Bounded Memory Folk Theorem*

Mehmet Barlo
Sabancı University

Guilherme Carmona
University of Surrey

Hamid Sabourian
University of Cambridge

May 2, 2019

Abstract

We show that the Folk Theorem holds for n-player discounted repeated games with bounded memory (recall) strategies. Our main result demonstrates that any payoff profile that exceeds the pure minmax payoff profile can be approximately sustained by a pure strategy finite memory subgame perfect equilibrium of the repeated game if the players are sufficiently patient. We also show that the result can be extended to any payoff profile that exceeds the *mixed* minmax payoff profile if players can randomize at each stage of the repeated game. Our results requires neither time-dependent strategies, nor public randomization, nor any communication. The type of strategies we employ to establish our result turn out to have new features that may be important in understanding repeated interactions.

Journal of Economic Literature Classification Numbers: C72; C73; C79

Keywords: Repeated Games; Memory; Bounded Rationality; Folk Theorem.

*We wish to thank Nathanaël Berestycki, George Mailath and Wojciech Olszewski for very helpful suggestions. Any remaining errors are, of course, ours.

1 Introduction

The extensive multiplicity of subgame perfect equilibrium (SPE) payoffs in repeated games, exemplified by the Folk Theorem, is due to players' ability to condition their behavior arbitrarily on the past (see **?**). Therefore, it is reasonable to expect, as suggested by **?**, that this multiplicity may be reduced if players have limited memory in the sense that they can condition their strategies only on the outcome of a limited number of past periods.

In **?**, we show that this intuition, however, does not hold when the set of actions in the stage game of the repeated game is sufficiently "large" so that each payoff profile is not isolated. In such games we prove that the Folk Theorem with SPE as the solution concept (henceforth, we shall refer to such Folk Theorems by FT) continues to hold with one period memory strategies where at each date players' behavior depends only on the outcome of the game in the previous period. The large action space assumption is critical in establishing this result because it allows players to encode the entire history of the past into the previous period's actions.

In the same study, we show that when the action spaces are not large, it is possible that no efficient payoff vector can be supported by a one period memory SPE strategy profile even if the discount factor is near one, validating the argument of **?** with one period memory strategies and finite actions. Hence, the question is whether or not the multiplicity of equilibrium payoffs prevails with finite actions and limited memory (not necessarily restricted to be one period). More specifically, does the FT depend critically on being able to recall the history of play all the way back to the beginning?

The current paper establishes that the FT for discounted repeated games continues to hold with time-independent bounded memory strategies even when the action sets are finite. Our main result displays that, when players are sufficiently patient, any feasible payoff vector that guarantees each player at least his pure strategy minmax payoff (individually rational payoffs) can be approximately sustained by a pure SPE strategy profile of the repeated game that at each stage recalls the outcomes of finite number of previous periods.[1] Furthermore,

[1]As it is the case for **?** FT results, our FT result with more than 2 players is established for generic games.

we show that the bound on the number of periods that the players need to recall to establish this result is uniform in the level of discounting, and depends only on the desired degree of payoff approximation.

With no memory restriction, the FT result with mixed (behavioral) strategies is stronger than that with pure strategies. This is because for any player the mixed strategy minmax payoff may be lower than the pure strategy minmax payoff. To complete the analysis of repeated games with finite memory, we extend our result for pure strategies to show that with three or more players, if players are sufficiently patient and are allowed to use behavioral strategies, then any payoff vector that guarantees each player at least his mixed minmax payoff profile can be approximately sustained by a behavioral SPE strategy profile that at each stage recalls the outcomes of finite number of previous periods. The following points need to be emphasized regarding our finite memory mixed FT result: First, it assumes that the players observe only the outcome of past randomizations (and not the randomization devices used in the past).[2] Second, it is obtained without introducing any public randomization or any external communication devices. Third, in contrast to our pure strategy result, the bound on the number of periods that the players need to recall to establish the mixed FT is not uniform in the level of discounting.

In this paper, memory refers to the number of past periods the players can recall.[3] An alternative way of imposing bounds on the memory is to limit the strategies to those that can be implemented by finite automata or more general Turing machines (see, for example, **?**, **?**, **?**).[4] With finite action spaces, assuming bounded memory as in our paper, is a stronger restriction than assuming that strategies can be represented by finite automata.[5] Effectively, while in both bounded memory and finite automata approaches, at each stage the players can have access to a finite amount of information about the past, the latter approach gives

[2]If randomization devices employed in the past are observable then the repeated game with mixed strategies is equivalent to one with a continuum of action space at each stage. Hence, it follows from Theorem 9 of **?** that the mixed FT holds with 1-period memory.

[3]Our definition of memory follows those by **?**, **?** and **?**. In the literature, strategies with such bounds on the memory are also referred to as bounded recall strategies.

[4]Another approach to memory limitation is one of modeling strategies as a neural network; see **?**.

[5]Every bounded memory strategy can be implemented by a finite automaton, whereas this is not the case the other way round.

the players the flexibility of choosing which information about the past can be retained (for example who has been the last deviator), whereas the former approach has no such flexibility as the information from distant past is permanently erased and players can only access recent information. This lack of flexibility makes it significantly more difficult to establish a FT type result with bounded memory than in the case with finite automata.

While one issue in repeated game literature concerns the multiplicity of equilibrium payoffs, another is about understanding the precise behavior that satisfies intertemporal incentives in repeated contexts. Our result is important not only because it shows that the FT does not depend on being able to recall the history of play all the way back to the beginning, but also because the kind of strategies/behavior needed to ensure intertemporal incentives with limited memory turn out to have new features that may be significant in understanding repeated interactions.

There are many reasons why one might be interested in results with limited memory. First, there is the bounded rationality aspect in which players can only recall a finite amount of public information concerning the past. For example, having access to past information can be costly for the individual and in equilibrium players may choose to recall a finite past. The results from psychological literature also indicate that people do not act on the entire history they observe and pay special attention to recent history. Second, in many institutional set-ups it is the convention to remove all the records after a certain number of years (possibly because of cost of storage of past records). Third, even if players use memory devices, such as pieces of paper or money, to keep track of the past,[6] these devices often become unusable after a certain number of years (e.g., the messages written become unreadable or coins fully depreciate). Finally, memory size may have implications for robustness of equilibria. For example, **?** and **?** show that private monitoring perturbations of public monitoring equilibria are robust if the equilibria have bounded recall.

To appreciate the difficulties and the novel behavioral features needed in establishing a FT with bounded memory, consider a typical pure "simple" strategy SPE profile used in proving a pure FT in n-player repeated games. Such a strategy profile is described by $n+1$ infinite paths $\pi^{(0)}, \pi^{(1)}, \ldots, \pi^{(n)}$ consisting of the equilibrium path of play $\pi^{(0)}$ and a

[6]In the (finite) automata representation of behavior, the states of the machine correspond to such devices.

punishment path $\pi^{(i)}$ for each player i (see **?**). The strategies are such that game begins with $\pi^{(0)}$ until some player deviates singly from $\pi^{(0)}$. At any stage, a single deviation by a player from any ongoing path triggers the punishment path for that player; otherwise, the game continues with the ongoing path.

In the first instance, it may seem that the problem of implementing such a simple profile with bounded memory is trivial if the memory size M is sufficiently large. In particular, if each of the $n+1$ paths has a finite cycle, then each can be distinguished and implemented as long as M is sufficiently large. Even when the paths are not finite, one can approximate the payoff corresponding to each path by a cyclical path. Therefore, finite memory should be sufficient to implement the paths approximately. But this is not enough. Strategies must also be such that after observing the outcomes of the previous M periods the following two critical properties hold: First, single player deviations can be detected. Second, the identity of the deviator is revealed. If either of the above two properties were not to hold, there might be incentives for some player to deviate and manipulate the path of future play.[7]

With 1-period memory it is easy to see how such simple strategies may violate the above properties. For example, consider any two action profiles a and b respectively belonging to two paths $\pi^{(i)}$ and $\pi^{(j)}$, for some i and j. Then the first property is violated if, for some player k, $a_k \neq b_k$ and $a_{-k} = b_{-k}$. This is because when $(b_k, a_{-k}) = b$ is observed, it is not clear if k has just deviated from $\pi^{(i)}$ and the punishment for k needs to be triggered or if the path $\pi^{(j)}$ is being followed and no deviation has occurred. Similarly, the second property is violated if, for a pair of players k and l, $a_l \neq b_l$, $a_k \neq b_k$ and $a_{-l,k} = b_{-l,k}$. This is because in this case when $(b_k, a_l, a_{-k,l}) = (b_k, a_l, b_{-k,l})$ is observed, it is not clear which of the two players k or l has deviated.

Does increasing the memory size help with ensuring that the above two properties hold? The next two examples show that these difficulties cannot be solved so easily even with large, but finite, memory.

Example 1: Consider a Prisoners' Dilemma in which at every date each player can either cooperate C or defect D. Suppose that the players are sufficiently patient and we want to implement a cycle path $\pi^{(0)} = \{\pi^t\}_{t=1}^{\infty}$ consisting of playing $((C,D),(D,C))$ repeatedly.

[7] In **?** we refer to simple strategies that satisfy the above two properties as "confusion-proof".

Assume that such a path yields for each player an average payoff strictly higher than the minmax payoff generated from playing (D, D).[8] The simple strategy that plays $\pi^{(0)}$ on the equilibrium path and plays (D, D) forever for any history inconsistent with the equilibrium path, is subgame perfect with unbounded memory. However, this strategy is not subgame perfect if players can remember at most an arbitrary but finite number M of past periods. To see this, consider any history with its last M entries (henceforth called the M-tail) equal to $(a^1, \pi^2, \ldots, \pi^M)$, for any $a^1 \neq \pi^1$. Then the simple strategy prescribes playing D for both players forever in the continuation game. But if $\pi^M = (D, C)$, then player 1 has the incentive to deviate. This is because if player 1 plays C instead of D at this history, the play returns to the equilibrium path in the next period, as $(\pi^2, \ldots, \pi^M, \pi^{M+1})$ would be recalled. In the case when $\pi^M = (C, D)$, by an analogous reasoning, player 2 has an incentive to deviate.

One way to overcome this difficulty may be to allow the play to continue along the equilibrium path even at some histories that are inconsistent with the equilibrium path. However, this alone is not sufficient. For example, consider a strategy profile that is otherwise identical to the above simple strategy profile except that it plays π^{M+1} at any history whose M-tail equals $(a^1, \pi^2, \ldots, \pi^M)$ for any a^1. In this case, if $\pi^M = (D, C)$, then player 2 will find it profitable to deviate from D to C at any history with its M-tail equal to $(a^2, a^1, \pi^2, \ldots, \pi^{M-1})$, for any $a^1 \neq \pi^1$ and any a_2. By doing so, he produces a history with its M-tail equal to $(a^1, \pi^2, \ldots, \pi^M)$ and brings the play back to the equilibrium path.[9] Thus, if we continue to change the strategy by allowing the play to return to the equilibrium path at these problematic histories, an inductive argument would imply that the play must be the equilibrium path after any possible history, a requirement clearly incompatible with subgame perfection.[10]

[8] Note that (C, D) and (D, C) may be the only efficient action profile in a Prisoner's Dilemma. One such example is:

	C	D
C	$2, 2$	$0, 5$
D	$5, 0$	$1, 1$

Here, playing $((C, D), (D, C))$ repeatedly induces the symmetric efficient payoff as the discount factor goes to 1.

[9] If $\pi^M = (C, D)$, player 1 has an incentive to deviate when $(a^2, a^1, \pi^2, \ldots, \pi^{M-1})$ is recalled.

[10] **?**, a predecessor to the current paper, considers the repeated Prisoners' Dilemma with bounded memory.

The above example shows that increasing the memory size by itself does not guarantee that the players can identify whether or not there has been a deviation. The next example shows that the problem of detecting the identity of the deviator can also not be easily resolved by having a large but finite memory.

Example 2: In this example there are three players, each player $i = 1, 2, 3$ has three (pure) actions α_i, β_i and γ_i in the stage game and the players discount the future by an arbitrarily small amount. Let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and suppose that the stage payoff u_i for each i is such that the (pure) action profile that minmaxes i is $m^i = (\beta_i, \alpha_{-i})$. Also, suppose that, for each $i = 1, 2, 3$, m^i is a Nash equilibrium of the stage game and $u_i(m^i) < u_i(m^j)$ for all $j = 0, \ldots, 3$, $j \neq i$, where $m^0 = (\gamma_1, \gamma_2, \gamma_3)$.[11] Then with no memory restriction the simple strategy profile defined by an equilibrium path $\pi^{(0)} = \{m^0, m^0, \ldots\}$ and a punishment path $\pi^{(i)} = \{m^i, m^i, \ldots\}$ for each $i = 1, 2, 3$, implements m^0 as a SPE.

