we define $m^\sigma(a \mid p)$ analogously.

Given any sequence of actions $a_1^1, \dots, a_1^t$ and a fixed control rule σ determining a unique recurrence class, we can mechanically calculate probabilities $\bar{p}^{t+1} = B(\cdot \mid \sigma_1, \bar{p}^t, a_1^t)$ (if this is not well defined, we set $\bar{p}^{t+1}$ to be an arbitrary element of the support of the process of beliefs $(p^t)_{t\geq 1}$) with $\bar{p}^1 = p^1$. These *simulated probabilities* need not coincide with the beliefs a Bayesian agent would have about player 1's types as his actions in the game could be derived from an arbitrary strategy s_1. For a control rule σ determining a unique recurrence class with support $\Theta \times \mathcal{P}$ and given any sequence $(a^t, \theta^t, \bar{p}^t(\sigma))_{t\geq 1}$, for $a \in A$ and $p \in \mathcal{P}$, we can compute the occupancy rate of actions conditional on simulated probabilities as

$$\bar{m}^\delta(a \mid p) = \frac{\sum_{t=1}^\infty \delta^{t-1} \mathbb{1}_{\{a^t = a, \bar{p}^t = p\}}}{\sum_{t=1}^\infty \delta^{t-1} \mathbb{1}_{\{\bar{p}^t = p\}}}.$$

[18] The problem of ensuring appropriate behavior from player 1 when the optimal rule pools is simple. This example is used just to illustrate the lemma.

[19] Letting $\bar{\pi} = \bar{\pi}^\sigma \in \Delta(\Theta \times \mathcal{P})$ be the stationary distribution for the Markov chain $(\theta^t, p^t)_{t\geq 1}$, given the control rule σ, with $\Theta \times \mathcal{P}$ the recurrence class of the Markov chain, it follows that

$$v_1^\infty(\sigma) = \sum_{(\theta,p)\in\Theta\times\mathcal{P}} u_1(\sigma(\theta \mid p), \theta)\bar{\pi}(\theta, p) \quad \text{and} \quad v_2^\infty(\sigma) = \sum_{(\theta,p)\in\Theta\times\mathcal{P}} u_2(\sigma(\theta \mid p))\bar{\pi}(\theta, p).$$

We define the stationary minmax value as the smallest payoff a player can attain when his rival chooses a fixed action and he chooses actions optimally. More formally,

$$\underline{v}_1 = \min_{a_2 \in A_2} \mathbb{E}_\pi[\max_{a_1 \in A_1} u_1(a, \theta)], \quad \underline{v}_2 = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a).$$

This definition of minmax value does not yield the lowest payoff one could impose on a player (Escobar and Toikka, 2013; Hörner et al., 2015), but it is simple to work with and fully satisfactory in our applications.[20] A vector $v \in \mathbb{R}^2$ is *strictly individually rational* if $v_i > \underline{v}_i$ for $i = 1, 2$.

The following theorem shows that the efficiency analysis performed in Section 4.1 is useful to understand equilibrium behavior.

Theorem 2 (Equilibrium Theorem). *Fix $\epsilon > 0$. For $\alpha, \alpha^1, \alpha^2 \in \mathbb{R}^2_{++}$, take control rules σ, σ^1, and σ^2 as in Lemma 2. Assume*

i. *All payoff vectors $v = v^\infty(\sigma), v^1 \equiv v^\infty(\sigma^1), v^2 \equiv v^\infty(\sigma^2)$ are strictly individually rational;*
ii. *$v_i^i < v_i < v_i^{-i}$, for $i = 1, 2$.*

Then, there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, the infinitely repeated game with discount factor δ has a perfect Bayesian equilibrium $s^* = (s_1^*, s_2^*)$ such that

a. *$\alpha \cdot v^\delta(s^*) \geq \rho^\alpha - 2\epsilon$; and*
b. *$\mathbb{P}_{s^*}\left[\max_{a \in A, p \in \mathcal{P}} |\bar{m}^\delta(a \mid p) - m^\sigma(a \mid p)| < \epsilon\right] \geq 1 - \epsilon$, where $\Theta \times \mathcal{P} \subseteq \Theta \times \Delta(\Theta)$ is the recurrence class of the process $(\theta^t, p^t)_{t \geq 1}$ generated by σ.*

This result characterizes approximately optimal equilibrium behavior when players are sufficiently patient.[21] The result assumes that we have player specific punishmenrs $v^1, v^2 \in \mathbb{R}^2$ so that when players are patient enough the target optimal payoff vector can be approximated. The first part of Theorem 2 shows that players' incentives can be aligned to attain total weighted payoffs arbitrarily close to ρ^α. Moreover, with sufficiently high probability, conditional on simulated beliefs, players equilibrium actions will approximate the frequencies induced by the approximately optimal rule σ. This means that the problem of determining approximately optimal equilibrium dynamics reduces to solving the dynamic programing problem AROE (4.6).

The construction of equilibrium strategies combines forgiveness and memory. If player 1 plays and action resulting in low current payoffs for player 2, player 2 keeps playing according to the efficient control σ_2 given simulated beliefs. But if the number of such actions becomes suspiciously high (which happens off-path), a punishment phase against player 1 is triggered.

[20] Our definition of minmax is restrictive because it only considers pure strategies. Furthermore, when player 2 is minmaxing 1, he could find optimal to use the information revealed by player 1 during the minmaxing phase. This introduces complexities beyond the scope of the paper. See Pęski and Toikka (2016).

[21] In contrast to two-player repeated games with complete information, our result requires the existence of player-specific punishments (Fudenberg and Maskin, 1986). In our problem, types are hidden and for some types the minmaxing action could actually yield high payoffs to the minmaxed player.

The proof of Theorem 2 revisits the review strategy idea from Radner (1981) and Townsend (1982). The proof builds strategies in which player 2 keeps checking whether the path of player 1's actions can be distinguished from the control rule σ_1. At each round, player 2 builds simulated beliefs $\bar{p}^t$ and checks whether the path of actions played by 1 is close to the path of action if player 1 were using the control rule σ_1. If this is not the case, a punishment phase is triggered. The proof shows that it is always in the interest of player 1 to choose a path of actions which is close to the one generated from the efficient control rule σ_1.

To formalize the construction of strategies, take $(a^1, \ldots, a^t) \in A^t$, $(p, a_1) \in \mathcal{P} \times A_1$, and define

$$N^t(p) = \sum_{t'=1}^{t} \mathbb{1}_{\{\bar{p}^{t'}=p\}}, \quad N^t(p, a_1) = \sum_{t'=1}^{t} \mathbb{1}_{\{(\bar{p}^{t'}, a_1^{t'})=(p,a_1)\}}, \quad \bar{m}^t(a_1 \mid p) = \frac{N^t(p, a_1)}{N^t(p)}.$$

The number $\bar{m}^t(a_1 \mid p)$ is the empirical frequency of player 1's actions conditional on $\bar{p}^t = p$.

For any decreasing sequence (b_k) converging to 0, we say that player 1 *passes the test* (b_k) given a history $(a^1, \ldots, a^t) \in A^t$ if

$$\max_{a_1 \in A_1} |m_1^{\sigma}(a_1 \mid p) - \bar{m}_1^t(a_1 \mid p)| \leq b_t$$

for all $p \in \mathcal{P}$. Given $T \geq 1$, a control rule σ and a sequence (b_k), construct the *decision problem of credible play* $(\sigma, (b_k), T)$ for player 1 as follows. For $t \leq T$, if player 1 has passed the test (b_k) in all previous rounds $t' = 1, \ldots, t-1$, then he can freely select his action a_1^t; otherwise, a_1^t is an action randomly drawn from the distribution $m_1(\cdot \mid \bar{p}^t)$. Player 2 is always forced to chose an action that matches σ_2 given the history. We define the *obedient strategy* for player 1 as $\hat{s}_1^t(\theta^1, \ldots, \theta^t, a^1, \ldots, a^{t-1}) = \sigma_1(\theta^t \mid \bar{p}^t)$ whenever he is allowed to choose actions. We will also define the *block-decision problem of credible play* $(\sigma, (b_k), T)^\infty$ as the infinite horizon problem in which a decision problem of credible play restarts after T rounds of play (with discount factor δ).

Lemma 3. *Let $\eta > 0$.*

a. *There exists a test (b_k) such that, for any initial belief $p^1 \in \Delta(\Theta)$*

$$P_{\hat{s}_1}[\textit{Player 1 passes the test } (b_k) \textit{ at } (a^1, \ldots, a^t) \textit{ for all } t] \geq 1 - \eta.$$

b. *There exists a test (b_k) and $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$ there exists $\bar{T}$ such that for all $T \geq \bar{T}$, for any strategy s_1 of player 1 in the block-decision problem of credible play $(\sigma, (b_k), T)^\infty$ given discount δ,*

$$\mathbb{P}_{s_1}\Big[\max_{a_1 \in A_1, p \in \mathcal{P}} |\bar{m}^\delta(a_1 \mid p) - m^\sigma(a_1 \mid p)| < \eta\Big] \geq 1 - \eta.$$

The first part of the lemma ensures that player 1 can pass the test using the obedient strategy $\tilde{s}_1$. The second part ensures that regardless of the strategy used by player 1, the occupancy rate of actions is close enough to the distribution of actions drawn from σ_1 given simulated beliefs.

To establish Theorem 2, we use this lemma to construct strategies delivering the desired weighted equilibrium payoffs ρ^{α}. Strategies are of the stick-and-carrot type (Fudenberg and Maskin, 1986). On the path of play, players choose actions mimicking the path of play in the equilibrium of the block-decision problem of credible play from Lemma 3. Any observable deviation by i triggers a punishment phase, in which player i is minmaxed during a fixed number of rounds, and then play proceed to a carrot phase in which players mimic the play of the game of credible play yielding payoffs v^i. Further deviations trigger new punishment phases.

The construction of equilibrium strategies is related to the so called quota mechanisms in Jackson and Sonnenschein (2007), Renault et al. (2013), Renou and Tomala (2015), and particularly Escobar and Toikka (2013). One difference between our construction and all previous papers is that in our model players observe actions, not reports or cheap-talk messages. The path of actions need not be a Markov chain, even when players follow (stationary Markov) control rules and, as a result, the equilibrium strategies in the game cannot be formulated by simply testing the transition rates between consecutive actions. To overcome this difficulty, we summarize the history of actions by constructing simulated beliefs $(\bar{p}^t)_{t\geq 1}$ from a dynamic programming formulation, and test actions conditional on those beliefs.

A second, more technical, difference is that in our model the process of beliefs need not be a finite Markov chain, let alone an irreducible Markov chain. To overcome this difficulty, we need to approximate the belief path using rounds of revelation. To achieve this, Lemma 2 approximates the efficient control rule by one that induces a unique recurrence class of beliefs using rounds of revelation along the path of play at an arbitrarily small efficiency loss.

4.3. **Games with Monotonic Efficient Control Rules.** We now provide a characterization of solutions to (4.6). This characterization uses lattice theory tools to show that solutions to AROE (4.6) are strictly increasing in types (and therefore types are separated).

We assume that A_1 and Θ are contained in $\mathbb{R}$ and write $A_1 = \{a^n \mid n = 1, \ldots, |A_1|\}$ and $\Theta = \{\theta^m \mid m = 1, \ldots, |\Theta|\}$ with $a^n < a^{n+1}$ and $\theta^m < \theta^{m+1}$. We extend the payoff function for player 1, u_1, to actions $a_1 \in \mathbb{R}$ and states $\theta \in \Theta$ so that $u_1(a_1, a_2, \theta)$ is twice continuously differentiable in $(a_1, \theta) \in \mathbb{R} \times \mathbb{R}$.

Definition 2. *We will say that u_1 has* strongly increasing differences *in (a_1, θ) if*

$$\min\Big\{\frac{\partial^2 u_1(a_1, a_2, \theta)}{\partial a_1 \partial \theta} \mid a_1 \in \mathbb{R}, a_2 \in A_2, \theta \in \mathbb{R}\Big\} > 0.$$

Proposition 2 shows conditions under which the optimal control rule is strictly increasing. Since actions are discrete, this property cannot be inferred by simply appealing to strong increasing differences. There are two forces behind this result. Separating rules (in particular, strictly increasing rules) make continuation beliefs more precise and therefore maximize continuation payoffs (Proposition 1). This effect is reinforced when the action set is rich because in this case the maximization of total period payoffs yield strictly increasing rules.

