COOPERATION DYNAMICS IN REPEATED GAMES OF ADVERSE SELECTION: TO FORGIVE IS NOT TO FORGET*

JUAN F. ESCOBAR† AND GASTÓN LLANES‡

ABSTRACT. We study repeated games with Markovian private information and characterize optimal equilibria as players become arbitrarily patient. We show that seemingly non-cooperative actions may occur in equilibrium and serve as signals of changes in private information. Players forgive such actions, and use the information they convey to adjust their continuation play. However, to forgive is not to forget: players keep track of the number of aggressions and enter into a punishment phase if that number becomes suspiciously high. Our model explains features of long-run relationships that are only barely understood, such as equilibrium defaults, unilateral price cuts, collusive price leadership, graduated sanctions, and restitutions. We also explore a model in which interactions are frequent and show how increasing the persistence of the process of types reduces informational frictions.

KEYWORDS: Repeated games, Markovian private information, signaling, tacit collusion, price leadership, price cuts, equilibrium defaults, graduated sanctions.

1. INTRODUCTION

> *In most sectors, and at most times each side had the choice of either aggressing or not aggressing the enemy. In the situation where each antagonist had this choice, the non-aggression of one was neither negative nor meaningless to the other; on the contrary, it was as positive and meaningful as the alternative act of aggression. Mutual aggression was in a real and obvious sense mutual communication; for when trench fighters fired against each other, neither doubted that the other's intent was to kill or injure. Similarly, the choice of non-aggression [. . .] was equally an act of communication.*
>
> *Ashworth (1980, p. 40)*

Date: September 16, 2015.

* Escobar acknowledges financial support from Conicyt (Fondecyt No. 1130435). Llanes acknowledges financial support from Conicyt (Fondecyt No. 1150326). Escobar and Llanes acknowledge financial support from the Institute for Research in Market Imperfections and Public Policy, MIPP, ICM IS130002, Ministerio de Economía, Fomento y Turismo.

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

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

During trench warfare in the First World War (1914-1918), frontline soldiers often refrained from attacking the enemy, provided that their restraint was reciprocated by soldiers on the other side. Army commanders were aware of the tendency towards non-aggresion and would order raids to correct the “offensive spirit” of the troops (Ashworth, 1980; Axelrod, 1984). Enemy soldiers would generally not be able to discern if aggressions were caused by opportunistic behavior or by military orders, and would be reluctant to return to the non-aggressive behavior. However, cooperation generally restarted after some time had elapsed, and in this way soldiers were successful at maintaining low levels of aggression for significant lengths of time.

Cooperative relationships often exhibit this type of dynamics. For example, firms trying to avoid price competition cycle between high and low prices (Markham, 1952; Bresnahan, 1987; Scherer and Ross, 1990), sovereign countries that default on their obligations are temporarily excluded from international capital markets but are eventually able to borrow again (Cole et al., 1995; Tomz, 2012), managers and unions enter into labor conflicts which are followed by periods with high production and generous bonuses (Li and Matouschek, 2013), and governments in self-enforcing trade agreements raise and lower their import tariffs, even though high tariffs may be detrimental to foreign partners (Bagwell and Staiger, 2005).

In this paper, we shed light on these phenomena by studying a discrete-time infinitely-repeated game with Markovian private information. Two players make perfectly observable decisions at each round. Player 1 is privately informed about his own payoffs, which evolve according to a finite Markov chain. Importantly, no communication is allowed, which implies that player 2 can learn about player 1’s types only by observing player 1’s actions.

We show that optimal equilibrium dynamics may allow for apparent cooperation breaks (such as aggressions, price cuts, debt defaults) that signal optimal continuation play. More generally, in our incomplete information model, equilibrium actions have an *informational content* that determines the most profitable course of play for the relationship. We show how optimal equilibria make use of endogenously generated information and explain behaviors (as in the case of live and let live) that are difficult to square with existing models.

Our main theoretical result is the characterization of a class of Pareto-optimal equilibria as players become arbitrarily patient. This characterization reduces the problem of determining the informational content of the informed player’s actions, as a function of history, to a dynamic programming equation defined on the set of total expected payoff functions. Our dynamic programing formulation is new to the literature and we provide several examples that prove its usefulness.

Section 2 illustrates our approach and results by studying a two-player two-action prisoners’ dilemma with incomplete information. We assume that player 1 has private information about his investment cost in a joint project. Investing is always a dominated action. When player 1’s cost is low, the socially desirable outcome is that both players invest, whereas when player 1’ cost is high it is socially optimal that no player invests. Player 1's cost evolves with positive persistence. The problem is subtle because player 2 does not observe player 1's type, nor can player 1 communicate his cost. We find welfare-maximizing equilibria in two steps.

In the first step, we relax incentive constraints by allowing players to commit to strategies at the beginning of the game to maximize the sum of expected payoffs. We reformulate this problem as a dynamic programming problem having as state variable the belief about player 1's type conditional on public information. Two interesting optimal dynamics arise when incentives are ignored. Under *reactive-signaling dynamics*, the informed party keeps signaling his type while player 2 imitates the behavior of the informed player. Under *time-off dynamics*, a failure to invest by the high cost player 1 triggers a *waiting phase*. During the waiting phase, both players refrain from investing during a fixed number of periods. Once the waiting phase is over, player 2 and the low-cost player 1 resume investments and a waiting phase is restarted when player 1's cost becomes high again.

In the second step, we prove that optimal dynamics can actually be mimicked when incentives are taken into account. We build strategies in the repeated game that keep track of the informed player's actions and test whether they are sufficiently likely to come from the underlying optimal decision rule.[1] Optimal play may require that the informed agent plays apparently hostile or aggressive actions. The uninformed agent will forgive such behaviors and continue to play according to the optimal rule. However, *to forgive is not to forget*: the uninformed agent keeps track of the number of aggressions, and players enter a punishment phase if that number becomes suspiciously high.

This simple model shows that adverse selection and imperfect communication can restrict the set of equilibrium payoffs to a strict subset of the set of equilibrium payoffs with perfect information. In the optimal equilibrium, the informed party has to shirk in some rounds, and has to incur in *costly gestures* or *let time pass by* to persuade the uninformed player to resume investments. These dynamics imply substantive welfare costs, and optimal equilibrium payoffs are bounded away from first-best payoffs, even as the discount factor goes to 1. If the costs of incomplete information are large enough, the optimal equilibrium consists of repetitions of the static Nash equilibrium.

Sections 3 and 4 extend the analysis to general games of one-sided incomplete information. We establish an upper bound for equilibrium payoffs by studying an average-reward optimality equation (AROE) for a hidden-state Markov decision problem. The AROE is a Bellman equation tailored to study undiscounted dynamic models. The hidden variable in the Markov decision process is player 1's type, which together with controls determine a distribution over actions. Once actions are observed, they are used to update beliefs. Beliefs about player 1's

[1] The process of actions is not a Markov process, so it is hard to perform tests based on it. We sidestep this difficulty by testing observed actions conditional on *simulated* public beliefs. See Section 4 for details.

type, given observed actions, are the state variable in the dynamic programming equation. At a more conceptual level, the AROE captures a basic trade-off between separating and pooling control rules. If player 1 pools given a public history, player 2 can better optimize his period payoffs. On the other hand, when different types of player 1 separate, continuation public beliefs are more precise and therefore the relationship gains from better information. We also show how strategies that forgive but do not forget can be designed to virtually attain the upper bound in the repeated game with low discounting.

Section 5 studies games with separating and monotonic dynamics. In these games, period payoffs have strictly increasing differences in actions and types, and player 1 has a set of actions which is sufficiently numerous. We show that player 1's actions are strictly increasing in his type and therefore he keeps signaling his current conditions. These results help explain a number of phenomena in long-run relationships that are only barely understood.

In Section 5.1, we characterize the optimal collusive scheme in a Bertrand game of differentiated products in which one of the firms has private information about its demand. Consistent with case studies (Marshall and Marx, 2013), in our model *unilateral price changes* occur on the path of play. Our model can be interpreted as a model of *collusive price leadership* (Stigler, 1947; Markham, 1951; Scherer and Ross, 1990), in which an uninformed firm follows the informed firm's price changes.[2] We show that the dynamics of price leadership may involve significant costs for leader and follower. When demand increases, the informed firm raises its price, and experiences a short-term loss until its price raise is matched by the follower. Likewise, the follower experiences a short-term loss when the leader lowers its price after a demand reduction.[3] Our model therefore provides concrete answers to some unsettled issues in industrial organization and antitrust.[4]

We also apply our results to provide a rationale to the commonly referred practice of *graduated sanctions* when collectively managing common-pool resources. As Ostrom's 1990 shows, in several successful long-run relationships, after a member breaks a norm, cheated partners mildly adjust their continuation actions. The use of severe punishments, like Nash reversion, is

[2] Collusive price leadership is relevant in many industries. Allen (1976), for example, documents collusive price leadership in the market of steam turbine generators in the 1960s and 1970s. In Section 5.1 we discuss additional empirical evidence.

[3] These short-term losses are significant in many industries. Clark and Houde (2013) study gasoline prices in Quebec, and find that a small price premium (2 cents or more per liter) for a few hours can result in a significant reduction in a station's sales for the day (around 35% to 50%).

[4] Green and Porter (1984), Abreu et al. (1986), and Rotemberg and Saloner (1986) study collusive equilibria with high and low price cycles, in which price movements are simultaneous across firms. Thus, there are no unilateral price changes or price leaders. Rotemberg and Saloner (1990) study collusive price leadership in a repeated Bertrand game, imposing exogenous constraints on strategies and on the timing of the game, and find that the leader always benefits more from price leadership than the follower. All these models have Pareto efficient equilibria if players are sufficiently patient. We study the optimal equilibrium without imposing exogenous restrictions, and show that price leadership can arise as an equilibrium outcome. This equilibrium may involve significant profit losses for both the leader and the follower, and as a result, may be inefficient even as the discount factor goes to 1.

the exception rather than the norm. As Dixit (2009) explains, this evidence is difficult to reconcile with existent theoretical frameworks. Our model fills this gap. In Section 5.2, we specialize our model to a repeated collective action game in which player 1 has private information about the benefits of a project. On the path of play, the lower the action by the informed player, the lower the expectation the uninformed player has about the relationship conditions, and therefore the lower player 2's action. Player 1 can also make restitutions that positively affect player 2's current payoffs and his continuation beliefs and actions.

Section 6 refines our analysis by studying the game in Section 2 as interactions become more frequent. Following a tradition initiated by Abreu et al. (1991), we observe that as interactions become more frequent not only the discount factor increases but also the process of hidden types becomes more persistent. We show that changing the persistence of the process of types has important effects on the dynamics of cooperation and equilibrium payoffs. In the limit, signaling becomes inexpensive compared to the benefits from more precise beliefs and, as a result, incomplete information has virtually no costs.

