Efficiently Breaking the Folk Theorem by Reliably Communicating Long Term Commitments

David K. Levine[1]

Abstract

The introduction of artificially intelligent algorithms in pricing decisions by firms has triggered a literature in industrial organization asking if the use of these algorithms will lead to collusive outcomes. In a simple repeated game environment it is shown that if algorithms can be reliably communicated or inferred the folk theorem breaks and the long-run outcome must be collusive.

[1] Department of Economics, RHUL and WUSTL

Acknowledgements: First version: June 10, 2023. I would like to thank Giacomo Calzolari, Patrick Chang, Hal Cole, Drew Fudenberg, Timothy Kehoe, John Mair, Matt Mitchell, Simon Mongey, Jose Penalva, Chris Phelan and the participants at the Timothy Kehoe Conference at the University of Minnesota. I gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL_0 and the Leverhulme Trust.

"Of course, the whole point of a Doomsday Machine is lost, if you keep it a secret! Why didn't you tell the world, EH?" Dr. Strangelove

1. Introduction

The introduction of artificially intelligent algorithms (AI) in pricing decisions by firms has triggered a literature in industrial organization asking if these algorithms will enable collusion between firms. The possibility of collusion between algorithms has long been established by folk theorems in the theory literature in which players are limited to choosing automata (algorithmic strategies).[2] Hence, provided response time is quick, firms can collude using AIs. There is, however, a deeper question that has been addressed with only partial success in the theory literature which is whether firms using AIs will always succeed in colluding, or whether they may only achieve some less mutually profitable equilibrium such as the Cournot equilibrium.[3] In this paper I consider the latter question for AI algorithms and the other use of programmed strategies.[4]

There is a simple intuition as to why in a repeated game between two players long-run outcomes should be efficient. If there is an existing status quo that is not efficient, each player has an incentive to make an offer to the other that improves utility for both. This is not a new idea, but the devil is in the details. In particular, what happens if the two commit to incompatible offers? Indeed, the standard model

[2] Rubinstein (1986) makes this point. He goes on to show that in the Prisoner's dilemma if players try to minimize the number of states used by their machines only a more limited set of outcomes (never-the-less including efficient ones) is attained. This latter result is generalized in Abreu and Rubinstein (1988).

[3] For evidence that they do succeed in colluding, see the experimental work of Calvano et al (2020), the simulation study of Asker, Fershtman and Pakes (2022) and the theoretical work of Cartea et al (2022).

[4] Empirical work showing that firms use programmed strategies includes Chen, Mislove and Wilson (2016), Brown and MacKay (2021), Leisten (2021) and Musolff (2022).

of Nash equilibrium in a repeated game can be viewed as a model in which players make simultaneous commitments, and, as might be expected since neither can respond to the other's commitment, the folk theorem holds for these games and there is no particular tendency to efficiency.

Reflection on the use of AIs by firms suggests that the standard repeated game model may not be adequate for addressing the issue of commitment. Specifically while an AI can respond quickly to opponent actions it must first be trained and this is an expensive and time-consuming procedure. Hence it is natural to think that while the response time of an AI is rapid, decisions about which AI to use are taken only occasionally. This is more broadly true whenever a decision maker delates authority, either to a computer program or to a bureacracy. Hence I distinguish between response time, which is short, and decision time which is long. As indicated this model applies not only to AIs but more broadly: for example, an organization, team, or sales force is trained to respond quickly, but that training is only updated occasionally.

In the context of decision time it is natural to think that decisions are taken asynchronously, that is, the two players are unlikely to simultaneously update their AIs or retrain their sales forces. The key insight of this paper is that when this is the case choice of an AI becomes a commitment, and players effectively take turns in making commitments. If these commitments are observed then I show that this breaks the folk theorem in a favorable way: in the long-run equilibrium play is efficient.

The role of time in this result needs emphasis. As indicated, there are two measures of patience and impatience in the model: reaction time and decision time. In calendar time reaction is fast, but the time between decisions is long. This leads to "folk-theoremesque" patience in terms of incentives - each player can provide the other with incentives. On the other hand, because decisions can be revised only infrequently (in calendar time) they represent a relatively long-term commitment. It means that in assessing the implications of a particular commitment each player is "relatively" myopic.

To understand the intuition of this main result, consider a player who is deciding how to design their AI or organization. Their opponent is commited to a particular automaton and will (probably) remain so for a long time into the future. Hence the player should design a best response to the current commitment of the opponent. However: only on-path play matters and the player is (largely) indifferent as to how to play off path as this will not matter until the opponent revises their commitment.[5] Hence off-path play should be designed to influence the opponent's play in the future when they revise their commitment. It should be in the form of an "offer" to the opponent that will provide them with incentives to "cooperate." This leads to play that in the long-run is efficient.

After illustrating this main idea with a simple example, I extend the model to a broad class of two-player *safety* games that includes public goods and duopoly games. There are three main results. First, long-run efficiency is shown to hold for observable commitments. By contrast, it is shown to fail and that instead the folk theorem holds with unobervable commitments. This reflects the fact, known certainly to Stanley Kubrick in 1964, that to be effective a commitment must be observed by the opponent. Finally, I modify the model to add an element of uncertainty similar to that in reputational models. This gives players an incentive to learn their opponent's unobserved commitment through active experimentation and observation. I show that in this case long-run efficiency does hold but with one additional key proviso: it

[5] This indifference to off-path play is important also in the evolutionary theory of cooperation, since players are indifferent between cooperating unconditionally and the equilibrium strategy of cooperating conditionally.

must be the case that players are restricted to use automata that are forgiving.

Literature Review

As indicated the study of automata in repeated games is not new, and originates in the work of Rubinstein (1986). That work also supposes that building automata is costly. This is modeled by assuming that players choose automata once and for all and try to minimize the number of states used by their automata. In contrast I model the cost of building automata by assuming that they are only built occasionally leading to rather different conclusions. The basic folk theorem result without costly choices of automata is reported in Rubinstein (1986) and has been extended to algorithmic learning procedures by Cartea et al (2022).

The idea that of players reacting quickly but planning slowly is not new, nor is the idea that the levels are important in the middle-run but reaction in the long-run. This is the idea in Levine (1981), but due to unresolved issues about observability the paper was never published. It is directly connected to the unpublished work of Salcedo (2015) and Lamba and Zhuk (2023) upon whose framework I build. Both study only the case of observable commitments and do so in a limited classes of games. Salcedo (2015) studies symmetric equilibria of a special class of symmetric games. Lamba and Zhuk (2023) also focus on symmetric games, albeit a price setting duopoly game. Lamba and Zhuk (2023) consider only automata that react to the previous period. I broaden the class of games to two player games that need only satisfy a safety condition described below and consider more general automata. My setup differs from theirs also smaller respects: neither assumes an adjustment cost and Salcedo (2015) allows automata of unlimited complexity and does not consider arbitrary initial conditions. Both restrict attention to Markov perfect equilibria. Lamba and Zhuk (2023) rule out cycles but do so by allowing a player's strategy space to depend on the (observed) strategy of the other player. I rule out cycles enogeneously through high costs, but players may cycle and incur the resulting costs. These small differences are not so important in the observable case, but my main goal is to extend the result to the unobservable case and I have chosen assumptions that make sense in both contexts.

