Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 176 (2018) 408–443

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Cooperation dynamics in repeated games of adverse selection ☆

Juan F. Escobar [a,*], Gastón Llanes [b]

[a] *Center for Applied Economics, Department of Industrial Engineering, University of Chile, Chile*
[b] *Pontificia Universidad Católica de Chile, Chile*

Received 4 April 2017; final version received 2 April 2018; accepted 6 April 2018
Available online 12 April 2018

Abstract

We study cooperation dynamics in repeated games with Markovian private information. After any history, signaling reveals information that helps players coordinate their future actions, but also makes the problem of coordinating current actions harder. In equilibrium, players may play aggressive or uncooperative actions that signal private information and partners tolerate a certain number of such actions. We discuss several applications of our results: We explain the cycles of cooperation and conflict observed in trench warfare during 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, and show that communication between cartel members may be socially efficient in a repeated Cournot game. Finally, we show that the welfare losses disappear as the persistence of the process of types increases and the interest rate goes to zero.

JEL classification: D01; D21; C72

Keywords: Repeated games; Private information; Signaling; Coordination; Collusion; Communication

☆ We thank Elton Dusha, Joe Harrington, Johannes Hörner, Romans Pancs, and several audiences for useful discussions. We are grateful to an associate editor and two referees for important comments and suggestions. Escobar acknowledges financial support from CONICYT (Fondecyt No. 1180723). 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.
* Corresponding author.
E-mail addresses: jescobar@dii.uchile.cl (J.F. Escobar), gaston@llanes.com.ar (G. Llanes).

https://doi.org/10.1016/j.jet.2018.04.004
0022-0531/

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 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 this paper, we shed light on this kind of phenomena by studying the scope for cooperation in a repeated game with private information. We assume the type profile follows an autonomous irreducible Markov chain where the evolution of types is independent across players. Values are private, actions are observable, and players cannot exchange cheap talk messages.[1] We show that the combination of private information and no 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 players to 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 result 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 a player is endogenously determined. Given any history of actions, a player 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 the player's actions, other players can move on to the next round with more precise beliefs about the type of player they will face. A second benefit from full revelation is that a player's payoff depends 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 not be optimal for the relationship: when a player is fully revealing his private information, it is harder for other players to predict his current action. The costs of revealing information at any given history are the losses that rival players incur when the 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 about types. 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 a 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

[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 illegal. Ashworth (1980) documents the communication problems faced by enemy troops trying to avoid confrontation during World War I, and Schelling (1960) explains that "an agreement on limits is difficult to reach … because communication becomes difficult between adversaries in a time of war."

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 players forgive apparently hostile actions – such as price reductions – but do not forget them. In equilibrium, each player keeps track of the number of actions played by others conditional on public beliefs, and (off-path) the relationship enters a punishment phase if the path of actions seems openly mischievous.

The equilibrium strategies exhibit dynamics that differ from those in previous literature. In an equilibrium with some information revelation, public beliefs determine the distribution over actions at any given history. Therefore, apparently uncooperative actions (such as price cuts and price wars in a collusion application) may occur on the path of play and be the optimal response of the relationship to incomplete information and no 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 formulation 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, etc. – and price discussions between competitors are illegal. We characterize an approximately 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 a price increase by one of the firms is followed by rivals. 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 local demand increases and the firm raises its price, it experiences a short-term loss until its price raise is matched by the rival. 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 when the discount factor is arbitrarily close to one.[2]

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 (2016) who show a strictly positive lower bound for the value of communication for the *cartel* in a repeated Bertrand game with private monitoring.

Our analysis is refined by studying a prisoners dilemma in which the length of the period parameterizes both the discount factor and the persistence of the process of types. This parameterization follows a tradition initiated by Abreu et al. (1991) for repeated games with moral imperfect monitoring. We show that as interactions become arbitrarily frequent and the interest rate goes to zero, signaling becomes inexpensive compared to the benefits from more precise beliefs and, as a result, incomplete information has virtually no costs.

In most repeated game models, it is never optimal to have players unilaterally choosing 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). Some recent exceptions are Mobius (2001) and Abdulkadiroğlu and Bagwell (2013), who assume that cooperative actions are sometimes unfeasible; Rahman (2014), who studies a collusion model with imperfect public monitoring in which unilateral price reductions may result in more informative signals; and Bernheim and Madsen (2017), who show a perfect monitoring repeated pricing game in which the best cartel arrangement is a mixed strategy path and price cuts occur with some probability. We view these results as complementary to ours.

Our results connect 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 paths of play are similar to those in 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 result that characterizes an approximately optimal equilibrium behavior in repeated games without communication. Further, our results identify new tradeoffs

[2] 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.

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

Fig. 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 $\beta > 0$.

and inefficiencies in repeated games with incomplete information, and can be applied to a variety of economic examples.[3]

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 a player's actions. For example, a low-price regime (or price war) may be triggered by a price cut, whereas returning to a high-price regime may require a unilateral price increase. 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 equilibrium path and there is little room for on-path renegotiation.[4]

The remainder of this paper is organized as follows. Section 2 provides examples that illustrate the model and results. Section 3 introduces the model. Section 4 presents the main theorems. Section 5 provides applications. Section 6 explores the model with frequent interactions and vanishing discount rates. Section 7 concludes. The Appendix provides further examples and proofs.

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 Fig. 1.

[3] 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 (2017) characterize the value of zero-sum games with Markovian private information.

[4] 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.

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

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

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. Fig. 2 shows the payoff matrix.

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.[5]

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 $\left(2\lambda(a-l) - (l-2b)(1-\lambda)\right)\frac{1}{2} > 0$ (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

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

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.[6] 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 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

[5] 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).

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

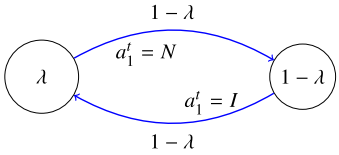


Fig. 3. Dynamics of beliefs $(p^t)_{t\geq 1}$ when player 1 uses the revealing strategy σ_1^R. The support of $(p^t)_{t\geq 1}$ is the set $\{\lambda, 1-\lambda\}$.

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 Fig. 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.

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).[7] In our construction of strategies, while player 2 can tolerate some failures (i.e., periods 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.[8] In other words, equilibrium strategies are so that player 2 forgives but does not forget failures.

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 $\big(2\lambda(a-l)-(l-2b)(1-\lambda)\big)\frac{1}{2}$. 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.[9] 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.

3. Model

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

[7] 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.

[8] 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.

[9] 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 repeated game without communication, Theorems 1 and 2 imply that transitions do matter to determine the equilibrium set. In Appendix A, we illustrate this observation by fulling solving for the limit equilibrium set in this game.

t.0 A randomization device χ^t is publicly realized
t.1 Player i is privately informed about $\theta_i^t \in \Theta_i$
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 i is $u_i(a, \theta_i)$. We will sometimes abuse notation and write $u_i(a, \theta)$ and $u(a, \theta) = (u_i(a, \theta_i))_{i=1}^n$. Players rank flows of payoffs according to $(1-\delta)\sum_{t\geq 1}\delta^{t-1}u_i(a^t, \theta_i^t)$, where $\delta < 1$ is the common discount factor. We assume that $|A_i| \geq |\Theta_i|$.

The realizations of the randomization device are independent across time and distributed according to the uniform distribution in $[0, 1]$. Each $(\theta_i^t)_{t\geq 1}$ is a Markov chain which is independent of the process $(\theta_{-i}^t)_{t\geq 1}$. The initial type of player i, θ_i^1, is drawn from a distribution $p_i^1 \in \Delta(\Theta)$. Player i's private types, $(\theta_i^t)_{t\geq 1}$, evolve with transition matrix P_i on Θ_i. We assume that the process of types has full support: for all $\theta_i, \theta_i' \in \Theta_i$, $P_i(\theta_i' \mid \theta_i) > 0$. Let $\pi_i \in \Delta(\Theta_i)$ be the stationary distribution for P_i.

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

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

where $u(a, \theta) = (u_1(a, \theta_1), \ldots, u_n(a, \theta_n))$. Let

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

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

The definitions of strategies and set of feasible payoffs differ from those used in stochastic games (Dutta, 1995; Hörner et al., 2011) 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 i decides based only on the sequence of actions, his own private types, and public randomizations in the game.

A strategy profile $s^* = (s_1^*, \ldots, s_n^*)$ 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 $\mathcal{E}(\delta, p^1) \subseteq \mathbb{R}^n$. It follows that $\mathcal{E}(\delta, p^1) \subseteq V(\delta, p^1)$ for all $\delta < 1$.

4. Equilibrium analysis

We will describe an 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.