Such a simple strategy profile has the two features that, when a deviator is identified, the punishment path for that player is implemented and that, after any history, the continuation path corresponds to one of the four paths $\pi^{(0)}, \ldots, \pi^{(3)}$. With finite memory, irrespective of how large the memory is, implementing m^0 as a SPE with strategies that have these two features is no longer feasible. To see this, fix the memory to be M and any strategy profile f with these features. By the second feature, at any history with its M-tail equal to $(\alpha, \alpha, \ldots \alpha)$ the continuation strategy prescribes playing a path $\pi^{(j)}$, for some $j = 0, \ldots, 3$. Consider any player $i \neq j$. Since f must play m^0 initially, by the first feature, if i deviates at date 1 by playing $a_i \neq m_i^0$, then f induces m^i at date 2. Also, if player i deviates again from m^i at date 2 by playing α_i instead of $m_i^i = \beta_i$, α will be observed and f would prescribe playing m^i again. Further, such deviations by i induce α again and thus, by induction, f also specifies playing m^i after a history consisting of (a_i, m_{-i}^0) followed by α played $(M-1)$ times. But then at such a history, player i can profitably deviate by playing α_i and inducing a history consisting of M consecutive α's. This is because his average continuation payoff from the deviation would be almost $u_i(m^j)$, whereas by not deviating he obtains $u_i(m^i)$.

The problem in the above example is that α could be the result of single deviation by 1

Example 1 is from this paper, which in turn attributes it to an anonymous referee.

[11] It is easy to construct an example with payoffs satisfying these properties.

from m^1, 2 from m^2 or 3 from m^3. Therefore, the history consisting of α played M times can be induced by any player through a sequence of deviations and cannot be attributed to deviations by any particular player. Hence, given that $u_i(m^i) < u_i(m^j)$ for all $j = 0, \ldots, 3$, $j \neq i$, there must be some profitable opportunities for some player to deviate.

The problems of detecting the latest deviation and the identity of the deviator clearly do not arise with unbounded memory because, for any history, one can use induction starting from the first period of the history to find the latest deviation. With bounded memory, such inductive reasoning, by definition, is not feasible. Therefore, to deal with these problems with limited memory one needs to ensure that all of the paths that the candidate strategy profile prescribes at each history are sufficiently distinct. This can be done if each action profile in each path is distinct from those in other paths by at least three components (e.g. **?**).[12] In fact, the richness assumption in **?** allows one to prove a Folk Theorem with bounded memory precisely because with rich action spaces, one can construct such paths at the cost of perturbing all the payoffs by a small amount. With finite action spaces, such an approach to making each path sufficiently distinct is clearly not possible.

Nevertheless, in this paper we show that the objective of making each path sufficiently distinct, so that deviations and the identity of deviators can be detected, can be achieved by ensuring that each path contains specific finite sequences of actions, henceforth referred to as signalling sequences. Each of these signalling sequences is carefully designed so that, once any of them is observed, the paths or deviations are identified and the players know how to play the continuation game without the need to know the entire past history. Furthermore, our construction has the feature that some of these signalling sequences (those that do not trigger player specific punishment) appear infinitely often along their respective path. Effectively, such signal sequences can be thought of as a set of rituals that have to be played every so often so that the players can coordinate their future play in an appropriate way to preserve the intertemporal incentives.

Introduction of the signalling sequences generates additional sets of issues. First, we also need to ensure that it is in the interest of the players to play these sequences. This

[12] Assuming such distinctness, **?** provides a characterization for the set of SPE outcomes of repeated games for the case of no discounting and finite number of pure actions.

makes the construction of signalling sequences and punishment paths needed to induce them rather intricate and complicated. At the same time, the signalling sequences must be almost costless. We achieve this by ensuring that, for each path, the proportion of times its signalling sequence occurs on the path is arbitrarily small. Since the lengths of these sequences are bounded, this is feasible by making any cycle path, and hence the memory, sufficiently long. Second, if the number of players n exceeds two, then any single player deviation from any signalling sequence can be detected by considering what the others are doing, whereas this is not feasible when $n = 2$. As a result, the proof of our FT for the case of pure strategies is somewhat different when $n = 2$ from the case when $n > 2$.

There are also additional difficulties specific to proving the mixed FT. First, when a player is being mixed minmaxed, other players may have to play a random strategy; but this is difficult to enforce because mixed strategies are not observable. The standard construction for dealing with this problem with unbounded memory is that of **?**. However, this method does not extend to the finite memory case (see section 5 for the intuition). Second, with mixed strategies the issue of distinguishing the signalling sequences from the rest of the paths (and from single player deviations from them) is more difficult because the signalling sequences need to be sufficiently distinct from any paths that happens with positive probability after any history.[13]

When the number of players exceeds 2, we deal with these two difficulties, and hence extend our pure FT result to the mixed case, by using an alternative approach based on **?** and by appealing to some concentration result from probability theory and by a careful design of the signalling sequences. Our pure FT with $n = 2$, however, does not extend in a similar way to a mixed FT. Section 5.1 discusses the issues that may arise in establishing a mixed FT when $n = 2$.

Our FT result is, however, an approximate one. With pure strategies, this is because with finite action spaces the path induced by any bounded memory strategy, after every history, must eventually enter a finite cycle. Since not all individually rational payoffs

[13]In our mixed FT randomization happens only when the players are being (approximately) minmaxed and hence this issue relates only to distinguishing the signalling sequences from the random outcomes that happen during the minmax phase.

can be implemented by finite cycles, it follows that the set of individually rational payoffs can at best be implemented approximately, with larger memory being needed to improve the approximation. In our mixed FT, the equilibrium paths we construct are still pure (randomization happens only off the equilibrium when players are being minmaxed), hence not all individually rational payoffs can be implemented.[14]

While our FT construction may require a large, but finite, memory (e.g. because the length of the cycles needed to implement individual rational payoffs is long or because of the need to make the signalling sequences almost costless), it is important to note that the aim of our construction is to demonstrate what can possibly be implemented as an equilibrium in most general settings; in specific cases, the construction can be made simpler and the memory needed can be quite small. For example, the grim-trigger strategy in the Prisoners' Dilemma that implements cooperation trivially requires 1-period memory. Also, as we mentioned before, Abreu-type simple strategy construction in which each action profile in each path is distinct from those in other paths by at least three components requires 1-period memory.

2 Related Literature

In contrast to our results, in some related literature, bounds on the memory result in significant reduction in the set of equilibria in repeated set-ups. However, these results require additional assumption(s) beyond bounded memory. For example, **?** show that in a dynamic model with one long-lived player facing a sequence of short-lived players and complete information, bounds on the memory can have a dramatic impact on the equilibrium set (only Nash equilibria of the stage game are consistent with limited memory). Their results, however, are critically dependent on players' ability to condition their behavior only on past

[14]Appealing to public randomization on the equilibrium paths would clearly implement any individually rational payoff exactly. But this is not sufficient to obtain an exact FT result. This is because, as we mentioned before, any non-signalling part of any path that happens with a positive probability must be sufficiently distinct from all the signalling sequences. Hence, any randomization on the equilibrium path would still be restricted by the need to differentiate itself from the signalling sequences.

actions of the other players (i.e., strategies are reactive).[15]

? consider the repeated Prisoners' Dilemma with imperfect public monitoring and finite memory. They show that, for some set of parameters, defection every period is the only strongly symmetric public perfect equilibrium with bounded memory (regardless of the discount factor), whereas the set strongly symmetric public perfect strategies with unbounded recall is strictly larger. The example considered by **?** does not satisfy the identifiability condition used in **?** to establish their FT results for repeated games with imperfect monitoring. By strengthening those identifiability conditions and by allowing asymmetric strategies, **?** obtain a FT result with bounded memory strategies for games with imperfect monitoring and finite action and outcomes spaces.

The construction in **?** faces similar difficulties as ours (and additional ones since they allow for imperfect public monitoring) but is simplified by the assumption that players can condition their play on calendar time, i.e. they use time-dependent bounded memory strategies. This feature allows them to divide play into a regular phase and a communication (signalling) phase that occur in a pre-specified set of dates and to use the outcomes in the latter phase to coordinate future plays. Since calendar time is unbounded and one of the reasons for limiting the analysis to bounded memory is to bound the set of objects on which the players can condition their behavior, in contrast to their work, in this paper we do not allow players to use time-dependent strategies. This means that we cannot have a prespecified set of dates for communication, as a result, we must ensure that players can understand from the play of the game when they are in the communication/signalling phase and when there are in the regular phase. This integration of the communication with the regular behavior in our setup makes the analysis quite intricate.

In addition to time-dependence, the results in **?** require the existence of a public correlating device (a continuum of public signals) and is somewhat weaker than the standard FT

[15]Anti FT type results have also been obtained with asynchronous choice in some cases. While some of these results do not involve memory restrictions in the context of some games (**?**), employing memory costs **?** establish that every finite memory Nash equilibrium must be a Markovian strategy profile. It would be interesting to see how in our setup our FT result with bounded memory would be affected if we were to introduce some degree of asynchronicity.

result for two-player games.[16] Furthermore, the equilibria that **?** construct also have the feature that the memory size becomes arbitrarily large as the discount factor goes to one, and that players are indifferent between several actions. In contrast, in our perfect monitoring set-up, we do not appeal to any public randomizing device and the standard FT conclusions holds irrespective of the number of players. Also, for our FT with pure strategies, the size of the memory needed to establish the FT result is independent of the discount factor and the players do not need to be indifferent between different actions in the equilibria that we construct – the equilibria can be made essentially strict (see also footnote 21).[17]

Independently and at the same time as us, **?** (henceforth, MO) have also considered the problem of establishing a FT without the use of a randomizing device for perfect monitoring set-ups with finite action spaces, with bounded memory and with equilibria that are essentially strict.[18] Their result is, however, a special case of our pure minmax FT. Specifically, they show that the FT holds with time-dependent bounded memory in games with more than two players.[19] Our pure minmax result is more general than theirs because we do not require players to condition their strategies on calendar time and because our FT also holds for two players.[20]

The aim of a great deal of the recent literature on FT is to provide a robustness check for the validity of equilibrium payoffs and/or behaviour by making it harder for players to coordinate on future play through the weakening of the perfect monitoring requirement. The main

[16]It shows that any payoff vector that Pareto dominate some static Nash equilibrium, rather than the usual minmax payoff vector, can be achieved as an equilibrium.

[17]The result of **?** for the case of perfect monitoring, however, is not a special case of our result because the existence of a rich set of public signals in their set-up allows them to establish an exact FT in which every feasible payoff profile exceeding the mixed minmax payoff can be sustained as an equilibrium payoff. Since we do not assume any public randomization, our FT results (both with pure or mixed strategies) establish that every individually rational payoff vector can be approximately sustained as an equilibrium payoff.

[18]Other works on repeated games with bounded (recall) memory include **?**, **?**, **?**, **?**, **?**, **?**, and **?**.

[19]While the details are different, their construction has announcement phases at prespecified set of dates that are inspired by the communication phases used in **?**.

[20]The proof of the FT with two players in our first version of the paper was rather cumbersome. We have simplified the proof as a result of conversations with George Mailath and Wojciech Olszewski. We would like to thank them for these very useful conversations and wish to mention that MO provide an independent proof of our 2 player result in their online addendum.

motivation of MO is within this tradition, as they are primarily interested in demonstrating that the perfect monitoring FT is behaviorally robust to almost-perfect almost-public private monitoring. As shown by MO, time-dependent bounded memory equilibria that are essentially strict is all that is required for this (see footnote 21 for more details). Therefore, their result is sufficient to establish that the above robustness exercise is valid for games with more than two players.

This paper is part another kind of robustness check in which the coordination on future play is made harder by restricting the set of strategies for reasons of memory, computation or other kind of limitations. Specifically, and in contrast to both **?** and **?**, we are interested in the robustness of the FT to finite bounds on the set of objects on which the players can condition their behavior (i.e. on the domain of the set of strategies) and wanted to treat time and history of past plays symmetrically. With this in mind, we did not want to take the time-dependence route as it allows for conditioning on an object that is unbounded.[21] Furthermore, as mentioned before, an important feature of bounded memory/recall formulation, in contrast to the finite automaton approach, is that the players do not have any flexibility as to which part of the past information they can retain. In particular, in this formulation at any stage, information from the distant past is permanently erased or lost. Time-dependence, on the other hand, always requires keeping track of the beginning of time. Thus, if the time index is a part of past information, time-dependence seems inconsistent with the bounded memory/recall formulation.[22] Hence, within the bounded memory/recall approach of this paper, the dispensability of using time index as a coordination device provides an important

[21] Our findings can, nevertheless, be used to demonstrate that the perfect monitoring FT with pure strategies is behaviorally robust to almost-perfect almost-public private monitoring (using similar arguments associated with Theorem 6 of MO). This requires (i) verifying that the strategy we use in the proof of our pure FT is patiently pseudo-strict, which follows easily from our analysis in Sections A.1.3 (two players) and A.2.2 (more than two players); and (ii) observing that the equilibrium we construct satisfies the property that the fraction of time the profile spends in a state at which some player has multiple myopic best responses arise is arbitrarily small.