I should also mention the work of Aumann and Sorin (1989) who give a result for special case of games of common interest. The idea behind this is different than in this paper or Salcedo (2015) and Lamba and Zhuk (2023). It is based on a reputational model and applies only to pure strategy equilibria. Like the unobservable case here forgiveness plays a key role. Also related to the result on the unobservable case is the paper of Jindani (2022) who shows that a particular learning process leads to an efficient long-run solution. That paper draws on the earlier work of Foster and Young (2003) on exogenously specified learning processes. Here, in the unobservable case, the learning process is endogenous. In the direction of endogeneous play there is a body of related work by Abreu and Pearce, for example Abreu and Pearce (2007). This also studies a game between two players. That work uses reputation and renegotation with binding contracts to introduce the key elements needed to break the folk theorem: commitment and foregiveness. They find a unique equilibrium payoff in contrast to the results here showing that the long-run must be efficient.

I should also mention the substantial literature on the evolution of strategies in a large population that enforce cooperation through punishment. Axelrod and Hamilton (1981), Binmore and Samuelson (1992), Johnson, Levine and Pesendorfer (2001), Dal Bo and Pujals (2015), and Juang and Sabourian (2021) are but a few examples. That literature is based on a different mechanism: cooperative strategies do well against each other so have an evolutionary advantage.

2. An Example

Before introducing the general model I want to talk through the simple example of the prisoners' dilemma game in which the actions are to cooperate or defect and the payoffs are given by

	cooperate	defect
cooperate	$2, 2$	$-1, 3$
defect	$3, -1$	$0, 0$

Table 2.1: PD game

Players have a common discount factor in the form of a discount rate over calendar time which we may normalize to one. I want to study a "folk theorem" type of environment in which players can observe and respond to each other quickly, but I want to distinguish between response time and decision time. The idea of quick response is modeled by taking the length of a period Δ to be short in calendar time so that the discount factor $e^{-\Delta}$ is close to one.

I now want to introduce the idea of decision time. The idea is that a player designs an AI or trains a team to respond in a certain way, but that the design decision or training takes place infrequently. We can think of a player as waking up, committing to a particular response, then falling asleep for a long time while the AI or team carries out the response plan. To be concrete, imagine that the commitment, that is the period of being asleep, lasts for $2T$ periods and that these overlap so that initially player 2 is committed for T periods and player 1 has just woken up. Think of this as a crude approximation to a Poisson process that wakes the players - that will be the formal model subsequently. Here the idea is that T is big in calendar time and specifically that $T = \tau/\Delta$ where calendar time τ is large.

For simplicity and concreteness, and indeed following the literature on evolution in repeated games, suppose that the possible commitments are responses to what the other player did last period. In technical terms these are automata. There are four such automata: play the same as the other player last period, that is, tit-for-tat, play the opposite of the other player last period, that is, anti-tit-for-tat sometimes called tat-for-tit, or to always cooperate, or to always defect. I suppose that player's automata are observable by the opponent. This is the starting point I will use for analyzing the unobservable case.

I am now going to talk through the case where initially player 2 is committed to alway defect. Player 1 as indicated, has just woken up and must decide which of four responses to implement for the next $2T$ periods. During the first T of those periods if player 1 plays tit-for-tat or always defect they get 0. If they play anti-tit-for-tat or always cooperate they get -1. If τ is large then the benefit of getting 0 over T periods rather than -1 is greater than any conceivable gain after T, so player 1 must commit to either tit-for-tat or always defect: for the next T periods it makes no difference.

The crucial point is this: tit-for-tat is in fact better than always defect. Consider what happens when player 2 wakes up. If player 1 is playing always defect then player 2 is in exactly the same position as player 1 at the beginning of the game and in periods $T+1$ to $2T$ both players will get 0. On the other hand if player 1 is playing tit-for-tat then player 2 will prefer to either playing tit-for-tat (with initial cooperation) or to always cooperate as both of these give 2, while anything else either cycles or gives 0. While in this game is possible to work out that the cycles give strictly less than 2 this will not be true in general games, so I am going to rule out cycles by fiat, modifying the usual repeated game setup by assuming an adjustment cost: if the current action is not the same as the previous period action then utility is reduced by $\phi^i F > 0$. If F is large, then without further calculation we can conclude that player 2 is going to cooperate and player 1 will get 2 between $T+1$ to $2T$, much better than the 0 from playing always defect, and since T is large, dominating anything after $2T$.

Continuing on in this way, we see that player 2 will also commit to tit-for-tat (with initial cooperation) and this will continue for the remainder of the game. In other words: starting with player 2 committed to always defect, the unique equilibrium is for player 1 to play tit-for-tat forever, and for player 2 to switch to tit-for-tat as soon as possible. Player 1 makes a good offer of cooperating to player 2 and as soon as player 2 is able they accept the offer and the outcome is efficient. Most important: this is the only equilibrium and the folk theorem has been broken in a favorable way.

3. The General Model

The simple example has a number of limitations, and I want to generalize it while preserving some of the simplicity. The general setting remains one of a infinitely repeated two player finite game. In each period $t = 1, 2, \ldots$ players $i \in \{1, 2\}$ choose observable actions $a_t^i \in A^i$ a finite set with M^i elements and receive utility $u^i(a_t^i, a_t^{-i})$. With respect to payoffs $u^i(a_t^i, a_t^{-i})$ I am going to assume a generic condition on payoffs I will state later. More important, I will limit the class of games to *safety games* like the prisoners' dilemma in which each player has a *safety action* $\mathbf{a}^i$ in the sense that this pure action guarantees that i gets a non-negative payoff and their opponent gets a non-positive payoff. This implies that both employing safety actions is a static Nash equilibrium in which each gets 0. It also implies that the individually rational payoff for each is 0. Specifically, a profile a is individually rational for i if $u^i(a) \geq 0$. There is also an adjustment cost: if $a_t^i \neq a_{t-1}^i$ then utility is reduced by $\phi^i F > 0$.

As noted, the prisoners' dilemma game is a safety game, where the safety action is to defect, as are the related public goods contribution games in which the individual gain from contributing is less than the cost and the safety action is to not donate. However, the class of games is much larger. It includes Cournot duopoly games with downward sloping demand and diminishing return to scale in which it is optimal for each player to produce to capacity when the other does and payoffs are profits minus the irrelevant constant of profits when both produce to capacity. Here the safety action is to produce to capacity. It also includes games with multiple equilibria, such as the 2x2 game with payoff matrix

$$\begin{bmatrix} 2,2 & -1,1 \\ 1,-1 & 0,0 \end{bmatrix}.$$

Since in the class of safety games the mutual minmax point is a static Nash equilibrium, when infinitely repeated the simple Friedman (1971) Nash threats folk theorem implies the Fudenberg and Maskin (1986) general folk theorem.