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 obtains. With communication, actions have no signaling content and the dynamics of cooperation are similar to those of games with complete information and changing types if players are sufficiently patient (Rotemberg and Saloner, 1986; Dutta, 1995). In those models, actions can perfectly respond to current conditions and there is no room (on the path of play) to observe hostile behaviors (as we do in reality). 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.[5]

Our results connect to the literature on repeated games with Markovian hidden types. Escobar and Toikka (2013), Renault et al. (2013), and Hörner et al. (2015) characterize optimal equilibria in games with communication. As explained above, dynamics in these models are very different from the ones in this paper. Athey and Bagwell (2001, 2008) characterize optimal equilibria in Bertrand games without communication, but their analysis exploits the special structure of their inelastic demand model. Hörner et al. (2010a) study equilibrium values in the zero-sum case.[6] Our contribution is to characterize optimal equilibrium in a fairly general class of games with hidden types and no communication.

[5] Price discussions between competitors are generally illegal. Ashworth (1980) documents the communication problems faced by enemy troops trying to avoid confrontation during World War I. When discussing limited war, Schelling (1960) explains that "an agreement on limits is difficult to reach ... because communication becomes difficult between adversaries in a time of war."

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

Other papers have also focused on defaults and cooperation cycles. 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.[7] 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.

We finally observe that in games with imperfect monitoring, players can also cycle between cooperative and uncooperative actions (Green and Porter, 1984; Abreu et al., 1986, 1990, 1991), but equilibrium dynamics differ significantly from the ones presented in our paper. 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. We show that in the case of adverse selection, regime changes depend on players' actions. For example, low-price regimes (price wars) may be triggered by price cuts, and returning to high-price regimes may require unilateral price rises. Abreu et al. (1991) studies a prisoners' dilemma with imperfect monitoring and shows that, under certain conditions, 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 (optimal equilibria sustains an informationally-constrained welfare optimum) and there is little room for renegotiation.

2. An Example

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). The investment may represent an advertising expenditure in a joint-advertising campaign, an investment in R&D in a research joint venture, or costly effort in a team of co-workers.

Stage payoffs are equal to investment revenues minus cost. If both players invest, each player obtains a revenue of a. If only one player invests, each player obtains a revenue of b. If no player invests, both players obtain zero revenues. Let $0 < b < a$. Player 1's investment cost in period t is $\theta^t \in \{l,h\}$, where $l < h$, and player 2's investment cost is l every period. Table 1 shows the game.

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.

[7] In those models, defaults are strategic while in our model defaults are mainly non-strategic.

	I	N
I	$a-l, a-l$	$b-l, b$
N	$b, b-l$	$0, 0$

$\theta^t = l$

	I	N
I	$a-h, a-l$	$b-h, b$
N	$b, b-l$	$0, 0$

$\theta^t = h$

TABLE 1. The game.

The cost parameter θ^t is realized at the beginning of period t and is privately known by player 1. Once player 1 privately observes his type, θ^t, players simultaneously choose actions. Actions are publicly and perfectly observed. Player 1's type evolves according to a Markov process with transition probabilities given by

$$\mathrm{P}[\theta^t = l \mid \theta^{t-1} = l] \;\; = \;\; \lambda, \quad \mathrm{P}[\theta^t = h \mid \theta^{t-1} = h] = \mu$$

where $\lambda + \mu \geq 1$ or, equivalently, the process of types has positive persistence. For simplicity, we assume that the initial type is drawn according to $\mathrm{P}[\theta^1 = l] = \lambda$. Players have a common discount factor $\delta < 1$ and maximize the discounted sum of period payoffs. This is a repeated game with Markovian incomplete information. Whereas player 1 knows that whole history of transpired types and play, player 2 can condition his behavior only on the history of actions.

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). Our focus in on optimal equilibria with incomplete information and no communication.

Before describing equilibrium behavior, let us characterize optimal dynamics ignoring incentive constraints. Observe that even when incentives are ignored and player 1 can make use of all available information (the history of play and his privately-held information), player 2 can only condition his behavior on public information (the history of play). To study this problem, we introduce controls. A control for player 1 is a pair $\sigma_1 = (\sigma_1(l), \sigma_1(h)) \in \{I, N\}^2$, whereas a control for 2 is simply $\sigma_2 \in \{I, N\}$. A control $\sigma = (\sigma_1(l), \sigma_1(h), \sigma_2)$ will be characterized by a triple XYZ. A control σ determines total period payoffs, given beliefs. If the control is INI and the public information determines $p^t = \mathbb{P}[\theta^t = l]$, then expected period payoffs are

$$p^t\big((1-\delta)2(a-l)\big) + (1-p^t)\big((1-\delta)(2b-l)\big).$$

But a control also determines player 2's continuation beliefs, given a current distribution on 1's types. For example, suppose the belief at time t is $\mathrm{P}(\theta^t = l) = p^t$. Given this belief, if player 1's control has $\sigma_1^t(l) = I$ and $\sigma_1^t(h) = N$ (that is, player 1 invests if and only if her type is l); then period $t+1$'s probabilities depend on the action of player 1 at period t:

$$\mathrm{P}_\sigma[\theta^{t+1} = l \mid a_1^t = I] = \lambda, \quad \mathrm{P}_\sigma[\theta^{t+1} = l \mid a_1^t = N] = 1 - \mu.$$

If player 1's control has $\sigma_1^t(l) = \sigma_1^t(h) = N$, period $t+1$ probabilities are given by

$$\mathrm{P}_\sigma[\theta^{t+1} = l \mid a_1^t = N] \quad = \quad p^t \lambda + (1 - p^t)(1 - \mu).$$

From player 2's perspective, the belief about player 1's type in $t+1$ depends on the control and the observation in t. We consider the belief held by player 2 at t about 1's type, $p^t = \mathrm{P}(\theta^t = l)$, as a state variable. Let $w(p)$ be the value for the problem of maximizing total discounted expected payoffs given beliefs p over all possible strategies. The above discussion leads to the following dynamic programming characterization for the value $w(p)$:

$$w(p) = \max\Big\{ w_{XYZ}(p) \mid X, Y, Z \in \{I, N\} \Big\}, \tag{2.1}$$

where

$$\begin{aligned}
w_{III}(p) &= p(1-\delta)2(a-l) + (1-p)(1-\delta)(2a-l-h) + \delta w(p\lambda + (1-p)(1-\mu)), \\
w_{IIN}(p) &= p(1-\delta)(2b-l) + (1-p)(1-\delta)(2b-h) + \delta w(p\lambda + (1-p)(1-\mu)), \\
w_{NNN}(p) &= (1-\delta)0 + \delta w(p\lambda + (1-p)(1-\mu)), \\
w_{NNI}(p) &= (1-\delta)(2b-l) + \delta w(p\lambda + (1-p)(1-\mu)), \\
w_{INI}(p) &= p((1-\delta)2(a-l) + \delta w(\lambda)) + (1-p)((1-\delta)(2b-l) + \delta w(1-\mu)), \\
w_{INN}(p) &= p((1-\delta)(2b-l) + \delta w(\lambda)) + (1-p)((1-\delta)0 + \delta w(1-\mu)), \\
w_{NII}(p) &= p((1-\delta)(2b-l) + \delta w(\lambda)) + (1-p)((1-\delta)(2a-l-h) + \delta w(1-\mu)), \\
w_{NIN}(p) &= p((1-\delta)0 + \delta w(\lambda)) + (1-p)((1-\delta)(2b-h) + \delta w(1-\mu)).
\end{aligned}$$

The optimal choice of control will trade-off current payoffs and the distribution over continuation beliefs. It is straightforward to see that controls *IIN*, *NNI*, *NII*, and *NIN* are never optimal. For example, under control *NNI* player 1 does not invest but player 2 does, which provides less period payoffs than the pooling control *NNN*. Since both controls determine the same distribution over continuation beliefs, control *NNI* cannot be optimal for any belief.

The following lemma summarizes some important properties of $w(p)$. All omitted proofs are in the Appendix.

Lemma 1. *$w(p)$ is nondecreasing, continuous, and convex.*

To understand the convexity property, fix beliefs $p = \mathrm{P}(\theta^t = l)$ yielding value $w(p)$. Suppose now that we are offered more detailed information about this probability: we are told that $\mathrm{P}(\theta^t = l) = q$ with probability π and $\mathrm{P}(\theta^t = l) = q'$ with probability $1-\pi$, such that $p = \pi q + (1-\pi)q'$. Now, value is $w(q)$ with probability π and $w(q')$ with probability $1-\pi$. Convexity implies that we always prefer to have more information: $w(p) = w(\pi q + (1-\pi)q') \leq \pi\, w(q) + (1-\pi)\, w(q')$. Intuitively, when information is revealed, the optimal control can be adjusted to yield better outcomes.

The solution to (2.1) can result in pooling dynamics, in which players play a fixed action profile in all rounds. Assuming $a-b < h/2$, the pooling control *III* is never optimal as the incremental social benefit of 1's investment is low compared to his high cost h.[8]. When solving (2.1), it is enough to focus on controls *NNN*, *INI*, and *INN*.

Lemma 2 shows that the optimal dynamics takes very simple forms. The second-best rule generates *reactive-signaling* dynamics if player 1 invests when his cost is low and does not invest when his cost is high, and player 2 imitates the action of player 1 in the previous period. Thus, player 1 *signals* his private information through his actions, and player 2 *reacts* to such information. The second-best rule generates *time-off* dynamics if player 1 invests only if he is in good standing and his cost is low, and player 2 invests if and only if player 1 is in good standing. Player 1 is in good standing if he invested in the previous period, or if he did not invest in the previous period, but was in good standing $\hat{\tau}+1$ periods before, where $\hat{\tau}$ is a natural number (possibly equal to 0).

Lemma 2. *If $a-b < h/2$ and $\lambda > \frac{l-2b}{2(a-b)-l}$, the second-best rule generates either reactive-signaling or time-off dynamics.*

The restriction to $\lambda > \frac{l-2b}{2(a-b)-l}$ ensures that the control *INI* is optimal at belief $p = \lambda$. The Lemma shows that if player 1 does not invest at belief λ, then player 1 will "signal" a change of type by investing to prompt player 2 to invest too, or, alternatively, player 2 will wait for $\hat{\tau}$ rounds to become optimistic about player 1's cost and resume investments. This result shows that, depending on the parameters, only one of the two paths prevails and rules out dynamics in which signaling can occur only after an exogenous number of rounds has transpired.

The choice between reactive-signaling and time-off dynamics depends on the comparison between the *miscoordination cost* and the *opportunity cost* of missed cooperation. The miscoordination cost is $l-2b$, which is the welfare loss suffered when only one player invests. Under reactive-signaling dynamics, players incur in a miscoordination cost every time player 1's type changes. Under time-off dynamics, players incur in a miscoordination cost when player 1's type goes from l to h, and when a waiting phase ends, if the type of player 1 is h. Observe that under time-off, players may not incur in the signaling cost when a waiting phase ends, because the type

[8] Another reason *III* is not optimal is that it does not improve information in the continuation nodes

of player 1 may be l when the waiting phase ends. The opportunity cost of missed cooperation is $2(a-l)$, and is the gain in welfare that would have accrued if both players had invested and player 1's cost was low. The cost of missed cooperation is incurred during a waiting phase in a time off rule. The total expected opportunity cost of missed cooperation depends on the optimal length of the waiting phase.