[22] A time-dependent strategy may of course be implementable as finite automata (for example the equilibrium strategies in MO are indeed described by finite automata). Furthermore, the inconsistency described above does not arise within the finite automaton approach, because here the players have the flexibility of retaining any finite piece of information at any time.

additional robustness check for the FT.

3 Notation and Definitions

The stage game: A *normal form game* G is defined by $G = \left(N, (A_i)_{i\in N}, (u_i)_{i\in N}\right)$, where $N = \{1, \ldots, n\}$ is a finite set of *players*, A_i is the set of player i's *actions* and $u_i : \prod_{j\in N} A_j \to \mathbb{R}$ is player i's *payoff function*. We assume that A_i is finite and $|A_i| \geq 2$ for all $i \in N$, where $|A_i|$ denotes the cardinality of A_i.

Let $A = \prod_{i\in N} A_i$ and $A_{-i} = \prod_{j\neq i} A_i$. We enumerate the set of action profiles by $A = \{a^1, \ldots, a^{|A|}\}$. We shall denote the maximum payoff in absolute value some player can obtain by $B = \max_{i\in N} \max_{a\in A} |u_i(a)|$.

The set of mixed action of player $i \in N$ is denoted by Δ_i. As above, we let $\Delta = \prod_{i\in N} \Delta_i$ and $\Delta_{-i} = \prod_{j\neq i} \Delta_i$. For each $i \in N$, the mixed extension of player i's payoff function is also denoted by u_i.

For any $i \in N$ denote, respectively, the *pure minmax payoff* and a *pure minmax profile* for player i by $v_i = \min_{a_{-i}\in A_{-i}} \max_{a_i\in A_i} u_i(a_i, a_{-i})$ and $m^i \in A$, where

$$m^i_{-i} \in \arg \min_{a_{-i}\in A_{-i}} \max_{a_i\in A_i} u_i(a_i, a_{-i}) \text{ and } m^i_i \in \arg \max_{a_i\in A_i} u_i(a_i, m^i_{-i}).$$

If G is a 2-player game, a *pure mutual minmax profile* is $\bar{m} = (m^2_1, m^1_2) \in A$.

Let $\mathcal{U} = \{u \in \text{co}\,(u(A)) : u_i \geq v_i \text{ for all } i \in N\}$ denote the set of *pure individually rational payoffs* and $\mathcal{U}^0 = \{u \in \text{co}(u(A)) : u_i > v_i \text{ for all } i \in N\}$.

Similar definitions apply to the case of mixed strategies. For any $i \in N$ denote, respectively, the *mixed minmax payoff* and a *mixed minmax profile* for player i by $\tilde{v}_i = \min_{\sigma_{-i}\in\Delta_{-i}} \max_{a_i\in A_i} u_i(a_i, \sigma_{-i})$ and $\mu^i \in \Delta$, where $\mu^i_{-i} \in \arg\min_{\sigma_{-i}\in\Delta_{-i}} \max_{a_i\in A_i} u_i(a_i, a_{-i})$ and $\mu^i_i \in \arg\max_{a_i\in A_i} u_i(a_i, \mu^i_{-i})$.

Let $\tilde{\mathcal{U}} = \{u \in \text{co}\,(u(A)) : u_i \geq \tilde{v}_i \text{ for all } i \in N\}$ denote the set of *mixed individually rational payoffs* and $\tilde{\mathcal{U}}^0 = \{u \in \text{co}(u(A)) : u_i > \tilde{v}_i \text{ for all } i \in N\}$.

The repeated game: The *infinitely repeated game* consists of an infinite sequence of repetitions of G. We denote the action of any player i in the repeated game at any date $t = 1, 2, 3, \ldots$ by $a^t_i \in A_i$. Also, let $a^t = (a^t_1, \ldots, a^t_n)$ be the profile of choices at t.

For any $t \geq 1$, a *t-stage history* is a sequence $h = (a^1, \ldots, a^t) \in A^t$ (the t-fold Cartesian product of A). The set of all t -stage histories is denoted by $H_t = A^t$. We represent the initial (empty) history by H_0. The set of all histories is defined by $H = \bigcup_{t \in \mathbb{N}_0} H_t$.[23] We also denote the length of any history $h \in H$ by $\ell(h)$.

Let $\Pi = A \times A \times \cdots = A^\infty$ be the set of (infinite) *outcome paths* in the repeated game. For any $a \in A$ and $k \in \mathbb{N}$, we denote a finite path consisting of a being played k times consecutively by $(a; k)$. Also, for two positive length histories $h = (a^1, \ldots, a^{\ell(h)})$ and $\bar{h} = (\bar{a}^1, \ldots, \bar{a}^{\ell(\bar{h})})$ in H we define the *concatenation of h and* $\bar{h}$ by $h \cdot \bar{h} = (a^1, \ldots, a^{\ell(h)}, \bar{a}^1, \ldots, \bar{a}^{\ell(\bar{h})})$.

For any non-empty history $h = (a^1, \ldots, a^{\ell(h)}) \in H$ and any integer m, define the *m-tail of h* by $T^m(h) = (a^{\max\{\ell(h)-m+1,1\}}, \ldots, a^{\ell(h)})$. We also adopt the convention that $T^0(h)$ is the empty history. For all $h \in H$ and all $k \in \mathbb{N}$ with $k \leq \ell(h)$, let $B^k(h) = (a^1, \ldots, a^{\ell(h)-k})$ denote the history obtained from h by removing the last k actions.

For all $i \in N$, a *pure strategy* for player i is a function $f_i : H \to A_i$ mapping histories into pure actions. The set of player i's strategies is denoted by F_i^p, and $F^p = \prod_{i \in N} F_i^p$ with a typical element $f = (f_1, \ldots, f_n)$. Given a strategy $f_i \in F_i^p$ and a history $h \in H$, we denote the *strategy induced by* f_i *at* h by $f_i|h$. Thus, $(f_i|h)(\bar{h}) = f_i(h \cdot \bar{h})$ for every $\bar{h} \in H$. We will use $(f|h)$ to denote $(f_1|h, \ldots, f_n|h)$ for every $f = (f_1, \ldots, f_n) \in F^p$ and $h \in H$.

Any strategy profile $f \in F^p$ induces an outcome path $\pi(f) = \{\pi^1(f), \pi^2(f), \ldots\} \in \Pi$ where $\pi^1(f) = f(H_0)$ and $\pi^t(f) = f(\pi^1(f), \ldots, \pi^{t-1}(f))$ for any $t > 1$.

We assume that all players discount the future payoffs by a common discount factor $\delta \in (0, 1)$. Thus, the *payoff in the repeated game* is given by $U_i(f, \delta) = (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(\pi^t(f))$. For any $\pi \in \Pi$, $t \in \mathbb{N}$, and $i \in N$, let $V_i^t(\pi, \delta) = (1-\delta) \sum_{r=t}^{\infty} \delta^{r-t} u^i(\pi^r)$ be the *continuation payoff* of player i at date t if the outcome path π is played. For simplicity, we write $V_i(\pi, \delta)$ instead of $V_i^1(\pi, \delta)$. Also, when the meaning is clear we shall not explicitly mention δ and refer to $U_i(f, \delta)$, $V_i^t(\pi, \delta)$ and $V_i(\pi, \delta)$ by $U_i(f)$, $V_i^t(\pi)$ and $V_i(\pi)$ respectively.

We also consider the case where players may choose mixed actions but observe only the realization of those mixed actions. For all $i \in N$, with some abuse of notation, we denote such a *behavior strategy* (also referred to as mixed strategy in this paper) for player i by a function $f_i : H \to \Delta_i$ mapping histories into mixed actions. The set of player i's strategies

[23] We use $\mathbb{N}_0$ and $\mathbb{N}$ to denote, respectively, the set of non-negative and positive integers.

is denoted by F_i^m, and $F^m = \prod_{i \in N} F_i^m$.

Given a strategy $f_i \in F_i^m$ and a history $h \in H$, the *strategy induced by f_i at h* is defined analogously as in the pure case and also denoted by $f_i|h$.

A behavior strategy $f \in F^m$ induces, for every period $t \in \mathbb{N}$, a probability distribution $\tilde{\pi}^t(f)$ over pure actions and a probability distribution $P_{f,t}$ over H_t as follows: $\tilde{\pi}^1(f)[a] = P_{f,1}(a) = f(H_0)[a]$ for all $a \in A = H_1$ and, for any $t > 1$, $h \in H_t$ and $a \in A$, letting $h = \bar{h} \cdot \bar{a}$ with $\bar{h} \in H_{t-1}$, $P_{f,t}(h) = P_{f,t-1}(\bar{h}) f(\bar{h})[\bar{a}]$ and $\tilde{\pi}^t(f)[a] = \sum_{h \in H_t : T^1(h) = a} P_{f,t}(h)$. Given a common discount factor $\delta \in (0,1)$, repeated game payoff of i when $f \in F^m$ is chosen is still denoted by $U_i(f, \delta) = (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} \sum_{a \in A} u_i(a) \tilde{\pi}^t(f)[a]$ for all $i \in N$ (when the meaning is clear we will refer to repeated game payoff by $U_i(f)$ without an explicit reference to δ).

We denote the repeated game described above for discount factor $\delta \in (0,1)$ by $G_p^\infty(\delta)$ if only pure actions are allowed and by $G_m^\infty(\delta)$ if mixed actions are allowed. Fix any $q \in \{m, p\}$. A strategy vector $f \in F^q$ is a *Nash equilibrium* of $G_q^\infty(\delta)$ if $U_i(f) \geq U_i(\hat{f}_i, f_{-i})$ for all $i \in N$ and $\hat{f}_i \in F_i^q$. Also, $f \in F^q$ is a *SPE* of $G_q^\infty(\delta)$ if $f|h$ is a Nash equilibrium for all $h \in H$. For all $M \in \mathbb{N}$, we say that $f \in F^q$ is a *M-memory strategy* if $f(h) = f(\bar{h})$ for all $h, \bar{h} \in H$ such that $T^M(h) = T^M(\bar{h})$. A strategy profile $f \in F^q$ is a *M-memory SPE* of $G_q^\infty(\delta)$ if f is a M-memory strategy and a SPE of $G_q^\infty(\delta)$.

4 Bounded memory Folk Theorem with pure strategies

In this section we restrict the analysis to the case of pure strategies. Our FT result for this case is the following:

Theorem 1 *Let G be a n-player game and suppose that either the interior of $\mathcal{U}$ in $\mathbb{R}^n$ is non-empty or $n = 2$ and $\mathcal{U}^0 \neq \emptyset$. Then, for all $\varepsilon > 0$, there exists $M \in \mathbb{N}$ and $\delta^* \in (0,1)$ such that, for all $u \in \mathcal{U}$ and $\delta \geq \delta^*$, there exists a M-memory pure SPE $f \in F^p$ of $G_p^\infty(\delta)$ such that $\|U(f, \delta) - u\| < \varepsilon$.*[24]

As we explained in the introduction, with finite action spaces, bounded memory and no randomization, the set of individually rational payoffs can at best be implemented approx-

[24]The non-emptiness of the interior of $\mathcal{U}$ in $\mathbb{R}^n$ implies the NEU condition used in **?**. In the proof we actually use the NEU condition.

imately. As a result, in the above FT, the size of the memory M needed depends on the degree of approximation ε. However, note that M is independent of the individual rational payoff u that is being implemented and the discount factor δ.

We next provide an intuition for the proof of Theorem 1. The proof itself can be found in Appendix A.

4.1 Intuition for the 2-player case

In 2-player games, the standard FT construction for sustaining an individually rational payoff vector $u \in \mathcal{U}$ as a SPE is a simple (pure) strategy profile that has the following structure: (i) it has an equilibrium path π that induces u and (ii) a common punishment path that starts with a punishment phase consisting of playing the mutual minmax $\bar{m}$ for some finite number of time T and then plays the equilibrium path π (see **?**).