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 $\alpha \in \mathbb{R}^n_{++}$, with $\sum_{i=1}^n \alpha_i = 1$, 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 $v^{\alpha,\delta} = v^\delta(s^{\alpha,\delta}) \in \mathbb{R}^n$ is an *efficient payoff vector*.[10]

Solutions to (4.1) can be found using dynamic programming tools. To see this, take the belief p_i^1 that a player $j \neq i$ has about player i's type at the beginning of the game. After player i's action is observed, the strategies also determine the belief p_i^2 that player $j \neq i$ has about the new type that player i has 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 weighted 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 $\Sigma_i = \{\sigma_i : \Theta_i \to A_i\}$ be a set of *controls* for player i and let $\Sigma = \Sigma_1 \times \ldots, \Sigma_n$.[11] An element $\sigma \in \Sigma$ is a *control profile*. Note that since types are independent, the belief about player i's types that players j and k have coincide (with $j \neq i \neq k$). The independence assumption also guarantees that the set of beliefs can be represented as the product set $\prod_{i=1}^n \Delta(\Theta_i)$. Let $p = (p_1, \ldots, p_n) \in \prod_{i=1}^n \Delta(\Theta_i)$ be a belief profile so that p_i is the belief that any player $j \neq i$ has about player i's type. Let $p_i(\theta_i)$ denote the θ_i-element of p_i. For $\sigma \in \Sigma$ and $p \in \prod_{i=1}^n \Delta(\Theta_i)$, we define the vector of expected period utilities $U(\sigma, p) \in \mathbb{R}^n$ by

$$U_i(\sigma, p) = \sum_{\theta \in \Theta} u_i(\sigma_1(\theta_1), \ldots, \sigma_n(\theta_n), \theta_i)\ p_1(\theta_1) \cdot p_2(\theta_2) \cdot \cdots \cdot p_n(\theta_n).$$

For $\alpha \in \mathbb{R}^n_{++}$, let $U^\alpha(\sigma, p) = \alpha \cdot U(\sigma, p) = \sum_{i=1}^n \alpha_i\ U_i(\sigma, p)$ be the ex-ante weighted sum of period payoffs given σ and beliefs p. We also define the *Bayes operator* $B_i(\cdot \mid \sigma_i, p_i, a_i) \in \Delta(\Theta_i)$ as

$$B_i(\theta_i' \mid \sigma_i, p_i, a_i) = \sum_{\{\theta_i \mid \sigma_i(\theta_i) = a_1\}} P(\theta_i' \mid \theta_i)\ \frac{p_i(\theta_i)}{\sum_{\{\hat{\theta}_i \mid \sigma_i(\hat{\theta}_i) = a_i\}} p_i(\hat{\theta}_i)} \tag{4.2}$$

whenever $\sigma_i(\hat{\theta}_i) = a_i$ for some $\hat{\theta}_i$ such that $p(\hat{\theta}_i) > 0$. In words, $B_i(\theta_i' \mid \sigma_i, p_i, a_i)$ is the probability that player $j \neq i$ assigns to $\theta_i^{t+1} = \theta_i'$ given that at the beginning of round t his belief about θ_i^t was p_i, player i uses the control $\sigma_i = \sigma_i(\theta_i^t)$, and player $j \neq i$ observed player i's action $a_i^t = a_i$. We write $B(\cdot \mid \sigma, p, a) = (B_i(\cdot \mid \sigma_i, p_i, a_i))_{i=1}^n$.

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

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

for all $p \in \prod_{i=1}^n \Delta(\Theta_i)$, with $p(\theta) = p_1(\theta_1) \ldots p_n(\theta_n)$. 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.

[10] 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 $\alpha \in \mathbb{R}^n_{++}$ is convex.

[11] The proof of Lemma 1 shows the restriction to pure strategies is without loss.

A *control rule* σ is such that for all $p \in \prod_{i=1}^{n} \Delta(\Theta_i)$, $\sigma(\cdot \mid p) \in \Sigma$. Using the control rule $\sigma^{\alpha,\delta}$, we can construct a (pure) strategy profile $s = s^{\alpha,\delta}$ from $\sigma^{\alpha,\delta}$ by setting[12]

$$s_i^t(a^1, \ldots, a^{t-1}, \theta_i^1, \ldots, \theta_i^t, \chi^1, \ldots, \chi^t) = \sigma_i^{\alpha,\delta}(\theta_i^t \mid p^t)$$

where p^t is the belief that players $j \neq i$ have about θ_i^t at the beginning of t and can be recursively computed as

$$p_i^{t+1}(\theta_i) = B_i(\theta_i \mid \sigma_i^{\alpha,\delta}(\cdot \mid p^t), p_i^t, a_i^t) \text{ for } t \geq 1.$$

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

Lemma 1. *Let $\alpha \in \mathbb{R}^n_{++}$ with $\sum_{i=1}^{n} \alpha_i = 1$. 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 explore 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 $p \in \prod_{i=1}^{n} \Delta(\Theta_i)$. Using this definition we can rewrite (4.3) as

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

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 $h^{\alpha}\colon \prod_{i=1}^{n} \Delta(\Theta_i) \to \mathbb{R}$, $w^{\alpha}\colon \prod_{i=1}^{n} \Delta(\Theta_i) \to \mathbb{R}$ 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)

$$h(p) + \rho = \max_{\sigma \in \Sigma} \left\{ U^{\alpha}(\sigma, p) + \sum_{a \in A} h\big(B(\cdot \mid \sigma, p, a)\big)\Big(\sum_{\theta \in \Theta, \sigma(\theta) = a} p(\theta)\Big) \right\} \tag{4.6}$$

for all $p \in \prod_{i=1}^{n} \Delta(\Theta_i)$. Let $\sigma^{\alpha}(\cdot \mid p) \in \Sigma$ be the control profile attaining the maximum in the dynamic programming problem (4.6) given p.

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

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

a. *The AROE (4.6) has a solution $(h^{\alpha}, \rho^{\alpha})$ and a control rule σ^{α} that attains the optimum.*

[12] This construction applies only for on-path histories. For off-path histories we define s arbitrarily.

[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 $\lim_{\nu \to \infty} w^{\alpha,\delta^\nu}(p) = \lim_{\nu \to \infty} w^{\alpha,\delta^\nu}(p^1)$.

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* $\bar{h}\colon \prod_{i=1}^n \Delta(\Theta_i) \to \mathbb{R}$ *is convex.*

c. *For any strategy* s, $\limsup_{\delta\to 1} \sum_{i=1}^n \alpha_i v_i^\delta(s) \leq \lim_{\nu\to\infty} w^{\alpha,\delta^\nu}(p^1) = \rho^\alpha$.

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 $p \in \prod_{i=1}^n \Delta(\Theta_i)$. As we show below, each of the two terms on the right-hand side of (4.6) is maximized either by a pooling or a separating rule.

A control rule σ is *separating* if for any belief $p \in \prod_{i=1}^n \Delta(\Theta_i)$ having positive probability in the path $(\theta^t, p^t)_{t\geq 1}$, types are separated: $\sigma_i(\theta_i \mid p) \neq \sigma_i(\theta_i' \mid p)$ for all $\theta_i \neq \theta_i'$ and all i. This means that all players' types can be perfectly inferred after observing their actions..

A separating control σ allows player i to fully *reveal* his type by setting a different action for each state of the world. The problem of maximizing player i's payoff $U_i(\sigma, p)$ typically results in a fully revealing strategy σ_i. A second benefit of perfect information revelation is that by fully separating his types in period t, player i makes continuation beliefs p_i^{t+1} more precise and therefore a player $j \neq i$ faces less uncertainty about θ_i^{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 $p_i', q_i' \in \Delta(\Theta_i)$ and $\lambda \in [0,1]$, $h(\lambda p_i' + (1-\lambda)q_i', p_{-i}) \leq \lambda h(p_i', p_{-i}) + (1-\lambda)h(q_i', p_{-i})$. If player i uses a separating rule σ_i in period t, he is fully resolving the uncertainty about θ_i^t at the end of round t and therefore maximizing $\sum_{a\in A} h\big(B_i(\cdot \mid \sigma_i, p_i, a_i), B_{-i}(\cdot \mid \sigma_{-i}, p_{-i}, a_{-i})\big)\big(\sum_{\theta\in\Theta, \sigma(\theta)=a} p(\theta)\big)$ over all $\sigma_i \in \Sigma_i$ keeping fixed σ_{-i}.

When player i pools, he does not reveal any information. The benefit of a pooling control is that it allows player $j \neq i$ to perfectly predict player i's current action. To see this, note that θ_i does not determine player j's current payoffs, and therefore the profile that maximizes player j's expected payoff $\max_{\sigma_i\in\Sigma_i} U_j(\sigma_i, \sigma_{-i}, p)$ will typically involve a pooling rule σ_i.

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.[14] The following result can be used to find those solutions in applications.

Proposition 1. *Consider* $p \in \prod_{i=1}^n \Delta(\Theta_i)$ *and a rule* $\bar{\sigma} = (\bar{\sigma}_1, \ldots, \bar{\sigma}_n)$ *such that for all* i *and all* $\theta_i \neq \theta_i'$, $\bar{\sigma}_i(\theta_i) \neq \bar{\sigma}_i(\theta_i')$ *and*

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

Then,

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

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

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.[15]

Finally, we use Theorem 1 to deduce an upper bound for the limit equilibrium set.[16]

Corollary 1.

$$\limsup_{\delta \to 1} \mathcal{E}(\delta, p^1) \subseteq \limsup_{\delta \to 1} V(\delta, p^1) \subseteq \bigcap_{\alpha \in \mathbb{R}^n_{++}} \left\{ v \in \mathbb{R}^n \mid \alpha \cdot v \leq \rho^{\alpha} \right\}.$$

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 a player losses credibility if his behavior does not match the efficient strategy profile. From an applied perspective, this implies that there exists an 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

$$p_i^{t+1} = B_i(\cdot \mid \sigma_i(\cdot \mid p^t), p_i^t, a_i^t) \quad \forall t \geq 1.$$

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

The construction of equilibrium strategies is subtle because, on the one hand, we want to allow player i to use his private information but, on the other, allowing him to freely choose actions may open up the room for opportunistic behavior. However, players $j \neq i$ can keep an account of the frequencies with which player i 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 consider rules that generate well-behaved paths of beliefs.