Proposition 2. *Assume that u_1 has strongly increasing differences in (a_1, θ). Let $\alpha \in \mathbb{R}^2_{++}$ be such that*

$$\alpha_1 u_1(a^{|A_1|-1}, a_2, \theta) + \alpha_2 u_2(a^{|A_1|-1}, a_2) > \alpha_1 u_1(a^{|A_1|}, a_2, \theta) + \alpha_2 u_2(a^{|A_1|}, a_2) \tag{4.8}$$

and

$$\alpha_1 u_1(a^1, a_2, \theta) + \alpha_2 u_2(a^1, a_2) < \alpha_1 u_1(a^2, a_2, \theta) + \alpha_2 u_2(a^2, a_2) \tag{4.9}$$

for all $a_2 \in A_2$ and all $\theta \in \Theta$ and $\alpha_1 u_1(a_1, a_2, \theta) + \alpha_2 u_2(a_1, a_2)$ is concave in $a_1 \in \mathbb{R}$. Define

$$c_1 = \max_{a_1 \in \mathbb{R}, a_2 \in A_2, \theta \in \mathbb{R}} \Big(-\alpha_1 \frac{\partial^2 u_1(a_1, a_2, \theta)}{\partial a_1^2} - \alpha_2 \frac{\partial^2 u_2(a_1, a_2)}{\partial a_1^2} \Big) \geq 0$$

and

$$c_2 = \min_{a_1 \in \mathbb{R}, a_2 \in A_2, \theta \in \mathbb{R}} \frac{\partial^2 u_1(a_1, a_2, \theta)}{\partial a_1 \partial \theta} > 0.$$

Assume that

$$\frac{2c_1}{\alpha_1 c_2} \max_{n=1,\ldots,|A_1|-1} \{a^{n+1} - a^n\} < \min_{m=1,\ldots,|\Theta|-1} \{\theta^{m+1} - \theta^m\}. \tag{4.10}$$

Then, any rule σ^α attaining the maximum in (4.6) is such that $\sigma^\alpha_1(\theta \mid p)$ is strictly increasing as a function of θ for all $p \in \Delta(\Theta)$ with $p(\theta) > 0$ for all $\theta \in \Theta$. Moreover, endowing $\Delta(\Theta)$ with the (partial) order $\geq_{\Delta(\Theta)}$ given by first-order stochastic dominance, and assuming that $P(\cdot \mid \theta') \geq_{\Delta(\Theta)} P(\cdot \mid \theta)$ for all $\theta' \geq \theta$, and $u_1(a, \theta)$ and $u_2(a)$ are supermodular (in (a, θ) and a respectively), then $\sigma^\alpha(\theta \mid p)$ is nondecreasing in (θ, p).

Equations (4.8)-(4.9) ensure that the optimal rule is not in the boundary. Provided the set of actions is rich enough, as imposed in (4.10), it follows that the optimal rule always separates types. Observe that the separating rule σ^α determines a unique ergodic class.

5. Applications

This section presents applications of our results and methods.

5.1. Live and Let Live.

In this section, we explore the issue of implicit cooperation between enemy combatants in the Western Front in World War I (Ashworth, 1980; Axelrod, 1984). In the Western Front, armies adopted mostly static positions along a trench line of 475 miles which ranged from the North Sea to the Swiss Alps. Trench warfare was different from traditional war in that "the same small units faced each other in immobile sectors for extended periods of time" (Axelrod, 1984, p. 77). Repeated interaction between enemy battalions allowed enemy soldiers to engage in cooperative attitudes and to limit the level of aggressions. Such behavior was known as live and let live.

Army commanders understood the potential for cooperation and tried to limit it by ordering raids and attacks on enemy trenches.[22] Enemy soldiers could not discern if such attacks were caused by military orders from high command or by opportunistic behavior.[23] Moreover, direct communication was difficult, if not impossible. As Ashworth (1980, p. 38) explains, "although verbally arranged truces occurred intermittently for the duration of the war [. . .] they were neither pervasive nor continuous." On the contrary, "such truces were mostly irregular and ephemeral, since being highly visible they were easily repressed by high command." For example, a British Divisional Commander issued a memo in 1917 stating that "any understanding with the enemy [. . .] is strictly forbidden [. . .] In the event of any infringement disciplinary action is to be taken" (Ashworth, 1980, p. 37). Yet, cooperation was prevalent and battalions were successful at maintaining low levels of aggression for significant lengths of time.

As this discussion suggests, cooperation between battalions arose under severe information asymmetries. We apply our general insights and results to shed light on this issue.[24] We consider a repeated game between two battalions. At each $t = 1, 2, \ldots,$ battalions 1 and 2 simultaneously decide S or NS (shoot or not). Battalion 1's private information is whether its army commanders have shown up or not and is represented by $\theta^t \in \{0, 1\}$, where $\theta^t = 0$ means that army commanders are absent. Payoffs are represented in Figure 5 .

	NS	S
NS	$R - \theta^t K,\ R$	$-C - \theta^t K,\ G$
S	$G,\ -C$	$0,\ 0$

Figure 5. A game between battalions.

We assume $C > G > R > 0$. These inequalities imply that when $\theta^t = 0$, playing S is a dominant action, but that the outcome (NS, NS) is socially desirable. In other words, when $\theta^t = 0$, the interaction between battalions is a prisoners dilemma.[25] The term $-\theta^t K$ captures the cost that

[22] In the British Army, for example, the lack of aggression was "both contrary to the spirit of the offensive, [...] and to an official British directive of 1915 which made active trench war mandatory," and a British training manual of 1916 stated that "the fostering of the offensive spirit [...] calls for incessant attention" (Ashworth, 1980, pp. 42-43).

[23] By attacks caused by opportunistic behavior we mean attacks that are not caused by army commanders orders but by the desire to have a short-run gain by inflicting losses on the enemy.

[24] Studying this well-documented example is interesting because it also yield insights about other episodes of limited war, such as the Korean War (Gorman, 1953) and the Cold War (Schelling, 1960).

[25] As Axelrod (1984) points out, "At any time, the choices are to shoot to kill or deliberately to shoot to avoid causing damage. For both sides, weakening the enemy is an important value because it will promote survival if a major battle is ordered in the sector. Therefore, in the short run it is better to do damage now whether the enemy is shooting back or not. This establishes that mutual defection is preferred to unilateral restraint [. . .], and that unilateral restraint by the other side is even better than mutual cooperation [. . .]. In addition, the reward for mutual restraint is preferred by the local units to the outcome of mutual punishment [. . .], since mutual punishment would imply that both units would suffer for little or no relative gain."

battalion 1 must pay if army commanders showed up and ordered raids ($\theta^t = 1$), but the battalion does not shoot. We assume that $2R - K < 0$ so that when $\theta^t = 1$, the outcome (S, S) maximizes the sum of stage payoffs.

Battalion 1's type evolves according to a Markov process with transition probabilities given by

$$\begin{aligned} \mathrm{P}[\theta^t = 0 \mid \theta^{t-1} = 0] &= \lambda, \\ \mathrm{P}[\theta^t = 1 \mid \theta^{t-1} = 1] &= \mu, \end{aligned}$$

where $\lambda + \mu \geq 1$. This means that the process of types has positive persistence. For simplicity, we assume that the initial type is drawn according to $\mathrm{P}[\theta^1 = 0] = \lambda$. Type θ^t is realized at the beginning of period t and is privately observed by player 1. Once player 1 observes his type, players simultaneously choose actions. Actions are publicly and perfectly observed. Players have a common discount factor $\delta < 1$ and their utility equals the discounted sum of period payoffs.

We focus on equilibrium strategies that maximize the sum of total payoffs. To do this, we first solve the AROE (4.6). The differential discounted function h maps distributions over $\{0, 1\}$ to real numbers. We simplify notation by keeping track of a single number $p \in [0, 1]$ representing the probability that $\theta = 0$ given public information. Thus, $h\colon [0, 1] \to \mathbb{R}$ is a convex function. Fixing p, the optimization problem on the right hand side of AROE (4.6) is defined over all controls $(\sigma_1(0), \sigma_1(1), \sigma_2) \in \{S, NS\}^3$. It is relatively simple to show that controls (NS, NS, S), (S, S, NS), (S, NS, NS), and (S, NS, S) are not optimal.[26] When $2R - K < G - C$ we can also rule out the control (NS, NS, NS). Indeed, the right hand side of AROE (4.6) at control (NS, NS, NS) equals

$$p(2R) + (1 - p)(2R - K) + h(p\lambda + (1 - p)\mu).$$

Evaluating the right hand side of (4.6) at (NS, S, NS) results in

$$p(2R) + (1 - p)(G - C) + ph(\lambda) + (1 - p)h(\mu).$$

Since h is convex, $ph(\lambda) + (1 - p)h(\mu) \geq h(p\lambda + (1 - p)\mu)$ and therefore control (NS, NS, NS) is not optimal. In the sequel, we rule out the control (NS, NS, NS) by assuming $2R - K < G - C$.

Lemma 4 characterizes optimal dynamics. We say that a control rule $\sigma\colon [0, 1] \to \{S, NS\}^3$ generates *reactive-signaling* dynamics if on the path, battalion 1 does not shoot when its type is $\theta^t = 0$ and shoots when its type is $\theta^t = 1$, whereas battalion 2 imitates the action of battalion 1 in the previous period. Thus, battalion 1 *signals* its private information through its actions, and battalion 2 *reacts* to such information. Given $\widehat{\tau} \in \{0, 1, 2, \dots\} \cup \{\infty\}$, we say that a control rule generates *time-off* dynamics if, on the path, battalion 1 does not shoot only if it is in good

[26] For example, control (NS, NS, S) gives less total period payoffs than (S, S, S). Since both controls determine the same distribution over continuation beliefs, control (NS, NS, S) cannot be optimal.

standing and its type is $\theta^t = 0$, and battalion 2 does not shoot if and only if battalion 1 is in good standing. Battalion 1 is in good standing if it did not shoot in the previous period, or if it shot in the previous period, but it was in good standing $\widehat{\tau} + 1$ periods before. Thus, a time-off control rule leads to a waiting phase of $\widehat{\tau}$ periods after an aggression by battalion 1.[27]

Let σ^* be the control rule solving AROE (4.6). The following result characterizes the optimal path.

Lemma 4. *If $\lambda < \frac{C-G}{2R+C-G}$, σ^* has both battalions playing S on the path of play. If $\lambda > \frac{C-G}{2R+C-G}$, σ^* generates either reactive-signaling or time-off dynamics (potentially, with $\widehat{\tau} = 0$ or $= \infty$).*

The restriction $\lambda > \frac{C-G}{2R+C-G}$ implies that control (NS, S, NS) is optimal at belief $p = \lambda$. If battalion 1 plays NS at $p = \lambda$ then in the next period the belief is λ and the optimal control continues to be (NS, S, NS). If battalion 1 plays S instead, it 'signals' a change in type, and in the next period the optimal control is either (NS, S, S) (in which case σ^* generates reactive-signaling dynamics) or (S, S, S) (in which case σ^* generates time-off dynamics).[28]

Regardless of the specific form that the solution to AROE (4.6) assumes, whenever both battalions have strictly positive limit-average payoffs, we can use Theorem 2 to deduce that such path can be an equilibrium outcome for the repeated game model when players are patient enough. Moreover, in this case, we can simply use the repetition of the static Nash equilibrium to punish observable defections.

The analysis of this repeated game model yields new insights about cooperation between battalions. First, alternating between periods of aggressions and periods of non-aggressions can be optimal for the battalions. These dynamics are consistent with those observed in the Western Front, where "many sectors were a mixture of war and peace, that is, of exchanges of peace as well as exchanges of aggression and these were more frequent than either very quiet or very active sectors" (Ashworth, p. 39).

Second, consistent with our equilibrium construction, soldiers under the live and let live system kept an account of the number of aggressions received from the other side. As Ashworth (1980) observes, "combatants generally had a good idea of what was, or was not, compatible with live and let live, and if one side deviated the other meted out punishments by returning to officially prescribed levels of aggression." Moreover, Ashworth (1980) notes that the rules "were not broken by the arrival of four to twelve grenades, which were regarded as routine, but if twelve were exceeded, 'the chances were', retaliation followed." This suggests that soldiers could have deemed sufficiently low numbers of aggressions as tolerable, which is similar to the combination of forgiveness and memory in the equilibrium strategies discussed after Theorem 2.

[27] Note that reactive signaling is *not* a particular case of time-off. A time-off control rule with $\widehat{\tau} = 0$ implies that battalion 1 always signals its type, but battalion 2 keeps playing NS.

[28] Lemma 4 rules out dynamics in which signaling can occur only after an exogenous number of rounds has transpired.

5.2. **Price Cuts and Price Leadership.** In this section, we study a model of tacit collusion with Bertrand competition, and show that price cuts and price leadership naturally arise in an equilibrium of the model.