Our FT construction with bounded memory involves modifying the above standard construction to deal with the issues that bounded memory raises. As illustrated by the examples in the Introduction, the identification of the ongoing path and whether or not there has been a single player deviation can be difficult with bounded memory. This implies that the equilibrium path and the punishment path need to be chosen carefully so that the above problems can be overcome when players observe only a fixed window of past outcomes. This issue will be dealt with by designing the equilibrium cycle appropriately. The key idea is to insert a signalling sequence of actions *regularly* in the equilibrium path. The purpose of this signalling sequence is that, once players have observed it, they can infer that the play is in the equilibrium path and can, therefore, ignore the part of the history that has occurred before. For such identification to be both possible and immune to single player deviations, the following must hold: the signalling sequence of actions must appear infinitely often on the equilibrium path, it should not appear anywhere else and no single player deviation, either from the equilibrium path or from the punishment path, should be able to escape the punishment phase.

Specifically, our construction of the bounded memory equilibrium strategy is as follows. Since the discount factor is close to 1, for any path, changing the order by which actions are played has an insignificant impact on the payoffs the players receive. Therefore, to approx-

imately implement the desired payoff profile u, all that matters is that each action profile is played a fraction of times sufficiently close to its coefficient in the convex combination of stage game payoffs yielding u. This irrelevance of the order allows us to define the equilibrium path $\pi = \{\pi^1, \pi^2, \ldots\}$ as the repetition of the cycle $\left((a^1; p^1), \ldots, (a^{|A|}; p^{|A|})\right)$ of length $K = \sum_{j=1}^{|A|} p^j$,[25] where, (i) a^1 is chosen to be such that it differs from the mutual minmax profile $\bar{m}$ in every coordinate (i.e. $a_i^1 \neq \bar{m}_i$ for all i), a^2 is set to equal $\bar{m}$ and all remaining actions are ordered arbitrarily; (ii) $p^1 \geq 2$ and $p^2 \geq 1$; and (iii) p^j/K is close to the coefficient of $u(a^j)$ in the convex combination yielding u, for all $j = 1, \ldots, r$.

The M-memory strategy profile that implements the above path can then be described as follows (see condition (18) in the Appendix for a formal definition): It begins at any date $t < p^1$ by playing the equilibrium action $\pi^t = a^1$ if no deviation from the equilibrium path $(\pi^1, \ldots, \pi^{t-1}) = (a^1; t-1)$ has occurred. It continues with playing the equilibrium path after any history of length greater or equal to p^1 if the M-tail of the history either contains p^1 consecutive occurrences of a^1 followed by the subsequent actions of the equilibrium path (if any) or consists of $M - t$ consecutive occurrences of $\bar{m}$ followed by the first t actions of the equilibrium cycle, for some $t = 0, 1, \ldots, p^1 - 1$. At any other history, the strategy profile prescribes playing $\bar{m}$.[26]

In this construction the sequence $(a^1; p^1)$ at the beginning of the equilibrium cycle is the required signalling phase described above. It appears infinitely often on the equilibrium path and it differs from the action profile $\bar{m}$ that is played during the punishment phase in every component.

To implement the equilibrium path by simply observing the first p^1 period of the path, the size of the memory has to be no less than the sum of K, the length of the equilibrium cycle, and p^1, the length of the signalling phase. Furthermore, since with the above strategy profile, deviations from the equilibrium path induce mutual minmax for M periods, we must also have $M \geq T$, the length of mutual minmax phase needed to deter deviations. Therefore,

[25]Recall that $A = \{a^1, \ldots, a^{|A|}\}$ and $(a; k)$ denotes the history consisting of the play of action profile a for k consecutive periods.

[26]Note that the above strategy profile does not satisfy Abreu's definition of simple strategy. This is because the punishment path is not unique: the number of times the mutual minmax action is to be played in response to a deviation depends on the number of times the mutual minmax appears before the punishment starts.

to obtain our result, the bound on the memory M needs to be no less than $\max\{K+p^1, T\}$.

Finally, the above construction is such that no single player deviation, either from the equilibrium path or from the punishment path, can escape the punishment phase. To see this, note first that because $a_i^1 \neq \bar{m}_i$, for all $i = 1, 2$, no single player can deviate from the mutual minmaxing phase and induce the signalling phase that is necessary to escape punishment. The same holds also regarding deviations from histories whose M-tail consists of $M-t$ consecutive occurrences of $\bar{m}$ followed by the first t actions of the equilibrium cycle, for some $t < p^1$, because a player deviating singly from a^1 will lead to an action different from both a^1 and $\bar{m}$. Last, consider any single-player deviation from the equilibrium path. Such a deviation does not result in a punishment phase only if the M-tail of the history after the deviation either contains p^1 consecutive occurrences of a^1 followed by the subsequent actions of the equilibrium path (if any) or, for some $t < p^1$, consists of $M-t$ consecutive occurrences of $\bar{m}$ followed by the first t actions of the equilibrium cycle. The latter cannot happen because the M-tail does not contain p^1 consecutive a^1's and hence the deviation could not be from the equilibrium path. Consider then the former case. In this case such a deviation is feasible only if the p^1-tail is $(a^1; p^1)$. Since $p^1 \geq 2$, both the action profile induced by the deviation and the action profile just before the deviation must be a^1. But, on the equilibrium path, only a^1 or $\bar{m}$ follow a^1. Since the deviation induces a^1, then it must be that the deviation is from $\bar{m}$. But $\bar{m}$ differs from a^1 in every coordinate, which implies that single-player deviation cannot produce such a history.

To illustrate our construction and explain why the assumption of $p^1 \geq 2$ and $p^2 \geq 1$ cannot be weakened, consider the repeated Prisoners' Dilemma described in the introduction (for example, with payoffs as described in footnote 8) and suppose that we want to construct a SPE with bounded memory that induces the payoffs that approximates the payoffs corresponding to playing the cycle $((C,D),(D,C))$ forever. Since in this example $\bar{m} = (D,D)$ and the profile that differs from $\bar{m}$ in every component is (C,C), our construction would then involve considering equilibrium strategies that play the cycle $\{((C,C);p^1),((D,D);p^2),((C,D);\ell),((D,C);\ell)\}$ repeatedly on the equilibrium path, for some $p^1 \geq 2, p^2 \geq 1$ and ℓ, and punish any deviation by playing (D,D) for M times followed by a return to the equilibrium path. For sufficiently large ℓ, relative to p^1 and p^2, such a

bounded strategy profile results in (C, D) and (D, C) being played almost all the time and in equal proportion on the equilibrium path.

In the above construction any $p^1 \geq 2$ and $p^2 \geq 1$ will suffice to achieve bounded memory SPE implementation. To see why p^2 has to be positive, suppose that $p^1 = 2$, $p^2 = 0$. Then, the signalling phase is $((C, C); 2)$ and the equilibrium cycle consists of just $(((C, C); 2), ((C, D); \ell), ((D, C); \ell))$. This implies that if the M-tail of a history is given by $(a^1, \ldots, a^{M-2}, (C, C), (C, C))$ for some sequence of action profiles $a^1, \ldots, a^{M-2}$, then the signalling phase $((C, C); 2)$ is observed and the players should play (C, D). But if player 2 deviates and plays C at this history, the next period M-tail of the resulting history would be $(a^2, \ldots a^{M-2}, (C, C), (C, C), (C, C))$. Since the signaling phase is observed again, such deviation does not trigger the punishment path.

To see why we need p^1 to be no less than 2, suppose that the equilibrium cycle is such that $p^1 = p^2 = 1$. Then the strategy profile recommends (C, D) at any history whose M-tail equals $(a^1, \ldots, a^{M-2}, (C, C), (D, D))$, for some $a^1, \ldots, a^{M-2}$. But if player 2 deviates at this history and plays C instead, the next period M-tail of the resulting history would be $(a^2, \ldots, a^{M-2}, (C, C), (D, D), (C, C))$. Since this history induces the signalling phase (C, C), such a deviation does not trigger the punishment path.

4.2 Intuition for the $n > 2$ case

With no bounds on memory and more than two players, to implement $u \in \mathcal{U}$ the standard FT calls for the use of a simple (pure) strategy consisting of an equilibrium path $\pi^{(0)}$ and n punishment paths $\pi^{(1)}, \ldots, \pi^{(n)}$ with the following property. The punishment path $\pi^{(i)}$ for player i consists of playing the minmax profile m^i for T periods followed by a path $\hat{\pi}^{(i)}$, referred to as the reward path corresponding to $\pi^{(i)}$; thus

$$\pi^{(i),t} = \begin{cases} m^i & \text{if } t \leq T \\ \hat{\pi}^{(i),t-T} & \text{otherwise.} \end{cases}$$

Therefore, the typical FT construction consists of three sets of sequences of action profiles: (i) the equilibrium path $\pi^{(0)}$, (ii) the minmax phase for each player i consisting of playing m^i a finite number of times T, and (iii) the reward paths $\hat{\pi}^{(i)}$ for each i. For convenience, we let $\hat{\pi}^{(i)}$ denote the equilibrium path when $i = 0$.

As in the above standard construction, the bounded memory strategy profiles we use to prove our FT are such that the incentives to play the equilibrium and reward paths are given by the threat of punishments, consisting of a sequence of the deviator's minmax action profile followed by the appropriate reward path. However, to identify each of the sequences described in (i)–(iii) and the appropriate action profile that has to be played, we add to the beginning of each of the above sequences a distinct signalling phase. As with the 2-player case, once players observe one of these signalling phases, they can identify what needs to be played and, therefore, can forget all that has happened before.

For example, each signalling phase could consist of a sequence $(s; l)$ where $s \in A$ is some fixed action profile and l is some number that is different for the different signalling phases. The idea is that when players observe a sequence of the form $(s; l)$ then, by counting the number of consecutive s's, which equals l in this sequence, they can identify the path to play.

The above, however, may not work as the players need to identify when the signalling phase starts and when it ends. Specifically, if $(s; l)$ is observed then the history is consistent with any signalling phase $(s; l')$ for all $l' \leq l$. To overcome this, we modify each signalling phase $(s; l)$ so that it is preceded and followed by another action, $s' \neq s$.

The addition of s' to the signalling phases is also not enough. First, we need to ensure that each signalling phase cannot be induced by single player deviations from another signalling phase. We deal with this problem by choosing s' to be such that it differs from s in every coordinate (i.e. $s_i \neq s'_i$ for all $i \in N$). Second, for reasons that will become clear later, we also need to assume that each signalling phase starts with two s''s and has at least two consecutive s's. Specifically, each signalling phase in our construction is described by $(s', s', (s; l), s')$ and we set l in each phase as follows: $l = i + 1$ for the minmax path of player i, $l = n + 2$ for the equilibrium path and $l = n + 2 + i$ for the reward path of player i.

As we discussed before, with δ close to 1, to approximately implement $u \in \mathcal{U}$, all that matters is that, on the equilibrium path, each action profile is played an appropriate fraction of times. The same holds for approximately implementing the payoffs corresponding to the reward paths. It may then seem that the simple strategy profile that we need is as follows:

(i) the equilibrium path $\pi^{(0)} = (\pi^{(0),1}, \pi^{(0),2}, \ldots)$ consists of the repetition of the following

type of cycle path

$$\left(s', s', (s; n+2), s', (a^1, ; p^{(0),1}), \ldots, (a^{|A|}; p^{(0),|A|})\right);$$

where $p^{(0),j}$ is chosen appropriately so that $\pi^{(0)}$ induce approximately u.

(ii) the reward path $\hat{\pi}^{(i)} = (\hat{\pi}^{(i),1}, \hat{\pi}^{(i),2}, \ldots)$, $i \in \{1, \ldots, n\}$, is the repetition of the cycle

$$\left(s', s', (s; n+i+2), s', (a^1; p^{(i),1}), \ldots, (a^{|A|}; p^{(i),|A|})\right);$$

where $p^{(i),j}$ is chosen appropriately so that $\pi^{(i)}$ induce approximately the appropriate reward payoff.

(iii) the punishment path $\pi^{(i)}$, $i \in \{1, \ldots, n\}$, is given by

$$\pi^{(i)} = \left(s', s', (s; i+1), s', (m^i; T), \hat{\pi}^{(i),1}, \hat{\pi}^{(i),2}, \ldots\right),$$

where T is chosen appropriately to deter single-player deviations.

Unfortunately, the problem is a great deal more complicated. An immediate issue is that we must ensure that the introduction of the signalling phases does not affect the incentives adversely. On all paths other than one's own punishment path, we can ensure that the players play the appropriate continuation path by the standard construction that invokes the punishment for the deviator after any single player deviation from such phases. The same is, however, not the case regarding the play of one's own punishment path.