Given a control rule σ and $\hat{T} \geq 1$, we build an *extended control rule that reveals* every $\hat{T}$ rounds as a policy $\hat{\sigma}^{\hat{T}} = (\hat{\sigma}_i^{\hat{T}})_{i=1}^n$ defined by

$$\hat{\sigma}_i^{\hat{T}}(\theta_i^t \mid p^t, \kappa^t) = \begin{cases} \sigma_i(\theta_i^t \mid p^t) & \text{if } \kappa^t < \hat{T} \\ \theta_i^t & \text{if } \kappa^t = \hat{T} \end{cases}$$

where $\kappa^t = \mod_{\hat{T}}(t)$, $p_i^{t+1} = B_i(\cdot \mid \hat{\sigma}_i^{\hat{T}}(\cdot \mid p^t), p_i^t, a_i^t)$, and we assume that $\Theta_i \subseteq A_i$ for all i.[17] In words, the extended control rule $\hat{\sigma}^{\hat{T}}$ is exactly like σ, but every $\hat{T}$ rounds, $\hat{\sigma}^{\hat{T}}$ reveals all

[15] In Appendix A, we provide an example in which even when σ^{α} separates some types, $\sigma^{\alpha} \neq \bar{\sigma}$. In the example, σ^{α} separates to generate more precise continuation beliefs.

[16] Recall that for a given a sequence of sets $(X_n)_{n \in \mathbb{N}}$, $x \in \limsup_{n \to \infty} X_n$ if and only if there exists a sequence $x_k \in X_{n_k}$, where $(n_k)_{k \in \mathbb{N}}$ goes to infinity, such that $x_k \to x$.

[17] This is without loss since we already assumed $|A_i| \geq |\Theta_i|$ for all i. This is the only part of the construction of equilibrium strategies where this assumption is used. The notation $\mod_T(x)$ refers to the modulo T congruence.

players' types and resets the updating process.[18] When the control rule σ never separates types, $\kappa^t \in \{1, \ldots, \hat{T}\}$ can be interpreted as the number of rounds that has transpired since the last round in which $\hat{\sigma}^{\hat{T}}$ perfectly revealed.

Note that a control rule σ that separates generates the same path as the extended control rule that reveals every $\hat{T}$ rounds, for all $\hat{T}$. In this case, the path of continuation beliefs belongs to the set $\prod_{i=1}^{n}\{P_i(\cdot \mid \theta_i) \mid \theta_i \in \Theta_i\}$, the support of the process $(\theta^t, p^t)_{t\geq 1}$ is $\Theta \times (\{p^1\} \cup \prod_{i=1}^{n}\{P_i(\cdot \mid \theta_i) \mid \theta_i \in \Theta_i\})$ and its unique recurrence class is $\Theta \times \prod_{i=1}^{n}\{P_i(\cdot \mid \theta_i) \mid \theta_i \in \Theta_i\}$. On the other hand, when the rule pools all types along the path, the path of the Markov chain $(\theta^t, p^t)_{t\geq 1}$ is typically countably infinite. In this case, the control rule and the extended control rule that reveals every $\hat{T}$ rounds generate different paths.

The following result shows that relaxing the optimality requirement to allow for approximate efficiency is enough to ensure the existence of an extended control rule that reveals and generates well behaved paths.

Lemma 2. *The following hold:*

a. *Any extended control rule $\hat{\sigma}^{\hat{T}}$ determines a unique recurrence class, that is, the process $(\theta^t, p^t, \kappa^t)_{t\geq 1}$ is a finite Markov chain and has a unique recurrence class*[19]*;*

b. *For all $\epsilon > 0$, and all $\alpha \in \mathbb{R}^n_{++}$, there exist an extended control rule that reveals $\hat{\sigma}^{\hat{T}}$, and $\overline{T} \in \mathbb{N}$ such that*

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

for all $T \geq \overline{T}$, and all p in the (finite) path of beliefs generated by $\hat{\sigma}^{\hat{T}}$ and p^1. Moreover, if σ^{α} is separating, we can take the rounds of revelation to be $\hat{T} = 1$.

When the control rule σ^{α} solving the AROE perfectly reveals types, this lemma is immediate and we can simply $\bar{T} = 1$. 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.[20] This rule generates an infinite belief path. We can modify the rule so that after a sufficiently large number of periods, player 1's 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 extended control rule $\hat{\sigma}^{\hat{T}}$ determining a unique recurrence class, the limit-average payoff

[18] Note that $\hat{\sigma}^{\hat{T}}$ is not a control rule since it conditions on the payoff irrelevant variable κ^t. This justifies the term "extended" in the definition.

[19] The process $(\theta^t, p^t, \kappa^t)_{t\geq 1}$ is a finite Markov chain and has a unique recurrence class if there exists a finite set $\mathcal{Q} \subseteq \prod_{i=1}^{n}\Delta(\Theta_i) \times \{1, \ldots, \hat{T}\}$ such that $(\theta^t, p^t, \kappa^t)_{t\geq 1} \subseteq \Theta \times \mathcal{Q}$ and a unique subset $\mathcal{Q}' \subseteq \mathcal{Q}$ such that for all $(\theta, p, \kappa) \in \Theta \times \mathcal{Q}'$, if the Markov chain visits (θ, p, κ), then in the next period it will stay in $\Theta \times \mathcal{Q}'$ with probability 1, and no proper subset of $\mathcal{Q}'$ has this property. See Stokey et al. (1989) for additional discussion.

[20] 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.

$$v_i^{\infty}(\hat{\sigma}^{\hat{T}}) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} u_i(\hat{\sigma}_1^{\hat{T}}(\theta_1^t \mid p^t, \kappa^t), \ldots, \hat{\sigma}_n^{\hat{T}}(\theta_n^t \mid p^t, \kappa^t), \theta_i^t)]$$

is well defined. This follows from Proposition 8.1.1 in Puterman (2005) after noticing that the limit above is the average reward from a stationary Markov decision rule over a finite state Markov process. We define $v^{\infty}(\hat{\sigma}^{\hat{T}}) = (v_i^{\infty}(\hat{\sigma}^{\hat{T}}))_{i=1}^n$.

For an extended control rule that reveals $\hat{\sigma}^{\hat{T}}$ that determines a unique recurrence class $\Theta \times \mathcal{Q}$ and $q = (p, \kappa) \in \mathcal{Q}$, define $m_i^{\sigma}(\cdot \mid q) \in \Delta(A_i)$ as the distribution over actions given q:

$$m_i^{\hat{\sigma}^{\hat{T}}}(a_i \mid q) = \sum_{\{\theta_i \in \Theta_i \mid a_i = \hat{\sigma}_i^{\hat{T}}(\theta_i \mid q)\}} p_i(\theta_i).$$

For $a \in A$, we define $m^{\hat{\sigma}^{\hat{T}}}(a \mid q)$ analogously.

Given any sequence of actions $a^1, \ldots, a^t \in A$ and $\hat{\sigma}^{\hat{T}}$, we can mechanically calculate probabilities $\bar{p}_i^{t+1} = B_i(\cdot \mid \hat{\sigma}_i^{\hat{T}}, \bar{p}_i^t, a_i^t)$ (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 $\bar{p}^1 = p^1$. These *simulated probabilities* need not coincide with the beliefs a Bayesian agent would have about player i's types as his actions in the game could be derived from an arbitrary strategy s_i. We can also compute the occupancy rate of actions conditional as

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

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

$$\underline{v}_i = \min_{a_{-i} \in A_{-i}} \mathbb{E}_{\pi_i}[\max_{a_i \in A_i} u_i(a, \theta)].$$

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 many applications.[21] A vector $v \in \mathbb{R}^n$ is *strictly individually rational* if $v_i > \underline{v}_i$ for $i = 1, \ldots, n$.

Let $\mathbb{F} \subseteq \mathbb{R}^n$ be the Pareto-frontier of the set $\cap_{\alpha \in \mathbb{R}^n_{++}} \{v \in \mathbb{R}^n_{++} \mid \alpha \cdot v \leq \rho^{\alpha}\}$ (see Corollary 1). $\mathbb{F}$ is the set of all limit feasible payoffs that are efficient. Let $V^c = \{v \in \mathbb{R}^n \mid v = \mathbb{E}_{\pi}[u(a, \theta)]$ for some $a \in A\}$ be the set of average payoffs that can be attained using pooling profiles. Let

$$W = \text{co}\Big(\mathbb{F} \cup V^c\Big)$$

where co denotes the convex hull. We also denote $\underline{W} = W \cap \{v \in \mathbb{R}^n \mid v_i > \underline{v}_i$ for all $i\}$. $\underline{W}$ is the set of payoffs in W that are strictly individually rational.

Definition 1. A vector $v \in \mathbb{R}^n$ allows player-specific punishments in $\underline{W}$ if there exists a collection of payoff profiles $(v^i)_{i=1}^n \subseteq \underline{W}$ such that for all i, $v_i^i < v_i$ and $v_i^j > v_i^i$ for all $j \neq i$.

[21] Our definition of minmax is restrictive because it only considers pure strategies. Furthermore, when player j is minmaxing i, player j could find optimal to use the information revealed by player i during the minmaxing phase. This introduces complexities beyond the scope of the paper. See Pęski and Toikka (2017).

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

Theorem 2 *(Equilibrium Theorem). Fix $\epsilon > 0$ and $\alpha \in \mathbb{R}^n_{++}$. Take the extended control rule $\hat{\sigma}^{\hat{T}}$ as in Lemma 2 part b. Assume that $v = v^{\infty}(\hat{\sigma}^{\hat{T}})$ is strictly individually rational and that allows for player-specific punishments in $\underline{W}$. 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 $s^* = (s_1^*, \ldots, s_n^*)$ such that*

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

This result characterizes behavior under an approximately optimal equilibrium when players are sufficiently patient. 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 q, players equilibrium actions will approximate the frequencies induced by the approximately optimal rule $\hat{\sigma}^{\hat{T}}$. This means that the problem of determining approximately optimal equilibrium dynamics reduces to solving the dynamic programing problem AROE (4.6) and in applications, we can use $\hat{\sigma}^{\hat{T}}$ to approximately describe the distribution over public histories in equilibrium.[22]

Theorem 2 assumes that we can build player-specific punishments $v^1, \ldots, v^n \in \mathbb{R}^n$ (Fudenberg and Maskin, 1986). Since $v = v^{\infty}(\hat{\sigma})$ is strictly individually rational and approximately efficient, the existence of player-specific punishments follows immediately when W has full rank. Checking that W has full dimension is relatively easy as W contains all average payoffs generated using pooling rules.