Two firms set prices $a_i \in A_i$ at each $t = 1, 2 \ldots$. Firms sell heterogeneous goods. The demand functions are given by

$$Q_1(a_1, a_2, \theta) = \theta - a_1 + z a_2, \quad Q_2(a_1, a_2) = 1 - a_2 + z a_1$$

with $0 < z < 1$. We normalize marginal costs to 0. Firm 1's demand shock is private information $\theta \in \{\underline{\theta}, \bar{\theta}\}$, with $\underline{\theta} < \bar{\theta}$. Players' utility functions equal revenues and take the form

$$\begin{aligned} u_1(a_1, a_2, \theta) &= Q_1(a_1, a_2, \theta)\, a_1, \\ u_2(a_1, a_2) &= Q_2(a_1, a_2) a_2. \end{aligned}$$

We assume that types follow a Markov chain P with $P(\theta' \mid \theta) > 0$ for all $\theta', \theta \in \{\underline{\theta}, \bar{\theta}\}$, with $P(\bar{\theta} \mid \bar{\theta}) \geq P(\bar{\theta} \mid \underline{\theta})$.[29]

We can apply Proposition 2 to characterize the welfare maximizing control rule σ^α, for $\alpha = (1, 1)$. Up to integer restrictions,

$$\sigma_2^\alpha(p) = \frac{1 + z\mathbb{E}_p[\theta]}{2(1 - z^2)}, \quad \sigma_1^\alpha(\theta \mid p) = \frac{\theta}{2} + z\sigma_2^\alpha(p).$$

Under the optimal control rule σ^α, firm 1 signals its type by choosing a higher price when its demand is high. When firm 1 chooses a high price in period t, then its demand is more likely to be high in period $t + 1$ and player 2's price is also higher. In this sense, a low price by firm 1 in t acts both as a signal of privately-observed demand conditions and as an invitation from firm 1 to firm 2 to switch to a low-price regime in $t + 1$. Likewise, a price increase by firm 1 in t is an invitation to switch to a high-price regime in $t + 1$. This mechanism matches the one described by Judge Posner in his decision on the High Fructose Corn Syrup case:

> *If a firm raises price in the expectation that its competitors will do likewise, and they do, the firm's behavior can be conceptualized as the offer of a unilateral contract that the offerees accept by raising their prices.*

Figure 6 illustrates a sample path of private costs and prices. Observe that σ^α is a rule determining a unique recurrence class and therefore Theorem 2 immediately applies.[30]

In contrast to other theoretical papers, such as Green and Porter (1984) and Abreu et al. (1986), in our setup unilateral price cuts actually *occur* and *are observed* in equilibrium, and apparent deviations can be seen as the result of firms using their private information to maximize total

[29] In the Supplementary Appendix, we explore a model in which both firms may have private information. In that model, both firms may become price leaders.

[30] Note that in this model, the repetition of the static equilibrium resulting in payoffs $(0, 0)$ can be used as a punishment.

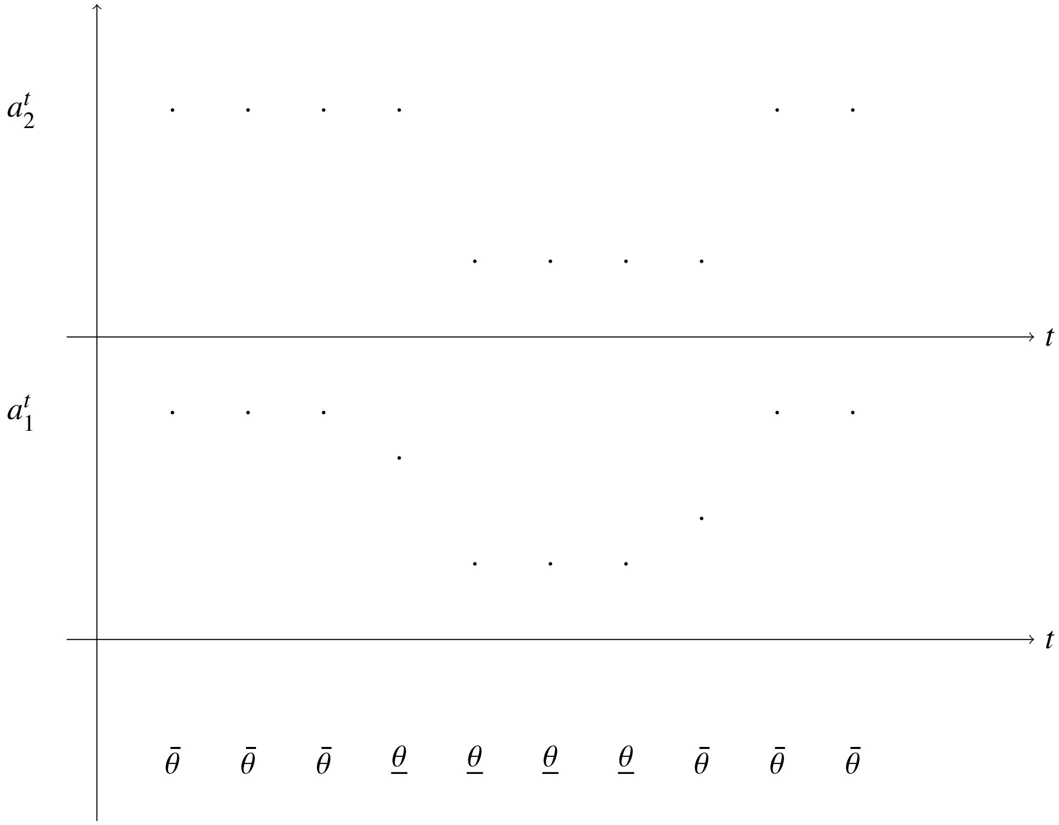


Figure 6. A sample path of private costs and prices.

profits and signal their continuation play. Rahman (2014) provides a complementary view in a repeated game model with imperfect monitoring. In such model, price cuts can be used to improve monitoring.[31]

Our model shows that price leadership may be an optimal strategy for firms when direct communication is not feasible. The informed firm becomes a price leader as whenever it raises its price in t, firm 2's price will be higher in $t + 1$. On the contrary, when firm 1 lowers its price in t, firm 2's price will be lower in $t + 1$. Thus, our model gives theoretical support to Stigler's (1947) observation that price leadership may be an efficient mechanism to transmit information, and to Markham's (1951) view that firms may use *price leadership in lieu of an overt agreement.* In our model, price leadership is an imperfect substitute to explicit communication. Indeed, if firms could freely communicate, firm 1 would send messages to firm 2 to coordinate their pricing decisions and firms would *simultaneously* raise or lower their prices. But without communication, firm 1's price is used as a signal of market conditions. Using firm 1's price as a substitute to communication entails a cost: Pricing decisions are uncoordinated and lack

[31] Collusion and price cuts can also arise in a mixed strategy equilibrium of a repeated Bertrand game (Bernheim and Madsen, 2014).

of communication does not allow the cartel to adjust prices to optimally assign demand after market conditions change.

Collusive price leadership has been extensively supported empirically.[32] Our model provides an explanation for price leadership in a natural repeated Bertrand game with incomplete information. Rotemberg and Saloner (1990) also study collusion and price leadership in a Bertrand model with incomplete information. Their model exhibits iid private information and for price leadership to emerge, within each round the informed firm must set its price before the uninformed one. Such sequentiality is not needed in our model. Furthermore, in Rotemberg and Saloner's (1990) model, price leadership entails no cost for the cartel as, within each round, production takes place after both firms has set prices. Empirical evidence supports the observation that unilateral price increases are costly for the cartel. For example, Clark and Houde (2013) study price leadership in gasoline markets in Quebec, and find that a small price premium for a few hours can result in a significant reduction in a station's sales for the day (up to 50%).

Our collusion model differs from the more standard analysis of Bertrand games with inelastic demand and incomplete information about costs. In Athey and Bagwell (2001), firms have iid private costs and, before choosing actions, can freely exchange messages. Athey and Bagwell (2008) and Escobar and Toikka (2013) extend the model to allow for Markovian private costs.[33] In all these works, firms can be arbitrarily close to the first best collusive outcome, in which only the lowest cost firm produces and fixed the consumers' reservation value. As Athey and Bagwell (2001) observe, communication can be dispensed with as prices can be used to signal costs (at an arbitrarily low cost) when firms are sufficiently patient. But this observation crucially depends on the assumption of inelastic demand. Our analysis shows that in more general Bertrand games, firms are bounded away from a perfectly collusive outcome when the exchange of messages is costly. Moreover, in the Bertrand models of Athey and Bagwell (2001, 2008) and Escobar and Toikka (2013), the path of collusive prices cannot be distinguished from the prices one would observe when firms' information is symmetric and players were patient (as in Rotemberg and Saloner, 1986). In contrast, our analysis not only shows that the costs of incomplete information can be substantive for a cartel, but also that asymmetric information has nontrivial implications for the dynamics of prices.[34]

[32] Nicholls (1951) describes price leadership in the cigarette industry; Stigler (1947) discusses price leadership in the steel, dynamite, anthracite, and airline industries; Allen (1976) shows evidence of collusive price leadership in the market of steam turbine generators; Mouraviev and Rey (2011) show that price leadership features in 16 out of 49 European Commission's cartel decisions as of July 2010; and Seaton and Waterson (2013) characterize price leadership in British supermarkets.

[33] Athey and Bagwell (2008) additionally study a model with perfectly persistent costs and prove that in the optimal equilibrium firms pool by fixing the monopoly price. Pęski (2014) shows that the pooling result does not survive to more general demand functions.

[34] Athey et al. (2004) study a repeated Bertrand game with iid cost and show that optimal equilibrium is in (on-path) pooling strategies when firms are restricted to use strongly symmetric strategies.

5.3. **The Social Value of Communication in Cartels.** Communication between cartel members can serve several roles. One role that communication has is to allow cartel members to better coordinate production. From a legal perspective, communication to share information about market conditions is typically seen as welfare enhancing (Carlton et al., 1996). Here, we confirm this intuition. We show that consumer surplus increases when firms communicate and therefore communication between cartel members has a pro-competitive effect.

Two firms set quantities $a_i \in A_i$ at each $t = 1, 2, \ldots$. Firms sell homogeneous products and the (inverse) demand is given by $\mathcal{P}(a_1 + a_2)$, where $\mathcal{P} > 0$ and it is strictly decreasing in $a_1 + a_2$. The marginal cost of firm 1 is $\theta \in \Theta$, whereas the marginal cost of firm 2 is $c > 0$. Firms's utility functions are

$$u_1(a_1, a_2, \theta) = \mathcal{P}(a_1 + a_2)a_1 - \theta a_1,$$
$$u_2(a_1, a_2) = \mathcal{P}(a_1 + a_2)a_2 - ca_2.$$

We assume that types follow a Markov chain with transitions $P(\theta' \mid \theta)$ for all $\theta, \theta' \in \Theta$. To simplify the analysis, we assume that $A_1 = A_2$ and $A_i = \{0, g, 2g, \ldots, (|A_i| - 1)g\}$, where $g > 0$. We define the monopoly quantity given any cost $\kappa \geq 0$ as

$$Q^M(\kappa) = \max_{q \in \{0, g, \ldots\}} \mathcal{P}(q)q - \kappa q.$$

Note that $Q^M(\kappa)$ decreases in κ. Assuming a sufficiently rich grid, we can assume that $Q^M(\theta)$ is strictly decreasing. Finally, we assume that the set of actions A_i is such that $Q^M(0) < \max\{a_i \mid a_i \in A_i\}$. We assume that no firm is always the most efficient one: $\min\{\theta \in \Theta\} < c < \max\{\theta \in \Theta\}$.

We focus on profiles that maximize the sum of firms' payoffs. If firms could communicate, only the firm having the lowest cost would produce the monopoly quantity $Q^M(\min\{\theta, c\})$ and total payoffs would be $\max_{q \in \{0, g, \ldots\}} \left\{\mathcal{P}(q)q - \min\{c, \theta\}q\right\}$. Theorem 4.1 in Escobar and Toikka (2013) implies that firms can attain monopoly profits on the path of play in the repeated game with communication.

When firms cannot communicate, the monopoly arrangement is not feasible. To characterize and approximately optimal path, assume that the belief that firm 2 has about θ is $p \in \Delta(\Theta)$ and consider the problem of maximizing the expected sum of firms' payoffs over all feasible rules:

$$\max_{\sigma_1 : \Theta \to A_1, \sigma_2 \in A_2} U^{(1,1)}(\sigma, p) := \sum_{\theta \in \Theta} \Big(\mathcal{P}(\sigma_1(\theta) + \sigma_2)(\sigma_1(\theta) + \sigma_2) - \theta\sigma_1(\theta) - c\sigma_2\Big)p(\theta) \quad (5.1)$$

Consider first the case $\mathbb{E}_p[\theta] := \sum_\theta \theta p(\theta) > c$. Then, for any solution of (5.1), $\sigma_2 = 0$. If not, $\sigma_2 > 0$. Take the alternative profile $\tilde{\sigma}_2 = \sigma_2 - g$ and $\tilde{\sigma}_1(\theta) = \sigma_1(\theta) + g$.[35] The difference in

[35] Note that $\sigma_1(\theta) \leq Q^M(0)$ and thus $\tilde{\sigma}_1(\theta) \in A_i$.

total expected payoffs would be

$$U^{(1,1)}(\tilde{\sigma}, p) - U^{(1,1)}(\sigma, p) = g\Big(\mathbb{E}_p[\theta] - c\Big) > 0.$$

Thus $\sigma_2 > 0$ cannot be optimal. It follows that the optimal solution is $\sigma_1(\theta) = Q^M(\theta)$ and $\sigma_2 = 0$ and total profits equal $(\mathcal{P}(Q^M(\theta)) - \theta)Q^M(\theta)$. In other words, firm 1 ends up producing even when it is less efficient than firm 2. Since σ is a separating rule, Proposition 1 implies that it solves the AROE given beliefs p. Intuitively, the cartel must decide production under uncertainty and let the firm that is more efficient on average to produce the monopoly quantity. Assuming that for all $p \in \{p^1\} \cup_{\theta \in \Theta} \{P(\cdot \mid \theta)\}$, $\mathbb{E}_p[\theta] < c$, the optimal control σ can actually be implemented as an equilibrium of the repeated game using Theorem 2.[36]

This analysis shows that the cartel gets lower payoffs when communication is not allowed. Perhaps surprisingly, consumers are also hurt by the lack of communication. To see this, note that with communication the quantity produced is $Q^M(\min\{\theta, c\})$. Without communication, the total quantity is $Q^M(\theta)$, where p is the belief that firm 2 has about θ. Since Q^M is decreasing, the quantity produced when the cartel communicates is always above the quantity produced when the cartel cannot communicate and the consumer's loss is smaller when the cartel can communicate than when it cannot. Intuitively, lack of communication distorts the cartel pricing and quantity decision as it cannot coordinate production efficiently. Communication improves not only the cartel's profits but also the consumers' surplus.