As in the example, players have a common discount factor in the form of a discount rate over calendar time which is normalized to one. The length of a period in calendar time is Δ so that the discount factor is $e^{-\Delta}$.

In the example I limited players commitments to one-period responses to the other player. I am now going to broaden the class of commitments to allow player i to choose within the class of B^i-state machines where $B^i \geq A^i$ is an integer. A B^i-*state machine* consists of a finite set[6] of states $\mathcal{B}^i$ with B^i elements together with mappings $\alpha^i : \mathcal{B}^i \to A^i$ and $\beta^i : \mathcal{B}^i \times A^{-i} \to \mathcal{B}^i$. The first mapping α^i says what action the machine will choose in state $b^i \in \mathcal{B}^i$ while the second says what state the machine will move to next period when the current state is b^i and the opponent plays a^{-i}. In the example the class of commitments consisted of the subset of 2-state machines for which $\beta^i(b^i, a^{-i}) = \beta^i(a^{-i})$. I refer to machines that depend only on the (finite length) past play of the other player as *reactive machines* as they only react to what

[6] Having a maximum finite number of states is a form of bounded rationality and rules out, for example, time varying strategies.

the other player did and not to their own past state or action. The class of *feasible machines* for player i is a fixed subset $\mathcal{D}^i$ of all B^i-state machines that includes at least the reactive machines and with the property that for any β^i every α^i is feasible: that is actions can be assigned to states in an arbitrary way. A *commitment* by player i consists of a choice of a machine $d^i \in \mathcal{D}^i$ together with an initial state $b^i \in \mathcal{B}^i$.

The model of coordinated overlapping commitments lasting $2T$ periods makes little sense, and I adopted it solely for illustrative purposes. As indicated I do want to assume that players wake up, take a decision, then fall asleep for a long period of time, but the natural way to think of this is as stochastic and asynchronous. Specifically, I want waking up to be triggered by independent Poisson events for the two players where the probability of that event for player i each period is $\Delta/(h^i\tau) > 0$. For convenience, normalize $h^i < 1$. As we do not imagine that Poisson events arrive at exactly the same time, if both receive a Poisson event in the same period a coin flip determines which one received the event "first."

A player might also like to make sure that their opponent's machine is in the "right" state prior to making a commitment, or if commitments are unobserved, to test their opponent's machine to see what it does. To model this I assume that after waking and prior to committing a player has N^i periods of *free play* in which they are not committed to any machine and can do what they like. For simplicity and because under the subsequent assumptions it will happen very rarely, I am going to assume that a player who wakes up during an opponents free play goes back to sleep. This reflects the reasonable idea that if your opponent is not yet committed but in the process of doing so it makes sense to wait and see what they are committed to before trying to make a commitment. It avoids the complicated (but rarely needed) reasoning: if I do this free play during my opponent's free play how will that alter their eventual commitment? I want to emphasize, however, that while implicitly players observe whether the other is awake the machines do not and cannot condition their play on Poisson events for the other player.

This structure makes sense, but it also leads to an analysis that is similar to the $2T$ coordinated overlapping commitments in the example: when player $-i$ wakes up the expected length of calendar time until the opponent wakes up is $h^{-i}\tau$ which I will assume is quite long. In this context it is useful to define the notion of a *switch*: this occurs when a player wakes up and finds that their opponent was the last to wake up.

Without loss of generality, we may continue to assume that the game begins with player 1 having just woken up and an initial condition which is a commitment (d^2, b^2) for player 2.

I am going to consider two assumptions about commitments. With *observable commitments* when a player wakes up they directly observe the commitment and current state of the opponent. This model is relevant if limited: if the commitment involves training a bureacracy to carry out a rule, it may be possible for the opponent to observe the training process. If the commitment involves coding an AI, it may be possible for the opponent to see the code used by the AI. With *unobservable commitments* the initial condition includes beliefs by player 1 in the form of a probability distribution over the initial condition for player 2 and subsequently a player can ony make observation based inferences about what commitment the opponent has made.

A *strategy* σ^i for a player in this game is a history dependent choice of commitment. The notion of equilibrium is sequential equilibrium. In the observable case the game has complete information so this reduces to subgame perfect equilibrium in which the relevant subgames begin with a player waking up and finding that their opponent has a particular commitment. Regardless of observability the game does have sequential equilibria. From Kreps and Wilson (1982) they exist for every time-truncated version of the game. Taking the limit as in Fudenberg and Levine (1983) there is a convergent subsequence of strategies and assessments that converge to a Nash equilibrium where play following is optimal with respect to the limit of the assessments. As consistent assessments cannot converge to a limit that is not consistent, this limit is a sequential equilibrium.

I will show that the folk theorem holds when commitments are not observed. The reason for this is that players can have point beliefs giving them no incentive to learn what their opponent might be doing. I want to also consider what happens if they do have an incentive to learn. This is not a new issue in equilibrium theory: point beliefs are instrumental also in the chain-store paradox, and it was to break the tyranny of point beliefs that the gang-of-four introduced committed types and the reputational model. Indeed: there are a number of ways of perturbing a model to get rid of point beliefs: players trembling, the logistic response used in quantal response equilibrium, global games, and so forth. These are all ways of forcing some uncertainty about what the opponent is doing. In the context here it is convenient to perturb beliefs directly as is sometimes done in the definition of trembling hand perfect equilibrium. Specifically, suppose that player i has a fixed distribution μ^i over feasible opponent commitments that puts weight at least $\underline{\mu}$ on each one. Let $\tilde{\mu}^i_t$ be a sequentially rational assessment for player i at wake-up time. I define an *ϵ-belief-perturbed equilibrium* as a best response in each sub-form following wake-up to the perturbed assessment $(1-\epsilon)\tilde{\mu}^i_t + \epsilon\mu^i$. Now when a player wakes up they no longer have point beliefs, so have an incentive to learn their opponent's commitment. Note that that the existence of equlibrium continues to hold in the perturbed model.

Assumptions

First, I want to make sure that N^i is sufficiently long to initalize the opponents machine. For initialization, observe that for a given initial condition b^{-i} some states in $\mathcal{B}^{-i}$ may be inaccessible, for example, if b^{-i} is an absorbing state. However, the machine must cycle in at most B^{-i} periods so an accessible state can be attained by an input of length at most B^{-i}. Hence I always assume

Assumption 3.1. *$N^i \geq B^{-i}$.*

As the order of limits is important but hard to parse it is useful to make the following definition

Definition 3.2. *F, τ are large and ΔF is small* is short-hand meaning: For the given game and fixed N^i there exist $\overline{\tau} > 0, \overline{F} > 0$ and for any $F > \overline{F}$ and $\tau > \overline{\tau}$ there exists $\underline{\Delta} > 0$ and for any Δ satisfying $0 < \Delta F < \underline{\Delta}$.