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

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

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

$$N^t(q, a) = \sum_{t'=1}^{t} \mathbb{1}_{\{(\bar{p}^{t'}, \kappa^t, a^{t'}) = (q, a)\}}, \quad N^t(q, a_{-i}) = \sum_{t'=1}^{t} \mathbb{1}_{\{(\bar{p}^{t'}, \kappa^t, a^{t'}_{-i}) = (q, a_{-i})\}},$$

[22] Note that Theorem 2 does not describe how the private histories of types are mapped to actions in equilibrium. Theorem 2 only characterizes the distribution over public histories. We also observe that Theorem 2 shows a particular approximately efficient equilibrium, and does not nail down behavior under another equilibria.

$$\bar{m}^t(a_i \mid q, a_{-i}) = \frac{N^t(q,a)}{N^t(q,a_{-i})}.$$

The number $\bar{m}^t(a_i \mid q, a_{-i})$ is the empirical frequency of player i's actions conditional on $(\bar{p}^t, \kappa^t) = q$ and $a^t_{-i} = a_{-i}$.

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

$$\max_{a_i \in A_i} |m_i^{\hat{\sigma}}(a_i \mid q) - \bar{m}^t(a_i \mid q, a_{-i})| \leq b_{N^t(q,a_{-i})}$$

for all $a_{-i} \in A_{-i}$ and all $q \in \mathcal{Q}$. Given $T \geq 1$, $\hat{\sigma} = \hat{\sigma}^{\hat{T}}$ and a sequence (b_k), construct the *game of credible play* $(\hat{\sigma}, (b_k), T)$ 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 in the support of $m_i^{\hat{\sigma}}(\cdot \mid \bar{p}^t, \kappa^t)$; otherwise, a_i^t is an action randomly drawn from the distribution $m_i^{\hat{\sigma}}(\cdot \mid \bar{p}^t, \kappa^t)$ at each $t' = t, \ldots, T$. We define the *obedient strategy* $\hat{s}_i$ for player i as $\hat{s}_i^t(\theta_i^1, \ldots, \theta_i^t, a^1, \ldots, a^{t-1}) = \hat{\sigma}_i(\theta_i^t \mid \bar{p}^t, \kappa^t)$ whenever he is allowed to choose actions. We will also define the *block game of credible play* $(\hat{\sigma}, (b_k), T)^\infty$ as the infinite horizon problem in which a game of credible play restarts after T rounds of play (with discount factor δ).

Lemma 3. *Let $\eta > 0$. There exists a test (b_k) such that the following hold:*

a. *For any i and any $s_{-i} \in S_{-i}$,*

$$\mathbb{P}_{(\hat{s}_i, s_{-i})}[\textit{Player } i \textit{ passes the test } (b_k) \textit{ at } (a^1, \ldots, a^t) \textit{ for all } t] \geq 1 - \eta.$$

b. *There exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$ there exists $\bar{T}$ such that for all $T \geq \bar{T}$, for any strategy profile s in the block game of credible play $(\hat{\sigma}, (b_k), T)^\infty$ given discount δ,*

$$\mathbb{P}_s\left[\max_{a \in A, q \in \mathcal{Q}} |\bar{m}^\delta(a \mid q) - m^{\hat{\sigma}}(a \mid q)| < \eta\right] \geq 1 - \eta.$$

The first part of the lemma ensures that player i can pass the test using the obedient strategy $\hat{s}_i$. The second part ensures that the occupancy rate of actions is close enough to the distribution of actions drawn from σ given simulated beliefs regardless of the strategies actually used.

To establish Theorem 2, we use this lemma to construct strategies that approximately result in 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 game of credible play. Lemmas 2 and 3 ensure that such path generates welfare close enough to the target ρ^α. 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 another game of credible play. Further deviations trigger new punishment phases.

The construction of equilibrium strategies is closely related to the 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.

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.[23] Enemy soldiers could not discern if such attacks were caused by military orders from high command or by opportunistic behavior.[24] 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.[25] 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 2's payoffs are common knowledge. 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 Fig. 4.

[23] 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).

[24] 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.

[25] 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).

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

Fig. 4. 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.[26] The term $-\theta^t K$ captures the cost that 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

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

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 $\mathrm{P}[\theta^1 = 0] = \lambda$.

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.[27] 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 $2R - K < G - C$.

[26] 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."

[27] 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.

Lemma 4 characterizes optimal dynamics. We say that a control rule $\sigma : [0,1] \to \{S, NS\}^3$ 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 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.[28]

Let σ^α be the control rule solving AROE (4.6) for $\alpha = (1/2, 1/2)$. The following result characterizes the optimal path.

Lemma 4. *If $\lambda < \frac{C-G}{2R+C-G}$, $\sigma^{(1/2,1/2)}$ has both battalions playing S on the path of play. If $\lambda > \frac{C-G}{2R+C-G}$, $\sigma^{(1/2,1/2)}$ 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 $\sigma^{(1/2,1/2)}$ generates reactive-signaling dynamics) or (S, S, S) (in which case $\sigma^{(1/2,1/2)}$ generates time-off dynamics).[29]

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.

[28] 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.

[29] 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 $a_i \in A_i \subseteq \mathbb{R}$ at each $t = 1, 2 \ldots$. Firms sell heterogeneous goods. The demand functions are given by

$$Q_i(a_i, a_j, \theta_i) = \max\{\theta_i - a_i + z a_j, 0\}$$

with $0 < z < 1$. Firms' marginal costs equal $c > 0$. Firm i's demand shock is private information $\theta_i \in \Theta_i \subseteq \mathbb{R}_+$. Players' utility functions take the form

$$u_i(a_i, a_j, \theta_i) = \Big(Q_i(a_i, a_j, \theta_i) - c\Big)\, a_i.$$

A higher demand in a given period makes more likely a higher demand in ensuing rounds: Endowing $\Delta(\Theta_i)$ with the first-order stochastic dominance order, $P_i(\cdot \mid \theta_i) \in \Delta(\Theta_i)$ increases as θ_i increases. Each player's action set is rich enough in the sense that $0, c \in A_i$. We finally assume that firm i can drive firm j's demand to 0: there exists $\bar{a}_i \in A_i$ such that $Q_j(\bar{a}_i, a_j, \theta_i) = 0$ for all $a_j \in A_j$ with $a_j \geq c$ and all $\theta_i \in \Theta_i$.[30] We write $A_i = \{a_i^1, \ldots, a_i^{|A_i|}\}$ and $\Theta_i = \{\theta_i^1, \ldots, \theta_i^{|\Theta_i|}\}$.

We begin our analysis by characterizing the controls that maximize current expected payoffs.

Lemma 5. *Fix $p \in \Delta(\{\underline{\theta}, \bar{\theta}\}) \times \Delta(\{\underline{\theta}, \bar{\theta}\})$ with $p_i(\theta_i) > 0$ for all $\theta_i \in \Theta_i$. Any solution $\bar{\sigma} \in \Sigma$ to*

$$\max_{\sigma \in \Sigma} \frac{1}{2} \sum_{\theta \in \Theta} \Big[\sigma_1(\theta_1) Q_1(\sigma_1(\theta_1), \sigma_2(\theta_2), \theta_1) + \sigma_2(\theta_2) Q_2(\sigma_1(\theta_1), \sigma_2(\theta_2), \theta_2)\Big] p_1(\theta_1) p_2(\theta_2)$$

is such that $\bar{\sigma}_i(\theta_i)$ is nondecreasing for $i = 1, 2$. When

$$\max_{k=2,\ldots,|A_i|} \{a_i^k - a_i^{k-1}\} \leq \frac{1}{4} \min_{k=2,\ldots,|\Theta_i|} \{\theta_i^k - \theta_i^{k-1}\} \tag{5.1}$$

for $i = 1, 2$, then $\bar{\sigma}_i(\theta_i)$ is strictly increasing. The set of solutions

$$\arg\max_{\sigma \in \Sigma} \frac{1}{2} \sum_{\theta \in \Theta} \Big[\sigma_1(\theta_1) Q_1(\sigma_1(\theta_1), \sigma_2(\theta_2), \theta_1) + \sigma_2(\theta_2) Q_2(\sigma_1(\theta_1), \sigma_2(\theta_2), \theta_2)\Big] p_1(\theta_1) p_2(\theta_2)$$

is nondecreasing in $p = (p_1, p_2) \in \Delta(\{\underline{\theta}, \bar{\theta}\}) \times \Delta(\{\underline{\theta}, \bar{\theta}\})$.[31]

This result is driven by the complementarity between prices and types. Noting that under (5.1) the grid of prices is rich enough, firms' types are separated and therefore any solution $\bar{\sigma}$ also solves the AROE equation (4.5) as a result of Proposition 1. We thus deduce that the optimal control rule σ^α, with $\alpha = (1/2, 1/2)$, separates types and induces a unique recurrence class $(\theta^t, p^t)_{t \geq 1}$. It is relatively simple to see that $\underline{v}_i = 0$ for $i = 1, 2$ and therefore the expected average payoff generated by $\sigma^{(1/2,1/2)}$, $v = v^\infty(\sigma^\alpha)$, is strictly individually rational. The set W has full dimension[32] and therefore v allows for player-specific punishments in $\underline{W}$. We use

[30] Given the demand, $\bar{a}_i < 0$.

[31] This means that if p and p' are such that p_i dominates in the first order stochastic dominance sense p_i' for $i = 1, 2$, and σ (resp. σ') solves the problem for p (resp. p'), then $\sigma_i(\theta_i) \geq \sigma_i'(\theta_i)$ for all θ_i.