First, once we introduce a signalling phase at the beginning of each punishment path, some player may have a profitable deviation in the minmax phase of his own punishment, if such deviation restarts the punishment path. For example, deviation by i at the beginning of the minmax phase of his own punishment path induces the outcome $(s', s', (s; i+1), s', (m^i; T), \hat{\pi}^{(i),1}, \hat{\pi}^{(i),2}, \ldots)$, whereas no deviation induces $((m^i; T-1), \hat{\pi}^{(i),1}, \hat{\pi}^{(i),2}, \ldots)$. If $(s', s', (s; i+1), s')$ generates a sufficiently high average payoff, then the deviation will be profitable. To deal with this problem, we modify the above simple strategy construction by assuming that deviations by a player from his own minmax action in his punishment path are ignored and the punishment path is not restarted. Such a change in the construction does not affect the incentives because there are no one-period gains to deviations during the minmax phase.

Second, some player may profitably deviate in the signalling phase of his own punishment path if such deviation restarts the signalling phase. For instance, if some player i obtains a high payoff by deviating from s' to some action a_i, he could perpetually deviate in the first period of the punishment path and obtain a path consisting in the repetition of (a_i, s'_{-i}) delivering him a higher payoff. Similarly, if some player i obtains a high payoff by deviating from s by playing some action a_i, then he could perpetually deviate in the third period of the punishment path and obtain a path consisting in the repetition of $(s', s', (a_i, s_{-i}))$ which could yield him a higher payoff.

We deal with this problem by specifying that when there is a deviation by a player in the signalling phase of his punishment path, the strategy prescribes the continuation of that particular signalling phase. But this by itself is not enough as we need to ensure that there is punishment to deter deviations during this phase (if s or s' were Nash equilibria of the stage game this would, of course, be unnecessary). We establish such deterrence by appropriately increasing the length of the minmax phase of the punishment path for each such deviation. Specifically, denoting the number of times that player i has deviated during the signalling phase of his punishment path by $\theta \in \{0, 1, \ldots, i+4\}$, the strategy profile requires that once the current signalling phase is over, the continuation path consists of playing $((m^i; (\theta+1)T), \hat{\pi}^{(i),1}, \hat{\pi}^{(i),2}, \ldots)$. Such construction implies that for every deviation during the signalling phase the length of the minmax phase increases by T.[27]

The above modification involving delayed punishments of deviations during the signalling phases of the punishment paths has two implications that are worth noting. First, each player i effectively has $i+5$ punishment paths indexed by $\theta \in \{0, 1, \ldots, i+4\}$.[28] We denote each of these by $\pi^{(i)}(\theta) = (s', s', (s; i+1), s', (m^i; (\theta+1)T), \hat{\pi}^{(i),1}, \hat{\pi}^{(i),2}, \ldots)$ and define the path $\pi^{(i)}(\theta)$ without its first $t-1$ elements by $\pi^{(i)}(\theta, t)$.

Second, ignoring one-period deviations by any player i during the signalling phases of

[27]The reason for having $\theta + 1$ instead of just θ is that player i needs to be punished even if he does not deviate in the signalling phase of his punishment path.

[28]By the number of punishment paths, we mean the number of distinct paths that a player can induce by a deviation, excluding the continuation path that occurs when the player does not deviate. Note also that our construction does not constitute a simple strategy profile because it will have, in addition to the equilibrium path, $\sum_{i=1}^{n}(i+5)$ punishments paths.

i's punishment path, as proposed above, means that the minmax phase starts after any sequences

$$((a_i^1, s'_{-i}), (a_i^2, s'_{-i}), (a_i^3, s_{-i}), \ldots, (a_i^{i+3}, s_{-i}), (a_i^{i+4}, s'_{-i})), \tag{1}$$

with $a_i^l \in A_i$ for all $l = 1, \ldots, i+4$, has been observed. Therefore, it follows that the signal for the punishment of player i are effectively all sequences satisfying (1) rather than just $(s', s', (s; i+1), s')$. To differentiate between any sequence described in (1) from the signalling phase $(s', s', (s; i+1), s')$, we shall call the former a generalized signalling phase for player i's punishment path.

Given the above, after any history $h = (a^1, \ldots, a^\tau)$, our M-period memory strategy profile f would satisfy the following conditions (the formal definition of the strategy is given in (28) in the Appendix):

(a) *(Equilibrium and reward path histories)* Suppose the t-tail of h is $(\hat{\pi}^{(i),1}, \ldots, \hat{\pi}^{(i),t})$, for some $i = 0, \ldots, n$ and $t \leq M$, and it includes the signalling phase $(s', s', (s; n+i+2), s')$ of $\hat{\pi}^{(i)}$, i.e. $n+i+5 \leq t$. Then f prescribes players to continue with $\hat{\pi}^{(i)}$.

(b) *(Punishment path histories)* Suppose that, for some $i = 1, \ldots, n$ and t such that $i+4 \leq t \leq M$, the t-tail of h has the following properties:

(i) the first $i+4$ elements of the t-tail is a generalized signalling phase of i as described in (1);

(ii) if $t \leq (\theta+1)T + i + 4$, where θ refers to the number of times that player i has deviated during the signalling phase (1), the remaining elements of the t-tail are such that the players other than i minmax i by playing m^i_{-i};

(iii) if $t > (\theta+1)T + i + 4$, in every period $i+4 < r \leq (\theta+1)T + i + 4$ of the t-tail all players other than i minmax i by playing m^i_{-i}, and the remaining elements of the t-tail correspond to the first $t - ((\theta+1)T + i + 4)$ elements of the path $\hat{\pi}^{(i)}$.

Then f requires the players to continue with $\pi^{(i)}(\theta, t+1)$.

(c) *(Histories involving deviations from (a)–(b))* Suppose case (b) does not apply and, for some $r \in \{\tau - M, \ldots, \tau\}$, $(a^1, \ldots, a^{r-1})$ satisfies the properties described in either (a) or (b) above, a^r involves a deviation by some player i from f as described in (a) and (b),

and $(a^{r+1}, \ldots, a^{\tau})$ is consistent with a generalized signalling phase for player i's punishment path. Then f prescribes $\pi^{(i)}(\theta, \tau - r + 1)$, where θ refers to the number of times that player i has deviated during $(a^{r+1}, \ldots, a^{\tau})$.

Conditions (a)–(c) describe the behavior after histories that have the following feature: For some $t \leq M$, its t-tail contains the entire signalling phase of one of the equilibrium or reward path, or an entire generalized signalling phase for a punishment path. In particular, (a)–(c) specify the appropriate path to be played once these signalling phases are observed and are followed by a sequence of actions in which there are either no deviations or only single-player deviations from the path corresponding to the signalling phase.

What if a complete generalized signalling phase does not appear in the M-tail of the history? The specification of what should be played at such histories cannot be arbitrary as the equilibrium should be such that it is not in the interest of any player to deviate *during* a generalized signalling phase of another player's punishment path. To deal with this case, we assume that if a complete generalized signalling phase does not appear in the M-tail of the history as in (a)–(c) and if, for some $t \leq M$, the t-tail of the history consists of a single-player deviation from s or s' by player i followed by an incomplete generalized signalling phase for the punishment of player i, then the strategy recommends players to continue with such signalling phase. For any other history, our construction prescribes playing the equilibrium path.[29] More formally, in addition to (a)–(c) above, we assume that the equilibrium strategies satisfy the following two conditions at every history $h = (a^1, \ldots, a^{\tau})$:

(d) *(Histories that involve deviations from incomplete signalling phases)* If none of the conditions (a)–(c) are satisfied and if for some $r \in \{\tau - M, \ldots, \tau\}$, a^r involves a deviation by some player i from s or s' and $(a^{r+1}, \ldots, a^{\tau})$ is consistent with a generalized signalling phase of player i's punishment path, then f prescribes $\pi^{(i)}(\theta, \tau - r + 1)$, where θ refers to the number of times that player i has deviated during $(a^{r+1}, \ldots, a^{\tau})$.[30]

[29]The specification of the continuation path here is somewhat arbitrary; all that is needed is that the play results in any of the equilibrium, reward or punishment paths.

[30]Unlike in the case of condition (c), there may be several values for r such that condition (d) holds. For example, if h satisfies none of the conditions (a)–(c) and $T^3(h) = ((a_i^1, s'_{-i}), (a_i^2, s'_{-i}), (a_i^3, s'_{-i}))$ for some $i \in N$ and $a_i^1, a_i^2, a_i^3 \neq s'_i$, then condition (d) is satisfied with $r = \tau$, $r = \tau - 1$ and $r = \tau - 2$. In our proof, we take the smallest such r (in the example in this footnote, f prescribes $\pi^{(i)}(\theta, t)$ with $\theta = 2$ and $t = 3$).

(e) *(Other histories)* If none of conditions (a)–(d) are satisfied and the last $0 \leq t < M$ periods corresponds to the first t periods of the equilibrium path $\pi^{(0)}$, then the strategy prescribes players to continue with $\pi^{(0)}$ (when $t = 0$ the strategy recommends the first action on the equilibrium path).[31]

To ensure that the above behavior described by (a)–(e) can be implemented when M is finite, however, several issues have to be addressed.

First, we need to set M to be large enough so that it is possible to distinguish between the different paths and phases. Specifically, let K be such that all individually rational payoffs can be approximately obtained by the average payoff of cycle paths of length K.[32] Also, note that the length of the longest signalling phase in the different punishment paths, the length of the longest minmax phase and the length of the longest signalling phase of the reward paths are respectively $n + 4$, $T(n + 5)$ and $2n + 5$. Then it follows that for the strategy profile to implement the punishment paths, the memory size has to be at least $(n + 4) + T(n + 5) + (2n + 5) + K$. We show in the appendix it suffices to have M greater than this bound to implement our strategy profile.

Second, even though the signalling phases of the different paths, including the generalized signalling phases as described by (1), are all different, this does not necessarily imply that, once they are observed, they can be used to identify the future path of play. For example, if the signalling phase $(s', s', (s; n+i+2), s')$ of $\hat{\pi}^{(i)}$ appears on $\hat{\pi}^{(j)}$, for $j \neq i$, then the strategy described above may not be well-defined. Furthermore, for these signalling phases to have the required property that once they are observed all previous history can be ignored, it should also be the case that they cannot be induced by a single player deviation from some other path. For example, if, for some $a_j \neq s'_j$, the sequence $(s', s', (s; n + i + 2), (a_j, s'_{-j}))$ appears on the reward path $\hat{\pi}^{(j)}$, then there may be an incentive for j to play s'_j on the path

[31]Note that at histories described in (e), it is possible that the path resulting from such a history fails to be the equilibrium path. For example, suppose that $T^{n+i+4}(h) = (s', s', (s; n + i + 2))$ for some $i \in N$ and that h does not satisfy (a)–(d). Then the strategy recommends the first action on the equilibrium path s'. The resulting history, denoted by h', satisfies $T^{n+i+5}(h') = (s', s', (s; n + i + 2), s')$, which equals the signalling phase of player i's reward path. At this point, player i's reward path will be played henceforth. This, of course, does not generate any problems as the strategy profile still implements a SPE path.

[32]Notice that K, and hence M, will depend on the degree of approximation.

$\hat{\pi}^{(j)}$ after $(s', s', (s; n+i+2))$, as such a deviation induces the signalling phase of $\hat{\pi}^{(i)}$.

The issue here is that we not only need the signalling phases to be distinct from each other, they also need to be appropriately distinct with respect to the equilibrium and reward paths, as well as with respect to minmax phases. We deal with these issues as follows.

By the same argument as before, the order by which the sequence of actions $\{a^1, \ldots, a^{|A|}\}$ is played on the equilibrium path and on each of the reward paths, as well as the number of times they are played on the path, do not matter as long as each action profile is played an appropriate fraction of times. This freedom to choose the order of the sequence $\{a^1, \ldots, a^{|A|}\}$ allow us to construct the equilibrium and the reward paths in such a way so that they are appropriately distinct from the signalling phases.

Specifically, we achieve this as follows. The first action profile a^1 is set to be equal to s and is followed by all the action profiles of the form (a_i, s_{-i}) for some $i \in N$ and $a_i \neq s_i$. These are followed by s', and then by action profiles of the form (a_i, s'_{-i}) for some $i \in N$ and $a_i \neq s'_i$. The remaining action profiles are ordered arbitrarily.[33] With this ordering, on the equilibrium and reward paths, s' and action profiles obtained by single player deviations from s' are never followed by s or by action profiles consisting of single player deviations from s, other than in the initial signalling phases. This ordering ensures that (i) for each $i = 0, \ldots, n$, the signalling phase of $\hat{\pi}^{(i)}$ appears only once on the cycle path of $\hat{\pi}^{(i)}$ and it does not appear on $\hat{\pi}^{(j)}$, for all $j \neq i$, (ii) the generalized signalling phase for each punishment path does not appear on $\hat{\pi}^{(j)}$, for all $j = 0, \ldots, n$ and (iii) no signalling phase can be induced from single player deviations from $\hat{\pi}^{(j)}$, for all $j = 0, \ldots, n$.