Athey and Bagwell (2001) show an example in which, for intermediate levels of patience, firms can better collude with communication. Our results apply even when firms are arbitrarily patient. Another role that communication has in cartels is to enhance monitoring (Whinston, 2008). As Awaya and Krishna (2014) show in a private monitoring Bertrand game with complete information, communication among firms allows them to set higher prices. Our finding is related to these ones, but here we show that communication also improves consumers welfare by reducing the price distortions that uncoordinated production induces.[37]

5.4. **Centralization, Communication, and Delays.** We now study the decision to centralize or decentralize decision making, which is affected by a trade off between coordination and the speed of adaptation. Centralization may allow an organization to coordinate agents on a jointly

[36] Since firm 2 never produces, its payoff equals the minmax, violating the conditions in Theorem 2. To deal with this difficulty, change the rule so that firm 2 produces g in every period and thus its payoff is strictly positive. When g is small enough, this entails an arbitrarily small loss. We also note that the restriction to $\mathbb{E}_p[\theta] < c$ –which is the main driver of the result that firm 2 does not produce–is rather strong and can be relaxed. In the more general model, the same basic intuition holds: Consumers are hurt by the lack of communication between firms.

[37] Our finding depends on the assumption that firms perfectly collude (either with or without communication). Shapiro (1986) shows in a static linear Cournot environment that communication can be detrimental for consumers's surplus. Gerlach (2009) also study the value of communication among cartel members for consumers in a Bertrand model with incomplete information. In his inelastic demand model, consumers' surplus vanish as firms become arbitrarily patient regardless of whether or not communication is available. He therefore emphasizes different mechanisms.

optimal set of choices, but it may also lead to delays in decision making (Simon, 1973; Radner, 1993; Bolton and Dewatripont, 1994; Van Zandt, 1999). As Roberts (2004, p. 235) explains: "Empowering those with the information to act upon it clearly speeds action [. . .]. There is no need to wait while the information is communicated up, absorbed, and analyzed, and then the decisions sent back down."

To model delays from direct communication, consider the coordination game of Section 2.1 and assume that the informed agent can send a cheap-talk message about his type, which reaches the uninformed party with a two-period lag. The timing within each period t is as follows. First, agent 1 learns θ^t and sends a message $m^t \in \Theta$ to agent 2. Second, agent 2 receives the message m^{t-2}. Third, agents choose their actions.

It is straightforward to extend our results to this new setting. First, given informational constraints, we find the optimal control rules as functions of the state. At any period, player 1 has private information over his type and over the message that has not reached player 2 yet. The state is now composed of a belief over the informed agent's type, together with a belief over the message sent in the last period. As before, the optimal rule may indicate player 1 to play according her private information –in addition to public beliefs–, but player 2's actions may only depend on public beliefs. Second, we show that it is possible to find strategies such that equilibrium play is arbitrarily close to optimal play.

Given a control $\sigma_1(\theta)$, an action a_1 in period $t-1$ and a message $m \in \Theta$ sent in period $t-2$, the belief over player 1's type in period t is

$$\mathcal{B}(\theta' \mid \sigma_1, a_1, m) = \sum_{\{\theta \mid \sigma_1(\theta) = a_1\}} P(\theta' \mid \theta) \, \frac{P(\theta \mid m)}{\sum_{\{\hat{\theta} \mid \sigma_1(\hat{\theta}) = a_1\}} P(\hat{\theta} \mid m)}$$

whenever $\sigma_1(\hat{\theta}) = a_1$ for some $\hat{\theta}$. As before, player 1 may signal her type with her actions. What changes is that player 1's message from two periods before conveys precise information about his type. This is why the rule for updating beliefs differs from the one in Section 4.1 (equation 4.2) .

The coordination game has two types ($\theta \in \{0, 1\}$) and two actions ($a_1 \in \{S, O\}$). Therefore, beliefs take simple forms. If the control is pooling (i.e., $\sigma_1(0) = \sigma_1(1) = a_1$), then

$$\mathcal{B}(\theta' \mid \sigma_1, a_1, m) = P(\theta' \mid 1) \, P(1 \mid m) + P(\theta' \mid 0) \, P(0 \mid m).$$

With a pooling control, player 1's actions do not transmit information, but her message two periods before does. For example, if player 1 is playing a pooling control, and sent a message $m^{t-2} = 1$, the probabilities that $\theta^t = 1$ and $\theta^t = 0$ are

$$\begin{aligned} \mathcal{B}(1 \mid \sigma_1, a_1, 1) &= P(1 \mid 1) \, P(1 \mid 1) + P(1 \mid 0) \, P(0 \mid 1) = \lambda^2 + (1-\lambda)^2, \\ \mathcal{B}(0 \mid \sigma_1, a_1, 1) &= P(0 \mid 1) \, P(1 \mid 1) + P(0 \mid 0) \, P(0 \mid 1) = 2 \, \lambda \, (1-\lambda). \end{aligned}$$

Given that $\lambda^2 + (1-\lambda)^2 > 2\,\lambda\,(1-\lambda)$ for $\lambda > 1/2$, a message $m^{t-2} = 1$ implies that $\theta^t = 1$ is more likely than $\theta^t = 0$.

On the other hand, if the control is separating (i.e., $\sigma_1(0) \neq \sigma_1(1)$), then

$$\mathcal{B}(\theta' \mid \sigma_1, a_1, m) = P(\theta' \mid \theta)\,\big|_{\{\theta \mid \sigma_1(\theta) = a_1\}} = \begin{cases} \lambda & \text{if } \sigma_1(\theta') = a_1, \\ 1-\lambda & \text{if } \sigma_1(\theta') \neq a_1. \end{cases}$$

With a separating control, actions perfectly signal player 1's type in the previous period, and therefore, the message from two periods before is redundant.

It simple to show that if the optimal control is separating for belief λ, then it is separating for belief $1-\lambda$. This is because the game with $\theta^t = 1$ is equal to the game with $\theta^t = 0$ (up to relabeling). Likewise, if the optimal control is pooling for belief $\lambda^2 + (1-\lambda)^2$, then it is pooling for the complementary belief $2\,\lambda\,(1-\lambda)$.

If a pooling control is optimal, then it is optimal to coordinate at t on the action that is optimal given period $t-2$'s message –that is, at time t, players play (S, S) when $m^{t-2} = 1$ and (O, O) when $m^{t-2} = 0$–, rather than playing a constant pooling profile (such as playing (S, S) or (O, O) for all t). To see this, suppose that players are coordinated on (S, S) for all t, and agent 2 receives a message $m^{t-2} = 0$. This means that the probability that $\theta^t = 0$ is larger than the probability than $\theta^t = 1$, which means that if players were to coordinate on playing (O, O) at time t, they would be more likely to obtain the joint payoff $1 + \alpha + \beta$ than if they were to play the constant rule (S, S).

We deduce that agents will optimally organize in one of two ways: They may play a pooling control and choose actions based on player 1's message from two periods before, or they may play a separating control, as in the signaling equilibrium of Section 2.1. The first case corresponds to an organization with centralized information and decision making. The second case corresponds to an organization with decentralized decision making.

By centralizing information, players avoid the cost of miscoordination that is incurred when player 1 signals a change in type. On the other hand, centralization means that players choose actions based on old information, which causes a two-period delay in adaptation. Moreover, this delay implies that player 1's type may have switched back to its original state by the time agents change their actions, in which case players would play actions that are not adapted to the state of the world for one more period.

In Section 2.1, we showed that a separating action profile yields a payoff of $\lambda(1 + \alpha + \beta)$ as $\delta \to 1$. In Section A.5 in the Appendix, we show that a pooling action profile in which players coordinate on player 1's message from two periods ago yields a payoff of $1 + \alpha\,(1 - 2\,(1-\lambda)\,\lambda) + \beta$ as $\delta \to 1$. Intuitively, by playing a pooling profile, players obtain a payoff of $1 + \beta$ in every period, and obtain α a proportion $1 - 2\,(1-\lambda)\,\lambda$ of times.

Centralization is preferred over decentralization when $\lambda < \frac{1+\alpha+\beta}{2\alpha}$. As the importance of coordination for player 1 (α) increases, decentralization becomes more valuable because it allows for faster adaptation to player 1's type. As the importance of coordination for player 2 (β) increases, centralization becomes more valuable because it allows players to avoid the costs of miscoordination caused by signaling type changes with actions. These results are consistent with the findings of McElheran (2014), which studies delegation of IT purchasing decisions and shows that a high value of adaptation (a large α) is associated with delegation, and a high value of coordination (a large β) is correlated with centralization.

Finally, note that in this simple model, it is easy to give agents incentives to play according to the optimal rule, given that agents obtain a payoff of 0 when they choose discoordinated actions. Our results show that our when direct communication has implicit costs, agents may optimally coordinate to play with limited communication.

6. Equilibrium as Interactions Become Frequent

Our limit results, Theorems 1 and 2, apply when $\delta \to 1$. As Abreu et al. (1991) point out, the limit $\delta \to 1$ can be interpreted saying that either interest (discount) rates are low or that players move frequently. In games with imperfect monitoring, Abreu et al. (1991) and Sannikov and Skrzypacz (2007) show that the two interpretations can lead to radically different results as when moves become more frequent not only the interest rates change but also the quality of the monitoring technology. In our perfect monitoring game of incomplete information, the impact of more frequent moves is also subtle as types are more likely to remain unchanged between two consecutive rounds. In this section, we explore these issues in a simple prisoners' dilemma.

Two players choose actions at each $t = D, 2D, \dots,$ where $D > 0$ is the period length. At each t, players play a prisoners dilemma, with the payoffs given in Figure 2. Monitoring is perfect, but only player 1 can observe $\theta^t \in \{l, h\}$ at the beginning of round t, with $l < h$. We parameterize both the discount factor and the transitions by D. The discount factor equals $\delta = \exp(-rD)$, where $r > 0$ is the discount rate per time unit. Transitions are given by

$$\mathbb{P}[\theta^t = l \mid \theta^{t-1} = l] = 1 - \phi D, \quad \mathbb{P}[\theta^t = h \mid \theta^{t-1} = h] = 1 - \chi D$$

with $\phi, \chi > 0$. We make explicit the dependence of the transition matrix and the Bayes operator on D by writing $P = P^D$ and $B = B^D$. Under this parametrization we can interpret our previous findings as taking the interest rate to 0 ($r \to 0$). Our interest now is in the limit $D \to 0$.

The formulation of the dynamic programming problem characterizing decision rules that maximize the sum of payoffs for $D > 0$ can be imported from Section 4. More explicitly, given a belief $p = \mathbb{P}[\theta^t = l]$, the value function for the problem of maximizing the sum of payoffs is

$$w^D(p) \quad = \tag{6.1}$$

$$\max_{\sigma\in\Sigma}\left\{(1-\exp(-rD))\,U^{(1,1)}(\sigma,p)+\ \exp(-rD)\sum_{a_1\in\{M,K\}}w^D(B^D(\cdot\mid\sigma_1,p,a_1))\sum_{\theta,\sigma_1(\theta)=a_1}p(\theta)\right\}.$$

The following result characterizes the solution to this problem when D and r are small.