Let $\beta = \frac{l-2b}{2(a-l)}$ measure the miscoordination cost relative to the opportunity cost of missed cooperation. The following lemma shows how optimal dynamics depend on the parameters of the model as the discount factor goes to 1.

Lemma 3. *Assume that $a-b < h/2$ and $1 < (\lambda+\mu)(1-\frac{\lambda}{2})$. There exists $\beta_0 \in]0, \frac{\lambda}{2(1-\lambda)}[$ (that does not depend on δ) such that*

i. *For $\beta < \beta_0$, there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, the optimal rule generates time-off dynamics;*
ii. *For $\beta \in]\beta_0, \frac{\lambda}{2(1-\lambda)}[$, there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, the optimal rule generates reactive-signaling dynamics;*
iii. *For $\beta > \frac{\lambda}{2(1-\lambda)}$, there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, the optimal rule generates a path in which no player ever invests.*

This lemma fully characterizes optimal dynamics under the added assumption that the process of types is persistent enough (so that $1 < (\lambda+\mu)(1-\frac{\lambda}{2})$).[9] When the miscoordination cost is small, time-off is optimal as even a short waiting period ensures that the benefits of mutual cooperation are realized. As β increases, the waiting period in time-off dynamics increases, and it is better to capture the benefits of cooperation by allowing player 1 to signal his changes in types. When the miscoordination cost is too high, there is no room for cooperation and players payoffs are 0.[10]

So far, we have ignored incentive problems. When thinking about the incentives player 1 would face to play according to optimal dynamics, one could be tempted to argue that those dynamics effectively “punish” a defection by player 1. For example, under reactive signaling, if player 1 does not invest when his cost is low, player 2 reacts by not investing in the next period, and this will provide incentives to player 1 to invest if and only if his cost is low. However, this type of argument can work only for some parameter values as reactive signaling and time-off dynamics arise to optimize the discounted sum of payoffs subject to informational constraint, but do not take into account incentives. It is therefore not obvious whether optimal dynamics payoffs can be attained when incentives are taken into account.

We now construct equilibrium strategies such that on-path play is arbitrarily close to second-best optimal dynamics. For concreteness, we take parameters such that, according to Lemma 3,

[9] In the Appendix we dispense with this restriction and provide a complete characterization of optimal dynamics

[10] Observe that the bounds in the Lemma improve upon the restrictions in Lemma 2 to have some cooperation.

optimal dynamics is reactive-signaling. Let $v_i \in \mathbb{R}$ be the limit average payoff accruing to player i under the reactive signaling rule and assume $v_i > 0$. This means that both players get more payoffs from the optimal dynamics than from the static Nash equilibrium resulting in payoffs $(0,0)$.

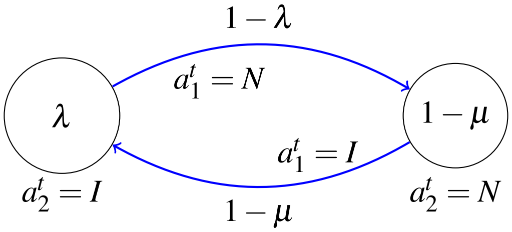


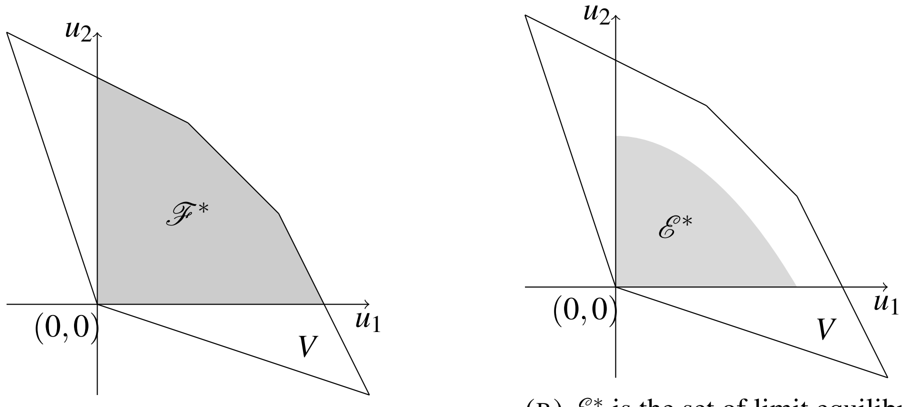
FIGURE 1. Dynamics of beliefs $(p^t)_{t\geq 1}$ when players use a rule resulting in reactive-signaling dynamics. The support of $(p^t)_{t\geq 1}$ is the set $\{\lambda, 1-\mu\}$.

Ensuring appropriate behavior by player 2 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 more subtle because player 2 cannot observe player 1's type, nor can he tell whether a failure to invest by player 1 is acceptable (because player 1's cost is high) or not. However, as play transpires, player 2 can keep checking whether player 1's behavior seems likely to have been generated from the reactive-signaling rule. Under reactive-signaling, the process of beliefs $(p^t)_{t\geq 1}$, with $p^t = \mathrm{P}[\theta^t = l \mid h^{t-1}]$ and h^{t-1} the public history up to and including round $t-1$, is Markovian, with transitions that can be drawn as shown in Figure 1. 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 state λ, 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 players' 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).[11] 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.[12] In other words, equilibrium strategies are so that player 2 forgives but does not forget failures.

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

[12] 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.



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

(B) $\mathscr{E}^*$ is the set of limit equilibrium payoffs in the game with incomplete information and no communication. It is strictly contained in $\mathscr{F}^*$. When signaling is too costly, as explained in Lemma 3, $\mathscr{E}^* = \{(0,0)\}$.

FIGURE 2. Sets of equilibrium payoffs for games with and without communication.

Our analysis has three main implications. First, as we explained above, informational constraints are key to determine optimal equilibrium dynamics. While incentive problems disappear as players become more patient, equilibria are bounded away from first-best payoffs. Indeed, with complete information (or with incomplete information and communication), players can attain average total payoffs equal to $2(a-l)\frac{1-\mu}{2-\lambda-\mu}$. Assuming the conditions under which reactive-signaling is optimal in Lemma 2, under incomplete information total payoffs are $\big(2\lambda(a-l)-(l-2b)(1-\lambda)\big)\frac{1-\mu}{2-\lambda-\mu}$. 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.[13] 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.[14]

Second, cooperation dynamics differ from those found in previous papers. For example, when monitoring is imperfect, punishments are triggered on the equilibrium path and cooperation may be resumed exogenously Abreu et al. (following a randomization device as in 1990) or not resumed at all Abreu et al. (as in 1991). In our model, actions have signaling content and cooperation is always resumed, either by taking a costly action (as in reactive signaling) or

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

[14] Awaya and Krishna (2014) study a repeated Bertrand game with imperfect private monitoring and show conditions under which the set of equilibrium payoffs without communication is strictly contained in the set of equilibrium payoffs with communication. They obtain a lower bound for the gap.

after a cooling-off period has elapsed (as in time-off dynamics). Since continuation payoffs are always close to optimal, virtually no value is burnt on the path of play and, in contrast to models with imperfect monitoring, there is little room for renegotiation.

Third, under reactive signaling, the on-path behavior of player 2 is identical to a tit-for-tat strategy. While tit-for-tat is an intuitive strategy and has received attention in the literature (Axelrod, 1984; Kalai et al., 1988; Kreps et al., 1982), no equilibrium framework exists under which it emerges as the optimal outcome. Our results fill this gap by showing how informational frictions make tit-for-tat a desirable strategy.

3. MODEL

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

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

We assume players know their payoffs. The period payoff function for player 1 is $u_1(a,\theta)$, whereas player 2's payoff is $u_2(a)$. We will sometimes abuse notation and write $u_i(a,\theta)$, even when player 2's payoff does not depend on θ. Players rank flows of payoffs according to $(1-\delta)\sum_{t\geq 1}\delta^{t-1}u_i(a^t,\theta^t)$, where $\delta < 1$ is the common discount factor. We assume that $|A_1| \geq |\Theta|$.[15]

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

A strategy for player 1 is a sequence of functions $s_1 = (s_1^t)_{t\geq 1}$ with $s_1^t \colon \Theta^t \times A^{t-1} \times [0,1]^t \to A_1$, whereas a strategy for the uninformed player 2 is $s_2 = (s_2^t)_{t\geq 1}$ with $s_2^t \colon A^{t-1} \times [0,1]^t \to A_2$. A strategy profile $s^* = (s_1^*, s_2^*)$ is a perfect Bayesian equilibrium if there exists a system of beliefs constructed from Bayes rule (when possible) such that s_i^* is sequentially rational (Fudenberg and Tirole, 1991). The set of perfect Bayesian equilibrium payoffs will be denoted $\mathscr{E}(\delta, p^1) \subseteq \mathbb{R}^2$.

A *decision rule* is a sequence $f = (f^t)_{t\geq 1}$ with $f^t = (f_1^t, f_2^t)$ and $f_1^t \colon A^{t-1} \times \Theta^t \times [0,1]^t \to \Delta(A_1)$ and $f_2^t \colon A^{t-1} \times [0,1]^t \to \Delta(A_2)$. A decision rule determines a possibly mixed action for

[15] A one-sided incomplete information model is considered for expositional simplicity. We discuss the case of two-sided incomplete information in the Conclusions.

each player i as a function of the publicly observed history of action profiles and randomizations, and possibly his own private history of realized types. Any decision rule f induces a probability distribution over histories. We can therefore define the vector of expected payoffs given a decision rule f as

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

Let $V(\delta,\lambda) = \{v = v^{\delta}(f) \in \mathbb{R}^2$ for some decision rule $f\}$ be the set of all (constrained) feasible payoffs that players can attain by employing arbitrary decision rules f. In passing, we note that $V(\delta, p^1) \subseteq \mathbb{R}^2$ is convex and compact and $\mathscr{E}(\delta, p^1) \subseteq V(\delta, p^1)$ for all $\delta < 1$.

Our definitions of decision rules and set of feasible payoffs differ from those encountered in studies of stochastic games (Dutta, 1995; Hörner et al., 2010b) and repeated games with incomplete information and communication (Escobar and Toikka, 2013; Hörner et al., 2015). In our model, player 2 decides based only on public information.

4. Analysis

We will characterize equilibrium play in two steps. In the first step, we provide a dynamic programming formulation for efficient decision rules. This characterization will employ some tools from undiscounted optimization problems with partially observed states. In the second step we show how such efficient rule can be approximated by equilibrium play of the infinitely repeated game.

4.1. **A Convenient Formulation for Feasible Payoffs.** A decision rule f is *efficient* if for some $\alpha \in \mathbb{R}^2_{++}$, with $\alpha_1 + \alpha_2 = 1$, f is a solution to

$$q(\alpha) = \max\{\alpha \cdot v^{\delta}(f) \mid f \text{ is a decision rule}\}. \tag{4.1}$$

We can construct an efficient payoff vector $v = v^{\alpha,\delta} = v^{\delta}(f^{\alpha,\delta}) \in \mathbb{R}^2$, where $f^{\alpha,\delta}$ solves (4.1). Since any such $v^{\alpha,\delta}$ solves the problem $\max\{\alpha \cdot v \mid v \in V(\delta, p^1)\}$, the set of efficient payoff vectors v that maximize payoffs given a direction $\alpha \in \mathbb{R}^2_{++}$ is convex.