[32] To see this, let $\hat{a}_i > c$ and compute the expected average payoff w^i assuming that i sets price $\hat{a}_i$ and j sets price $a_j = c$ so that $w_i^i > 0$ and $w_j^i = 0$. It follows that $\{(0, 0), w^1, w^2\} \subseteq W$ and therefore W has full rank.

Theorem 2 to build an equilibrium s^* resulting in payoffs arbitrarily close to $\rho^{(1/2,1/2)}$ such that the observed behavior under s^* is close to the observed behavior under $\sigma^{(1/2,1/2)}$.

Under strategy s^*, when firm i chooses a high price in period t, then firms' prices are higher in $t+1$ (keeping fixed the type of firm j in t). A price increase by firm i in t is seen as an invitation to switch to a high-price regime in $t+1$.[33] 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.*

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 signal 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.[34]

In our model, price cuts and price leadership are imperfect substitutes to explicit communication. Indeed, if firms could freely communicate, firms would exchange messages to coordinate their pricing decisions and firms would *simultaneously* raise or lower their prices. But without communication, prices are used as a signal of market conditions. Using prices to substitute communication entails a cost: Pricing decisions are uncoordinated and lack 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 (Nicholls, 1951; Stigler, 1947; Allen, 1976; Mouraviev and Rey, 2011; Seaton and Waterson, 2013). 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%).[35]

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.[36] In all these works, firms can be arbitrarily close to the first best collusive outcome, in which only

[33] This type of dynamics also arises in the alternating-move Bertrand model in Maskin and Tirole (1988).

[34] Collusion and price cuts can also arise in a mixed strategy equilibrium of a repeated Bertrand game (Bernheim and Madsen, 2017).

[35] Note that price leadership could also arise with private information even if firms did not collude.

[36] 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.

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.[37]

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

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

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 $\kappa \in \Theta \cup \{c\}$ as

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

Note that $Q^M(\kappa)$ decreases in κ. We assume that $Q^M(\kappa)$ is strictly decreasing and 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 approximately 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 an 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:

[37] 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.

$$\max_{\sigma_1:\ \Theta\to A_1,\sigma_2\in A_2} U^{(1,1)}(\sigma, p) := \sum_{\theta\in\Theta}\Big(\mathcal{P}(\sigma_1(\theta)+\sigma_2)(\sigma_1(\theta)+\sigma_2) - \theta\sigma_1(\theta) - c\sigma_2\Big)p(\theta) \tag{5.2}$$

Assume $\mathbb{E}_p[\theta] := \sum_\theta \theta p(\theta) < c$. Then, for any solution of (5.2), $\sigma_2 = 0$. If not, $\sigma_2 > 0$. Take the alternative profile $\tilde{\sigma}_2 = \sigma_2 - g$ and $\tilde{\sigma}_1(\theta) = \sigma_1(\theta) + g$.[38] The difference in 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 ex-ante more efficient produce the monopoly quantity. Assuming that $\mathbb{E}_p[\theta] < c$ for all $p \in \{p^1\} \cup_{\theta\in\Theta} \{P(\cdot \mid \theta)\}$, the optimal control rule $\sigma^{(1/2,1/2)}$ is separating and can be implemented as an equilibrium of the repeated game using Theorem 2.[39]

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 $Q^M(\theta)$. 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 (2016) 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.[40]

6. Equilibrium as interactions become frequent and discount rates vanish

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

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

[39] 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 can also construct player specific punishments as in Section 5.2.

[40] 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. Our finding depends on the assumption that firms perfectly collude (either with or without communication). Our results are thus complementary to those in Shapiro (1986), who explores the value of communication in static oligopoly games.

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 game of incomplete information, the impact of more frequent moves is also subtle as types are more likely to remain unchanged between two rounds. In this section, we illustrate these differences in a simple prisoners' dilemma.

Two players choose actions at each $t = D, 2D, \ldots$, where $D > 0$ is the period length. At each t, players play a prisoners dilemma, with the payoffs given in Fig. 2. 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 $r \to 0$ for a fixed D. One may also ask what happens when $D \to 0$ keeping fixed the interest rate r fixed. In this section, we show that when D is small enough, we can approximate the full information payoffs when r is chosen sufficiently small.

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

$$w^D(p) = \max_{\sigma \in \Sigma} \left\{ (1 - e^{-rD}) U^{(1,1)}(\sigma, p) + e^{-rD} \sum_{a_1 \in \{I,N\}} w^D\left(B^D(\cdot \mid \sigma_1, p, a_1)\right) \sum_{\theta, \sigma_1(\theta) = a_1} p(\theta) \right\}. \tag{6.1}$$

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

Proposition 2. *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 D is small enough, and that the incentive costs are modest if we can also pick r to be sufficiently small.[41] Intuitively, when D is small, 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). Note that first best payoffs (with full information and perfect commitment) converges to $2(a - l)\frac{\chi}{\phi + \chi}$ as $D \to 0$ – the payoff attained in the game with frequent moves and low interest rate.

[41] This is also related to Skrzypacz and Toikka's (2015) finding that when trade is more frequent, the increase in the persistence of the process of types is detrimental for incentives in mechanism design.

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 players forgive but do 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 could extend our results to allow for action-dependent transitions. Another extension would be to allow for restricted or costly communication or communication 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.[42] 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 2 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

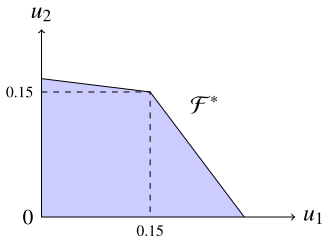
This Appendix consists of two parts. Appendix A provides some examples, and Appendix B provides proofs.

Appendix A. Examples

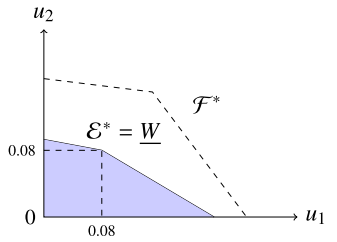
A.1. Limit equilibrium payoff set

Consider the example in Section 2.2, and suppose that $a = 1$, $b = 0.325$, $l = 0.7$, and $h = 4$. Fig. 5a shows the set of limit equilibrium payoffs in the game with complete information or incomplete information and communication ($\mathcal{F}^*$), which contains all feasible payoffs above the minmax vector $(0, 0)$. The point $(0.15, 0.15)$ shown in the graph corresponds to the strategy profile by which players play $\{I, I\}$ when the state is $\theta^t = l$ and $\{N, N\}$ when the state is $\theta^t = h$.

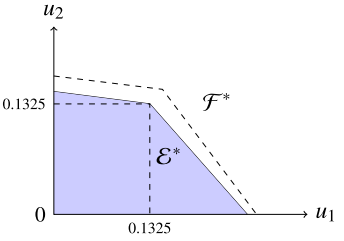
[42] This is similar to the model in Sandleris (2008), but in that model the game has a finite horizon and shocks are drawn once.



(A) Complete information or incomplete information with communication



(B) Incomplete information without communication ($\lambda = 0.6$)



(C) Incomplete information without communication ($\lambda = 0.9$)

Fig. 5. Limit sets of equilibrium payoffs.

Fig. 5b shows the set of limit equilibrium payoffs in the game with incomplete information and no communication ($\mathcal{E}^*$) for $\lambda = 0.6$. This set coincides with the set of individually rational payoffs $\underline{W}$ because as $\delta \to 1$ incentive issues disappear. The point $(0.08, 0.08)$ shown in the graph corresponds to a reactive signaling control profile: players attempt to coordinate on $\{I, I\}$ while the state is $\theta^t = l$ and on $\{N, N\}$ while the state is $\theta^t = h$. When the state changes from l to h, player 1 signals this change by playing N, when the state changes from h to l, player 1 signals by playing I. This means that players are playing $\{I, I\}$ a proportion $\frac{1}{2}\lambda$, $\{N, I\}$ a proportion $\frac{1}{2}(1-\lambda)$, $\{N, N\}$ a proportion $\frac{1}{2}\lambda$ of time, and $\{I, N\}$ a proportion $\frac{1}{2}(1-\lambda)$ of time.

To the left of $(0.08, 0.08)$ in the upper frontier of $\mathcal{E}^*$, players are mixing between the reactive signaling profile and a pooling control profile in which players play $\{I, I\}$ regardless of the state (such control profile yields $a - \frac{1}{2}l - \frac{1}{2}h = -1.35$ for player 1 and $a - l = 0.3$ for player 2). To the right of $(0.08, 0.08)$, players mix between the reactive signaling profile and a separating control in which player 1 plays I when the state is l and N when the state is h (such control profile yields $\frac{1}{2}(a-l) + \frac{1}{2}b = 0.3125$ for player 1 and $\frac{1}{2}(a-l) + \frac{1}{2}(b-l) = -0.0375$ for player 2).

Fig. 5c shows $\mathcal{E}^*$ for $\lambda = 0.9$. As λ approaches 1, the limit set of the game without communication converges to the limit set with communication.

A.2. Signaling even when payoffs do not change

Consider a game with two players $i = 1, 2$ and three states $\theta^t \in \{1, 2, 3\}$. The state is private information of player 1. Payoffs are given in Table 1.

Table 1
Game payoffs.