Proposition 3. *The following hold:*

a. *There exists $\bar{D}>0$ such that for all $D<\bar{D}$ and all $p\in[\chi D,1-\phi D]$, the right-hand side of (6.1) has a unique solution $\bar{\sigma}$, with $\bar{\sigma}_1(l\mid p)=I$ and $\bar{\sigma}_1(h\mid p)=NI$. Moreover, $w^D(p)\to 2(a-l)\frac{\chi}{\phi+\chi}$ as $D\to 0$.*
b. *For all $\epsilon>0$, there exists $\hat{D}\in]0,\bar{D}[$ such that for $D<\hat{D}$ we can find $\bar{r}(=\bar{r}(D))$ such that the game played every D units of time with discount rate $r<\bar{r}(D)$ has an equilibrium attaining payoffs within distance ϵ of $(a-l)\frac{\chi}{\phi+\chi}(1,1)'$.*

This result shows that a separating rule (that generates a reactive-signaling path) is optimal whenever the game is played frequently, and that the incentive costs are modest. Intuitively, when the game is played frequently, the costs of signaling a change of type is small (it is incurred once) compared to the benefit of perfectly revealing information (which results in almost perfect information for several rounds of interaction).[38] This implies that as interactions become more frequent, it becomes more likely that players can attain the full benefits of cooperation without incurring significant signaling costs. Indeed, as $D\to 0$, first best payoffs converges to $2(a-l)\frac{\chi}{\phi+\chi}$ –the payoff attained in the game with frequent moves.

7. Conclusions and Extensions

Oftentimes, economic agents in a long-run relationship can only partially know the conditions under which their partners are making decisions. Moreover, communicating tough or favorable conditions is difficult because such protocols are either incomplete or non-existent (Schelling, 1960; Marschak and Radner, 1972; Whinston, 2008). Communication may also be difficult because economic shocks may materialize only after some other player has already made a decision. We explore optimal equilibria in this type of environment. Our exercise uncovers new tradeoffs arising in dynamic models of incomplete information –how much information is revealed is endogenously determined and the uninformed player forgives but does not forget apparently hostile actions. We show that the cooperation paths are quite rich and novel, and provide applications that shed light on phenomena that were previously unexplained.

Some extensions to our model are simple. We have worked with a one-sided incomplete information game to emphasize the forces in the model, but extending the results to allow for two-sided incomplete information is relatively simple.[39] We could also extend our results to allow for restricted or costly communication (in the direction of Section 5.4) or communication

[38] The costs of signaling are $O(D)$ whereas the benefits are $O(1)$.

[39] Details are available in a Supplementary Appendix.

only once the stage game has been played (but before the subsequent type is realized). Our setup can also be used to explore equilibria in a dynamic model of sovereign default, in which a country faces privately observed (economical or political) shocks that may make defaults socially attractive (Cole et al., 1995; Sandleris, 2008). In such model, a government decision of whether or not to pay its debt would affect others' beliefs about fundamentals and their willingness to lend or invest in the future.[40] A more challenging question is to explore the equilibrium set when the discount factor is not arbitrarily close to 1, possibly allowing for imperfect monitoring. We suspect that when δ is not close to 1, our insights (about whether and how information is revealed and about how strategies balance forgiveness and memory) will also show up, but additional incentive constraints may introduce new tradeoffs. Another interesting extension is to explore the continuous time limit model in Section 6, keeping constant the interest rate $r > 0$. Keeping fixed the interest rate r, the review blocks used in Proposition 3 become arbitrarily long and therefore the informed player need not have incentives to play an obedient strategy. These extensions are left for future research.

Appendix A. Proofs

This Appendix contains proofs for all the results in the main text.

A.1. **Proofs for Section 4.1.**

Proof of Lemma 1. The result is the standard dynamic programming formulation of partially observed Markov decision processes (Arapostathis et al., 1993). A minor subtlety arises due to the fact that our control variables are mixed strategies which, in contrast to what is typically addressed in the literature, involve private randomizations. To address this, note that a decision rule can be equivalently written as $s = (s_i^t)$ with $s_i^t \colon A^{t-1} \times \Theta_i^t \times [0,1]^t \times [0,1] \to A_i$. In other words, we can reformulate a behavior strategy by assuming that $a_i^t = s_i^t(a^1, \ldots, a^{t-1}, \theta_i^1, \ldots, \theta_i^t, \chi^1, \ldots, \chi^t, \chi_i^t)$ where χ_i^t is only used by player i. We can expand the set over which the maximization (4.1) is performed by allowing rules where all players at t condition on the whole vector $(\chi_1^t, \ldots, \chi_N^t)$. This relaxed efficiency problem admits a dynamic programming formulation in which, without loss, public randomizations are not used. Since the solution of the relaxed problem is feasible for (4.1), we deduce that $q(\alpha) = w^{\alpha,\delta}(\lambda)$. □

Proof of Theorem 1. We use the so-called vanishing discount approach. Parts a and b follow from Platzman (1980) or Theorem 11 in Hsu et al. (2006). It is enough to note that the hidden Markov process $(\theta^t)_{t \geq 1}$ has full support and note that, for example, Assumption 2 in Hsu et al. (2006) holds. To deduce c, we use part (d) Corollary on p.369 in Platzman (1980). □

[40] This is similar to the model in Sandleris (2008), but in that model the game has a finite horizon and shocks are drawn once.

Proof of Proposition 1. Consider the problem

$$\max_{\sigma\in\Sigma}\sum_{a_1\in A_1} h(B(\cdot\mid\sigma_1,p,a_1))\sum_{\theta\in\Theta,\sigma_1(\theta)=a_1} p(\theta)$$

with $h\colon\Delta(\Theta)\to\mathbb{R}$ convex. The solution is any separating rule (in particular, $\bar{\sigma}(\cdot\mid\bar{p})$ in the text solves this problem). To see this, notice that the problem can be reformulated as the problem of choosing a Bayes-consistent belief distribution over beliefs with the purpose of maximizing a convex function (Gentzkow and Kamenica, 2011). The value of that problem equals the concave hull of the objective and is attained by a distribution putting appropriate weights over delta-Dirac beliefs. □

A.2. **Proofs for Section 4.2.**

Proof of Lemma 2. Let $\bar{\sigma}^{\alpha}$ be the control rule solving the AROE given α. In particular, there exists $\bar{T}\in\mathbb{N}$ such that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\bar{\sigma}^{\alpha},\bar{p}^1}[\alpha\cdot u(a^t,\theta^t)]\geq\rho^{\alpha}-\epsilon/2$$

for all $T\geq\bar{T}$, and all $\bar{p}_1\in\{p_1\}\cup\Big(\cup_{\theta\in\Theta}\{P(\cdot\mid\theta)\}\Big)$. Let $Q^t(p_1)\subseteq\Delta(\Theta)$ be the finite set of beliefs having positive probability under $\bar{\sigma}^{\alpha}$ at round t given p_1. Let

$$Q\equiv\Big(\bigcup_{\bar{p}_1\in\{p_1\}\cup\Big(\cup_{\theta\in\Theta}\{P(\cdot\mid\theta)\}\Big)}Q^{\bar{T}}(\bar{p}_1)\Big)\setminus\Big(\bigcup_{t=1}^{\bar{T}-1}\bigcup_{\bar{p}_1\in\{p_1\}\cup\Big(\cup_{\theta\in\Theta}\{P(\cdot\mid\theta)\}\Big)}Q^{t}(\bar{p}_1)\Big)$$

be the set of beliefs that can be reached at time $\bar{T}$ (for some initial belief $\bar{p}_1\in\{p_1\}\cup(\cup_{\theta\in\Theta}\{P(\cdot\mid\theta)\})$) but that cannot be reached before. For $p\in Q$, define the control rule $\sigma_1(\cdot\mid p)\colon\Theta\to A_1$ such that $\sigma_1(\theta\mid p)\neq\sigma_1(\theta'\mid p)$ for $\theta\neq\theta'$ and $\sigma_2(p)$ arbitrary. For $p\notin Q$, take $\sigma(\cdot\mid p)\equiv\bar{\sigma}^{\alpha}(\cdot\mid p)$. Intuitively, the control rule σ is similar to σ^{α} but at beliefs $p\in Q$, σ perfectly reveals player 1's type. By construction, σ determines a unique recurrence class, with a set of beliefs in

$$\bigcup_{t=1}^{\bar{T}}\bigcup_{\bar{p}_1\in\{p_1\}\cup\Big(\cup_{\theta\in\Theta}\{P(\cdot\mid\theta)\}\Big)}Q^{t}(\bar{p}_1).$$

Moreover, for any $n\in\mathbb{N}$,

$$\frac{1}{\bar{T}}\sum_{t=\bar{T}n+1}^{\bar{T}(n+1)}\mathbb{E}_{\sigma,p_1}[\alpha\cdot u(a^t,\theta^t)]\geq\frac{1}{\bar{T}}\sum_{t=\bar{T}n+1}^{\bar{T}(n+1)}\mathbb{E}_{\sigma^{\alpha},p_1}[\alpha\cdot u(a^t,\theta^t)]-\frac{\epsilon}{4}\geq\rho^{\alpha}-\frac{3}{4}\epsilon$$

and therefore for all p_1 and all $T\geq\bar{T}$, $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\sigma,p^1}[\alpha\cdot u(a^t,\theta^t)]\geq\rho^{\alpha}-\epsilon$. □

Proof of Lemma 3. Let us first prove a. Since the control rule σ determine a unique recurrence class (Definition 3), there exists an irreducible transition matrix $\bar{P}$ for the joint process of states and beliefs, $(\theta^t, p^t)_{t\geq 1} \in \Theta\times\mathcal{P}$ and a unique stationary distribution $\bar{\pi}$ on $\Theta\times\mathcal{P}$. Using Blackwell's (1957) construction, we can extend the Markov chain (θ^t, a^t) to the negative numbers $t \in \mathbb{Z}$, and compute the invariant measure $\bar{\pi}(\theta, p) = \mathrm{P}\{\theta_0 = \theta, \mathrm{P}[\theta_0 = \cdot \mid (a^t)_{t\leq 0}] = p(\cdot)\}$. In particular, for any $(\theta, p) \in \Theta \times P$,

$$\bar{\pi}(\theta \mid p) = \mathrm{P}\Big[\theta^0 = \theta \mid p = \mathrm{P}[\theta^0 = \cdot \mid (a^t)_{t\leq 0}]\Big] = p(\theta). \tag{A.1}$$

For any sequence $(\theta^t, p^t)_{t\geq 1}$, we define the empirical transition matrix $\bar{P}^t$ on $\Theta \times \mathcal{P}$ as

$$\bar{P}^t\left((\theta', p') \mid (\theta, p)\right) = \frac{|\{t' \leq t-1 \mid (\theta^{t'}, p^{t'}) = (\theta, p), (\theta^{t'+1}, p^{t'+1}) = (\theta', p')\}|}{|\{t' \leq t-1 \mid (\theta^{t'}, p^{t'}) = (\theta, p)\}|}.$$

and the empirical measures

$$\bar{\pi}^t(\theta, p) = \frac{1}{t}\sum_{t'=1}^{t} \mathbb{1}_{(\theta^{t'}, p^{t'})=(\theta,p)} \quad \bar{\pi}^t(p) = \sum_{\theta\in\Theta} \bar{\pi}^t(\theta, p) = \frac{1}{t}\sum_{t'=1}^{t} \mathbb{1}_{p^t = p}.$$

Finally, for $(\theta, p) \in \Theta \times \mathcal{P}$ define $N^t(\theta, p) = \sum_{t'=1}^{t} \mathbb{1}_{(\theta^{t'}, p^{t'})=(\theta,p)}$.

Our first observation is that there exists a constant $c_1 > 0$ (depending on $\bar{P}$ and $\bar{\pi}$) such that for any $t \geq 1$ and an empirical transition matrix $\bar{P}^t$ on $\Theta \times \mathcal{P}$ sufficiently close to $\bar{P}$,

$$\|\bar{\pi}^t - \bar{\pi}\| \leq c_1\|\bar{P}^t - \bar{P}\| + c_1\frac{1}{t}$$

where $\|\cdot\|$ is the supreme norm. To see this inequality, we borrow the following two formulas from Lemma B.2 in Escobar and Toikka (2013)

$$\bar{\pi}^t = \left(I - \bar{P}^t + E\right)^{-1}(\mathbb{1} + e^t), \quad \bar{\pi} = \left(I - \bar{P} + E\right)^{-1}\mathbb{1}$$

where $\|e^t\| \leq \frac{|\Theta||\mathcal{P}|}{t}$ and note that the map $\bar{P}' \mapsto \left(I - \bar{P}' + E\right)^{-1}$ is Lipschitz in a neighborhood of $\bar{P}$. Moreover, since $\bar{\pi}(\theta, p) > 0$ for all $\theta \in \Theta$ and all $p \in P$, without loss we can take c_1 such that

$$\|\frac{\bar{\pi}^t(\theta, p)}{\bar{\pi}^t(p)} - \bar{\pi}(\theta \mid p)\| \leq c_1\|\bar{P}^t - \bar{P}\| + c_1\frac{1}{t}$$

for all $(\theta, p) \in \Theta \times \mathcal{P}$. Combining this observation with (A.1) we deduce that for all $p \in \mathcal{P}$

$$\|\bar{\pi}^t(\cdot \mid p) - p(\cdot)\| \leq c_1\|\bar{P}^t - \bar{P}\| + c_1\frac{1}{t} \tag{A.2}$$