The crucial fact here is that after picking τ we pick $\underline{\Delta}$. The importance of ΔF is this: after waking there is a period of free play of at most $\max N^i$ periods and after that both players are committed to particular machines. These machines will jointly cycle after some fixed additional time, at most $B^1 B^2$ periods. Hence there is a *short-run epoch* of up to $\overline{N} \equiv \max_i N^i + B^1 B^2$ periods during which very little can be said about how players play, followed by a cycle. As adjustment costs can occur during this epoch from the point of view of average present value the utility contribution of this epoch is proportional to ΔF. After the short-run epoch ends there is a *middle-run epoch* of cycles that continue for roughly the calendar time between Poisson events which is proportional to τ. When τ is very large the future after τ matter very little. The order of limits allow us then to choose ΔF sufficiently small that the short-run epoch matter much less than that distant future.

For clarity I give a formal definition of short-run, middle-run and long-run

Definition 3.3. An *epoch* is continguous sequence of periods. The *short-run* is an epoch begining with one player waking up until the completion of the first cycle of committed machines or until some player wakes up. The *middle-run* is from the completion of the first cycle of committed machines until some player wakes up. The *long-run* is the infinite epoch following the medium run. According to this definition there may be several short-runs before a medium run, but every medium run is followed by a long-run.

To summarize: players wake up according to an exogenous Poisson process. This is infrequent in calendar time and this is crucial to the results because it enables "overlapping commitments." An alternative would be to have a cost of attention and endogenize the waking up but this is beyond the reach of this paper. Adjustment costs are assumed to be large enough to rule out cycles. There is evidence of cycling in some empirical work such as Brown and MacKay (2021) and Chen, Mislove and Wilson (2016) but this also is beyond the reach of this paper. I should also emphasize that while the assumption of observability makes sense as indicated in the introduction it will not generally hold and I provide two results concerning unobservability, one negative and one positive.

I will also use a generic assumption on payoffs. First I define several constants.

Definition 3.4. The *scale* of the stage game $\Gamma > 0$ is the largest utility difference between any two profiles.

The mixed action set $\mathcal{A}^i$ for player i are the mixed strategies that are divisible by an integer less than or equal to B^1, B^2. The *grain* of the stage game is

$$\rho \equiv \min_{i\in\{1,2\}} \min_{a^i\neq\tilde{a}^i\in A^i, \alpha^{-i}\neq\tilde{\alpha}^{-i}\in\mathcal{A}^{-i}} |u^i(a^i,\alpha^i) - u^i(\tilde{a}^i,\tilde{\alpha}^{-i}|$$

.

I can now state the generic assumption on payoffs. This is a strong no-ties condition.

Assumption 3.5. $\rho > 0$.

Note that this is sufficient for the results, but not necessary, as it does not hold in the example, although of course it does for arbitrarily small perturbations.

Finally, as the goal is to confront equilibrium with efficiency it is useful to say what sort of efficiency is under consideration.

Definition 3.6. A profile a is *constrained efficient* if it is Pareto efficient among all pure profiles that are individually rational for both players.

4. The Main Results

There are three main results concerning the observable and unobservable case respectively.

Observable Theorem. *If F, τ are large and ΔF is small then with observable commitments and any initial condition*

(i) every sequential equilibrium converges to some $\hat{a}$ in the sense that after at most two switches middle-run play on the equilibrium path is always $\hat{a}$.

(ii) the limit $\hat{a}$ is constrained efficient

This first result shows that observable commitments break the folk theorem a good way by leading to long run constrained efficiency. Notice that this does not break the usual version of the folk theorem which refers to average present value payoffs: about these we can say very little.[7] Rather, the Friedman (1971) folk

[7] If, as in Salcedo (2015) and Lamba and Zhuk (2023), we assume the initial condition is endogenous, and, following the model here, assume that it is chosen by the one player who is initially asleep, then the method of proof of the Observable Theorem implies that constrained efficiency begins immediately and hence average present value payoffs are efficient.

theorem, which is relevant for safety games that are repeated in the ordinary sense, has an obvious corollary: for any pure profile $\hat{a}$ that Pareto dominates a static Nash equilibrium for all sufficiently large discount factors (small Δs in this context) there is a subgame perfect equilibrium in which $\hat{a}$ is always played on the equilibrium path. That is: in the long-run anything can happen. That version breaks with observable commitments. By contrast it remains unchanged with unobservable commitments.

Unobservable Theorem. *For any pure profile $\hat{a}$ that is individually rational then with unobservable commitments there exists an initial condition such that if F, τ are large and ΔF is small then there is a sequential equilibrium in which $\hat{a}$ is always played along the equilibrium path.*

As this result is not central to the paper but included to show that the model without commitment is a folk theorem environment this is proven in Appendix II.

The reason the folk theorem holds in the unobservable case is that players have can have point beliefs and consequently no reason to try to learn their opponent's commitment. If we instead examine ϵ-belief-perturbed equilibrium they will not have point beliefs and will have an incentive to learn. In this case we can partially retrieve the result of the Observable Theorem.

Learning Theorem. *Suppose that all feasible machines are reactive. For any $1 > \epsilon > 0$ if F, τ are large and ΔF is small then for any initial condition and any ϵ belief-perturbed equilibrium*

(i) every sequential equilibrium converges to some $\hat{a}$ in the sense that after at most two switches middle-run play on the equilibrium path is always $\hat{a}$.

(ii) the limit $\hat{a}$ is constrained efficient.

This says that the result of the Observable Theorem goes through with one crucial additional assumption: all feasible machines must be reactive. The need for this assumption is discussed in Section 7 but the idea is that if unforgiving machines such as grim trigger are perceived as possible then there will be a reluctance to conduct the experimention needed to learn.

I turn now to the details of the proofs.

5. Preliminaries

Before proving the main theorems it is useful to develop some key results concerning optimal play. It is convenient first to define an additional constant

Definition 5.1. $\lambda_\tau = (1/h^1) + (1/h^2) - (1/(h^1 h^2 \tau))$

Recall that $\overline{N} \equiv \max_i N^i + B^1 B^2$ is the maximum length of the short-run epoch.

The idea is that average expected present value can be computed by computing it separately for each epoch and providing separate bounds for each epoch. Specifically the following bounds are proven in Appendix I

Lemma 5.2. *There exist constants* $\overline{\zeta}, \underline{\zeta} > 0$ *for all* $\tau \geq 1$, $F \geq \Gamma$ *(recall that* Γ *is the scale of payoffs) and* $\Delta \leq 1/(\lambda_1 2\overline{N})$ *such that*

(short-run) $\overline{\Gamma}_S$ *the* importance of the short-run *defined as the greatest difference in average expected present value over all short-run periods between any two different strategies satisfies* $\overline{\Gamma}_S \leq \overline{\zeta}\Delta F$.