	A	B	C
A	5, 5	0, 0	0, 0
B	4, 4	4, 4	0, 0
C	0, 0	0, 0	0, 0

$\theta^t \in \{1, 2\}$

	A	B	C
A	0, 0	0, 0	0, 0
B	0, 0	0, 0	0, 0
C	0, 0	0, 0	5, 5

$\theta^t = 3$

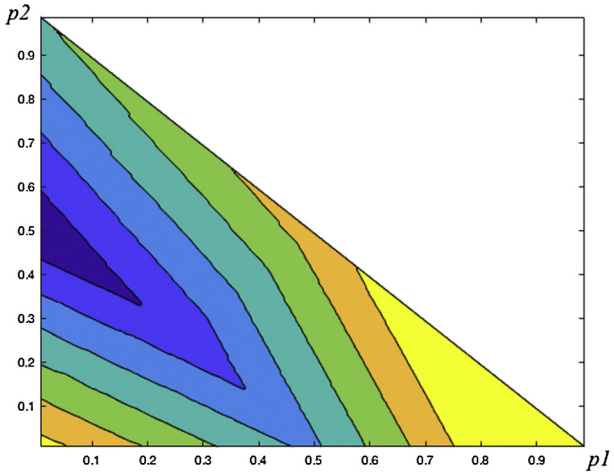


Fig. 6. Level curves for the value function. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

The game is a coordination game: From a static point of view, it is optimal to coordinate on $\{A, A\}$ when the state is 1 or 2, and to coordinate on $\{C, C\}$ when the state is 3. Transition probabilities are given by

$$P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.3 & 0.6 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

We solve the game numerically for $\delta = 0.999$. Fig. 6 shows level curves for the value function $v(p_1, p_2)$, which gives the optimal discounted sum of the sum of utilities as a function of the beliefs over states $\{1, 2\}$. Cooler colors (tending to blue) indicate smaller values for the value function. Warmer colors (tending to yellow) indicate larger values. The figure shows that value is largest when players are very sure that the state is 1 or 3, which is the expected outcome.

Gameplay evolves so that player 1 signals its type and player 2 plays A when the likelihood of state 1 is high, and C when the likelihood of states 2 and 3 is high. If player 2 is sure that state was 1 in $t - 1$, for example, then at t she believes that the state is 1 with probability 0.8, 2 with probability 0.1 and 3 with probability 0.1. Our calculations show that in this case, it is optimal for player 1 to play A if the state is 1, B if the state is 2 and C if the state is 3, while player 2 plays A.

The interesting feature of this example is that player 1 optimally signals a change from state 1 to state 2, even though these states have the same payoffs. The reason is that a change to state 2

makes a transition to state 3 very likely, which induces players to coordinate on playing C, even before the transition to state 3 has happened.

Appendix B. Proofs

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

B.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 strategy profile 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. The only caveat is that our dynamic programming equation is multilinear in beliefs as types are independent. To apply the vanishing discount approach, we extend equation (4.3) to allows for any belief $p \in \Delta(\Theta)$ and note that any solution must also be a solution when we restrict $p \in \prod_{i=1}^n \Delta(\Theta_i)$. 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). □

Proof of Proposition 1. Consider the problem

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

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 (Aumann and Maschler, 1995; 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. □

B.2. Proofs for Section 4.2

Proof of Lemma 2. Let $Q^t(p) \subseteq \prod_{i=1}^n \Delta(\Theta_i)$ be the finite set of beliefs having positive probability under $\bar{\sigma}^\alpha$ at round t given $p^1 = p$ when $\hat{\sigma}^{\hat{T}}$ is used (so that $Q^1(p) = \{p\}$). To prove a, note that all elements in

$$\cup_{t=1}^{\hat{T}} \left(\Theta \times \cup_{\theta \in \Theta} Q^t(P(\cdot \mid \theta)) \times \{t\} \right)$$

are visited infinitely often by the Markov chain $(\theta^t, p^t, \kappa^t)_{t\geq 1}$. Since the finite Markov chain $(\theta^t, p^t, \kappa^t)_{t\geq 1}$ visits the set $\Theta \times \prod_{i=1}^{n} \cup_{\theta_i \in \Theta_i} \{P_i(\cdot \mid \theta_i)\} \times \{1\}$ with probability 1 at rounds $\hat{T} + 1, 2\hat{T} + 1, \ldots$, the set $\cup_{t=1}^{\hat{T}} (\Theta \times \cup_{\theta\in\Theta} Q^t(P(\cdot \mid \theta)) \times \{t\})$ is its unique recurrence class.

To prove b, let σ^α be the control rule solving the AROE given α. In particular, there exists $\hat{T} \in \mathbb{N}$ such that

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\sigma^\alpha, p}[\alpha \cdot u(a^t, \theta^t)] \geq \rho^\alpha - \epsilon/4 \tag{B.1}$$

for all $T \geq \hat{T}$, and all $p \in \{p^1\} \cup \left(\cup_{\theta\in\Theta} \{P(\cdot \mid \theta)\}\right)$. Choose $\hat{T}$ large enough so that

$$\frac{1}{\hat{T}} \max_{a,a'\in A,\theta,\theta'\in\Theta} \{\alpha \cdot (u(a, \theta) - u(a', \theta'))\} < \frac{\epsilon}{4}. \tag{B.2}$$

Build the extended control rule that reveals every $\hat{T}$ rounds, $\hat{\sigma}^{\hat{T}}$, from σ^α. Note that for any $n \in \mathbb{N}$, and any $p, \bar{p}^1 \in \{p^1\} \cup \left(\cup_{\theta\in\Theta} \{P(\cdot \mid \theta)\}\right)$

$$\begin{aligned}
&\frac{1}{\hat{T}} \sum_{t=\hat{T}n+1}^{\hat{T}(n+1)} \mathbb{E}_{\hat{\sigma}^{\hat{T}}, p}[\alpha \cdot u(a^t, \theta^t) \mid p_{\hat{T}n+1} = \bar{p}^1] - \frac{1}{\hat{T}} \sum_{t=\hat{T}n+1}^{\hat{T}(n+1)} \mathbb{E}_{\sigma^\alpha, p}[\alpha \cdot u(a^t, \theta^t) \mid p_{\hat{T}n+1} = \bar{p}^1] \\
&= \frac{1}{\hat{T}} \Big(\mathbb{E}_{\hat{\sigma}^{\hat{T}}, p}[\alpha \cdot u(a^{\hat{T}(n+1)}, \theta^{\hat{T}(n+1)}) \mid p_{\hat{T}n+1} = \bar{p}^1] \\
&\quad - \mathbb{E}_{\sigma^\alpha, p}[\alpha \cdot u(a^{\hat{T}(n+1)}, \theta^{\hat{T}(n+1)}) \mid p_{\hat{T}n+1} = \bar{p}^1] \Big) \\
&> -\epsilon/4
\end{aligned}$$

where the equality follows since at rounds $t = \hat{T}n + 1, \ldots, \hat{T}(n + 1) - 1$, $\hat{\sigma}^{\hat{T}}$ and σ^α prescribe the same actions. Using (B.1)

$$\begin{aligned}
&\frac{1}{\hat{T}} \sum_{t=\hat{T}n+1}^{\hat{T}(n+1)} \mathbb{E}_{\hat{\sigma}^{\hat{T}}, p}[\alpha \cdot u(a^t, \theta^t) \mid p_{\hat{T}n+1} = \bar{p}^1] \\
&\geq \frac{1}{\hat{T}} \sum_{t=\hat{T}n+1}^{\hat{T}(n+1)} \mathbb{E}_{\sigma^\alpha, p}[\alpha \cdot u(a^t, \theta^t) \mid p_{\hat{T}n+1} = \bar{p}^1] - \frac{\epsilon}{4} \geq \rho^\alpha - \frac{1}{2}\epsilon
\end{aligned}$$

and therefore

$$\frac{1}{\hat{T}} \sum_{t=\hat{T}n+1}^{\hat{T}(n+1)} \mathbb{E}_{\hat{\sigma}^{\hat{T}}, p}[\alpha \cdot u(a^t, \theta^t)] \geq \rho^\alpha - \frac{1}{2}\epsilon. \tag{B.3}$$

Now, pick $\overline{T}$ such that

$$\hat{T}/T \max_{a,\theta} |\alpha \cdot u(a, \theta)| < \epsilon/4 \text{ and } \frac{\hat{T}}{T} \lfloor \frac{T}{\hat{T}} \rfloor (\rho^\alpha - \frac{1}{2}\epsilon) > \rho^\alpha - \frac{3}{4}\epsilon \tag{B.4}$$

for all $T > \overline{T}$ (here $\lfloor x \rfloor = \max\{y \in \mathbb{N} \mid y \leq x\}$). As a result, for all $T > \overline{T}$

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\hat{\sigma}\hat{T},p}[\alpha\cdot u(a^t,\theta^t)]\\
&=\frac{1}{T}\Big(\sum_{n=0}^{\lfloor T/\hat{T}\rfloor-1}\sum_{t=\hat{T}n+1}^{\hat{T}(n+1)}\mathbb{E}_{\hat{\sigma}\hat{T},p}[\alpha\cdot u(a^t,\theta^t)]+\sum_{t=1+\hat{T}\lfloor T/\hat{T}\rfloor}^{T}\mathbb{E}_{\hat{\sigma}\hat{T},p}[\alpha\cdot u(a^t,\theta^t)]\Big)\\
&=\frac{\hat{T}}{T}\Big(\sum_{n=0}^{\lfloor T/\hat{T}\rfloor-1}\frac{1}{\hat{T}}\sum_{t=\hat{T}n+1}^{\hat{T}(n+1)}\mathbb{E}_{\hat{\sigma}\hat{T},p}[\alpha\cdot u(a^t,\theta^t)]\Big)+\frac{1}{T}\sum_{t=1+\hat{T}\lfloor T/\hat{T}\rfloor}^{T}\mathbb{E}_{\hat{\sigma}\hat{T},p}[\alpha\cdot u(a^t,\theta^t)]\\
&\geq\frac{\hat{T}}{T}\lfloor\frac{T}{\hat{T}}\rfloor(\rho^{\alpha}-\frac{1}{2}\epsilon)-\frac{\hat{T}}{T}\max_{a,\theta}|\alpha\cdot u(a,\theta)|\\
&>\rho^{\alpha}-\epsilon
\end{aligned}
$$

where the first inequality follows from (B.3) and the second inequality follows from (B.4). This proves the result. □