Now, ignore the moves of player 0 and assume that player 1's actions are never modified. Use Lemma B.1 in Escobar and Toikka (2013) to show that there exists a decreasing sequence $(d_k)_k$ converging to 0 such that

$$\mathrm{P}_{\hat{s}_1}[\|\bar{P}^t(\cdot \mid (p, \theta)) - \bar{P}(\cdot \mid (p, \theta))\| < d_{N^t(p,\theta)} \quad \forall t \geq 1, \forall(\theta, p)] \geq 1 - \frac{\eta}{2}. \tag{A.3}$$

Fix $0 < \psi < \min_{\theta,p} \bar{\pi}(\theta, p)$ and use Theorem 1.10.2 in Norris (1997) to find $\bar{t}$ such that

$$\mathrm{P}_{\hat{s}_1}[N^t(p, \theta) \geq t(\bar{\pi}(\theta, p) - \psi),\ \forall t \geq \bar{t}] \geq 1 - \frac{\eta}{2}. \tag{A.4}$$

Define $c_2 = \min_{\theta,p} \bar{\pi}(\theta, p)(> 0)$ and the sequence $(b_k)_k$ by $b_k = c_1|\Theta|\left(d_{k(c_2-\psi)} + \frac{1}{k}\right)$ for all $k \geq \bar{t}$ (for $k < \bar{t}$, $b_k = 2$). From (A.2), (A.3), and (A.4)

$$\mathrm{P}_{\hat{s}^1}[\|\bar{\pi}^t(\cdot \mid p) - p(\cdot)\| \leq \frac{1}{|\Theta|} b_t \quad \forall t \geq 1, p \in \mathcal{P}] \geq 1 - \eta.$$

Note that for any element of the event above, player 1 passes the test (b_k) because

$$\max_{a_1 \in A_1} \|m^t(a_1 \mid p) - m(a_1 \mid p)\| \leq |\Theta| \|\bar{\pi}^t(\cdot \mid p) - \bar{\pi}(\cdot \mid p)\| \leq b_t$$

and therefore P[1 passes test (b_k) at $(a^1, \ldots, a^t)$, $\forall t] \geq 1 - \eta$. It follows that if we introduce the possibility that player 0 changes player 1's actions after failing a test, the lower bound for the probability above remains unaltered.

We now prove b. There exists $\bar{T} \geq 1$ such that for any $T \geq \bar{T}$, and any strategy s_1 for player 1 in the credible reporting game $(\sigma, (b_k), T)$,

$$\mathrm{P}_{s_1}[\|m^T(\cdot \mid p) - m(\cdot \mid p)\| \leq \eta, \forall p \in \mathcal{P}] \geq 1 - \eta.$$

This observation follows by noticing that regardless of the strategy s_1 used by 1, if at any given round player 1 fails the test, the continuation actions are drawn from $m(\cdot \mid p)$ (see Lemma B.5 in Escobar and Toikka (2013)). Therefore, with sufficiently high probability, for any strategy s^1, player 1 passes a relaxed test at the end of the block given the history of actions $(a^1, \ldots, a^T)$. □

Proof of Theorem 2. Take $\eta > 0$ small enough in Lemma 3 such that the expected average payoff for player 2 over the course of a game of credible play is within ϵ of v_2 and, for any sequential best response $s_1 = s_1^\delta$ of player 1 in the block-game of credible play, $v_1^\delta(s_1^\delta) \geq v_1 - \epsilon$. In particular, $\alpha \cdot v^\delta(s_1^\delta) \geq \alpha \cdot v - \epsilon \geq \rho^\alpha - 2\epsilon$, where the last inequality follows from Lemma 2.

We now construct the equilibrium strategy profile s^* as follows. Players start in a *cooperative phase* by choosing actions as in the equilibrium of the games of credible play $(\sigma^\alpha, (b_k), T)^\infty$. Any observable deviation by player i triggers a *stick phase* in which the players play minmax against i during L periods. Any deviation by a player restart a minmax phase of L rounds against that player. After the L rounds of minmax against i, a *carrot phase* is started in which players choose actions as in the equilibrium of the game of credible play $(\sigma^{\alpha^i}, (b_k), T)^\infty$. Deviations restart the minmax phase and so on.

Let $\epsilon > 0$ be small enough such that for some $\gamma \in]0, 1[$

$$v_i^{-i} - v_i^i > 2\epsilon, \quad (1 - \gamma) > \frac{2\epsilon}{v_i^i - \underline{v}_i}, \quad \gamma\left(v_i^{-i} - v_i^i - 2\epsilon\right) > (1 - \gamma)\left(\underline{v}_i - m + \epsilon\right)$$

for $i = 1, 2$. Take $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$ the games $(\sigma^{\alpha'}, (b_k), T)^\infty$, for $\alpha' = \alpha, \alpha^1, \alpha^2$, have discounted equilibrium payoffs $U^{\alpha'}(\delta)$ within distance ϵ of the target payoffs $v^{\alpha'}$. Define the length of the stick phase as $L(\delta) = \max\{d \in \mathbb{N} \mid d \leq \frac{\ln(\gamma)}{\ln(\delta)}\}$ and note that $\delta^L \to \gamma$. Lemma

6.1 in Escobar and Toikka (2013) shows that discounted payoffs during the L periods of the stick phase against i are bounded above by $(1-\delta^L)(\underline{v}_i+\epsilon)$ for δ sufficiently large.

Now, consider the incentives in the carrot phase

$$v_i - \epsilon \geq (1-\delta)M + (\delta - \delta^{L+1})(\underline{v}_i + \epsilon) + \delta^{L+1}(v_i^i + \epsilon)$$

The incentives of player i in the stick phase against $j \neq i$ can be written

$$(1-\delta^L)m + \delta^L(v_i^j - \epsilon) \geq (1-\delta)M + (\delta - \delta^{L+1})(\underline{v}_i + \epsilon) + \delta^{L+1}(v_i^i + \epsilon)$$

Finally, the incentives of player i in the carrot phase against j can be written as

$$v_i^j - \epsilon \geq (1-\delta)M + (\delta - \delta^{L+1})(\underline{v}_i + \epsilon) + \delta^{L+1}(v_i^i + \epsilon)$$

Taking the limit as $\delta \to 1$ in all these inequalities, by construction of ϵ and γ, we deduce the existence of a critical discount factor such that all incentive constraints hold. □

A.3. **A Proof for Section 4.3.**

Proof of Proposition 2. Consider first a solution $\sigma^* \in \Sigma$ to the problem

$$\max_{\sigma\in\Sigma} \sum_{\theta\in\Theta} \Big(\alpha_1 u_1(\sigma_1(\theta), \sigma_2, \theta) + \alpha_2 u_2(\sigma_1(\theta), \sigma_2)\Big)p(\theta)$$

Since $p(\theta) > 0$ for all θ, $\sigma_1^*(\theta) \in \arg\max_{a_1\in A_1}\{\alpha_1 u_1(a_1, \sigma_2^*, \theta) + \alpha_2 u_2(a_1, \sigma_2^*)\}$. Fix $\theta^m \in \Theta$ with $m < |\Theta|$ and $a^n = \sigma_1^*(\theta)$ with $2 \leq n \leq |A_1| - 1$. By concavity, the derivative

$$\frac{\partial}{\partial a_1}\Big(\alpha_1 u_1 + \alpha_2 u_2\Big)(a^{n-1}, \sigma_2^*, \theta^m)$$

is nonnegative. Now,

$$\begin{aligned}\frac{\partial}{\partial a_1}\Big(\alpha_1 u_1 + \alpha_2 u_2\Big)(a^{n+1}, \sigma_2^*, \theta^{m+1}) &= \frac{\partial}{\partial a_1}\Big(\alpha_1 u_1 + \alpha_2 u_2\Big)(a^{n-1}, \sigma_2^*, \theta^m) \\ &+ \int_{a^{n-1}}^{a^{n+1}} \frac{\partial^2}{\partial a_1^2}\Big(\alpha_1 u_1 + \alpha_2 u_2\Big)(y, \sigma_2^*, \theta^m)dy + \alpha_1 \int_{\theta^m}^{\theta^{m+1}} \frac{\partial^2}{\partial a_1 \partial\theta} u_1(a^{n-1}, \sigma_2^*, y)dy \\ &\geq |a^{n+1} - a^n|(-c_1) + \alpha_1 c_2(\theta^{m+1} - \theta^m)\end{aligned}$$

is positive under (4.10). It follows that $\sigma_1^*(\theta^{m+1} \mid p) \geq a^{n+1} > \sigma_1^*(\theta^m \mid p)$. To deduce the second part of the Proposition, use the results in Van Zandt and Vives (2007) for monotone comparative statics in Bayesian games. □

A.4. **A Proof for Section 5.1.**

Proof of Lemma 4. Under the assumptions $\lambda > \frac{C-G}{2R+C-G}$, the AROE (4.6) is maximized by control (NS, S, NS) when $p = \lambda$. This follows from the fact that under this restriction on parameters,

(NS, S, NS) maximizes $\max_\sigma U^{(1/2,1/2)}(\lambda)$ and, from Proposition 1, (NS, S, NS) also solves (4.6).

Now, if at belief $p = 1 - \mu$, control (NS, S, S) is optimal for the right hand side of AROE (4.6), then σ generates reactive-signaling dynamics and the result holds. So, suppose that (NS, S, S) is not optimal at $p = 1 - \mu$. This means that either (S, S, S) or (NS, S, NS) are optimal at $p = 1 - \mu$. If (NS, S, NS) is optimal, then σ generates time-off dynamics with $\widehat{\tau} = 0$. If (S, S, S) is optimal at $1 - \mu$, then it must result in higher total payoffs than (NS, S, S) for all $p > (1 - \mu)$.[41] When the control (S, S, S) is employed, the path of beliefs increases as time passes by. If after some belief in the path, (NS, S, NS) is optimal, then the optimal control rule generates time-off dynamics with finite $\widehat{\tau}$. If not, (S, S, S) is played along the path and the optimal control rule generates time-off dynamics with $\widehat{\tau} = \infty$. □

A.5. **A Proof for Section 5.4.** In this section, we show how to obtain the discounted sum of payoffs when players play a pooling profile and choose actions that fit player 1's message from two periods before. That is, at time t, players play (S, S) when $m^{t-2} = 1$ and (O, O) when $m^{t-2} = 0$.

Given the symmetry of payoffs with respect to beliefs, the only relevant information for constructing payoffs is agent 1's announcements of a change in type. Let S_{XY} represent a state of past announcements such that agent 1 announced a change in type in the last period if $X = 1$ (otherwise $X = 0$) and announced a change in type in the current period if $Y = 1$ (otherwise $Y = 0$). There are four states to be considered: S_{00}, S_{01}, S_{10}, and S_{11}. The value functions for each state are:

$$\begin{aligned}
W_{00} &= (1-\delta)(1+\alpha+\beta) + \delta(\lambda W_{00} + (1-\lambda) W_{01}),\\
W_{01} &= (1-\delta)(1+\beta) + \delta(\lambda W_{10} + (1-\lambda) W_{11}),\\
W_{10} &= (1-\delta)(1+\beta) + \delta(\lambda W_{00} + (1-\lambda) W_{01}),\\
W_{11} &= (1-\delta)(1+\alpha+\beta) + \delta(\lambda W_{10} + (1-\lambda) W_{11}).
\end{aligned}$$

Solving the above system of equations and taking the limit as $\delta \to 1$ yields $1+\beta+\alpha(1-2(1-\lambda)\lambda)$, which is the result mentioned in the text.

A.6. **A Proof for Section 6.**

Proof of Proposition 3. Lemma 4 shows that the optimal equilibrium follows either reactive-signaling or time-off dynamics. Let $W_{RS}(D)$ be given the reactive signaling rule. Let $w^{\tau}_{TO}(D)$ be the average value when a time-off control rule is used, given a punishment $\tau \in \{0, 1, 2\} \cup \{\infty\}$.

[41] To see this, let $h_\sigma(p)$ be the right hand side of AROE (4.6) given a control σ. Note that $h_{(S,S,S)}(0) = h_{(NS,S,S)}(0)$, $h_{(S,S,S)}(1-\mu) > h_{(NS,S,S)}(1-\mu)$, and $h_{(S,S,S)}(p)$ is convex whereas $h_{(NS,S,S)}(p)$ is linear. These three conditions imply that $h_{(S,S,S)}(p) > h_{(NS,S,S)}(p)$ for all $p > 1 - \mu$.

The limit of the value of playing reactive-signaling when $D \to 0$ is

$$\lim_{D\to 0} w_{RS}(D) = 2(a-l)\frac{\chi}{\phi+\chi}.$$

The limit of the value of playing time-off for a given τ when $D \to 0$ is

$$\lim_{D\to 0} \Big(\max_{\tau\in\{0,1,2\ldots\}} w_{TO}^{\tau}(D) \Big) = 2(a-l)\frac{\chi}{\phi+\chi}.$$

Now, we can also compute the derivatives and deduce that

$$\lim_{D\to 0} \frac{\partial w_{RS}}{\partial D}(D) \in \mathbb{R} \quad \lim_{D\to 0} \max_{\tau\in\{0,1,2,\}\cup\{\infty\}} \frac{\partial w_{TO}^{\tau}(D)}{\partial D} = -\infty.$$

It follows that there exists $\hat{D}$ such that for all $D < \hat{D}$, a reactive-signaling control has greater value than an optimally chosen time-off control. This proves part a of the proposition.