(middle-run flow) $\gamma(\Delta, \tau)$ *the* value of a steady state flow *defined as the average expected present value during a middle run with a steady state yielding a single unit of utility each period satisfies* $\gamma(\Delta, \tau) \geq \underline{\zeta}$.

(middle-run cycle) $\overline{\xi}_M$ *the* value of a cycle *defined as the greatest averaged expected present value for player* i *during a middle run that has a non-trivial cycle for player* i *satisfies* $\overline{\xi}_M \leq \Gamma - \underline{\zeta} F$.

(long-run) $L(\Delta, \tau, \sigma)$ *the* long run value *defined as the average expected present value after the next wake-up of a single unit of utility each period satisfies* $1/(1+\underline{\zeta}\tau) \geq L(\Delta, \tau, \sigma) \geq \underline{\zeta}/\tau$.

(reversal) $\underline{\delta}^i_R$*the* importance of a reversal *defined as the expected discount factor from commitment until a reversal before player i wakes up again satisfies* $\underline{\delta}^i_R \geq \underline{\zeta}/(1+\overline{\zeta}\tau)$.

These bounds imply the following key result that holds regardless of assumptions about observability.

Theorem 5.3. *If* F, τ *are large and* ΔF *is small and a player* i *has beliefs that are a point mass on* (d^{-i}, b^{-i}) *that player must commit to a machine that plays a constant action* $\hat{a}^i$ *in the middle-run. If this machine yields a steady state* $\hat{a}$ *against* (d^{-i}, b^{-i}) *then* $\hat{a}$ *must yield the highest utility among all steady states that are feasible with respect to* (d^{-i}, b^{-i}).

Proof. The automaton that plays the safety action no matter what yields at least 0 utility each period. Suppose instead that a machine is chosen that does not result in a constant action in the middle-run. By Lemma 5.2 (middle-run cycle) during the middle-run this gives utility at most $\Gamma - \underline{\zeta}F$ per period, from (short-run) $\overline{\zeta}\Delta F$ during the short-run and from (long-run) $\Gamma/(1+\underline{\zeta}\tau)$ during the long-run, so for small ΔF and large τ is negative. This shows that a constant action must be chosen.

Next, observe that the middle-run gain of the best middle-run steady state $\hat{a}$ and any other steady state a by (middle-run flow) is at least $\rho\underline{\zeta}$. By contrast the short-run loss from $\hat{a}$ is at most $\overline{\zeta}\Delta F$ and long-run loss at most $\Gamma/(1+\underline{\zeta}\tau)$ so again for small ΔF and large τ the total loss is less than $\rho\underline{\zeta}$ so that a cannot be optimal. □

6. Observable Commitments

To help in following the proof, I reiterature the statement of the Observable Theorem.

Observable Theorem. *If F, τ are large and ΔF is small then with observable commitments and any initial condition*

(i) every sequential equilibrium converges to some $\hat{a}$ in the sense that after at most two switches middle-run play on the equilibrium path is always $\hat{a}$.

(ii) the limit $\hat{a}$ is constrained efficient

Proof. We know from Theorem 5.3 that when the commitment decision by i is made the choice is between different a^i that will be constant in the middle-run. The same will be true of the opposing player when a reversal occurs. Hence after the first reversal the middle-run must be a steady state where both players play a constant action.

Assume that the first reversal has occured. If i's commitment allows it is possible that the best choice for $-i$ causes a cycle for i: as shown in the proof of Theorem 5.3 this is no good, it would be better to offer $-i$ a steady state: this can be done, for example, by the strategy of just playing the chosen a^i no matter what.

What steady states a might be offered by i given a current steady state $\overline{a}$? Let $\hat{a}$ maximize i's stage game utility over pure profiles subject to $-i$ getting at least $u^{-i}(\overline{a})$ and zero. Consider the reactive machine for player i that responds to $\overline{a}^{-i}$ with $\overline{a}^i$, to $\hat{a}^{-i}$ with $\hat{a}^i$ and respond to anything else with the safety action. Hence for player $-i$ the constant action $\overline{a}^{-i}$ results in the steady state $\overline{a}$, the constant action $\hat{a}^{-i}$ results in the steady state $\hat{a}$ and any other constant action a^{-i} results in the steady state $(\mathbf{a}^i, a^{-i})$. Of these by construction $\hat{a}$ is best and so is chosen. Also $\hat{a}$ - as it maximizes

i's utility subject to $-i$ getting at least the utility from $\overline{a}$ and is individually rational - is constrained efficient.

In the middle-run clearly no result better than $\hat{a}$ is possible. If $u^i(a) = u^i(\hat{a})$ then the genericity condition says the two must be the same, so the only alternative is to choose a steady state a with $u^i(a) < u^i(\hat{a})$. This loses at least $\rho\underline{\zeta}$ per period in the middle-run by Lemma 5.2 (middle-run flow). As the short-run gain by (short-run) is at most $\overline{\zeta}\Delta F$ and the long-run gain by (long-run) at most $\Gamma/(1+\overline{\zeta}\tau)$ for large F, τ this is no good. Hence $\hat{a}$ will be played in the middle-run.

Finally, an offer $\tilde{a}$ by i might be made that would be accepted and give less utility than $\hat{a}$. By (middle-run flow) this would lose at least $\rho\underline{\zeta}$ per period in the middle-run following reversal, so including the short-run and long-run after reversal choosing large F and τ as in the previous paragraph we can assure that the loss would be at least $\rho\underline{\zeta}/2$ following reversal. However, potentially the alternative offer might incur less cost and provide more benefit in the short-run prior to reversal: by (short-run) this is at most $\overline{\zeta}\Delta F$. The loss that offsets this gain must be discounted by no more than $\underline{\zeta}/(1+\overline{\zeta}\tau)$ by Lemma 5.2 (reversal) as it occurs only following reversal. Hence the loss in average present value is at least $\rho\underline{\zeta}^2/(2(1+\overline{\zeta}\tau))$. Here is where the order of limits is crucial: recall that ΔF is chosen after τ. Hence it may be chosen so small that the loss after reversal is greater than the gain in the immediate short-run. Hence the offer should be $\hat{a}$.

Once the steady state on the equilibrium path is constrained efficient, there is nowhere to go: there is no “better offer” that can be made to the opposing player, so they keep that middle-run steady state. □

7. Learning and Forgiving

Recall the definition of an ϵ-belief-perturbed equilibrium. Each player i has a fixed distribution μ^i over feasible opponent commitments that puts weight at least $\underline{\mu}$ on each one. Let $\tilde{\mu}_t^i$ be a sequentially rational assessment for player i at wake-up time. Then each player must play a best response in each sub-form following wake-up to the perturbed assessment $(1-\epsilon)\tilde{\mu}_t^i + \epsilon\mu^i$. My goal is to show that the conclusion of the Observable Theorem holds for ϵ-belief-perturbed equilibria.