To characterize efficient decision rules, we introduce some notation. Let $\Sigma_1 = \{\sigma_1 \colon \Theta \to A_1\}$ be a set of *controls* for player 1 and let $\Sigma = \Sigma_1 \times A_2$. Let $p \in \Delta(\Theta)$ be a belief about player 1's type given public information, and let $p(\theta)$ denote the θ-element of p. For $\sigma \in \Sigma$ and $p \in \Delta(\Theta)$, we define the vector of expected period utility $U(\sigma, p) \in \mathbb{R}^2$ as

$$U_1(\sigma, p) = \sum_{\theta\in\Theta} u_1(\sigma_1(\theta), \sigma_2, \theta)\; p(\theta)$$

and $U_2(\sigma, p) = \sum_{\theta\in\Theta} u_2(\sigma_1(\theta), \sigma_2) p(\theta)$. For $\alpha \in \mathbb{R}^2_{++}$, we consider the ex-ante weighted sum of period payoffs $U^{\alpha}(\sigma, p) = \alpha \cdot U(\sigma, p) = \sum_{i=1}^{2} \alpha_i\, U_i(\sigma, p)$ given a control profile $\sigma \in \Sigma$ and beliefs $p \in \Delta(\Theta)$. We also define the Bayes operator $B(\cdot \mid \sigma_1, p, a_1) \in \Delta(\Theta)$ as

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

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

For $\alpha \in \mathbb{R}^2_{++}$ with $\alpha_1 + \alpha_2 = 1$, consider the only solution to the Bellman equation

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

for all $p \in \Delta(\Theta_i)$.[16] As in the analysis of Section 2, this equation captures how, given beliefs p, a control determines 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. Any σ such that $\sigma(\cdot \mid p) \to \Sigma$, for $p \in \Delta(\Theta)$, will be a (Markov) *control rule*. Using the control rule $\sigma^{\alpha,\delta}$, we can construct a (non-randomized) decision rule $f = f^{\alpha,\delta}$ from $\sigma^{\alpha,\delta}$ by setting

$$\begin{aligned} f_1^t(a^1, \ldots, a^{t-1}, \theta^1, \ldots, \theta^t, \chi^1, \ldots, \chi^t) &= \sigma_1^{\alpha,\delta}(\theta^t \mid p^t), \\ f_2^t(a^1, \ldots, a^{t-1}, \chi^1, \ldots, \chi^t) &= \sigma_2^{\alpha,\delta}(p^t), \end{aligned}$$

where p^t is the belief that player 2 has about θ^t at the beginning of t. Observe that the sequence $(p^t)_{t \geq 1}$ can be recursively computed as

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

given the initial belief p^1.

Lemma 4. *Let $\alpha \in \mathbb{R}^2_{++}$. The following hold:*

a. *The value of the maximization problem (4.1) is $q(\alpha) = w^{\alpha,\delta}(\lambda)$.*
b. *The decision rule $f = f^{\alpha,\delta}$ 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 will characterize equilibrium behavior when players are patient. It will be useful to consider efficient decision rules and payoffs as $\delta \to 1$. We define the *differential discounted value* function as

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

[16] Existence and uniqueness of solution $w^{\alpha,\delta}$ follows from standard arguments, (see Stokey and Lucas, 1989).

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

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

Just to set ideas, assume that there exist subsequences $(h^{\alpha,\delta^{\nu}})_{\nu\geq 0}$ and $(w^{\alpha,\delta^{\nu}})_{\nu\geq 0}$ pointwise converging to functions $h^{\alpha}\colon \Delta(\Theta)\to\mathbb{R}$ and $w^{\alpha}\colon \Delta(\Theta)\to\mathbb{R}$. That is, $h^{\alpha}(p)=\lim_{\nu\to\infty}h^{\alpha,\delta^{\nu}}(p)$ and $w^{\alpha}(p)=\lim_{\nu\to\infty}w^{\alpha,\delta^{\nu}}(p)$ for all p. Therefore $\rho^{\alpha}=\lim_{\nu\to\infty}w^{\alpha,\delta^{\nu}}(p^1)$ does not depend on p^1. 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)

$$h(p)+\rho=\max_{\sigma\in\Sigma}\Big\{u^{\alpha}(\sigma,p)+\sum_{a_1\in A_1}h\big(B(\cdot\mid\sigma_1,p,a_1)\big)\big(\sum_{\theta\in\Theta,\sigma_1(\theta)=a_1}p(\theta)\big)\Big\} \tag{4.6}$$

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

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

Theorem 1 (Optimality Theorem)**.** *Fix $\alpha\in\mathbb{R}^2_{++}$. 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 in (4.6).*

b. *For any converging subsequence $h^{\alpha,\delta^{\nu}}\to\bar{h}$ as $\nu\to\infty$, we can take $\rho=\lim_{\nu\to\infty}w^{\alpha,\delta^{\nu}}(p^1)$ that does not depend on p^1, and obtain a pair $(\bar{h},\rho)$ that solves the AROE (4.6). The function $\bar{h}\colon\Delta(\Theta)\to\mathbb{R}$ is convex.*

c. *For any decision rule f, $\limsup_{\delta\to 1}\sum_{i=1}^{2}\alpha_i v_i^{\delta}(f)\leq\rho=\lim_{\nu\to\infty}w^{\alpha,\delta^{\nu}}(p^1)=\rho^{\alpha}$.*

This result shows that studying (4.6) is useful to determine optimal payoffs and behavior when players are patient. The first part 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 obtained by solving problems with discount factores that go to 1. The third part formally establishes that the solution $\rho\in\mathbb{R}$ to (4.6) provides a tight upper bound for the value of the discounted problem, as the discount factor goes to 1.

The AROE (4.6) is central to our analysis. The right-hand side of (4.6) captures the trade-off that an optimal control σ solves as a function of current beliefs $p\in\Delta(\Theta)$. An optimal rule takes into account current period payoffs and the distribution over beliefs in the subsequent round. Since the differential value h is convex in $p\in\Delta(\Theta)$, more precise beliefs always improve continuation payoffs. The benefit of a perfectly separating rule is that player 1 can use his private information to maximize his own ex-post total payoffs and improve continuation beliefs. On the other, the cost of a separating rule is that player 2's payoff is maximized when player 1's behavior can be perfectly predicted.[17] More generally, solutions to (4.6) will involve a complex mix of trade-offs, and explicit formulas are in general unfeasible.[18]

The following result shows that when rules that separate types maximize current weighted payoffs, they also maximize total undiscounted weighted payoffs.

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

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

Then,

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

4.2. **Equilibrium Analysis.** In this section, we investigate the conditions under which the undiscounted optimal dynamics can be approximated by an equilibrium of our repeated game.

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

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

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

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

Definition 1. *A control rule σ* determines a unique recurrence class *if the process $(\theta^t, p^t)_{t \geq 1}$ is a finite Markov chain having a unique recurrence class.*[19]

This is arguably an important restriction over the Markov process $(\theta^t, p^t)_{t \geq 1}$. On the one hand, the path of the Markov chain $(\theta^t, p^t)_{t \geq 1}$ could be countable. This case arises when the

[17] The solution to the maximization problem $\max_{\sigma \in \Sigma} U_2(\sigma, p)$ will typically be a pooling rule, in which player 2 can perfectly predict the action player 1 will employ.

[18] Problem (4.6) is similar to a bandit problem with Markovian hidden state (Keller and Rady, 1999). Separating rules maximize *exploration*. Propositions 1 and 3 show conditions under which the standard exploration vs exploitation dilemma (Bergemann and Valimaki, 2006) does not arise.

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

rule pools all types along the path and the initial belief does not coincide with the stationary distribution of the transition matrix P. Exploring the ergodicity properties of $(\theta^t, p^t)_{t\geq 1}$ in hidden Markov models is a question dating back to Blackwell (1951). Interesting recent developments exist, but they do not apply to a model like ours in which the observation variable is endogenous.[20] As the following result shows, any separating control rule determines a unique recurrence class.

Proposition 2. *Assume that the control rule σ is such that for any belief $p \in \Delta(\Theta)$ having positive probability in the path $(\theta^t, p^t)_{t\geq 1}$, types are separated: $\sigma(\theta \mid p) \neq \sigma(\theta' \mid p)$ for all $\theta \neq \theta'$. Then, σ determines a unique recurrence class.*

This result follows since when types are separated, continuation beliefs come from the set $\{P(\cdot \mid \theta) \mid \theta \in \Theta\}$. In this case, the support of the process $(\theta^t, p^t)_{t\geq 1}$ is $\Theta \times (\{p^1\} \cup \{P(\cdot \mid \theta) \mid \theta \in \Theta\})$ and its only recurrence class is $\Theta \times \{P(\cdot \mid \theta) \mid \theta \in \Theta\}$.

As we already discussed, a control rule σ^α solving the AROE need not determine a unique recurrence class. However, the following result shows that relaxing the optimality requirement to allow for approximate optimality is enough to ensure the existence of a control rule determining a unique recurrence class.

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

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

When the control rule σ^α solving the AROE separates types, this lemma is immediate. To intuitively understand this result, consider the game in Section 2 and assume that the optimal rule is such that player 1 pools by playing I for any belief $p > (1-\mu)$.[21] This rule generates an infinite belief path. We can modify the rule so that after a sufficiently large number of periods, player 1's rule separates his types. This will, again, generate a new belief path that can be truncated after some time by changing the rule so that player 1's types are separated again. The modified rule determines a unique recurrence class and incurs an arbitrarily small loss in welfare.

[20] Recent results by Van Handel (2009) and Tong and Van Handel (2012) apply to models in which the observation variable (the action in our case) is assumed to have full support. To adapt those results to our setup one would need to restrict attention to perfectly mixed controls, which are suboptimal in our model. While it is true that an optimal control rule can be approximated by perfectly mixed controls, the process of beliefs would have an infinite support. Our proof for Theorem 2 cannot be extended to accommodate infinite beliefs.