To prove b, we follow steps close to those in the proof of Theorem 2. The definition of game of credible reporting remains unaltered for any given D. We will prove that for a proper choice of parameters, we can replicate Lemma 3. We construct the sequence b_k from the definition of d_k (see proof of Lemma 3) by picking $0 < \psi < \lim_{D\to 0} \bar{\pi}^D(\theta, p)$, with $\bar{\pi}^D$ the stationary distribution given D, and $b_k = c_1|\Theta|(d_{k(c_2-\psi)} + \frac{1}{k})$. Conditions (A.2) and (A.3) follow immediately for any D. Condition (A.4) is also immediate, just notice that the choice of $\bar{t}$ depends on D so $\bar{t} = \bar{t}(D)$. This completes the first part of Lemma 3. To see the second part, construct $\bar{T} = \bar{T}(D)(> \bar{t}(D))$ so that for any strategy s_1 $\mathbb{P}_{s_1}^{D}[\|m^T(\cdot \mid p) - m^D(\cdot \mid p)\| \leq \epsilon \quad \forall p \in P^D] \geq 1-\epsilon$. Note that for the game of credible play $(\bar{\sigma}, (b_k^D), T)$, with $T \geq \bar{T}(D)$, Player 1 can obtain a payoff at least $(a-l)\frac{\chi}{\phi+\chi} - \epsilon$. By construction, Player 2's payoff is within ϵ of $(a-l)\frac{\chi}{\phi+\chi}$. Fixing $\tau, T \geq \bar{T}(D)$, we can find $\bar{r}(D)$ such that for all $r < \bar{r}(D)$, for any best response s_1 in the block-game of credible play, Player 1 obtains a payoff at least $(a-l)\frac{\chi}{\phi+\chi} - \epsilon$. Taking $D \leq \bar{D}$ and $r \leq \bar{r}(D)$ (sufficiently small if needed), by definition equilibrium payoffs in the game played every D units of time with discount rate r are bounded above by $2\,(a-l)\frac{\chi}{\phi+\chi} + \epsilon$. Observable deviations from the path of play of the block-credible reporting game are punished by Nash reversion. Provided $\bar{r}(D)$ is chosen sufficiently small, the result follows. □

References

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

Abreu, D., D. Pearce, and E. Stacchetti (1986): "Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory*, 39, 251–269.

——— (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1063.

Acemoglu, D. and A. Wolitzky (2014): "Cycles of Conflict: An Economic Model," *The American Economic Review*, 104, 1350–1367.

Allen, B. T. (1976): "Tacit Collusion and Market Sharing: The Case of Steam Turbine Generators," *Industrial Organization Review*, 4, 48–57.

Arapostathis, A., V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus (1993): "Discrete Time Controlled Markov Processes with Average Cost Criterion: A Survey," *SIAM Journal on Control and Optimization*, 31, 282–344.

Arrow, K. (1985): "Informational Structure of the Firm," *The American Economic Review*, 303–307.

Ashworth, T. (1980): *Trench Warfare, 1914-1918: The Live and Let Live System*, New York: Holmes and Meier.

Athey, S. and K. Bagwell (2001): "Optimal Collusion with Private Information," *RAND Journal of Economics*, 32, 428–465.

——— (2008): "Collusion with Persistent Cost Shocks," *Econometrica*, 76, 493–540.

Athey, S., K. Bagwell, and C. Sanchirico (2004): "Collusion and price rigidity," *Review of Economic Studies*, 71, 317–349.

Awaya, Y. and V. Krishna (2014): "On Tacit versus Explicit Collusion," Working paper, Penn State.

Axelrod, R. (1984): *The Evolution of Cooperation*, New York: Basic Books.

Bagwell, K. and R. Staiger (2005): "Enforcement, Private Political Pressure, and the General Agreement on Tariffs and Trade/World Trade Organization Escape Clause," *The Journal of Legal Studies*, 34, 471–513.

Bergemann, D. and J. Valimaki (2006): "Bandit Problems," *Yale University*.

Bernheim, B. and E. Madsen (2014): "Price Cutting and Business Stealing in Imperfect Cartels," Working paper, National Bureau of Economic Research.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, vol. 1, 93–102.

——— (1957): "The Entropy of Functions of Finite-State Markov Chains," in *Trans. First Prague Conf. Inf. Theory, Statistical Decision Functions, Random Processes*, 13–20.

Bolton, P. and M. Dewatripont (1994): "The firm as a communication network," *Quarterly Journal of Economics*, 109, 809–839.

Carlton, D., R. Gertner, and A. Rosenfield (1996): "Communication Among Competitors: Game Theory and Antitrust," *George Mason Law Review*, 5, 423.

Clark, R. and J.-F. Houde (2013): "Collusion with Asymmetric Retailers: Evidence from a Gasoline Price-Fixing Case," *American Economic Journal: Microeconomics*, 5, 97–123.

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

Dutta, P. (1995): "A Folk Theorem for Stochastic Games," *Journal of Economic Theory*, 66, 1–32.

Escobar, J. and J. Toikka (2013): "Efficiency in Games with Markovian Private Information," *Econometrica*, 81, 1887–1934.

Fudenberg, D. and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54, 533–554.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT Press.

Gale, D. and R. Rosenthal (1994): "Price and Quality Cycles for Experience Goods," *The RAND Journal of Economics*, 590–607.

Gensbittel, F. and J. Renault (2015): "The Value of Markov Chain Games with Incomplete Information on both Sides," *Mathematics of Operations Research*.

Gentzkow, M. and E. Kamenica (2011): "Bayesian Persuasion," *American Economic Review*, 101.

Gerlach, H. (2009): "Stochastic Market Sharing, Partial Communication and Collusion," *International Journal of Industrial Organization*, 27, 655–666.

Gorman, P. F. (1953): "Limited War: Korea, 1950," mimeo, Harvard University.

Green, E. and R. Porter (1984): "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87–100.

Hörner, J., T. Sugaya, S. Takahashi, and N. Vieille (2010): "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem," Working paper, Cowles Foundation.

——— (2011): "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica*, 79, 1277–1318.

Hörner, J., S. Takahashi, and N. Vieille (2015): "Truthful Equilibria in Dynamic Bayesian Games," Working paper, Yale University.

Hsu, S.-P., D.-M. Chuang, and A. Arapostathis (2006): "On the Existence of Stationary Optimal Policies for Partially Sbserved MDPs under the Long-Run Average Cost Criterion," *Systems & Control Letters*, 55, 165–173.

Jackson, M. O. and H. F. Sonnenschein (2007): "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241–258.

Keller, G. and S. Rady (1999): "Optimal Experimentation in a Changing Environment," *The Review of Economic Studies*, 66, 475–507.

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

LIU, Q. AND A. SKRZYPACZ (2014): "Limited Records and Reputation Bubbles," *Journal of Economic Theory*, 151, 2–29.

MARKHAM, J. (1951): "The Nature and Significance of Price Leadership," *American Economic Review*, 41, 891–905.

MARSCHAK, J. AND R. RADNER (1972): *Economic Theory of Teams*, Yale University Press.

MARSHALL, R. AND L. MARX (2013): *The Economics of Collusion: Cartels and Bidding Rings*, MIT Press.

MCELHERAN, K. (2014): "Delegation in Multi-Establishment Firms: Evidence from IT Purchasing," *Journal of Economics & Management Strategy*, 23, 225–258.

MOURAVIEV, I. AND P. REY (2011): "Collusion and Leadership," *International Journal of Industrial Organization*, 29, 705–717.

NICHOLLS, W. H. (1951): *Price policies in the cigarette industry*, Vanderbilt University Press, Nashville TN.

NORRIS, J. (1997): *Markov Chains*, Cambridge University Press.

PĘSKI, M. (2014): "Repeated Games with Incomplete Information and Discounting," *Theoretical Economics*, 9, 651–694.

PĘSKI, M. AND J. TOIKKA (2016): "Value of Persistent Information," .

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

PLATZMAN, L. K. (1980): "Optimal Infinite Horizon Undiscounted Control of Finite Probabilistic Systems," *SIAM Journal on Control and Optimization*, 18, 362–380.

PUTERMAN, M. L. (2005): *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, vol. 414, Wiley.

RADNER, R. (1981): "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," *Econometrica*, 49, 1127–1148.

——— (1993): "The organization of decentralized information processing," *Econometrica*, 61, 1109–1146.

RAHMAN, D. (2014): "The Power of Communication," *The American Economic Review*, 104, 3737–3751.

RENAULT, J., E. SOLAN, AND N. VIEILLE (2013): "Dynamic Sender Receiver Games," *Journal of Economic Theory*, 148, 502–534.

RENOU, L. AND T. TOMALA (2015): "Approximate Implementation in Markovian Environments," *Journal of Economic Theory*, 159, 401–442.

ROBERTS, J. (2004): *The Modern Firm: Organizational Design for Performance and Growth*, Oxford University Press.

ROTEMBERG, J. AND G. SALONER (1986): "A Supergame-Theoretic Model of Price Wars during Booms," *The American Economic Review*, 76, 390–407.

——— (1990): "Collusive Price Leadership," *The Journal of Industrial Economics*, 93–111.

Sandleris, G. (2008): "Sovereign Defaults: Information, Investment and Credit," *Journal of International Economics*, 76, 267–275.

Sannikov, Y. and A. Skrzypacz (2007): "Impossibility of Collusion Under Imperfect Monitoring with Flexible Production," *American Economic Review*, 97, 1794–1823.

Schelling, T. (1960): *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

Scherer, F. M. and D. Ross (1990): "Industry Market Structure and Economic Performance," .

Seaton, J. S. and M. Waterson (2013): "Identifying and characterising price leadership in British supermarkets," *International Journal of Industrial Organization*, 31, 392–403.

Shapiro, C. (1986): "Exchange of Cost Information in Oligopoly," *The Review of Economic Studies*, 53, 433–446.

Simon, H. A. (1973): "Applying information technology to organization design," *Public Administration Review*, 33, 268–278.

Stigler, G. (1947): "The Kinky Oligopoly Demand Curve and Rigid Prices," *The Journal of Political Economy*, 432–449.

Stokey, N. and E. Lucas, R. with Prescott (1989): *Recursive Methods in Economic Dynamics*, Cambridge: Harvard University Press.

Tong, X. and R. Van Handel (2012): "Ergodicity and Stability of the Conditional Distributions of Nondegenerate Markov Chains," *The Annals of Applied Probability*, 22, 1495–1540.

Townsend, R. (1982): "Optimal Multi-Period Contracts and the Gain from Enduring Relationships under Private Information," *Journal of Political Economy*, 90, 1166–1186.

Van Handel, R. (2009): "The Stability of Conditional Markov Processes and Markov Chains in Random Environments," *The Annals of Probability*, 1876–1925.

Van Zandt, T. (1999): "Real-time decentralized information processing as a model of organizations with boundedly rational agents," *Review of Economic Studies*, 66, 633–658.

Van Zandt, T. and X. Vives (2007): "Monotone Equilibria in Bayesian Games of Strategic Complementarities," *Journal of Economic Theory*, 134, 339–360.

Whinston, M. (2008): *Lectures on Antitrust Economics*, The MIT Press.

Appendix B. Supplementary Appendix

B.1. **Two-Sided Incomplete Information.** We now extend the model to allow for two-sided incomplete information. All our main results hold, but the notation becomes more cumbersome. We also provide an application to a two-sided incomplete information Bertrand example.

B.1.1. *Model.* The model is as the one in the main text, but now the timing within each round t is as follows

t.0 A randomization device χ^t is publicly realized
t.1 Player i is privately informed about $\theta_i^t \in \Theta$
t.2 Players choose actions $a_i^t \in A_i$ simultaneously
t.3 Players observe the action profile chosen $a^t \in A$

The period payoff function for player i is $u_i(a, \theta_i)$. We sometimes abuse notation and write $u_i(a, \theta)$. Players rank flows of payoffs according to $(1 - \delta) \sum_{t \geq 1} \delta^{t-1} u_i(a^t, \theta^t)$, where $\delta < 1$ is the common discount factor. We assume that $|A_i| \geq |\Theta_i|$ for $i = 1, 2$.

The initial type of player i, θ_i^1, is drawn from a distribution $p_i^1 \in \Delta(\Theta_i)$. Player i's private types, $(\theta_i^t)_{t \geq 1}$, evolve according to a Markov chain (p_i^1, P_i), where $p_i^1 \in \Delta(\Theta)$ and P_i is a transition matrix on Θ. Both Markov chains are independent. We assume that the process of types has full support. This means that for all $\theta, \theta' \in \Theta$, $P_i(\theta' \mid \theta) > 0$. Let $\pi_i \in \Delta(\Theta)$ be the stationary distribution for P_i.