The problem is: belief perturbation is not good enough. Suppose that instead of the trigger automaton used in the proof of the Unobservable Theorem there are also automata that respond not only to opponent play last period, but also to own play last period. One such automaton is the grim-trigger machine. This looks first to see if the player themselves played the safety action last period and if so plays the safety action, otherwise it plays as the trigger machine. The point is that unless you conform against a grim-trigger machine you will be punished with the safety action forever. If we replace the trigger-machines in the proof of the Unobservable Theorem with grim-trigger machines it goes through unchanged, but is now robust to small ϵ belief perturbations: if ϵ is small the chance of triggering the safety action is so large that it is sub-optimal to experiment. Nobody tests the doomsday machine on purpose.

By contrast with grim-trigger machines, reactive machines are forgiving: after a fixed period of time they ignore what the opposing player did and the fact is that there no benefit from threatening to punish forever rather than to merely punish enough: there is a reason nobody has produced a doomsday machine. Given this, it makes sense assume that players are *forgiving* in the sense that are restricted to using reactive machines and that this is common knowledge (along with the length of history that these machines are limited to). The importance of being forgiving has been documented in other contexts: it is crucial to the results of Aumann and Sorin (1989) on common interest games, and the evolutionary advantage of forgiveness is indicated both in the simulations of Axelrod and Hamilton (1981) and in the theory of Fudenberg and Maskin (1990). It is also found in the laboratory work of Fudenberg, Dreber and Rand (2012).

For convenience I reiterature the statement of the Learning Theorem before proving it.

Learning Theorem. *Suppose that all feasible machines are reactive. For any* $1 > \epsilon > 0$ *if* F, τ *are large and* ΔF *is small then for any initial condition and any* ϵ *belief-perturbed equilibrium*

(i) every sequential equilibrium converges to some $\hat{a}$ *in the sense that after at most two switches middle-run play on the equilibrium path is always* $\hat{a}$.

(ii) the limit $\hat{a}$ *is constrained efficient.*

In short: with belief-perturbed equilibrium and reactive machines the folk theorem is again broken and there is again long-run efficiency.

Proof. First I examine what happens if a player chooses to learn. A key fact about reactive machines is that they are not only forgiving, but they are steady state machines in the sense that given a constant opponent action after a fixed period of time they respond with a steady state. As cycles are still a very bad idea a player is interested in which steady states are offered in the middle-run. This can easily be determined by running each constant sequence long enough and seeing what the opponent does.[8]

[8] In general, as shown by Levine and Szentes (2006), determining an opponent machine by testing it is problematic. This is not the case for reactive machines as testing it against every sufficiently long sequence shows exactly which machine it is. However, such extensive testing is not needed simply to determine the steady states.

Once this is done beliefs are point beliefs and the proof of the Observable theorem remain valid: the folk theorem is again broken and there is again long-run efficiency.

Should a player choose to learn? As ϵ is fixed choosing small enough Δ means the cost of testing is not prohibitive relative to the benefit. Specifically, since it is assumed that $F \geq \Gamma$ the cost of testing a constant action for player i is at most $(B^{-i}+1)F\Delta$. Moreover, if there is an action that has not been tested for which an opponent choice could yield a better payoff than the best known steady state there is at least an $\epsilon\underline{\mu}$ probability that this steady state is available according to the perturbed beliefs. As shown in the proof of Observable Commitments, the gain if this better steady state is available is at least $\rho\underline{\zeta}/2$. In other words, the expected gain from testing is at least $\epsilon\underline{\mu}\rho\underline{\zeta/2}$. Hence if ΔF is sufficiently small it is optimal to conduct the test and there is long-run efficiency. □

Appendix I

Recall that $\lambda_\tau = (1/h^1) + (1/h^2) - (1/(h^1 h^2 \tau))$ so that $\lambda_\tau \Delta/\tau$ is the probability that some player wakes up each period. Note that $\lambda_1 \leq 1$.

Lemma. *there exist constants $\overline{\zeta}, \underline{\zeta} > 0$ for all $\tau \geq 1$, $F \geq \Gamma$ and $\Delta \leq 1/(\lambda_1 2\overline{N})$ such that*

(short-run) $\Gamma_S \leq \overline{\zeta}\Delta F$

(middle-run flow) $\gamma(\Delta,\tau) \geq \underline{\zeta}$

(middle-run cycle) $\overline{\xi}_M \leq \Gamma - \underline{\zeta}F$

(long-run) $1/(1+\underline{\zeta}\tau) \geq L(\Delta,\tau,\sigma) \geq \underline{\zeta}/\tau$

Proof. **Short-run.** Recall that $\overline{\Gamma}_S$ is the greatest difference in average expected present value over all short-run periods between any two different strategies. The greatest difference between any two strategies in any individual period is $\Gamma + F$. If we compute the average expected present value assuming that each wakeup event triggers $\overline{N}$ periods of such a difference this overcounts the actual periods since after free play a new wakeup event could occur before the current short-run concludes. As we are interested in an upper bound on Γ_S we compute accordingly an upper bound on the loss

$$\begin{aligned}\overline{\Gamma}_S &\leq (1-e^{-\Delta})\left(1+\sum_{t=1}^{\infty}(\lambda_\tau\Delta/\tau)e^{-\Delta t}\right)\overline{N}(\Gamma+F)\\ &= \left(1-e^{-\Delta}+e^{-\Delta}\lambda_\tau\Delta/\tau\right)\overline{N}(\Gamma+F)\\ &\leq 2\left(1+\lambda_1\right)\overline{N}\left(\Delta F\right)\end{aligned}$$

giving the first result.

Middle-run flow

Recall that $\gamma(\Delta, \tau)$ is the average expected present value during a middle run with a steady state yielding a single unit of utility each period. This is

$$\gamma(\Delta, \tau) = (1 - e^{-\Delta}) \sum_{t=0}^{\infty} e^{-\Delta t}(1 - \lambda_\tau \Delta/\tau)^t$$

$$= \frac{\tau}{\frac{e^{-\Delta}\lambda_\tau\Delta}{1-e^{-\Delta}} + \tau}$$

$$\geq \frac{1}{\lambda_1 \frac{1/(\lambda_1 2\overline{N})}{1-e^{-1/(\lambda_1 2\overline{N})}} + 1}$$

where the final step uses the fact that $\Delta/(1 - e^{-\Delta})$ is increasing in Δ. This gives the middle-run flow result.

Middle-run cycle

Recall that $\overline{\xi}_M$ is the greatest average expected present value for player i during a middle run that has a non-trivial cycle for player i. We may take $\overline{N}$ as a bound on the length of the cycle and assume that the loss from at least one switch F occurs at the end of the cycle. Hence, the average expected present value is at most Γ minus a lower bound on the probability that the cycle is not interupted by a wake up event $(1 - \overline{N}\lambda_\tau\Delta/\tau)$. This gives

$$\overline{\xi}_M \leq \Gamma - (1 - \overline{N}\lambda_\tau\Delta/\tau)(1 - e^{-\Delta})e^{-\overline{N}\Delta}F/\overline{N} \leq \Gamma - (1 - \overline{N}\lambda_1\Delta)e^{-\overline{N}\Delta}F/\overline{N}$$

$$\leq \Gamma - \frac{e^{-(1/2)}}{2\overline{N}}F$$

giving the middle-run cycle result.