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

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

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

are well defined. This follows from Proposition 8.1.1 in Puterman (2005) after noticing that the limits are average rewards from a stationary Markov decision rule over a finite state Markov process. Letting $\bar{\pi} = \bar{\pi}^\sigma \in \Delta(\Theta \times \mathscr{P})$ be the stationary distribution for the Markov chain $(\theta^t, p^t)_{t\geq 1}$, given the control rule σ, with $\Theta \times \mathscr{P}$ the recurrence class of the chain, it follows that

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

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

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

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

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

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

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

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

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

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

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

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

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

Then, there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, the infinitely repeated game with discount factor δ has a perfect Bayesian equilibrium $s^* = (s^*_1, s^*_2)$ such that

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

This result characterizes approximately optimal equilibrium behavior. It shows that provided players are patient enough, players' incentives can be aligned to attain total weighted payoffs arbitrarily close to ρ^α. Moreover, with sufficiently high probability, conditional on simulated beliefs, players equilibrium actions will approximate the frequencies induced by the approximately optimal rule σ. In other words, given observed actions, equilibrium behavior cannot be distinguished from optimal dynamics.[22]

The proof of Theorem 2 proceeds by constructing strategies in which player 2 forgives but does not forget. To do that, we revisit the review strategy idea by Radner (1981) and Townsend (1982) and build strategies in which player 2 keeps checking whether player 1's actions can be distinguished from the rule σ_1. The details of our formulation are closely related to the quota mechanisms in Jackson and Sonnenschein (2007), Renault et al. (2013), and particularly Escobar and Toikka (2013). One conceptual difference is that in our model players cannot explicitly communicate and therefore we cannot formulate the problem as a mechanism design one.[23]

For a given sequence of actions $(a^1, \ldots, a^t) \in A^t$ and $(p, a_1) \in \mathscr{P} \times A_1$, define

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

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

The first step is to introduce an artificial player, player 0, who decides player 2's actions (as a function of the history) and can also decide whether he lets player 1 to choose his action or whether player 0 itself chooses player 1's action at any given round. We will fix the behavior of player 0 and analyze the incentives player 1 has when deciding actions. Player 0 can always

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

[23] A more technical difference is that in our model we have to approximate the decision rule by an approximately optimal decision rule determining well-behaved Markov chains. This was already addressed in Lemma 5.

"interpret" an action by player 1 through the control $\sigma_1^\alpha(\cdot \mid \bar{p}^t)$ given the *simulated* probability $\bar{p}^t$. Indeed, given a history of actions $(a^1,\dots,a^t)$, player 0 computes the simulated probabilities $\bar{p}^1 = p^1$ and recursively define $\bar{p}^{t+1}(\cdot) = B(\cdot \mid \sigma_1, p^t, a_1^t)$. If $a_1^t \notin A_1(\bar{p}^t)$, we assign $\bar{p}^{t+1} = p_0$ where $p_0 \in \mathscr{P}$ is arbitrary.

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

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

for all $p \in \mathscr{P}$. Given $T \ge 1$, a rule σ and sequence (b_k), the game of credible play $(\sigma,(b_k),T)$ is constructed as follows. For $t \le T$, if player 1 has passed the test (b_k) in all previous rounds $t' = 1,\dots,t-1$, then he can freely select his action a_1^t; otherwise, player 0 chooses a_1^t by randomly drawing an action according to the distribution $m(\cdot \mid \bar{p}^t)$. We define the *obedient strategy* for player 1 as $\hat{s}_1^t(\theta^1,\dots,\theta^t,a^1,\dots,a^{t-1}) = \sigma_1(\theta^t \mid \bar{p}^t)$ whenever he is allowed to choose actions. We will also define the block-game of credible play $(\sigma,(b_k),T)^\infty$ as the infinite horizon problem in which a game of credible play restarts after T rounds of play (with discount factor δ).

Lemma 6. *Let* $\eta > 0$.

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

$$\mathrm{P}_{\hat{s}_1}[\textit{Player 1 passes the test } (b_k) \textit{ at } (a^1,\dots,a^t) \textit{ for all } t] \ge 1-\eta.$$

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

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

To establish Theorem 2, we use this lemma to construct strategies delivering the desired weighted equilibrium payoffs ρ^α. Strategies are of the stick-and-carrot type (Fudenberg and Maskin, 1986). On the path of play, players choose actions mimicking the path of play in the equilibrium of the block-game of credible play from Lemma 6. Any observable deviation by i triggers a punishment phase, in which player i is minmaxed by a number of rounds, and then play proceed to a carrot phase in which players mimic the play of the game of credible play yielding payoffs v^i.

5. Games with Separating and Monotonic Dynamics

We now provide a characterization of solutions to (4.6). We assume that A_1 and Θ are contained in $\mathbb{R}$ and write $A_1 = \{a^n \mid n = 1,\dots,|A_1|\}$ and $\Theta = \{\theta^m \mid m = 1,\dots,|\Theta|\}$ with $a^n < a^{n+1}$ and $\theta^m < \theta^{m+1}$. We extend the payoff function for player 1, u_1, to actions $a_1 \in \mathbb{R}$ and states $\theta \in \Theta$ so that $u_1(a_1,a_2,\theta)$ is twice continuously differentiable in $(a_1,\theta) \in \mathbb{R}\times\mathbb{R}$.

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

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

Proposition 3 shows conditions under which the optimal control rule is strictly increasing. Since actions are discrete, this property cannot be inferred by simply appealing to strong increasing differences. There are two forces behind this result. Separating rules (in particular, strictly increasing rules) make continuation beliefs more precise and therefore maximize continuation payoffs (Proposition 1). This effect is reinforced when the action set is rich because in this case the maximization of total period payoffs yield strictly increasing rules.

Proposition 3. *Assume that u_1 has strongly increasing differences in (a_1,θ). Let $\alpha \in \mathbb{R}^2_{++}$ be such that*

$$\alpha_1 u_1(a^{|A_1|-1},a_2,\theta)+\alpha_2 u_2(a^{|A_1|-1},a_2) > \alpha_1 u_1(a^{|A_1|},a_2,\theta)+\alpha_2 u_2(a^{|A_1|},a_2) \qquad (5.1)$$

and

$$\alpha_1 u_1(a^1,a_2,\theta)+\alpha_2 u_2(a^1,a_2) < \alpha_1 u_1(a^2,a_2,\theta)+\alpha_2 u_2(a^2,a_2) \qquad (5.2)$$

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

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

and

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

Assume that

$$\frac{2c_1}{\alpha_1 c_2} \max_{n=1,\ldots,|A_1|-1} \{a^{n+1}-a^n\} < \min_{m=1,\ldots,|\Theta|-1} \{\theta^{m+1}-\theta^m\}. \qquad (5.3)$$

Then, any rule σ^α attaining the maximum in (4.6) is such that $\sigma_1^\alpha(\theta \mid p)$ is strictly increasing as a function of θ for all $p \in \Delta(\Theta)$ with $p(\theta) > 0$ for all $\theta \in \Theta$. Moreover, endowing $\Delta(\Theta)$ with the (partial) order $\geq_{\Delta(\Theta)}$ given by first-order stochastic dominance, and assuming that $P(\cdot \mid \theta') \geq_{\Delta(\Theta)} P(\cdot \mid \theta)$ for all $\theta' \geq \theta$, and $u_1(a,\theta)$ and $u_2(a)$ are supermodular (in (a,θ) and a respectively), then $\sigma^\alpha(\theta \mid p)$ is nondecreasing in (θ, p), where .

Equations (5.1)-(5.2) ensure that the optimal rule is not in the boundary. Provided the set of actions is rich enough, as imposed in (5.3), it follows that the optimal rule always separates types. Observe that the separating rule σ^α determines a unique ergodic class. Some applications follow.

5.1. **Collusion with Bertrand Competition.** Tacit collusion is a prominent feature of many industries, as documented, for example, by Bresnahan (1987) for the American automobile market, and by Blume et al. (2002) for the European industrial sugar market. In this section, we study a model of tacit collusion with Bertrand competition. Two firms set prices $a_i \in A_i$ at each $t = 1, 2 \ldots$. Firms sell heterogeneous goods. The demand functions are given by

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

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

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

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

We can apply Proposition 3 to characterize the welfare maximizing control rule σ^{α}, for $\alpha = (1, 1)$. Up to integer restrictions,

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

Under the optimal control rule σ^{α}, firm 1 signals its type by choosing a higher price when its demand is high. When firm 1 chooses a high price in period t, then its demand is more likely to be high in period $t + 1$ and player 2's price is also higher. In this sense, a low price by firm 1 in t triggers a price war in $t + 1$, in which firm 2's price is low and firm 1's prices are also low. The price war is over only once firm 1's price raises. Observe that σ^{α} is a rule determining a unique recurrence class and therefore Theorem 2 applies.

As Marshall and Marx (2013) explain, during the period 2000-2005, the European Commission classified 9 out of the 22 major industrial cartels as showing evidence of "of frequent bargaining problems and deviations by cartel members, occurring throughout the cartel period."[24] These "deviations" are also highlighted by Genesove and Mullin (2001) in the study of the sugar cartel. In contrast to other theoretical papers, in our setup equilibrium price cuts actually occur and apparent deviations can be seen as the result of firms using their private information to maximize total profits and signal their continuation play.[25]

Our model also explains collusive price leadership: the informed firm becomes a price leader as whenever it raises its price in t, firm 2's price will be higher in $t + 1$. Thus, our model is one of the few[26] that gives theoretical support to Stigler's 1947 observation that price leadership may

[24] See also Bernheim and Madsen (2014).

[25] Models of price dispersion could also be interpreted as generating equilibrium price cuts. See Bernheim and Madsen (2014) for an application of this idea in a collusion context. Price cuts could also improve monitoring in collusion models with imperfect public monitoring (Rahman, 2014).

[26] 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.

be an efficient mechanism to transmit information, and to Markham's 1951 view that firms may use "price leadership in lieu of an overt agreement." Several papers document collusive price leadership. Allen (1976) shows evidence of collusive price leadership in the market of steam turbine generators, and Marshall et al. (2008) discuss evidence of collusive price leadership in vitamins, rubber chemicals, sorbates, monochloroacetic acid and organic peroxides, polyester staple, high-pressure laminates, amino acids, carbonless paper, cartonboard, and graphite electrodes. Mouraviev and Rey (2011) show that price leadership features in 16 out of 49 European Commission's cartel decisions as of July 2010.

The present 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.[27] In all these works, firms can be arbitrarily close to the first best collusive outcome, in which only the lowest cost firm produces and fixed the consumers' reservation value. As Athey and Bagwell (2001) observe, communication can be dispensed with as prices can be used to signal costs (at an arbitrarily low cost). 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.[28]

The linearity restrictions imply that average prices do not depend on whether firms can actually communicate freely. To see this, suppose that θ^t realizes independently across time. With communication, firm's prices would equal

$$\sigma_2^{com}(\theta) = \frac{1+z\theta}{2(1-z^2)}, \quad \sigma_1^{com}(\theta) = \frac{\theta}{2} + z\sigma_2^{com}(\theta).$$

Taking expectations with respect to θ, it follows that $\mathbb{E}[\sigma_i^{\alpha} - \sigma_i^{com}] = 0$ for $i = 1,2$. On the other hand, with communication prices fully react to θ^t. Exploring the impact of communication among colluding firms on consumers' welfare in more general Bertrand games seems like a promising research questions.

[27] 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.