Proof of Lemma 3. We prove a. Use Lemma B.1 in Escobar and Toikka (2013) (which provides a rate of convergence in Glivenko–Cantelli theorem) to deduce the existence of a test (b_k) such that regardless of the strategy followed by $-i$, whenever player i's actions are not changed

$$
\mathrm{P}_{\hat{s}_i,s_{-i}}\Big[\max_{\theta_i\in\Theta_i}|\bar{m}^t(\theta_i\mid q,a_{-i})-p_i(\theta_i)|<\frac{b_{N^t(q,a_{-i})}}{|\Theta|}\quad\forall q\in\mathcal{Q},a_{-i}\in A_{-i},\forall t\Big]>1-\eta
$$

where $\bar{m}^t(\theta_i\mid q,a_{-i})=\frac{\sum_{t'=1}^{t}\mathbb{1}_{(\theta_i^t,\vec{p}^t,\kappa^t,a_{-i}^t)=(\theta_i,q,a_{-i})}}{\sum_{t'=1}^{t}\mathbb{1}_{(\vec{p}^t,\kappa^t,a_{-i}^t)=(q,a_{-i})}}$ is the empirical frequency of player i's types. Note that whenever i is obedient

$$
\bar{m}^t(a_i\mid q,a_{-i})-m^{\hat{\sigma}}(a_i\mid q)=\sum_{\theta_i\in\Theta_i\ \mathrm{st}\ \hat{\sigma}_i(\theta_i|q)=a_i}\Big(\bar{m}^t(\theta_i\mid q,a_{-i})-p_i(\theta_i)\Big).
$$

As a result,

$$
\mathrm{P}_{\hat{s}_i,s_{-i}}\Big[\max_{a_i\in A_i}|\bar{m}^t(a_i\mid q,a_{-i})-m_i^{\hat{\sigma}}(a_i\mid q)|<b_{N^t(q,a_{-i})}\quad\forall q\in\mathcal{Q},a_{-i}\in A_{-i},\forall t\Big]>1-\eta.
$$

Now, if player i actually plays the game of credible play, then in the event above he will pass the test and his actions will not be modified.

To prove b, define $c_k=2\max_{1\leq j\leq k}jb_k/k+1/k$ and use Lemma B.5 in Escobar and Toikka (2013) to deduce in the game of credible play $(\hat{\sigma},(b_k),T)$, for all strategy profile s,[43]

$$
\begin{aligned}
&\mathrm{P}_s\Big[\max_{a_i\in A_i}|\bar{m}^T(a_i\mid q,a_{-i})-m_i^{\hat{\sigma}}(a_i\mid q)|<c_{N^t(q,a_{-i})}\quad\forall i=1,\ldots,n,q\in\mathcal{Q},a_{-i}\in A_{-i}\Big]\\
&>1-\eta.
\end{aligned}
$$

It follows that there exists $\overline{T}$ such that for $T>\overline{T}$, in the game of credible play $(\hat{\sigma},(b_k),T)$, for all strategy profile s,

[43] This result says that regardless of the strategy profile used, all players will pass the relaxed test c_k. Intuitively, when a player fails the test, the ensuing path of actions is generated using the target rule $\hat{\sigma}_i$ (which by definition passes the test b_k).

$$\mathrm{P}_s\Big[\max_{a\in A}|\bar{m}^T(a\mid q)-m^{\hat{\sigma}}(a\mid q)|<\eta/2\quad\forall q\in Q\Big]>1-\eta. \tag{B.5}$$

Now, for any such $T\geq\overline{T}$, pick $\bar{\delta}<$ such that for all $\delta>\bar{\delta}$, and every path $a^1,\ldots,a^T$,

$$\max_{a\in A}|\bar{m}^{T,\delta}(a\mid q)-\bar{m}^T(a\mid q)|<\eta/2$$

where

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

is the discounted finite horizon occupancy rate. As a result, for any block game of credible play $(\hat{\sigma},(b_k),T)^\infty$, for all strategy profile s, and for any element of the event in (B.5),

$$\begin{aligned}&\max_{a\in A}|\bar{m}^{\delta}(a\mid q)-m^{\hat{\sigma}}(a\mid q)|\\&=\max_{a\in A}\Big|\frac{\sum_{n=1}^{\infty}\Big(\frac{\sum_{t=nT+1}^{T(n+1)}\delta^{t-1}\mathbb{1}_{a^t=1,q^t=q}}{\sum_{t=nT+1}^{n(T+1)}\delta^{t-1}\mathbb{1}_{q^t=q}}-m^{\hat{\sigma}}(a\mid q)\Big)\sum_{t=nT+1}^{n(T+1)}\delta^{t-1}\mathbb{1}_{q^t=q}}{\sum_{t=1}^{\infty}\delta^{t-1}\mathbb{1}_{q^t=q}}\Big|\\&<\eta\end{aligned}$$

This completes the proof. □

Proof of Theorem 2. Before constructing the equilibrium strategies, we use Lemmas 2 and 3 to build a test (b_k) and $\bar{T}$ such that for all $T\geq\bar{T}$ there exists $\bar{\delta}$ such that for all $\delta>\bar{\delta}$, player i can get a payoff which is at least $v_i^\infty(\hat{\sigma})-\epsilon$ by playing obediently in the block game of credible play $(\hat{\sigma},(b_k),T)^\infty$ (this holds regardless of the strategies used by $-i$). In particular, in any equilibrium s^δ of $(\hat{\sigma},(b_k),T)^\infty$, player i's equilibrium payoff satisfies $v_i^\delta(s^\delta)\geq v_i^\infty(\hat{\sigma})-\epsilon$. Thus,

$$\alpha\cdot v^\delta(s^\delta)\geq\alpha\cdot v^\infty(\hat{\sigma})\geq\rho^\alpha-2\epsilon.$$

This means that any equilibrium of the block game of credible play $(\hat{\sigma},(b_k),T)^\infty$ results in total weighted payoffs that are at most 2ϵ below the target ρ^α.

Now, since $v^\infty(\hat{\sigma})$ allows for player-specific punishments in $\underline{W}$, we can find $(v^i)_{i=1}^n\subseteq\underline{W}$ such that $v_i>v_i^i$ and $v_i^j>v_i^i$ for $j\neq i$. Since $v^i\in W$, we can use Lemmas 2 and 3 to find a (perhaps randomized) extended control rule $\hat{\sigma}^i$ and build a block game of credible play $(\hat{\sigma}^i,(b_k),T^i)^\infty$ in which players get equilibrium payoffs arbitrarily close to v^i for all δ sufficiently large.

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 restarts 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 $(\hat{\sigma}^i,(b_k),T^i)^\infty$. Deviations restart the minmax phase and so on.

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

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

for $i, j = 1, \ldots, n$ with $i \neq j$. Define the length of the stick phase as $L = L(\delta) = \max\{d \in \mathbb{N} \mid d \leq \frac{\ln(\gamma)}{\ln(\delta)}\}$ 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. Define $M = \max_{i=1,\ldots,n, a\in A, \theta_i \in \Theta_i} u_i(a,\theta)$ and $m = \min_{i=1,\ldots,n, a\in A, \theta_i\in\Theta_i} u_i(a,\theta)$.

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.

Since the path of play of the strategy profile s^* coincides with the equilibrium s^δ of $(\hat{\sigma}, (b_k), T)^\infty$, it follows that $\alpha \cdot v^\delta(s^*) \geq \rho^\alpha - \epsilon$. Lemma 3 part b also implies that s^* satisfies the second part of the Theorem. This concludes the proof. □

B.3. 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 $p > (1-\mu)$.[44] 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 $\widehat{\tau} = \infty$. □

B.4. A proof for Section 5.2

Proof of Lemma 5. Since $p_i(\theta_i) > 0$, $\bar{\sigma}_i(\theta_i)$ is a solution to

$$\max_{a_i \in A_i} \sum_{\theta_j \in \Theta_j} \Big(u_i(a_i, \bar{\sigma}_j(\theta_j), \theta_i) + u_j(\bar{\sigma}_j(\theta_j), a_i, \theta_j) \Big) p_j(\theta_j).$$

The fact that $\bar{\sigma}_i$ is nondecreasing follows since $u_i(a_i, a_j, \theta_i)$ has increasing differences in (a_i, θ_i).

[44] 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 $p > 1-\mu$.