There is still the issue of appropriate distinctness of the signalling phases from the minmax ones. Since the signalling phases consist of two action profiles s and s' that are distinct in every component, it follows trivially that the signalling phases, including the generalized ones, cannot occur when all players are minmaxing a specific player and, furthermore, the former sequences cannot be induced by single player deviations from a minmax phase. However, in our construction, we assume that a deviation by any player i from his minmax profile

[33]For example, when $n = 3$ and $A_i = \{\alpha, \beta\}$ for all $i \in N$, a possible ordering respecting the above properties would be $a^1 = s = (\alpha, \alpha, \alpha)$, $a^2 = (\beta, \alpha, \alpha)$, $a^3 = (\alpha, \beta, \alpha)$, $a^4 = (\alpha, \alpha, \beta)$, $a^5 = s' = (\beta, \beta, \beta)$, $a^6 = (\alpha, \beta, \beta)$, $a^7 = (\beta, \alpha, \beta)$, $a^8 = (\beta, \beta, \alpha)$.

m^i is ignored and the future play is not affected by such a deviation. This means that we must also ensure that signalling phase, including the generalized ones, cannot be induced by single player deviations from sequences $((a_i^1, m_{-i}), \ldots, (a_i^\tau, m_{-i}))$ that involve single player deviations by player i from his own minmax phase. Our requirement that each signalling phase contains at least two consecutive s''s and two consecutive s's at the beginning of these phases deals with this issue.

To see the role of at least two consecutive s's in the signalling phases, suppose that instead of assuming that the signalling phases of the punishment path of each i has $i+1$ consecutive s's, we have i consecutive s's. This means that the signalling phase of player i's punishment is given by $(s', s', (s; i), s')$. Consider then a 3-player game with $m^3 = (s_3, s'_{-3})$, a history $h = ((s'; 2), (s; 3), s', s', s', (s'_1, s_{-1}), s')$ and $M \geq 10$. Since $s' = (s'_3, , m^3_{-3})$, $(s'_1, s_{-1}) = (s_2, m^3_{-2})$ and the signalling phase for player i's punishment is $((s'; 2), (s; i), s')$, it follows that h consists of the signalling phase for player 3's punishment, followed by (s'_3, m^3_{-3}) being played twice, followed by (s_2, m^3_{-2}) and followed by s', the first action of the signalling phase of player 2's punishment path. Hence, by part (c) of our construction above, the strategy prescribes continuing with punishing player 2 by playing $\big((s; 2), s', (m^2; T), \hat{\pi}^{(2),1}, \hat{\pi}^{(2),2}, \ldots\big)$. But $T^4(h) = ((s'; 2), (s'_1, s_{-1}), s')$ is a generalized signalling phase of player 1's punishment. Thus, part (b) of our construction also applies. Therefore, the strategy also recommends $\big((m^1; 2T), \hat{\pi}^{(1),1}, \hat{\pi}^{(1),2}, \ldots\big)$.

The problem here arises because $s' = (s'_3, m^3_{-3})$ and $(s'_1, s_{-1}) = (s_2, m^3_{-2})$. Hence, single-player deviations from m^3 can induce both s' and single-player deviations from s, and, as a result, the continuation strategy after history h is not well-defined.

Having s played $i + 1$ times in the signalling phase of i's punishment solves the above problem as follows. In this case the signalling phase of player 1 is $(s', s', (s; 2), s')$. This means that if player 1 deviates from s during his signalling phase this is preceded and succeeded by s and s' or the reverse. Since it cannot be the case that *both* s and s' can be induced by a player deviating from his own minmax profile, it follows that deviations by player 1 from his own signalling phase are not consistent with phases involving another player deviating from his own minmax phase. Hence, the problem described above does not arise.

Similarly, to see the role of having two consecutive s''s at the beginning of the signalling

phases, suppose that instead of assuming that the signalling phases of the punishment path of each i is $(s', s', (s; i+1), s')$, we assume that it consists of $(s', (s; i+1), s')$ with only one s' at the beginning of these phases. Consider a 3-player game with $m^1 = (s'_1, s_{-1})$, a history $h = (s', (s; 2), s', (s; 3), (s_2, s'_{-2}))$ and $M \geq 8$. Since $s = (s_1, m^1_{-1})$, $(s_2, s'_{-2}) = (s'_3, m^1_{-3})$ and the signalling phase for player i's punishment is $(s', (s; i+1), s')$, it follows that h consists of the signalling phase for player 1's punishment, followed by (s_1, m^1_{-1}) being played three times, followed by (s'_3, m^1_{-3}). Hence, by part (c) of our construction above, the strategy prescribes $\pi^{(3)}$. But $T^5(h) = (s', (s; 3), (s_2, s'_{-2}))$ is a generalized signalling phase of player 2's punishment. Thus, part (b) of our construction also applies. Therefore, the strategy also recommends $\big((m^2; 2T), \hat{\pi}^{(2),1}, \hat{\pi}^{(2),2}, \ldots\big)$.

The problem here arises because $s = (s_1, m^1_{-1})$ and $(s_2, s'_{-2}) = (s'_3, m^1_{-3})$. Hence, single-player deviations from m^1 can induce both s and single-player deviations from s'; as a result, the continuation strategy after history h is not well-defined.

Having two s''s at the beginning of the signalling phases solves this problem. Then the signalling phase of player 2 would be $((s'; 2), (s; 3), s')$. But such phase is consistent with the signalling phase of player 1 followed by 1's minmax phase only if *both* s and s' could be induced by player 1 deviating from his own minmax profile.[34] Since s and s' are distinct in every component, this is not feasible, hence, the problem described above does not arise.

4.3 Comparing the proof for $n > 2$ to $n = 2$

Before concluding the discussion of the proof of Theorem 1, note that the proof of the Theorem for the $n = 2$ case cannot be applied to $n > 2$ case because the common punishment (involving mutual minmaxing) used in the former cannot be applied to the latter. Also, with bounded memory, a common punishment may be necessary to prove a FT result when $n = 2$ because it is not always possible to detect the identity of a deviator in a two player settings.[35]

[34]This is because in this case the number of s''s at the beginning of the signalling phase of player 2 is different from that at the end of the signalling phase of player 1.

[35]For example, since in the proof of the $n > 2$ case the signalling phase of the punishment phases of players 1 and 2 are respectively $(s', s', (s; 2), s')$ and $(s', s', (s; 3), s')$ it follows that if the proof is applied to the $n = 2$ case, then when the history under consideration is given by $(s', s', (s; 2), (s'_1, s_2))$, it is not clear if 2 has deviated from the signalling phase of the punishment path of 1 or if 1 has deviated from the signalling

Hence, our proof for $n > 2$ case also cannot be applied to $n = 2$ case because the former uses a separate punishment for each player.

5 Bounded memory Folk Theorem with mixed strategies

The set of mixed individually rational payoffs $\tilde{\mathcal{U}}$ contains that with pure strategies $\mathcal{U}$. This section demonstrates that the bounded memory FT holds for $\tilde{\mathcal{U}}$ in the case of three or more players who can choose behavioral strategies.

Theorem 2 *Suppose that $n > 2$ and the interior of $\mathcal{U}$ in $\mathbb{R}^n$ is non-empty. Then, for all $\varepsilon > 0$, there exist $\delta^* \in (0,1)$ such that, for all $u \in \tilde{\mathcal{U}}$ and $\delta \geq \delta^*$, there exists $M \in \mathbb{N}$ and a M-memory behavior SPE $f \in F^m$ of $G_m^\infty(\delta)$ such that $\|U(f,\delta) - u\| < \varepsilon$.*[36]

The proof can be found in Appendix B. In the rest of the section, we discuss the additional difficulties that arise in establishing our mixed bounded memory FT and provide an intuition for the method we use to overcome these difficulties, and thereby, to prove the result.

With or without restrictions on memory, the key difficulty in showing a mixed FT as compared with a pure FT is to provide incentives so that, in the minmax phase of a given player i, all players other than i play their part of i's mixed minmax action profile. The standard proof of the unbounded memory mixed FT is that of **?**. They solve the problem by making each player $j \neq i$ indifferent between all the actions in the support of his part of i's mixed minmax action profile. In turn, this indifference is achieved by making the continuation payoff that occurs after the minmax phase dependent on the sequence of (pure) action profiles that were actually played in the minmax phase.

?'s approach of making each player $j \neq i$ exactly indifferent by adjusting future continuation payoffs appropriately works because any feasible payoff can be exactly implemented by

phase of the punishment path of 2.

[36]If the interior of $\mathcal{U}$ in $\mathbb{R}^n$ is non-empty, then so is the interior of $\tilde{\mathcal{U}}$ in $\mathbb{R}^n$. However, the converse does not hold. Theorem 2 also holds if the interior of $\tilde{\mathcal{U}}$ in $\mathbb{R}^n$ is non-empty and $\max_{a \in A} u_i(a) > v_i$ for all i. For more, see footnote 41.

some strategy profile. But with bounded memory restrictions this is not possible as feasible payoffs can, in general, only be approximately implemented. For this reason, we use an alternative approach based on **?**.

In **?** approach, the construction of the equilibrium strategy profile is similar to that in **?** except that the punishment path for any player i is such that the minmax phase for i involves playing action profiles that *approximately* minmax player i for a finite number of periods T, and the reward phase following the minmax phase involve random reward paths that reward other players if they have *approximately* minmaxed player i during i's minmax phase. Specifically, at the end of the minmax phase of player i a statistical test is performed to determine whether or not each of the other players have played sufficiently close to their part of i's minmax. For each player other than i, passing the test yield him a higher continuation payoff than the one he obtains if he fails the test, thus providing incentives to pass the test. Furthermore, the test is constructed to have the following two properties. First, each player other than i passes it with probability arbitrarily close to 1 if he plays his part of i's minmax in each period of the minmax phase. Second, i's payoff in his minmax phase is close to his minmax payoff whenever all other players pass the test. Thus, in summary, while it may not be optimal for all players other than i to play their part of i's minmax in each period of i's minmax phase, these players will play in such a way that each of them will pass the statistical test with a probability close to one and player i's payoff in his minmax phase will be close to his minmax payoff with a probability also close to one.

To obtain a mixed bounded memory FT, we modify the construction for pure strategies in the previous section so that the minmax phase and the reward phase are as in **?**. This modification, however, introduces a number of difficulties and new issues because of the finite memory restriction and discounting.[37]

First, in our pure strategy FT the length of the minmax phase T was chosen independently of δ. When the minmax phase and the reward phase are modified as in **?**, however, we cannot ensure that this is the case, or more generally that $\lim_{\delta \to 1} \delta^T = 1$. Since the memory needed to implement the strategy profile has to exceed T, this in turn implies, in contrast to the result with pure strategies, that the minimum memory needed to obtain a mixed FT is no

[37] **?** mixed FT result is for the case of finitely repeated game with no discounting.

longer uniform in δ.

The length of the minmax phase T cannot be chosen independently of δ because the set of feasible payoffs cannot, in general, be exactly implemented with bounded memory and because with the modification the continuation payoff in the reward phases depends on the outcome of the statistical test that is conducted at the end of the minmax phases. Specifically, these two features create difficulties in providing incentives for each player $i \in N$ in the reward paths that follow i's minmax phase if T does not depend on δ. To see this, note first that the statistical test after the minmax phase makes the reward path following the minmax phase random and not unique. Let $\underline{u}_i$ and $\bar{u}_i$ be, respectively, player i's lowest and highest reward payoff after i's minmax phase, and let $\zeta > 0$ be the probability that $\bar{u}_i$ is the reward payoff. Assuming that any deviation by i from his reward phase is followed by a punishment path that consist of a minmax phase of T periods for i followed by a random reward path that depend on the outcome of the test (in the proof, the punishment also involves an initial signalling phase; but to simplify the discussion here we will ignore these signalling phases in the computation below), the lowest reward payoff $\underline{u}_i$ must then satisfy the following incentive condition:

$$\begin{aligned} \underline{u}_i &\geq (1-\delta)B_i + (1-\delta)\sum_{t=1}^{T}\delta^t\tilde{v}_i^t + \delta^{T+1}(\zeta\bar{u}_i + (1-\zeta)\underline{u}_i) \\ &= (1-\delta)B_i + (1-\delta)\sum_{t=1}^{T}\delta^t\tilde{v}_i^t + \delta^{T+1}(\underline{u}_i + \zeta\varphi). \end{aligned} \tag{2}$$

where B_i denotes the maximum payoff from one period deviation, $\tilde{v}_i^t$ denotes i's expected payoff in period t of his minmax phase and $\varphi = \bar{u}_i - \underline{u}_i \geq 0$. But if T is independent of δ, or more generally if $\lim_{\delta\to 1}\delta^T = 1$, the limit of the right hand side of (2) is $\underline{u}_i + \zeta\varphi$ as δ goes to 1. Hence (2), may fail to hold for sufficiently large δ if ζ and φ do not vanish.