A strategy for player i is a sequence of functions $s_i = (s_i^t)_{t \geq 1}$ with $s_i^t \colon \Theta_i^t \times A^{t-1} \times [0, 1]^t \to \Delta(A_i)$. We can therefore define the vector of expected payoffs $v^\delta(s)$ given s, the set of all feasible payoffs $V(\delta, \lambda)$, and the set of equilibrium payoffs $\mathcal{E}(\delta, p^1) \subseteq V(\delta, p^1)$ as we did in the main text.

B.1.2. *Efficient Payoffs.* A strategy s is *efficient* if for some $\alpha \in \mathbb{R}^2_{++}$, s is a solution to

$$q(\alpha) = \max\{\alpha \cdot v^\delta(s) \mid s \text{ is a strategy profile }\}. \tag{B.1}$$

To characterize the solutions to this problem, we introduce some notation. Let $\Sigma_i = \{\sigma_i \colon \Theta_i \to A_i\}$. Let $p_i \in \Delta(\Theta_i)$ be a belief about player i's type given public information. For $\sigma \in \Sigma$ and $p = (p_1, p_2) \in \Delta(\Theta_1) \times \Delta(\Theta_2)$, we define the vector of expected period utility $U(\sigma, p) \in \mathbb{R}^2$ as

$$U_i(\sigma, p) = \sum_{\theta \in \Theta} u_i(\sigma_1(\theta_1), \sigma_2(\theta_2), \theta_i)\; p_1(\theta_1) p_2(\theta_2)$$

For $\alpha \in \mathbb{R}^2_{++}$, we consider the ex-ante weighted sum of period payoffs $U^\alpha(\sigma, p) = \alpha \cdot U(\sigma, p)$. We also define the Bayes operator $B_i(\cdot \mid \sigma_i, p_i, a_i) \in \Delta(\Theta_i)$ as

$$B_i(\theta_i' \mid \sigma_i, p_i, a_i) = \sum_{\{\theta_i \mid \sigma_i(\theta_i) = a_i\}} P_i(\theta_i' \mid \theta_i) \; \frac{p_i(\theta_i)}{\sum_{\{\hat{\theta}_i \mid \sigma_i(\hat{\theta}_i) = a_i\}} p_i(\hat{\theta}_i)}. \tag{B.2}$$

For $\alpha \in \mathbb{R}^2_{++}$, consider the only solution to the Bellman equation

$$w^{\alpha,\delta}(p) = \max_{\sigma\in\Sigma} \left\{(1-\delta)U^{\alpha}(\sigma,p)+\delta\sum_{a\in A} w^{\alpha,\delta}\Big(B_1(\cdot \mid \sigma_1,p_1,a_1),B_2(\cdot \mid \sigma_2,p_2,a_2)\Big) \sum_{\theta\in\Theta,\sigma(\theta)=a} p_1(\theta_1)p_2(\theta_2)\right\} \tag{B.3}$$

for all $p \in \Delta(\Theta_1)\times\Delta(\Theta_2)$. Take $\sigma^{\alpha,\delta}(\cdot \mid p)$ as the control profile attaining the maximum in (4.3) as a function of beliefs p. Any σ such that $\sigma(\cdot \mid p) \to \Sigma$, for $p \in \Delta(\Theta)$, will be a (Markov) *control rule*. Using the control rule $\sigma^{\alpha,\delta}$, we can construct a (non-randomized) strategy profile $s = s^{\alpha,\delta}$ from $\sigma^{\alpha,\delta}$ as we did in the main text. That this dynamic programming formulation (4.3) provides a solution to the problem of efficient payoffs given weighs $\alpha \in \mathbb{R}^2_{++}$ is obvious given the analysis in the main text.

We can also take the limit $\delta \to 1$ to deduce the *average reward optimality equation* (AROE)

$$h(p)+\rho =$$
$$\max_{\sigma\in\Sigma}\left\{\alpha_1U_1(\sigma,p)+\alpha_2U_2(\sigma,p)+\sum_{a\in A} h(B_1(\cdot \mid \sigma_1,p_1,a_1),B_2(\cdot \mid \sigma_2,p_2,a_2))\Big(\sum_{\theta\in\Theta,\sigma(\theta)=a} p_1(\theta_1)p_2(\theta_2)\Big)\right\}$$

for all $p = (p_1,p_2) \in \Delta(\Theta_1)\times\Delta(\Theta_2)$. Let $\sigma^{\alpha}(\cdot \mid p) \in \Sigma$ be the control profile attaining the maximum in the dynamic programming problem (4.6) given $p \in \Delta(\Theta)$.

Theorem 1 can be easily extended to this more general setup without any changes.

B.1.3. *Equilibrium Theorem.* We now extend our main equilibrium theorem.

A control rule σ together with the initial beliefs p^1 recursively determine a belief process $(p^t)_{t\geq 1}$ by

$$p^{t+1}(\theta) = B_1(\theta_1 \mid \sigma_1,p^t_1,a^t_1)\times B_2(\theta_2 \mid \sigma_2,p^t_2,a^t_2) \quad \forall t\geq 1.$$

Given any control rule σ, the joint process $(\theta^t,p^t)_{t\geq1}$ is Markovian, with $p^1 \in \Delta(\Theta_1)\times\Delta(\Theta_2)$ and $\theta^1 \in \Theta_1\times\Theta_2$ given.

The following definition will be useful to test suspicious behavior.

Definition 3. *A control rule σ* determines a unique recurrence class *if the process $(\theta^t,p^t)_{t\geq1}$ is a finite Markov chain having a unique recurrence class.*

A separating solution σ^α to (4.6) determines a unique recurrence class. The following result shows that relaxing the optimality requirement to allow for approximate efficiency is enough to ensure the existence of a control rule determining a unique recurrence class.

Lemma 5. *For all $\epsilon > 0$, and all $\alpha \in \mathbb{R}^2_{++}$, there exists a control rule σ, and $\bar{T} \in \mathbb{N}$ such that*

a. *σ determines a unique recurrence class; and*
b. $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\sigma,p}[\alpha\cdot u(a^t,\theta^t)] \geq \rho^\alpha - \epsilon$ *for al $T \geq \bar{T}$, and all p in the (finite) path of beliefs generated by σ and p_1. Moreover, when σ^α is a separating rule, we can take $\sigma = \sigma^\alpha$.*

Moreover, if σ^α is separating, we can take $\sigma = \sigma^\alpha$.

The proof of this lemma is analogous to the proof of Lemma 2 (which applies to the one-sided case).

For any control rule σ determining a unique recurrence class, the limit-average payoffs

$$v_1^\infty(\sigma) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} u_1(\sigma(\theta^t \mid p^t), \theta_1^t)], \quad v_2^\infty(\sigma) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} u_2(\sigma(\theta^t \mid p^t), \theta_2^t)]$$

are well defined. Letting $\bar{\pi} = \bar{\pi}^\sigma \in \Delta(\Theta \times \mathcal{P})$ be the stationary distribution for the Markov chain $(\theta^t, p^t)_{t\geq 1}$, given the control rule σ, with $\Theta \times \mathcal{P}$ the recurrence class of the Markov chain, it follows that

$$v_1^\infty(\sigma) = \sum_{(\theta,p)\in\Theta\times\mathcal{P}} u_1(\sigma(\theta \mid p), \theta)\bar{\pi}(\theta, p) \quad \text{and} \quad v_2^\infty(\sigma) = \sum_{(\theta,p)\in\Theta\times\mathcal{P}} u_2(\sigma(\theta \mid p))\bar{\pi}(\theta, p)$$

We define $v^\infty(\sigma) = (v_i^\infty(\sigma))_{i=1,2}$.

Fix a control rule σ determining a unique recurrence class $\Theta \times \mathcal{P}$. Define $m_1^\sigma(\cdot \mid p) \in \Delta(A_1)$ as the distribution over actions given a belief $p \in \mathcal{P}$:

$$m_1^\sigma(a_1 \mid p) = \sum_{\{\theta\in\Theta \mid a_1=\sigma_1(\theta\mid p)\}} p(\theta).$$

For $a \in A$ and $p \in \mathcal{P}$, we define $m^\sigma(a \mid p)$ analogously.

Given any sequence of actions $a_1^1, \ldots, a_1^t$ and a fixed control rule σ determining a unique recurrence class, we can mechanically calculate probabilities $\bar{p}_i^{t+1} = B(\cdot \mid \sigma_i, \bar{p}_i^t, a_i^t)$ (if this is not well defined, we set $\bar{p}_i^{t+1}$ to be an arbitrary element of the support of the process of beliefs $(p_i^t)_{t\geq 1}$) with $\bar{p}_i^1 = p_i^1$. The definitions of distribution over actions $m^\sigma(a \mid p)$ and occupancy rates $\bar{m}^\delta(a \mid p)$ are analogous to the one-sided case. The definitions of minmax values can also be extended in the obvious way.

Theorem 3 (Equilibrium Theorem, Two-Sided Incomplete Information)**.** *Fix $\epsilon > 0$. For $\alpha, \alpha^1, \alpha^2 \in \mathbb{R}^2_{++}$, take control rules σ, σ^1, and σ^2 as in Lemma 5. Assume*

i. *All payoff vectors $v = v^\infty(\sigma), v^1 \equiv v^\infty(\sigma^1), v^2 \equiv v^\infty(\sigma^2)$ are strictly individually rational;*
ii. *$v_i^i < v_i < v_i^{-i}$, for $i = 1, 2$.*

Then, there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, the infinitely repeated game with discount factor δ has a perfect Bayesian equilibrium $s^* = (s_1^*, s_2^*)$ such that

a. *$\alpha \cdot v^\delta(s^*) \geq \rho^\alpha - 2\epsilon$; and*
b. *$\mathbb{P}_{s^*}\left[\max_{a\in A, p\in\mathcal{P}} |\bar{m}^\delta(a \mid p) - m^\sigma(a \mid p)| < \epsilon\right] \geq 1 - \epsilon$, where $\Theta \times \mathcal{P} \subseteq \Theta \times \Delta(\Theta)$ is the recurrence class of the process $(\theta^t, p^t)_{t\geq 1}$ generated by σ.*

This theorem is proved testing actions conditional on beliefs. To formulate the test, we introduce some terminology. For any decreasing sequence (b_k) converging to 0, we say that player i *passes the test* (b_k) given a history $(a^1, \ldots, a^t) \in A^t$ if

$$\max_{a_i \in A_i} |m_i^{\sigma}(a_i \mid p) - \bar{m}_i^t(a_1 \mid p)| \leq b_t$$

for all $p \in \mathcal{P}$. Given $T \geq 1$, a rule σ and sequence (b_k), the game of credible play $(\sigma, (b_k), T)$ is constructed as follows. For $t \leq T$, if player i has passed the test (b_k) in all previous rounds $t' = 1, \ldots, t-1$, then he can freely select his action a_i^t; otherwise, player 0 chooses a_i^t by randomly drawing an action according to the distribution $m_i(\cdot \mid \bar{p}^t)$. We define the *obedient strategy* for player i as $\hat{s}_i^t(\theta_i^1, \ldots, \theta_i^t, a^1, \ldots, a^{t-1}) = \sigma_i(\theta_i^t \mid \bar{p}^t)$ whenever he is allowed to choose actions. We will also define the *block-game of credible play* $(\sigma, (b_k), T)^{\infty}$ as the infinite horizon problem in which a game of credible play restarts after T rounds of play (with discount factor δ). This test has similar properties to those of the test in the main text. In particular, the test allows player i to pass the test with high probability regardless of the strategy used by $-i$ just by using the obedient strategy. The proof is similar to that of Lemma 3 and therefore omitted.

B.1.4. *A Bertrand Example.* We now revisit the Bertrand example in Section 5.2, but assume that both firms have private information. More concretely, the demand functions are given by

$$Q_i(a, \theta_i) = \theta_i - a_i + z a_{-i}$$

where $\theta_i \in \{\underline{\theta}_i, \bar{\theta}_i\}$. We assume that each θ_i^t follows a Markov chain $P_i(\theta_i' \mid \theta_i)$, with $P_i(\bar{\theta}_i \mid \bar{\theta}_i) \geq P(\bar{\theta}_i \mid \underline{\theta}_i)$. We focus on equilibria that maximizes the sum of payoffs. It is relatively simple to see that up to integer constraints, the solution to the AROE takes the form

$$\sigma_i(\theta_i \mid p) = \frac{\theta_i}{2} + \frac{z}{2(1-z^2)}\Big(\mathbb{E}_{p_j}[\theta_j] + z\mathbb{E}_{p_i}[\theta_i]\Big)$$

with $p = (p_1, p_2) \in \Delta(\{\bar{\theta}_1, \underline{\theta}_1\}) \times \Delta(\{\bar{\theta}_2, \underline{\theta}_2\})$. This means that on the path of play, both firms signal their types. The larger the price fixed by i in t, the larger both firms' prices in $t+1$. In this model, both firms can become price leaders. Yet, if the types of one of the firms is independent across time, then that firm will not be a price leader as its price in t does not convey relevant information for $t+1$.