Long-run

Recall that $L(\Delta, \tau, \sigma)$ is the average expected present value of a unit utility flow after the next wake-up and we want both an upper and lower bound.

$$L(\Delta,\tau,\sigma) = e^{-N^i\Delta}\sum_{t=0}^{\infty} e^{-\Delta t}(\lambda_\tau\Delta/\tau)\left(1-\lambda_\tau\Delta/\tau\right)^t$$

$$= \frac{e^{-N^i\Delta}\lambda_\tau\Delta/\tau}{1-e^{-\Delta}+e^{-\Delta}\lambda_\tau\Delta/\tau}.$$

The lower bound is given by

$$L(\Delta,\tau,\sigma) = \frac{e^{-N^i\Delta}\lambda_\tau\Delta/\tau}{1-e^{-\Delta}+e^{-\Delta}\lambda_\tau\Delta/\tau} \geq \frac{e^{-N^i\Delta}\lambda_\tau\Delta/\tau}{\Delta+\lambda_\tau\Delta/\tau} = \frac{e^{-N^i\Delta}\lambda_\tau/\tau}{1+\lambda_\tau/\tau} \geq e^{-1/(2\lambda_1)}\lambda_1/\tau$$

and the upper bound

$$L(\Delta,\tau,\sigma) = \frac{e^{-N^i\Delta}\lambda_\tau\Delta/\tau}{1-e^{-\Delta}+e^{-\Delta}\lambda_\tau\Delta/\tau} \leq \frac{e^{-\Delta}\lambda_\tau\Delta/\tau}{1-e^{-\Delta}+e^{-\Delta}\lambda_\tau\Delta/\tau}$$

$$= \frac{1}{(\tau/\lambda_\tau\Delta)(e^{\Delta}-1)+1}$$

as $\Delta \leq 1$ we have $e^{\Delta}-1 \geq e\Delta$ so

$$\leq \frac{1}{(e/\lambda_1)\tau+1}.$$

Reversal

Recall that $\underline{\delta}_R^i$ is the expected discount factor from commitment until a reversal before player i wakes up again. In period t this is the probability $-i$ wakes up $\Delta/(h^{-i}\tau)$ times the probability that neither player had woken up before, and we add up over periods to get

$$\underline{\delta}_R^i = \sum_{t=0}^{\infty} e^{-\Delta t}(\Delta/(h^{-i}\tau))\left((1-\Delta/(h^1\tau))(1-\Delta/(h^2\tau))\right)^t$$

$$= \frac{\Delta/(h^i\tau)}{1 - e^{-\Delta} + \Delta e^{-\Delta}/(h^1\tau) + \Delta e^{-\Delta}/(h^2\tau) - (\Delta e^{-\Delta}/\tau)^2/(h^1h^2)}.$$

We are interested in a lower bound, and since $\Delta e^{-\Delta}/\tau \leq 1$

$$\underline{\delta}_R^i \geq \frac{\Delta/(\max\{h^1,h^2\}\tau)}{\Delta + 4\Delta e^{-1/(\lambda_1 2\overline{N})}/(\min\{h^1,h^2\}\tau)} = \frac{1}{\max\{h^1,h^2\}}\frac{1}{\tau + 4e^{-1/(\lambda_1 2\overline{N})}/\min\{h^1,h^2\}}$$

proving the reversal result. □

Appendix II: Unobervable Commitments

As a kind of sanity check I show that without observable commitments the model is indeed a folk-theorem model. To help in following the proof, I reiterature the statement of the Unobservable Theorem.

Unobservable Theorem. *For any pure profile $\hat{a}$ that is individually rational then with unobservable commitments there exists an initial condition such that if F, τ are large and ΔF is small then there is a sequential equilibrium in which $\hat{a}$ is always played along the equilibrium path.*

Proof. What is a sequential equilibrium in this context? In each subform after player i wakes that player has an assessment in the form of a probability distribution over the commitment pairs (d^{-i}, b^{-i}) of the opponent. The set of player strategies has not changed, and there are two requirements of sequentiality: first that in each subform the "subgame" induced by the assessment the strategies are a Nash equilibrium and second that the assessments satisfy a consistency requirement. In this setting the consistency requirement is rather simple: there are two kinds of commitment pairs by $-i$: those that are consistent with the history of play and past assessments those that are not. Those that are not must be assigned zero probability. The requirement for the remainder is that if a player wakes several times in a row and the history had positive probability in the previous assessment the relative probabilities within the currently consistent set must not be changed.

Define the *$\hat{a}$-trigger machine* to be the reactive machine that reacts to $\hat{a}^{-i}$ with $\hat{a}^i$ and plays $\mathbf{a}^i$ otherwise: by assumption there is a such a machine. The initial condition is that player 2 is committed to the $\hat{a}$-trigger plan and this is assessed to be the case by player 1 with probability 1.

Next define a subform to be *normal* if play is consistent with each player using a feasible commitment that results during the middle run with the steady state of $\hat{a}$ and prior to the next waking of either player the player who least recently committed has always played $\hat{a}^i$ or $\mathbf{a}^i$. A *normal history* is one in which every subform has been normal. In any subform following a normal history the assessment is a point mass on an opponent commitment that is consistent with the history and would respond to any other action other than $\hat{a}^i$ or $\mathbf{a}^i$ with $\mathbf{a}^{-i}$, and that plays a constant middle-run action of $\tilde{a}^{-i}$ with $u^i(\mathbf{a}^i, \tilde{a}^{-i}) \leq u^i(\hat{a})$. Since i only played $\hat{a}^i$ or $\mathbf{a}^i$ during a normal history because for any given state process any mapping α^{-i} to actions is feasible such a commitment exists. The assessment is by construction consistent. Define a strategy for normal histories to use $\hat{a}^i$ during free play, then commit to the $\hat{a}$-trigger machine with initial condition $\hat{a}^{-i}$. For all other histories pick some sequential equilibrium, which one does not matter. Notice that the proposed assessments are certainly consistent and the path of play for these strategies is always $\hat{a}$ as required by the theorem. The point is to prove that when F, τ are large and ΔF at standard histories no player wants to deviate.

For normal histories assessments are always point masses so Theorem 5.3 applies so that players must commit to a middle-run constant best response. According to their beliefs they can attain the middle-run steady state $\hat{a}$, steady states of the form $(a^i, \mathbf{a}^{-i})$ or steady states of the form $(\mathbf{a}^i, \tilde{a}^{-i})$. If $u^i(\mathbf{a}^i, \tilde{a}^{-i}) = u^i(\hat{a})$ then by the generic payoff assumption $(\mathbf{a}^i, \tilde{a}^{-i}) = \hat{a}$ so the final case is redudant and in all cases $\hat{a}$ is best. The $\hat{a}$-trigger machine gives this steady state, and given the proposed equilibrium strategies the opponent $-i$ is actually using the $\hat{a}$-trigger machine. Under these circumstances can i do better than the $\hat{a}$-trigger machine?