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

5.2. **Graduated Sanctions in Collective Action Games.** Case studies show that punishments are not drastic but rather gradual. In many groups, players "who violate operational rules are likely to be assessed graduated sanctions" (Ostrom, 1990, p. 94) and are even given opportunities to make restitutions. As Dixit (2009) and Abreu et al. (2005) argue, this evidence contrasts with the more standard theories of repeated games with perfect and imperfect monitoring (Abreu, 1988; Green and Porter, 1984; Abreu et al., 1986, 1991). Our set-up provides a rationale for graduated sanctions in repeated games with incomplete information.[29]

We specialize our model to a repeated collective action game. At each t, players simultaneously choose actions $a_i \in A_i \subseteq \mathbb{R}_{++}$. Player 1's type θ belongs to a set $\Theta \subset \mathbb{R}_{++}$. An action is interpreted as a contribution to the team (or group), whereas player 1's private type determines how much player 1 benefits from the contributions. More concretely, we assume

$$u_1(a,\theta) = \theta a_1 a_2 - a_1^2 \quad \text{and} \quad u_2(a,\theta) = a_1 a_2 - a_2^2.$$

The terms $-a_i^2$ in the utility functions capture the costs of contributing, while $\theta a_1 a_2$ and $a_1 a_2$ are the benefits obtained by each player from the complementary contributions. The benefit that player 1 obtains from the project is privately known. We assume that the transition $P(\cdot \mid \theta)$ first-order stochastically increases in θ.

Proposition 3 implies that, under the boundary restrictions and $\max\{a^{n+1} - a^n\} \leq 4 \min\{a_2 \in A_2\} \min\{\theta^{m+1} - \theta^m\}$, the rule maximizing the sum of utilities $\sigma^* = (\sigma_1(\theta \mid p), \sigma_2(p))$ is separating and increases in (θ, p). Given a belief $p \in \Delta(\Theta)$, consider two actions for player 1 $\bar{a}_1 > \hat{a}_1$. The corresponding continuation beliefs are $\bar{p} > \hat{p}$. This means that the actions that player 2 will take are, respectively, $\bar{a}_2 > \hat{a}_2$. More generally, conditional on p^t, the action chosen by player 2 in $t+1$ is strictly increasing as a function of the action chosen by player 1 in t, a_1^t. This means that the dynamics generated in our model of incomplete information exhibit graduated sanctions (players always contribute positive amounts) that fit the size of previous contributions. Player 1 also "makes restitutions" by taking higher actions that positively affect player 2's continuation beliefs and actions.

6. Equilibrium as Interactions Become Frequent

Our limit results, Theorems 1 and 2, apply when $\delta \to 1$. As Abreu et al. (1991) point out, the limit $\delta \to 1$ can be interpreted saying that either interest (discount) rates are low or that players move frequently. In games with imperfect monitoring, Abreu et al. (1991) show that the two interpretations can lead to radically different results as when moves become more frequent not only the interest rates change but also the quality of the monitoring technology. In our perfect monitoring game of incomplete information, the impact of more frequent moves is also subtle as

[29] Other works explain graduated sanctions using repeated extensive-form games (Mailath et al., 2004).

types are more likely to remain unchanged between two consecutive rounds. In this subsection, we explore these issues 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 game as in Section 2, with the payoffs given in Table 1. Monitoring is perfect, but only player 1 can observe $\theta^t \in \{l, h\}$ at the beginning of round t, with $l < h$. We parameterize both the discount factor and the transitions by D. The discount factor equals $\delta = \exp(-rD)$, where $r > 0$ is the discount rate per time unit. Transitions are given by

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

with $\phi, \chi > 0$. The initial type is drawn from a probability distribution such that $\mathbb{P}(\theta^1 = l)$. We 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 findings in Section 2 as taking the interest rate to 0 ($r \to 0$). Our interest now is in the limit $D \to 0$.

The formulation of the dynamic programming problem characterizing decision rules that maximize the sum of payoffs for $D > 0$ can be imported from Sections 2 and 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

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

The following result characterizes the solution to this problem when D is small.

Proposition 4. *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 $\bar{\sigma}$. Such solution is such that $\sigma_1(l \mid p) = I$ and $\sigma_1(h \mid p) = NI$. Moreover, $w^D(p) \to 2(a - l)\frac{\chi}{\phi + \chi}$ as $D \to 0$.*
b. *For all $\varepsilon > 0$, 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 (resulting in reactive signaling dynamics) is optimal whenever the game is played frequently, and that the incentive costs are modest. Intuitively, when the game is played frequently, the costs of signaling a change of type is small (it is incurred once) compared to the benefit of perfectly revealing information (which results in almost perfect information for several rounds of interaction).[30] This implies that as interactions become more frequent, it becomes more likely that players can attain the full benefits of cooperation without incurring significant signaling costs. Indeed, as shown in Section 2, if players can communicate

[30] The costs of signaling are $\mathscr{O}(D)$ whereas the benefits are $\mathscr{O}(1)$.

average total payoffs equal $2(a-l)\frac{1-\mu}{2-\lambda-\mu}$ which, as $D \to 0$, converges to $2(a-l)\frac{\chi}{\phi+\chi}$ –the payoff attained in the game with frequent moves.

This proposition, together with Section 2, show that the effects of reducing the interest rate toward zero are different from those of making the interactions more frequent. When $r \to 0$, the dynamics of cooperation can be time-off, whereas when $D \to 0$ the dynamics of cooperation are reactive signaling. This finding resonates well with those for games with complete information but imperfect monitoring (Abreu et al., 1991; Sannikov and Skrzypacz, 2007). While a related point is made in a mechanism design problem by Skrzypacz and Toikka (2015), we seem to be the first ones explicitly studying the differences between low interest rates and frequent interactions in a repeated game model of incomplete information.[31]

7. Conclusions

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 either because such protocols are non-existent or incomplete (Schelling, 1960; Marschak and Radner, 1972), or because those conditions materialize only after some other player has already made a decision. We explore the design of optimal equilibria in this type of environment, and show that the dynamics of cooperation are quite rich and novel, and shed light on phenomena that were previously unexplained.

Our results help explain some of the dynamics commonly observed in cooperative relations. First, we explain why economic agents sometimes find it optimal to forgive hostile or aggressive conducts from other agents in a long-run relationship. Second, we explain why to forgive is not to forget: most agents have a limit to the number of aggressions they are willing to tolerate, and the cooperative relationship may end if that limit is surpassed. Third, we show that restarting cooperation after an aggressive conduct has been observed may require costly gestures from the infringing party, or that agents may have to spend a cooling-off period until cooperation is sufficiently likely to be successful again. Our model shows that these behaviors may arise as an efficient way of transmitting information about the likelihood of successful cooperation. Finally, we show that incomplete information may have a significant effect on welfare, even when players are very patient (i.e., have a large valuation for future payoffs), and that in some cases, it may even preclude a cooperative relationship (i.e., the optimal equilibrium consists of repetitions of a static Nash equilibria).

Our model also explains why firms sometimes engage in unilateral price changes, and gives theoretical support to Stigler's 1947 observation that price leadership may arise optimally as a

[31] In Skrzypacz and Toikka's 2015 model when trade is more frequent, the increase in the persistence of the process of types is detrimental for incentives. In our model, the increase in the persistence of the process helps as signaling a type has more benefits. Cardaliaguet et al. (2015) study a limit model in the zero-sum case and characterize the value using a Hamilton-Jacobi equation.

way to transmit information between competitors. Our model also explains the rationale behind the commonly observed practice of graduated sanctions and restitutions in common-pool resource settings (Ostrom, 1990).

Some extensions to our model would be relatively simple to execute. We have worked with a one-sided incomplete information game to emphasize the forces in the model, but extending the results to allow for two-sided incomplete information entails no challenge.[32] We could also extend our results to allow for restricted communication or communication only once the stage game has been played (but before the subsequent type is realized). It would also be interesting to explore the equilibrium set when the discount factor is not arbitrarily close to 1, possibly allowing imperfect monitoring.[33] We leave these extensions for future research.

Appendix A. Proofs

Proof of Lemma 1. Convexity and continuity follow since w is the maximum of convex and continuous functions. To see that w is nondecreasing, we prove that whenever w is nondecreasing, so is the function

$$\max\{w_{III}(p), w_{NNN}(p), w_{INI}(p), w_{INN}(p)\}.$$

Observe first that w_{III}, w_{NNN} and w_{INI} are nondecreasing if w is nondecreasing. Write

$$w_{INN}(p) = \delta w(1-\mu) + p\big((1-\delta)(2b-l) + \delta w(\lambda) - \delta w(1-\mu)\big)$$

and note that this function is nondecreasing whenever $(1-\delta)(2b-l) + \delta w(\lambda) - \delta w(1-\mu)$ is nonnegative. But if this last term is negative, then

$$w_{INN}(p) \leq \delta w(1-\mu) \leq \delta w(p\lambda + (1-p)(1-\mu)) = w_{NNN}(p),$$

and therefore the maximization cannot be attained by the rule *INN*. It follows that $\max\{w_{III}(p), w_{NNN}(p), w_{INI}(p), w_{INN}(p)\}$ is nondecreasing. □

Proof of Lemma 2. We begin by showing there exist $p_1 \in \left(0, \frac{1}{2}\right)$, $p_2 \in (0,1)$, and $p_3 \in [0,1)$ such that: (a) $w_{INI} < w_{INN}$ for $p < p_1$ and $w_{INI} > w_{INN}$ for $p > p_1$, (b) $w_{INI} < w_{NNN}$ for $p < p_2$ and $w_{INI} > w_{NNN}$ for $p > p_2$, and (c) $w_{NNN} < w_{INN}$ for $p < p_3$ and $w_{NNN} > w_{INN}$ for $p > p_3$. We will show how to obtain (b). The other results are obtained in a similar way.

[32] Extending the dynamic programming formulation to a two-sided incomplete information model is immediate. Escobar and Toikka (2013) and Renou and Tomala (2013) design protocols to handle incentives in dynamic models of two-sided Markovian private information. Those arguments can be readily applied to our setup under ergodicity restrictions.

[33] Our dynamic programming characterization (4.3) still provides an upper bound for the equilibrium set or when monitoring is imperfect. Frankel (2015) studies "discounted quotas" and shows their optimality properties in mechanism design problem with iid types and arbitrary patience.

Recall that:

$$\begin{aligned} w_{INI}(p) &= p((1-\delta)2(a-l)+\delta w(\lambda))+(1-p)((1-\delta)(2b-l)+\delta w(1-\mu)), \\ w_{NNN}(p) &= \delta w(p\lambda+(1-p)(1-\mu)). \end{aligned}$$

It is straightforward to see that $w_{INI}(0) < w_{NNN}(0)$ and $w_{INI}(1) > w_{NNN}(1)$. Also, $w_{INI}(p)$ is linear and $w_{NNN}(p)$ is convex (given that $w(p)$ is convex). Thus, the two functions intersect exactly once. Thus, there exists $p_2 \in (0,1)$ such that $w_{INI} > w_{NNN}$ for $p > p_2$ and $w_{INI} < w_{NNN}$ for $p < p_2$.

If joint investment is ever to take place, then $w_{INI}(\lambda) \geq \max\{w_{NNN}(\lambda), w_{INN}(\lambda)\}$. A sufficient condition for $w_{INI}(\lambda) > w_{INN}(\lambda)$ is $\lambda \geq \frac{1}{2}$. By convexity,

$$w(p\lambda+(1-p)(1-\mu)) \leq pw(\lambda)+(1-p)V(1-\mu).$$

Thus, if $\lambda 2(a-l)+(1-\lambda)(2b-l) > 0$, then $w_{INI}(p) > w_{NNN}(p)$. A sufficient condition is $\lambda \geq \frac{l-2b}{2(a-b)-l}$. Thus, if $\lambda \geq \max\left\{\frac{1}{2}, \frac{l-2b}{2(a-b)-l}\right\}$, the optimal rule dictates play of *INI* when beliefs are $p = \lambda$.