But ζ is determined by the probability that players other than i pass the test after T periods, and hence, for any finite T, it is positive even as $\delta \to 1$. Also, for any fixed memory size we cannot ensure that player i receives the same payoff in all reward phases that follow his minmax phase even as $\delta \to 1$, i.e. we cannot ensure that φ vanishes. This is because these reward phases are chosen to incentivize other players to (approximately) minmax player i during i's minmax phase, hence, they involve different payoffs for other players depending

on whether they pass the tests or not. Given the need for inducing different payoffs for the other players and given that with a fixed memory it is not feasible to implement all feasible payoffs exactly, it follows that we cannot guarantee that i's continuation payoff is exactly the same in all the reward phases following i's minmax phase for any fixed memory.[38]

To solve the above problem, we will allow for $\lim_{\delta\to 1}\delta^T$ to be less than 1 and ensure (2) as follows: Since $\underline{u}_i$ is individually rational, choose $\varepsilon > 0$ such that $\underline{u}_i > \tilde{v}_i + \varepsilon$. Then, by making the statistical test sufficiently stringent, we have that $(1-\delta^T)(\tilde{v}_i+\varepsilon) > (1-\delta)\sum_{t=1}^{T}\delta^t\tilde{v}_i^t$. Then it follows that we can fix $c \in (0,1)$ such that

$$\underline{u}_i > (1-c)(\tilde{v}_i+\varepsilon) + c\underline{u}_i. \tag{3}$$

Next choose any $\varphi > 0$ sufficiently small so that

$$\underline{u}_i > (1-c)(\tilde{v}_i+\varepsilon) + c(\underline{u}_i+\varphi). \tag{4}$$

Finally, choose $t(\delta)$ such that $\lim_{\delta\to 1}\delta^{t(\delta)} = c$ and set $T = t(\delta)$; then it follows from (4) that we can find $\bar{\delta}$ sufficiently close to 1 so that (2) holds for all $\delta \geq \bar{\delta}$.[39]

The second difficulty in using **?** approach in our setting involves ensuring that in player i's minmax phase each player $j \neq i$ passes the test with probability arbitrarily close to 1 provided that he plays his part of i's minmax in each period. **?**'s setup with no discounting allows him to use **?** approachability results to establish such result. The **?** approachability result could be used in our setting with discounted payoff provided that T could be chosen independently of δ. Since for reasons explained before, we assume that T depends on δ such that $\lim_{\delta\to 1}\delta^{t(\delta)} < 1$, we cannot appeal to **?**'s approachability results to show that each $j \neq i$ passes his statistical test with a probability close to 1 by playing his part of i's minmax in

[38]The above problem arises because with finite memory one can not implement any arbitrary feasible payoff exactly. We can, of course, make φ vanish by letting the memory size become arbitrary large and then let δ go to 1. However, there is no guarantee that by taking the double limit one can find a sufficiently large memory size M and a sufficiently large discount factor δ such that φ is small enough for (2) to hold.

[39]While condition (2) is satisfied as long as $\lim_{\delta\to 1}\delta^{t(\delta)} \in (0,1)$, in the proof of Theorem 2, we assume that this limit is close to 1. We make this additional assumption in order to ensure that each player (a) prefers to punish than to be punished (condition (45) in the proof) and (b) finds it optimal to pass the statistical test in another player's minmax phase with a sufficiently high probability (condition (44) in the proof).

each period of i's minmax phase. Instead, we use **?** concentration results to establish the same conclusion in our setting.[40]

The third difficulty in using **?** approach in our setting arises because players may randomize during the minmax phases. This implies that some of the signalling phases may occur during the minmax phase (this possibility could not occur in the construction with pure strategies because in this case the minmax phase involves playing the same action profile a finite number of periods whereas the signalling phase involves playing the sequence of action profiles $(s', s', (s; l), s')$ for some l with $s'_i \neq s_i$ for all $i \in N$). To avoid this difficulty we modify the strategy profile so that whenever the sequence of actions played during i's minmax phase is such that a signalling sequence is "close" to happen, then all players other than i take actions to prevent such signalling sequence to occur. For example, if the signalling sequence is $(s', s', (s; i+1), s')$ then every j takes an action s'_j if the last $i+1$ action profiles are $(s', s', (s; i-1))$. While this specification of the strategy prevents signalling phases to occur during the minmax phases of any player i, it implies, in contrast with **?**, that players may not play "close" to i's minmax after some histories. This creates two potential issues that need to be dealt with. First, if such histories occur too frequently, then there is no guarantee that i's payoff in his minmax phase is close to his minmax payoff whenever all other players pass their statistical tests. To prevent such histories to occur frequently (more precisely, to make the fraction of times that such histories occur during the T periods of minmaxing sufficiently small), we make the signalling sequences sufficiently long by assuming that the signalling phase of player i's minmax phase is $((s'; 2), (s, Q+i+1), s')$ for some integer Q. By making Q sufficiently large we can make i's payoff in his minmax phase close to his minmax payoff, whenever all other players pass their statistical tests. Second, the statistical test that is performed at the end of the minmax phase of any player i has to be modified so that it only compares the behavior of all players other than i at histories at which they were not required to take actions to prevent the signalling sequence of i occurring.

Modifying the signalling phase and the minmax phases as proposed in the previous paragraph, however, in turn raises a fourth difficulty. In our pure bounded memory FT, each player was deterred from deviation from the signalling phase of his punishment path by the

[40]See **?** for more details and an approachability results for the discounting case.

threat of increasing the length of the minmax phase following the signalling phase by T periods for every deviation. This approach however does not work given the modifications we have proposed above. To see this, suppose that we fix Q and modify the strategy as proposed in the previous two paragraphs but still assume that each player is deterred from deviation from the signalling phase of his punishment path by the threat of increasing the length of the minmax phase as in the pure strategy case. Next, consider a player $i \in N$ and a history such that the play is in the last stage of the signalling sequence of i's punishment path and he has deviated $\theta - 1$ times from his signalling phase for some $\theta \in \{0, 1, \ldots, Q + i + 4\}$. Then the continuation strategy specifies a minmax phase of length θT followed by a reward path. A one-period deviation at this history induces a minmax phase of length $(\theta + 1)T$ followed by the reward path. For any θ', let $\tilde{v}_i(\theta', Q)$ refer to the average (across periods) expected payoff from the minmax phase with length $\theta' T$. Then, denoting the one-period gain from deviation by g_i and the average expected payoff from the reward phases for player i by u_i, to deter deviation the following must hold:

$$(1 - \delta^{\theta T+1})\tilde{v}_i(\theta, Q) + \delta^{\theta T+1} u_i - (1 - \delta^{(\theta+1)T+1})\tilde{v}_i(\theta + 1, Q) - \delta^{(\theta+1)T+1} u_i \geq (1 - \delta) g_i.$$

Since the limit of the RHS of the last expression as δ tends to 1 is zero, to ensure that the incentive to playing according to the signalling phase holds for large δ, we need the limit of the LHS to be positive:

$$(1 - \delta^{\theta T+1})(\tilde{v}_i(\theta, Q) - \tilde{v}_i(\theta + 1, Q)) + \delta^{\theta T+1}(1 - \delta^T)(u_i - \tilde{v}_i(\theta + 1, Q)) > 0.$$

While the difference $\tilde{v}_i(\theta, Q) - \tilde{v}_i(\theta + 1, Q)$ can be made arbitrarily small by making T and Q large, the last inequality may not hold because the difference $\tilde{v}_i(\theta, Q) - \tilde{v}_i(\theta + 1, Q)$ may not be zero and because θ can be chosen to be as large as $Q + i + 3$, and hence, $\delta^{\theta T+1}$ can become arbitrarily small for large values of Q (recall that $\lim_{\delta \to 1} \delta^T = c < 1$ and $\lim_{\delta \to 1} \delta^{\theta T+1} = c^\theta$ where θ can be chosen to be as large as $Q + i + 3$).

We solve the above difficulty by changing the punishment path for deviating from one's own signalling phase as follows. Since the interior of $\mathcal{U}$ is non-empty, there exists an action profile $\bar{a}^{(i)} \in A$ such that $u_i(m^i) < \max_{a_i \in A_i} u_i(a_i, \bar{a}^{(i)}_{-i}) = u_i(\bar{a}^{(i)})$.[41] The change involves

[41] If we assume that there exists an action profile $\bar{a}^{(i)} \in A$ such that $u_i(m^i) < \max_{a_i \in A_i} u_i(a_i, \bar{a}^{(i)}_{-i}) =$

keeping the length of the minmax phase constant at T independently of the number of deviations, and then introducing a deterrence phase of length $(\gamma+1)(Q+i+4)$, for some γ, after the signalling phase and before the start of the minmax phase. The deterrence phase consists of $Q+i+4$ (the length of the signalling phase) consecutive parts with each part consisting of $\gamma+1$ consecutive periods. In the first γ periods of any part $\tau=1,\ldots,Q+i+4$ of the deterrence phase the action profile $\bar{a}^{(i)}$ is played if player i has not deviated in the τ-th period of his signalling phase and m^i is played if player i has deviated in the τ-th period of his signalling phase. To avoid a signalling phase to occur, γ is chosen to be less than Q and, after such γ actions in the $(\tau+1)$-th period of the part, each player $j\neq i$ is required to play s'_j whereas player i plays a static best-reply to s'_{-i}. Since $u_i(m^i)<u_i(\bar{a}^{(i)})$, by choosing γ sufficiently large, we ensure that player i does not have any incentive to deviate during his own signalling phase.[42]

5.1 Difficulties in extending Theorem 2 to $n=2$

The proof of the Folk Theorem in Theorem 2 does not apply to the case with two players because s and s' need to be different from each other in at least 3 components; otherwise, confusing instances may arise (the reasoning is the same as that described in Subsection 4.3 for the pure strategy case). More generally, with two players detecting the identity of the deviator may not be possible. Hence, a common punishment path seems to be needed to deter deviations. In the pure FT with $n=2$, such common punishment consists of playing the pure mutual minmax profile for a finite number of times. Therefore, one solution would be to construct a strategy profile similar to that in Theorem 1 with $n=2$ except that the common punishment path involves playing action profiles that are *approximately* close to mixed mutually minmax profile for a finite number of periods T (henceforth called mutual minmax phase), a statistical test at the end of the mutual minmax phase that determines

$u_i(\bar{a}^{(i)})$, then, as mentioned in footnote 36, in Theorem 2 we could weaken the assumption that the interior of $\mathcal{U}$ in $\mathbb{R}^n$ is non-empty to the interior of $\tilde{\mathcal{U}}$ in $\mathbb{R}^n$ being non-empty.

[42]This procedure may also be used in the proof of Theorem 1 with $n>2$ to replace the feature adopted in the proof of increasing the length of the minmax phase in response to deviation made by player in the signaling phase of his own punishment path with this procedure.

whether or not each player have played sufficiently close to his part of the mutual minmax and a continuation strategy that returns to the equilibrium path if both players have passed the test and restarts the punishment phase otherwise. Thus one applies the method used in Theorem 2 for $n > 2$ to the proof of Theorem 1 with $n = 2$ by having a statistical test as in Theorem 2 at the end of the mutual minmax phase. Such test, however, seems to require players to identify the beginning of the mutual minmax phase. As in the proof of Theorem 2, we could deal with this issue by having a signalling phase/sequence at the beginning of the punishment path before the (common) mutual minmax phase, say $((s';2),(s;\ell),s')$ for some ℓ.[43] But the introduction of such signalling sequences introduces additional issues in addition to those in the proof of Theorem 2.