One possibility is to choose an alternative strategy that plays the same way, for example, always play $\hat{a}$: this results in the same future, so no gain. Another is to induce a middle-run cycle when $\mathbf{a}^{-i}$ is played. By the usual argument this is a bad idea.

The final possibility is to incur a short-run loss in hopes of a better future: by playing differently now, perhaps the opponent can be convinced that there is a better deal when there is another reversal. The short run lasts at most $\overline{N}$ periods and gains at most Γ per period, since it must play differently, must lose at least F in one of those periods, so for $\Delta \leq 1$ and $F \geq \Gamma$ at least $e^{-\overline{N}}\Delta\Gamma$ in average present value. Hence if an alternative strategy is to be profitable would have to garner that profit by convincing the opponent after the next switch not to continue playing the steady state $\hat{a}$ but instead switch to some more desirable steady state (from i's point of view). The problem is that whatever i does $-i$ (who is in fact using the $\hat{a}$-trigger machine) is only going to play $\hat{a}^{-i}$ or $\mathbf{a}^{-i}$ so the subform will continue to be normal and $-i$ is going to assess that any action other than $\hat{a}^{-i}$ or $\mathbf{a}^{-i}$ is going to be responded to with $\mathbf{a}^i$. Hence the only possible middle run steady states will be $\hat{a}$, $(a^i, \mathbf{a}^{-i})$ and $\hat{a}$ is by assumption strictly better for $-i$ so they will choose that.

This leaves the issue of whether there might be some future short-run gain that pays for the current short-run loss. Recall that $\overline{N} \equiv \max_i N^i + B^1 B^2$ is the greatest number of periods in a short-run epoch so that the length in calendar time is at most $\overline{N}\Delta$. The payoff gain in each period is at most the scale of the game Γ. Hence, in average present value this gain is at most $\Gamma\overline{N}\Delta$ at the beginning of that future short-run. However, by Theorem 5.2 the long-run bound implies that this is discounted by at most $1/(1+\underline{\zeta}\tau)$, so in first period average present value no greater than $\Gamma\overline{N}\Delta/(1+\underline{\zeta}\tau)$. For fixed $\Gamma, \overline{N}$ and sufficiently large τ this is smaller than the short-run average present value initial loss of at least $e^{-\overline{N}}\Delta\Gamma$. Hence this is not optimal either. □

References

Abreu, Dilip and David Pearce (2007): "Bargaining, Reputation, and Equilibrium Selection in Repeated Games with Contracts," *Econometrica* 75: 653-710.

Abreu, D., and A. Rubinstein (1988): "The structure of Nash equilibrium in repeated games with finite automata," *Econometrica*: 1259-1281.

Asker, J., C. Fershtman and A. Ariel Pakes (2022): "The impact of AI design on pricing," *Journal of Economics and Management Strategies*, forthcoming.

Aumann, R.J. and S. Sorin (1989): "Cooperation and bounded recall," *Games and Economic Behavior* 1: pp.5-39.

Axelrod, R. and W. D. Hamilton (1981): "The evolution of cooperation," *Science.*

Binmore, Kenneth G., and Larry Samuelson (1992): "Evolutionary stability in repeated games played by finite automata," *Journal of Economic Theory* 57: 278-305.

Bowles, Samuel and Jung-Kyoo Choi (2013): "Coevolution of farming and private property during the early Holocene" *Proceedings of the National Academy of Science*, doi:10.1073/pnas.1212149110.

Brown, Z.Y. and A. MacKay (2021): "Competition in pricing algorithms," National Bureau of Economic Research.

Calvano, E., G. Calzolari, V. Denicolo, V. and S. Pastorello (2020): "Artificial intelligence, algorithmic pricing, and collusion," *American Economic Review*, 110: 3267-3297.

Cartea, A., P. Chang, J. Penalva and H. Waldon, H. (2022): “The Algorithmic Learning Equations: Evolving Strategies in Dynamic Games,” SSRN.

Chen, L., A. Mislove and C. Wilson (2016): “An empirical analysis of algorithmic pricing on amazon marketplace,” *Proceedings of the 25th international conference on World Wide Web*: 1339-1349.

Dal Bó and Pujals (2015): “The Evolutionary Robustness of Forgiveness and Cooperation,” mimeo.

Foster, D. P. and H. P. Young, H. P. (2003): “Learning, hypothesis testing, and Nash equilibrium,” *Games and Economic Behavior* 45: 73-96.

Friedman, James W. (1971): “A non-cooperative equilibrium for supergames,” *Review of Economic Studies* 38: 1-12.

Fudenberg, Drew, Anna Dreber and David G. Rand (2012): “Slow to anger and fast to forgive: Cooperation in an uncertain world,” *American Economic Review* 102: 720-749.

Fudenberg, D. and D. K. Levine (1983): “Subgame-Perfect Equilibria of Finite- and Infinite-Horizon Games," *Journal of Economic Theory* 31: 251-258.

Fudenberg, Drew, and Eric Maskin (1986): “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information,” *Econometrica* 54: 533-554.

Fudenberg, Drew, and Eric Maskin (1990): “Evolution and Cooperation in Noisy Repeated Games,” *American Economic Review*, 80 : 274-279.

Juang, W-T. and Sabourian, H. (2021): "Rules and Mutation - A Theory of How Efficiency and Rawlsian Egalitarianism/Symmetry May Emerge," mimeo Cambridge.

Jindani, S. (2022): "Learning efficient equilibria in repeated games," *Journal of Economic Theory* 205.

Johnson, P., D. K. Levine and W. Pesendorfer (2001): "Evolution and Information in a Gift Giving Game," *Journal of Economic Theory* 100: 1-22.

Klein, Timo (2021): "Autonomous algorithmic collusion: Q-learning under sequential pricing," *RAND Journal of Economics* 52: 538–558.

Kreps, D. and R. Wilson (1982): "Sequential Equilibria," *Econometrica* 50: 863-94.

Lamba, Rohit and Sergey Zhuk (2023): "Pricing with Algorithms," Penn State University.

Leisten, M., 2021. Algorithmic competition, with humans. working paper.

Levine, David K. (1981): "Long Run Collusion in a Partially Myopic Industry," MIT PhD Dissertation.

Levine, D. K. and B. Szentes (2006): "Can A Turing Player Identify Itself?" *Economics Bulletin* 1: 1-6

Musolff, L., 2022. Algorithmic pricing facilitates tacit collusion: Evidence from e- commerce. In Proceedings of the 23rd ACM Conference on Economics and Computation (pp. 32-33).

Rubinstein, A. (1986): “Finite automata play the repeated prisoner’s dilemma,” *Journal of Economic Theory* 39: 83-96.

Salcedo, Bruno (2015): “Pricing Algorithms and Tacit Collusion,” Pennsylvania State University.