Now suppose that current play is at *INI*, and player 1 plays *N*. Next period, beliefs change to $p = 1-\mu$. Optimal play at $p = 1-\mu$ depends on the comparison between w_{INI}, w_{INN}, and w_{NNN}. There are three cases.

First, if $p_1 < 1-\mu$ and $p_2 < 1-\mu$, then the optimal play at $p = 1-\mu$ is *INI*. Thus, the optimal rule is to play *INI* for any belief p, which is a particular case of time off with $\hat{\tau} = 0$.

Second, if $1-\mu < p_2$ and $1-\mu < p_3$, then the optimal play at $p = 1-\mu$ is *INN*. In this case, player 1 plays *I* if her type is l and plays *N* if her type is h. Next period, beliefs change to either λ or $1-\mu$, and optimal play is either *INI* or *INN*. Thus, the optimal rule is reactive signaling.

Third, if $p_3 < 1-\mu < p_2$, then the optimal play at $p = 1-\mu$ is *NNN*. Player 1's action does not reveal her type. Next period, beliefs are updated to $p' = (1-\mu)\lambda + \mu(1-\lambda)$. If $p' < p_2$, play continues at *NNN*. Play continues at *NNN* until $\mathrm{P}(\theta^t = l) > p_2$, in which case play switches to *INI*. Thus, the optimal rule is time off with $\hat{\tau} > 0$. Note that if $p_2 > \frac{1-\mu}{2-\lambda-\mu}$, then play will not switch back to *INI* and the optimal play will be *NNN* forever. A sufficient condition for $p_2 < \frac{1-\mu}{2-\lambda-\mu}$ is $\frac{1-\mu}{1-\lambda} > \frac{l-2b}{2(a-l)}$. □

Proof of Lemma 3. Follows from Lemma 7 below. □

Lemma 7. *Reactive signaling leads to positive welfare if and only if $\beta < \beta_{RS} = \frac{\lambda}{2(1-\lambda)}$, and time off leads to positive welfare if and only if $\beta < \beta_{TO} = \frac{1-\mu}{(1-\lambda)(2-\lambda-\mu)}$. Optimal cooperation dynamics depend on parameters as follows: (1) if $\beta_{TO} \leq \beta_{RS}$, there exists a threshold $\beta_0 < \beta_{TO}$ such that time off is better than reactive signaling for $\beta < \beta_0$ and reactive signaling is better than time off for $\beta_0 < \beta < \beta_{RS}$, and (2) if $\beta_{TO} > \beta_{RS}$, then either (a) there exist thresholds β_1 and β_2, with $\beta_1 < \beta_2 < \beta_{RS}$, such that time off is better than reactive signaling for $\beta < \beta_1$ and*

$\beta_2 < \beta < \beta_{TO}$, and reactive signaling is better than time off for $\beta_1 < \beta < \beta_2$, or (b) time off is better than reactive signaling for all $\beta < \beta_{TO}$.

Proof of Lemma 7. At time 0, the expected values of the RS and TO rules when $\delta \to 1$ are:

$$\begin{aligned} w_{RS} &= \frac{(1-\mu)\,(\lambda - 2\,\beta\,(1-\lambda))}{2-\lambda-\mu}, \\ w_{TO} &= \frac{P(\tau) - \beta\,(1-\lambda)}{(\tau+1)(1-\lambda)+P(\tau)}, \end{aligned}$$

where

$$P(\tau) = \frac{(1-\mu)\left(1-(\lambda+\mu-1)^{\tau+1}\right)}{2-\lambda-\mu}$$

is the probability that $\theta^{t+\tau+1} = l$, given that $\theta^t = h$. See Appendix B for details on how to obtain these expressions.

We begin by showing that $\hat{\tau}$ (the optimal length of the waiting phase in TO) is nondecreasing in β. The second derivative of w_{TO} with respect to T and β is

$$\frac{\partial^2 w_{TO}}{\partial\tau\,\partial\beta} = \frac{(1-\lambda)\,\left((1-\lambda)+\left(1-P(\tau)\right)\,\log(\lambda+\mu-1)\right)}{\left((\tau+1)(1-\lambda)+P(\tau)\right)^2},$$

which is positive. Therefore, $\hat{\tau}$ is nondecreasing in β.

We now show that value decreases with β for both RS and TO. The derivatives of w_{RS} and w_{TO} with respect to β are:

$$\begin{aligned} \frac{\partial w_{RS}}{\partial\beta} &= -\frac{2\,(1-\lambda)\,(1-\mu)}{2-\lambda-\mu}, \\ \frac{w_{TO}}{\partial\beta} &= -\frac{1-\lambda}{(\tau+1)(1-\lambda)+P(\tau)}. \end{aligned}$$

Notice also that the derivative of w_{RS} is constant with respect to β, but the derivative of w_{TO} is nondecreasing in β (that is, it decreases in absolute value as β increases), because $\hat{\tau}$ is nondecreasing in β. This means that w_{RS} is linear and w_{TO} is convex with respect to β.

It is easy to see that for β sufficiently large, both w_{RS} and w_{TO} are negative, thus RS and TO are dominated by a pooling rule in which players never invest. In particular, $w_{RS} > 0$ iff $\beta < \beta_{RS}$ and $w_{TO} > 0$ iff $\beta < \beta_{TO}$, where $\beta_{RS} = \frac{\lambda}{2\,(1-\lambda)}$, and $\beta_{TO} = \frac{1-\mu}{(1-\lambda)\,(2-\lambda-\mu)}$.

Now we proceed to compare RS and TO. First, note that for RS to be optimal, $w_{RS} \geq w_{TO}$ when $T = 0$ (if RS is worse than TO when $T = 0$, then it is also worse at the optimal $\hat{\tau}$). This implies that a necessary condition for RS to be optimal is that $\beta \geq \frac{1-\mu}{2\mu-1}$. Thus, TO dominates RS for small β.

Suppose that $\beta_{TO} < \beta_{RS}$. Given continuity of w_{TO} and w_{RS} with respect to β, and the convexity of w_{TO}, there exists exactly one point in which the two lines cross. Thus, there exists a threshold $\beta_0 < \beta_{TO}$ such that TO dominates RS for $\beta < \beta_0$ and RS dominates TO for $\beta_0 < \beta < \beta_{RS}$.

Finally, suppose that $\beta_{TO} > \beta_{RS}$. Given continuity of w_{TO} and w_{RS} with respect to β, and the convexity of w_{TO}, we know that close to β_{TO} (i.e., for large β) TO must dominate RS. There are two cases: (a) w_{RS} may be above w_{TO} for intermediate values of β, or (b) w_{RS} may lie below w_{TO} for all β. □

Proof of Lemma 4. The result is the standard dynamic programming formulation of partially observed Markov decision processes (Arapostathis et al., 1993). A minor subtlety arises due to the fact that our control variables are mixed strategies which, in contrast to what is typically addressed in the literature, involve private randomizations. To address this, note that a decision rule can be equivalently written as $f = (f_i^t)$ with $f_i^t \colon A^{t-1} \times \Theta_i^t \times [0,1]^t \times [0,1] \to A_i$, where the last component of an element in the range only determines the action of player i. In other words, $a_i^t = f_i^t(a^1, \ldots, a^{t-1}, \theta_i^1, \ldots, \theta_i^t, \chi^1, \ldots, \chi^t, \chi_i^t)$ where χ_i^t is only used by i. We can expand the set over which the maximization (4.1) is performed by allowing rules where all players at t condition on the whole vector $(\chi_1^t, \ldots, \chi_N^t)$. This relaxed efficiency problem admits a dynamic programming formulation in which, without loss, public randomizations are not used. Since the solution of the relaxed problem is feasible for (4.1), we deduce that $q(\alpha) = w^{\alpha,\delta}(\lambda)$. □

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

Proof of Proposition 1. Consider the problem

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

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

Proof of Lemma 5. Let $\bar{\sigma}^{\alpha}$ be the control rule solving the AROE given α. In particular, there exists $\bar{T} \in \mathbb{N}$ such that

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

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

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

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

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

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

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

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

Proof of Lemma 6. Let us first prove a. Since the control rule σ determine a unique recurrence class (Definition 1), there exists an irreducible transition matrix $\bar{P}$ for the joint process of states and beliefs, $(\theta^t,p^t)_{t\geq1} \in \Theta \times \mathscr{P}$ and a unique stationary distribution $\bar{\pi}$ on $\Theta \times \mathscr{P}$. Using Blackwell's 1957 construction, we can extend the Markov chain (θ^t,a^t) to the negative numbers $t \in \mathbb{Z}$, and compute the invariant measure $\bar{\pi}(\theta,p) = \mathrm{P}\{\theta_0 = \theta, \mathrm{P}[\theta_0 = \cdot \mid (a^t)_{t\leq0}] = p(\cdot)\}$. In particular, for any $(\theta,p) \in \Theta \times P$,

$$\bar{\pi}(\theta \mid p) = \mathrm{P}\Big[\theta^0 = \theta \mid p = \mathrm{P}[\theta^0 = \cdot \mid (a^t)_{t\leq0}]\Big] = p(\theta). \tag{A.1}$$

For any sequence $(\theta^t,p^t)_{t\geq1}$, we define the empirical transition matrix $\bar{P}^t$ on $\Theta \times \mathscr{P}$ as

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

and the empirical measures

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Theorem 2. Take $\eta > 0$ small enough in Lemma 6 such that the expected average payoff for player 2 over the course of a game of credible play is within ε of v_2 and, for any sequential best response $s_1 = s_1^\delta$ of player 1 in the block-game of credible play, $v_1^\delta(s_1^\delta) \geq v_1 - \varepsilon$. In particular, $\alpha \cdot v^\delta(s_1^\delta) \geq \alpha \cdot v - \varepsilon \geq \rho^\alpha - 2\varepsilon$, where the last inequality follows from Lemma 5.

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

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

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

for $i = 1, 2$. Take $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$ the credible reporting games $(\sigma^{\alpha'}, (b_k), T)^\infty$, for $\alpha' = \alpha, \alpha^1, \alpha^2$, have discounted equilibrium payoffs $U^{\alpha'}(\delta)$ within distance ε of the target payoffs $v^{\alpha'}$. Define the length of the stick phase as $L(\delta) = \max\{d \in \mathbb{N} \mid d \leq \frac{\ln(\gamma)}{\ln(\delta)}\}$ and note that $\delta^L \to \gamma$. Lemma 6.1 in Escobar and Toikka (2013) shows that discounted payoffs during the L periods of the stick phase against i are bounded above by $(1-\delta^L)(\underline{v}_i + \varepsilon)$ for δ sufficiently large.