To see the second part, fix $\theta_i = \theta_i^k$ with $k < |\Theta_i|$ and take $\tilde{a}_i = \max\{a_i < \bar{\sigma}_i(\theta_i^k)\}$ (which is well defined since $0, c \in A_i$). Since the objective function in the optimization problem is concave, it must be the case that

$$\frac{\partial}{\partial a_i} \sum_{\theta_j \in \Theta_j} \Big(u_i(\tilde{a}_i, \bar{\sigma}_j(\theta_j), \theta_i) + u_j(\bar{\sigma}_j(\theta_j), \tilde{a}_i, \theta_j) \Big) p_j(\theta_j) \geq 0$$

for otherwise $\bar{\sigma}_i(\theta_i^k)$ would not be optimal. Now, take $\hat{a}_i = \min\{a_i > \bar{\sigma}_i(\theta_i^k)\}$. Using the fundamental theorem of calculus, it follows that

$$\begin{aligned}
&\frac{\partial}{\partial a_i} \sum_{\theta_j \in \Theta_j} \Big(u_i(\hat{a}_i, \bar{\sigma}_j(\theta_j), \theta_i^{k+1}) + u_j(\bar{\sigma}_j(\theta_j), \hat{a}_i, \theta_j) \Big) p_j(\theta_j) \\
&= \frac{\partial}{\partial a_i} \sum_{\theta_j \in \Theta_j} \Big(u_i(\tilde{a}_i, \bar{\sigma}_j(\theta_j), \theta_i) + u_j(\bar{\sigma}_j(\theta_j), \tilde{a}_i, \theta_j) \Big) p_j(\theta_j) \\
&\quad + \int_{\theta_i^k}^{\theta_i^{k+1}} \frac{\partial^2}{\partial a_i \partial \theta_i} \sum_{\theta_j \in \Theta_j} \Big(u_i(\tilde{a}_i, \bar{\sigma}_j(\theta_j), y) \Big) p_j(\theta_j) dy \\
&\quad + \int_{\tilde{a}_i}^{\hat{a}_i} \frac{\partial^2}{\partial a_i^2} \sum_{\theta_j \in \Theta_j} \Big(u_i(y, \bar{\sigma}_j(\theta_j), \theta_i) \Big) p_j(\theta_j) dy \\
&\geq 0 + \min\{\theta_i^{l+1} - \theta_i^l\} - 4 \max\{a_i^{l+1} - a_i^l\} \\
&\geq 0.
\end{aligned}$$

As a result, $\bar{\sigma}_i(\theta_i^{k+1}) \geq \hat{a}_i > \bar{\sigma}_i(\theta_i^k)$, which proves the result. □

B.5. A proof for Section 6

Proof of Proposition 2. 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 $w_{TO}^{\tau}(D)$ be the average value when a time-off control rule is used, given a punishment $\tau \in \{0, 1, 2\} \cup \{\infty\}$. 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.

The proof of b follows by replication the steps of the proof of Theorem 2. Details are available upon request. □

References

Abdulkadiroğlu, A., Bagwell, K., 2013. Trust, reciprocity, and favors in cooperative relationships. Am. Econ. J. Microecon. 5, 213–259.

Abreu, D., Milgrom, P., Pearce, D., 1991. Information and Timing in Repeated Partnerships. Econometrica 59, 1713–1733.

Abreu, D., Pearce, D., Stacchetti, E., 1986. Optimal cartel equilibria with imperfect monitoring. J. Econ. Theory 39, 251–269.

Abreu, D., Pearce, D., Stacchetti, E., 1990. Toward a theory of discounted repeated games with imperfect monitoring. Econometrica 58, 1041–1063.

Acemoglu, D., Wolitzky, A., 2014. Cycles of conflict: an economic model. Am. Econ. Rev. 104, 1350–1367.

Allen, B.T., 1976. Tacit collusion and market sharing: the case of steam turbine generators. Ind. Organ. Rev. 4, 48–57.

Arapostathis, A., Borkar, V.S., Fernández-Gaucherand, E., Ghosh, M.K., Marcus, S.I., 1993. Discrete time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Optim. 31, 282–344.

Arrow, K., 1985. Informational structure of the firm. Am. Econ. Rev., 303–307.

Ashworth, T., 1980. Trench Warfare, 1914–1918: The Live and Let Live System. Holmes and Meier, New York.

Athey, S., Bagwell, K., 2001. Optimal collusion with private information. Rand J. Econ. 32, 428–465.

Athey, S., Bagwell, K., 2008. Collusion with persistent cost shocks. Econometrica 76, 493–540.

Athey, S., Bagwell, K., Sanchirico, C., 2004. Collusion and price rigidity. Rev. Econ. Stud. 71, 317–349.

Aumann, R., Maschler, M., 1995. Repeated Games with Incomplete Information. The MIT Press.

Awaya, Y., Krishna, V., 2016. On communication and collusion. Am. Econ. Rev. 106, 285–315.

Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, New York.

Bagwell, K., Staiger, R., 2005. Enforcement, private political pressure, and the general agreement on tariffs and trade/world trade organization escape clause. J. Leg. Stud. 34, 471–513.

Bergemann, D., Valimaki, J., 2006. Bandit Problems. Yale University.

Bernheim, B., Madsen, E., 2017. Price cutting and business stealing in imperfect cartels. Am. Econ. Rev. 107, 387–424.

Carlton, D., Gertner, R., Rosenfield, A., 1996. Communication among competitors: game theory and antitrust. George Mason Law Rev. 5, 423.

Clark, R., Houde, J.-F., 2013. Collusion with asymmetric retailers: evidence from a gasoline price-fixing case. Am. Econ. J. Microecon. 5, 97–123.

Cole, H., Dow, J., English, W., 1995. Default, settlement, and signalling: lending resumption in a reputational model of sovereign debt. Int. Econ. Rev. 36, 365–385.

Dutta, P., 1995. A folk theorem for stochastic games. J. Econ. Theory 66, 1–32.

Escobar, J., Toikka, J., 2013. Efficiency in games with Markovian private information. Econometrica 81, 1887–1934.

Fudenberg, D., Maskin, E., 1986. The Folk theorem in repeated games with discounting or with incomplete information. Econometrica 54, 533–554.

Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press.

Gale, D., Rosenthal, R., 1994. Price and quality cycles for experience goods. Rand J. Econ., 590–607.

Gensbittel, F., Renault, J., 2015. The value of Markov chain games with incomplete information on both sides. Math. Oper. Res. 40, 820–841.

Gentzkow, M., Kamenica, E., 2011. Bayesian persuasion. Am. Econ. Rev. 101.

Gerlach, H., 2009. Stochastic market sharing, partial communication and collusion. Int. J. Ind. Organ. 27, 655–666.

Gorman, P.F., 1953. Limited War: Korea, 1950. Harvard University. Mimeo.

Green, E., Porter, R., 1984. Noncooperative collusion under imperfect price information. Econometrica 52, 87–100.

Hörner, J., Sugaya, T., Takahashi, S., Vieille, N., 2011. Recursive methods in discounted stochastic games: an algorithm for $\delta \to 1$ and a Folk theorem. Econometrica 79, 1277–1318.

Hörner, J., Takahashi, S., Vieille, N., 2015. Truthful equilibria in dynamic Bayesian games. Econometrica 83, 1795–1848.

Hsu, S.-P., Chuang, D.-M., Arapostathis, A., 2006. On the existence of stationary optimal policies for partially observed MDPs under the long-run average cost criterion. Syst. Control Lett. 55, 165–173.

Jackson, M.O., Sonnenschein, H.F., 2007. Overcoming incentive constraints by linking decisions. Econometrica 75, 241–258.

Keller, G., Rady, S., 1999. Optimal experimentation in a changing environment. Rev. Econ. Stud. 66, 475–507.

Liu, Q., 2011. Information acquisition and reputation dynamics. Rev. Econ. Stud. 78, 1400–1425.

Liu, Q., Skrzypacz, A., 2014. Limited records and reputation bubbles. J. Econ. Theory 151, 2–29.

Markham, J., 1951. The nature and significance of price leadership. Am. Econ. Rev. 41, 891–905.

Marschak, J., Radner, R., 1972. Economic Theory of Teams. Yale University Press.

Marshall, R., Marx, L., 2013. The Economics of Collusion: Cartels and Bidding Rings. MIT Press.
Maskin, E., Tirole, J., 1988. A theory of dynamic oligopoly, II: price competition, kinked demand curves, and edgeworth cycles. Econometrica 56, 571–599.
Mobius, M., 2001. Trading Favors. Tech. rep. Harvard University.
Mouraviev, I., Rey, P., 2011. Collusion and leadership. Int. J. Ind. Organ. 29, 705–717.
Nicholls, W.H., 1951. Price Policies in the Cigarette Industry. Vanderbilt University Press, Nashville TN.
Pęski, M., 2014. Repeated games with incomplete information and discounting. Theor. Econ. 9, 651–694.
Pęski, M., Toikka, J., 2017. Value of persistent information. Econometrica 85, 1921–1948.
Phelan, C., 2006. Public trust and government betrayal. J. Econ. Theory 130, 27–43.
Platzman, L.K., 1980. Optimal infinite horizon undiscounted control of finite probabilistic systems. SIAM J. Control Optim. 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.
Rahman, D., 2014. The power of communication. Am. Econ. Rev. 104, 3737–3751.
Renault, J., Solan, E., Vieille, N., 2013. Dynamic sender receiver games. J. Econ. Theory 148, 502–534.
Renou, L., Tomala, T., 2015. Approximate implementation in Markovian environments. J. Econ. Theory 159, 401–442.
Rotemberg, J., Saloner, G., 1986. A supergame-theoretic model of price wars during booms. Am. Econ. Rev. 76, 390–407.
Rotemberg, J., Saloner, G., 1990. Collusive price leadership. J. Ind. Econ., 93–111.
Sandleris, G., 2008. Sovereign defaults: information, investment and credit. J. Int. Econ. 76, 267–275.
Sannikov, Y., Skrzypacz, A., 2007. Impossibility of collusion under imperfect monitoring with flexible production. Am. Econ. Rev. 97, 1794–1823.
Schelling, T., 1960. The Strategy of Conflict. Harvard University Press, Cambridge, MA.
Scherer, F.M., Ross, D., 1990. Industry Market Structure and Economic Performance.
Seaton, J.S., Waterson, M., 2013. Identifying and characterising price leadership in British supermarkets. Int. J. Ind. Organ. 31, 392–403.
Shapiro, C., 1986. Exchange of cost information in oligopoly. Rev. Econ. Stud. 53, 433–446.
Skrzypacz, A., Toikka, J., 2015. Mechanisms for repeated trade. Am. Econ. J. Microecon. 7, 252–293.
Stigler, G., 1947. The kinky oligopoly demand curve and rigid prices. J. Polit. Econ., 432–449.
Stokey, N., Lucas, E., with Prescott, R., 1989. Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge.
Townsend, R., 1982. Optimal multi-period contracts and the gain from enduring relationships under private information. J. Polit. Econ. 90, 1166–1186.
Whinston, M., 2008. Lectures on Antitrust Economics. The MIT Press.