Now, consider the incentives in the carrot phase

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

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

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

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

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

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. $\square$

Proof of Proposition 3. Consider first a solution $\sigma^* \in \Sigma$ to the problem

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

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

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

is nonnegative. Now,

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

is positive under (5.3). It follows that $\sigma_1^*(\theta^{m+1} \mid p) \ge a^{n+1} > \sigma_1^*(\theta^m \mid p)$. To deduce the second part of the Proposition, use the results in Van Zandt and Vives (2007) for monotone comparative statics in Bayesian games. $\square$

Proof of Proposition 4. Lemma 2 shows that the optimal equilibrium follows either reactive-signaling or time-off dynamics. The limit of the value of playing reactive-signaling when $D \to 0$ is

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

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

$$\lim_{D\to 0} w_{TO}(\tau) = \frac{\chi\,(\tau+1)\,2\,(a-l) + \phi\,(2b-l)}{(\phi+\chi)\,(\tau+1)},$$

and the derivative of this expression with respect to τ is

$$-\frac{\phi\,(2b-l)}{(\phi+\chi)\,(\tau+1)^2} > 0,$$

Thus, as $D \to 0$, $\hat{\tau} \to \infty$, and limit of the value of playing time-off when $D \to 0$ is

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

which is equal to the limit value of playing reactive signaling. This result is very intuitive. As $D \to 0$, the process of types becomes perfectly persistent, and the probability of a type change is equal to 0. In the first period of play, the probability that player 1 has low cost is $\chi/(\phi+\chi)$. Thus, the value of playing either reactive signaling or time off is $2\,(a-l)\,\frac{\chi}{\phi+\chi}$.

In order to compare the two rules, we compare the derivatives of the limit value with respect to D, as $D \to 0$ from the right. Simple calculations show

$$\lim_{D\to 0} \frac{\partial w_{RS}}{\partial D} = (-\chi\,(a-l) + (2\chi + r)\,(2b-l))\,\frac{\chi}{\phi+\chi}, \quad \lim_{D\to 0} \frac{\partial w_{TO}}{\partial D} = -\infty.$$

Both derivatives are negative, but the derivative corresponding to time off is larger in absolute value. Thus, to the left of $D=0$, reactive signaling has greater value than time off. This proves part a of the proposition.

To prove b, we follow steps close to those in the proof of Theorem 2. The definition of game of credible reporting remains unaltered for any given D. We will prove that for a proper choice of parameters, we can replicate Lemma 6. We construct the sequence b_k from the definition of d_k (see proof of Lemma 6) by picking $0 < \psi < \lim_{D\to 0} \bar\pi^D(\theta, p)$, with $\bar\pi^D$ the stationary distribution given D, and $b_k = c_1|\Theta|(d_{k(c_2-\psi)} + \frac{1}{k})$. Conditions (A.2) and (A.3) follow immediately for any D. Condition (A.4) is also immediate, just notice that the choice of $\bar t$ depends on D so $\bar t = \bar t(D)$. This completes the first part of Lemma 6. To see the second part, construct $\bar T = \bar T(D)(> \bar t(D))$ so that for any strategy s_1 $\mathbb{P}^D_{s_1}[\|m^T(\cdot \mid p) - m^D(\cdot \mid p)\| \le \varepsilon \quad \forall p \in P^D] \ge 1-\varepsilon$. Note that for the game of credible play $(\bar\sigma, (b^D_k), T)$, with $T \ge \bar T(D)$, Player 1 can obtain a payoff at least $(a-l)\frac{\chi}{\phi+\chi} - \varepsilon$. By construction, Player 2's payoff is within ε of $(a-l)\frac{\chi}{\phi+\chi}$. Fixing $\tau, T \ge \bar T(D)$, we can find $\bar r(D)$ such that for all $r < \bar r(D)$, for any best response s_1 in the block-game of credible play, Player 1 obtains a payoff at least $(a-l)\frac{\chi}{\phi+\chi} - \varepsilon$. Taking $D \le \bar D$ and $r \le \bar r(D)$ (sufficiently small if needed), by definition equilibrium payoffs in the game played every D units of time with discount rate r are bounded above by $2(a-l)\frac{\chi}{\phi+\chi} + \varepsilon$. Observable deviations from the path of play of the block-credible reporting game are punished by Nash reversion. Provided $\bar r(D)$ is chosen sufficiently small, the result follows. □

REFERENCES

ABREU, D. (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 383–396.

ABREU, D., D. BERNHEIM, AND A. DIXIT (2005): "Self-Enforcing Cooperation with Graduated Punishments," Working paper, Princeton University.

ABREU, D., P. MILGROM, AND D. PEARCE (1991): "Information and Timing in Repeated Partnerships," *Econometrica*, 59, 1713–1733.

ABREU, D., D. PEARCE, AND E. STACCHETTI (1986): "Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory*, 39, 251–269.

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

ACEMOGLU, D. AND A. WOLITZKY (2014): “Cycles of Conflict: An Economic Model,” *The American Economic Review*, 104, 1350–1367.

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

ARAPOSTATHIS, A., V. S. BORKAR, E. FERNÁNDEZ-GAUCHERAND, M. K. GHOSH, AND S. I. MARCUS (1993): “Discrete Time Controlled Markov Processes with Average Cost Criterion: A Survey,” *SIAM Journal on Control and Optimization*, 31, 282–344.

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

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

ATHEY, S. AND K. BAGWELL (2001): “Optimal Collusion with Private Information,” *RAND Journal of Economics*, 32, 428–465.

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

ATHEY, S., K. BAGWELL, AND C. SANCHIRICO (2004): “Collusion and price rigidity,” *Review of Economic Studies*, 71, 317–349.

AWAYA, Y. AND V. KRISHNA (2014): “On Tacit versus Explicit Collusion,” Working paper, Penn State.

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

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

BERGEMANN, D. AND J. VALIMAKI (2006): “Bandit Problems,” *Yale University*.

BERNHEIM, B. AND E. MADSEN (2014): “Price Cutting and Business Stealing in Imperfect Cartels,” Working paper, National Bureau of Economic Research.

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

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

BLUME, C., N. STRAND, AND E. FÄRNSTRAND (2002): “Sweet Fifteen: The Competition on the EU Sugar Markets,” Swedish competition authority report 2002:2, Stockholm, Sweeden.

BRESNAHAN, T. (1987): “Competition and Collusion in the American Automobile Industry: The 1955 Price War,” *The Journal of Industrial Economics*, 35, 457–482.

CARDALIAGUET, P., C. RAINER, D. ROSENBERG, AND N. VIEILLE (2015): “Markov Games with Frequent Actions and Incomplete Information,” *Mathematics of Operations Research*.

CLARK, R. AND J.-F. HOUDE (2013): “Collusion with Asymmetric Retailers: Evidence from a Gasoline Price-Fixing Case,” *American Economic Journal: Microeconomics*, 5, 97–123.

COLE, H., J. DOW, AND W. ENGLISH (1995): “Default, Settlement, and Signalling: Lending Resumption in a Reputational Model of Sovereign Debt,” *International Economic Review*, 36, 365–385.

DIXIT, A. (2009): “Governance Institutions and Economic Activity,” *The American Economic Review*, 3–24.

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

ESCOBAR, J. AND J. TOIKKA (2013): “Efficiency in Games with Markovian Private Information,” *Econometrica*, 81, 1887–1934.

FRANKEL, A. (2015): “Discounted Quotas,” Working paper, Chicago Booth.

FUDENBERG, D. AND E. MASKIN (1986): “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information,” *Econometrica*, 54, 533–554.

FUDENBERG, D. AND J. TIROLE (1991): *Game Theory*, MIT Press.

GALE, D. AND R. ROSENTHAL (1994): “Price and Quality Cycles for Experience Goods,” *The RAND Journal of Economics*, 590–607.

GENESOVE, D. AND W. MULLIN (2001): “Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case,” *American Economic Review*, 91, 379–398.

GENSBITTEL, F. AND J. RENAULT (2015): “The Value of Markov Chain Games with Incomplete Information on both Sides,” *Mathematics of Operations Research*.

GENTZKOW, M. AND E. KAMENICA (2011): “Bayesian Persuasion,” *American Economic Review*, 101.

GREEN, E. AND R. PORTER (1984): “Noncooperative Collusion under Imperfect Price Information,” *Econometrica*, 52, 87–100.

HÖRNER, J., D. ROSENBERG, E. SOLAN, AND N. VIEILLE (2010a): “On a Markov Game with One-Sided Incomplete Information,” *Operations Research*, 58, 1107–1115.

HÖRNER, J., T. SUGAYA, S. TAKAHASHI, AND N. VIEILLE (2010b): “Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem,” Working paper, Cowles Foundation.

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

HÖRNER, J., S. TAKAHASHI, AND N. VIEILLE (2015): “Truthful Equilibria in Dynamic Bayesian Games,” Tech. rep., Yale University.

HSU, S.-P., D.-M. CHUANG, AND A. ARAPOSTATHIS (2006): “On the Existence of Stationary Optimal Policies for Partially Sbserved MDPs under the Long-Run Average Cost Criterion,” *Systems & Control Letters*, 55, 165–173.

JACKSON, M. O. AND H. F. SONNENSCHEIN (2007): "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241–258.

KALAI, E., D. SAMET, AND W. STANFORD (1988): "A Note on Reactive Equilibria in the Discounted Prisoner's Dilemma and Associated Games," *International Journal of Game Theory*, 17, 177–186.

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

KREPS, D., P. MILGROM, J. ROBERTS, AND R. WILSON (1982): "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic Theory*, 27, 245–252.

LI, J. AND N. MATOUSCHEK (2013): "Managing Conflicts in Relational Contracts," *American Economic Review*, 103, 2328–2351.

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

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

MAILATH, G., V. NOCKE, AND L. WHITE (2004): "When the Punishment Must Fit the Crime: Remarks on the Failure of Simple Penal Codes in Extensive-Form Games," Working paper, University of Pennsylvania.

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

——— (1952): *Competition in the Rayon Industry*, vol. 1, Cambridge, MA: Harvard University Press.

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

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

MARSHALL, R. C., L. M. MARX, AND M. E. RAIFF (2008): "Cartel Price Announcements: The Vitamins Industry," *International Journal of Industrial Organization*, 26, 762–802.

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

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

OSTROM, E. (1990): *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press.

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

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

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

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

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

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

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

RENOU, L. AND T. TOMALA (2013): "Approximate Implementation in Markovian Environments," Working paper, HEC Paris.

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

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

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

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

SCHERER, F. M. AND D. ROSS (1990): "Industry Market Structure and Economic Performance," .

SKRZYPACZ, A. AND J. TOIKKA (2015): "Mechanisms for Repeated Trade," *American Economic Journal: Microeconomics*.

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

STOKEY, N. AND E. LUCAS, R. WITH PRESCOTT (1989): *Recursive Methods in Economic Dynamics*, Cambridge: Harvard University Press.

TOMZ, M. (2012): *Reputation and International Cooperation: Sovereign Debt across Three Centuries*, Princeton University Press.

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

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

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

VAN ZANDT, T. AND X. VIVES (2007): "Monotone Equilibria in Bayesian Games of Strategic Complementarities," *Journal of Economic Theory*, 134, 339–360.