First, we must ensure that no signalling phase occurs during the mutual minmax phase. As mentioned in the previous subsection, to avoid a similar difficulty in the proof of Theorem 2 with $n > 2$ the equilibrium strategy profile is chosen such that whenever the sequence of actions played during i's minmax phase is such that a signalling sequence is "close" to happen, then all players other than i take actions to prevent such signalling sequence to occur. Specifically, in Theorem 2 the signalling phase chosen for i's minmax phase is $(s', s', (s; Q + i + 1), s')$ and if the last $Q + i + 2$ action profiles are $(s', s', (s; Q + i))$, then players other than the i take action profile s'_{-i} to prevent such signalling sequence to occur and i does a best response to s'_{-i}. With two players, such an approach requires the actions taken by the players to prevent the occurrence of the signalling sequence during the minmax phase to be best responses to each other. This is not an issue if the one shot game G has a Nash equilibrium in pure strategies because s' could be set to be equal to the Nash equilibrium and then one could appeal to the method used in the proof of Theorem 2 to ensure that the signalling phase do not occur during the mutual minmax phase. If the one shot game, however, does not have a pure Nash equilibrium then the method used in Theorem 2 need to be modified in order to ensure that signalling phases do not happen

[43]Such a signalling sequence for the mutual minmax phase would be different from the proof of our pure FT with $n = 2$ which requires players to play the pure mutual minmax profile by default (specifically, the players only play an action profile different from the pure mutual minmax profile if some specific histories have occurred).

during the mutual minmax phase.

Second, with a signalling phase before the mixed mutual minmax phase, we need to ensure that the players do not have an incentive to deviate during the signalling phase. In the proof of the FT with $n > 2$ in Theorem 2, we dealt with this problem by doing the following after any single player deviation by player i from the signalling phase of the punishment path of i: (i) continue with the signalling phase and consider any single player deviation by i from the signalling phase as being part of the generalised signalling phase, and (ii) punish the deviations after the signalling phase and before the start of the minmax phase by introducing a deterrence phase consisting of $Q + i + 4$ (the length of the signalling phase) parts of length $(\gamma + 1)$ for some γ. The deterrence phase in (ii) however also needs to be such that the signalling phase do not occur during the deterrence phase. To ensure this in the proof of Theorem 2 we assume that the last period of each part of the deterrence phase is such that all players other than i play s'_{-i} and player i plays a static best-reply to s'_{-i}. With $n = 2$, we face the same issue as that described in the previous paragraph.

With $n = 2$, a consequence of (i) is that if $(\bar{a}^1, \ldots, \bar{a}^{\ell+3})$ is a signalling phase then any sequence $(a^1, \ldots, a^{\ell+3}) \in A^{\ell+3}$ such that for each $1 \leq t \leq \ell + 3$, there exists $i \in \{1, 2\}$ such that $a_i^t = \bar{a}_i^t$ is a generalised signalling phase. This implies that at any history each player can always induce the signalling phase of the mutual minmax phase singly on his own. But then the strategy profile must be such that for each player the continuation payoff after any history is no less than what he can obtain starting from the signalling phase of the common mutual minmax. While this is not an issue at any history on the equilibrium path, ensuring that such a property holds at histories during the minmax phase is not easy to establish. For instance, if failure of passing the test results in restarting the punishment phase, then when during a mutual minmax phase a player is certain that the statistical test will fail before the end of the phase, it may be in the player's interest to deviate and restart the punishment phase sooner by playing the signalling phase rather than wait for the minmax phase to end.

If the length of the mutual minmax phase T is sufficiently long, deviations of this kind towards the end of the mutual minmax phase can be deterred provided we modify the strategy profile such that if the test passes then the continuation payoff is u and if the test fails then the continuation payoff is some $u' \in \tilde{\mathcal{U}}^0$ such that $u' < u$. This is because with large T

then the players would not want to deviate towards the end of the mutual minmax phase even when he is almost certain that the statistical test will fail by restarting the signalling sequence as no deviation guarantees at least u' whereas deviation results in the minmax phase of T periods before obtaining at least u'. However, such a modification of the strategy profile clearly may not be enough to deter deviations early in the mutual minmax phase.

Another approach to detering deviations during the minmax phase that involves restarts the signalling phase is to modify the construction of the play during the minmax phase. In paricular, the play in the minmax phase in Theorem 2 is based on following Gossner's approach of constructing a T-period horizon game, fixing any equilibrium of this finite game and playing the equilibrium in the minmax phase. By modifying the T-period game so that each player has an additional option of ending the T-period game early by starting the signalling phase, it may be possible to construct an equilibrium of the game such that no player would want to exercise the option of starting the signalling phase.

Third, in Theorem 2 with $n > 2$, an essential part of why during the minmax phase of any player i it is in the interest of every other player to approximately minmax player i is the following: in the T-period horizon game used in the construction of the play during the minmax phase of i, each player other than i can pass the test with a high probability by (mixed) minmaxing i irrespective of what others do; hence this must also be true in every equilibrium of the finite game and thus for the minmax phase of each player i. With $n = 2$ and a common punishment, each player can only pass the test and be rewarded with a high continuation payoff when both players pass the test. But then a player cannot guarantee that he can pass the test with a high probability. Indeed, if one player plays in the mutual minmax phase in such a way that he will fail the test with probability one, then the other player has an incentive to play static best-replies in each period and, thus, in general will not pass the test as well.

There may be several ways of dealing with this problem. One possible route may be to show that the T-period horizon game has an equilibrium in which both players pass the test with a high probability. We conjecture that this may be true because in such a game it seems that if one player is passing the test with sufficiently high probability then it is a best response for the other player to choose a strategy that passes the test with a high

probability. Another route is to have different continuation payoffs depending on who has passed the test. For example, the continuation payoff would be (approximately) u if both pass the test, and $u^i \in \tilde{\mathcal{U}}^0$ if i passes the test and $j \neq i$ fails and $u' \in \tilde{\mathcal{U}}^0$ if no player passes the test with $u_i \geq u_i^i > u_i^j \geq u_i'$.[44]

Given the current length of this paper and given the above issues we have not generalized Theorem 2 to the case of two players. We conjecture that such a generalisation is feasible, however, at this stage such an extension remains an open question.

A Proof of Theorem 1 (pure Folk Theorem)

For all $x \in \mathbb{R}^n$, let $||x|| = \max_{i=1,\ldots,n} |x_i|$. Since $\mathcal{U}$ is compact, it suffices to show that for all $\varepsilon > 0$ and all $u \in \mathcal{U}$, there exist $M \in \mathbb{N}$ and $\delta^* \in (0,1)$ such that for all $\delta \geq \delta^*$, there exists a M-memory SPE $f \in F^p$ of $G^\infty(\delta)$ with $\|U(f,\delta) - u\| < \varepsilon$. Furthermore, since $\mathcal{U}$ equals the closure of $\mathcal{U}^0$, we only need to show that the above holds for any $u \in \mathcal{U}^0$. Therefore, in the rest of this appendix, we show that for all $\varepsilon > 0$ and $u \in \mathcal{U}^0$, there exist $M \in \mathbb{N}$ and $\delta^* \in (0,1)$ such that for all $\delta \geq \delta^*$, there exists a M-memory SPE $f \in F^p$ of $G^\infty(\delta)$ with $\|U(f,\delta) - u\| < \varepsilon$.

A.1 2-player case

In this subsection, for convenience, we normalize payoffs so that $u_i(\bar{m}) = 0$ for both $i = 1, 2$.

Fix any $\varepsilon > 0$ and $u \in \mathcal{U}^0$. Let $0 < \eta < \min_{i=1,2}(u_i - v_i)$, $0 < \gamma < \min\{\eta/3, \varepsilon/2\}$ and $0 < \xi < 1$ be such that $2\xi < \eta - 2\gamma$.

Order $A = \{a^1, \ldots, a^{|A|}\}$ so that $a_i^1 \neq \bar{m}_i$ for all i, and $a^2 = \bar{m}$. Also, for any $k \in \mathbb{N}$, let

$$\mathcal{U}_k = \left\{ w \in \mathbb{R}^N : w = \sum_{a \in A} \frac{p^a u(a)}{k} \text{ for some } (p^a)_{a \in A} \text{ such that } \right.$$
$$\left. p^a \in \mathbb{N} \text{ for all } a, p^1 \geq 2, p^2 \geq 1 \text{ and } \sum_{a \in A} p^a = k \right\}.$$

[44] If such an approach works it may also help to solve some of the other difficulties mentioned above. For example, in Theorem 2, the deterrence phase after the signalling sequence is assumed in order to deter deviations during the signalling phase. Such a deterrence phase may not be needed if the continuation payoffs after the minmax phase could be made to depend on the number of deviations during the signalling phase (in addition to being dependent on the outcome of the tests).

Using an analogous argument to **?**, it follows that $\mathcal{U}_k$ converges to $\text{co}(u(A))$ in the Hausdorff distance. Therefore, there must exist $K \in \mathbb{N}$ such that

$$\text{co}(u(A)) \subseteq \cup_{x \in \mathcal{U}_K} B_\gamma(x), \tag{5}$$

where $B_\gamma(x)$ denotes the open ball of radius γ around x. Let $p^1, \ldots, p^{|A|}$ be such that $p^k \geq 0$ for all $1 \leq k \leq r$, $p^1 \geq 2$, $p^2 \geq 1$, $\sum_{k=1}^{|A|} p^k = K$ and

$$\left\| \sum_{k=1}^{|A|} \frac{p^k u(a^k)}{K} - u \right\| < \gamma. \tag{6}$$

Note that (5) implies that such a sequence $p^1, \ldots, p^{|A|}$ exists. Let $u' = \sum_{k=1}^{|A|} p^k u(a^k)/K$ and π consist of repetitions of the cycle $((a^1; p^1), \ldots, (a^{|A|}; p^{|A|}))$.

Let $T \in \mathbb{N}$ and $M \in \mathbb{N}$ be such that

$$T > K\left(\frac{B}{\xi} + 1\right), \text{ and} \tag{7}$$

$$M \geq 2T + K. \tag{8}$$

Also, let $\delta^* \in (0,1)$ be such that for all $\delta \in [\delta^*, 1)$

$$\min\left\{\xi \frac{\delta^K - \delta^T}{1 - \delta^K}, \xi\delta^T \frac{1 - \delta^{T+1}}{1 - \delta}, \delta^M(B + \xi)\right\} > B, \tag{9}$$

$$\sup_{(x^1,\ldots,x^K) \in [-B,B]^K} \left| \frac{1-\delta}{1-\delta^K} \sum_{k=1}^{K} \delta^{k-1} x^k - \frac{1}{K} \sum_{k=1}^{K} x^k \right| < \gamma. \tag{10}$$

Note that such $\delta^* \in (0,1)$ exists because the limit of the left hand side (9) and (10) as $\delta \to 1$ are, respectively, $\min\{\xi(T-K)/K, \xi(T+1), B+\xi\}$ and 0, and because, due to (7) and $0 < \xi < 1$, $\min\{\xi(T-K)/K, \xi(T+1), B+\xi\} > B$.

Fix any $\delta \geq \delta^*$. We will prove that there is a M-memory SPE f with $||U(f,\delta) - u|| < \varepsilon$.

Note that

$$||V(\pi,\delta) - u|| \leq ||V(\pi,\delta) - u'|| + ||u' - u|| < 2\gamma < \varepsilon, \tag{11}$$

where the second inequality follows from (10) and (6) and the third from the assumption that $\gamma < \varepsilon/2$. So it suffices to show that there is a (pure) M-memory SPE $f \in F^p$ with $\pi(f) = \pi$.

Before defining the strategy profile f, note the following properties of u' and $V^t(\pi, \delta)$. First, for all $i = 1, 2$,

$$u'_i > u_i - \gamma > v_i + \eta - \gamma > v_i + 2\xi, \text{ and} \tag{12}$$

$$V_i^t(\pi, \delta) > u'_i - \gamma > u_i - 2\gamma > v_i + \eta - 2\gamma > v_i + 2\xi \text{ for all } t \in \mathbb{N}. \tag{13}$$

(The first inequality in (13) follows from (10), the first in (12) and the second in (13) from (6), the second in (12) and the third in (13) since $\eta < u_i - v_i$ and the last inequality in both (12) and (13) because $2\xi < \eta - 2\gamma$).

Second, the following claim must hold.

Claim 1 *For all $i = 1, 2$, $t \in \mathbb{N}$ and $\delta \geq \delta^*$, $V_i^t(\pi, \delta) \geq \delta^T V_i(\pi, \delta)$.*

Proof. Fix any $i = 1, 2$, $t \in \mathbb{N}$ and $\delta \geq \delta^*$. Then, $V_i^t(\pi) = (1-\delta)\sum_{l=k}^{K} \delta^{l-k} u_i(\pi^l) + \delta^{K-k+1} V_i(\pi) \geq -B(1-\delta^{K-k+1}) + \delta^{K-k+1} V_i(\pi)$ for some $1 \leq k \leq K$. Hence, since $k \geq 1$, it follows that $V_i^t(\pi) \geq -B(1-\delta^K) + \delta^K V_i(\pi)$.

Therefore, it suffices to show that $(\delta^K - \delta^T) V_i(\pi) \geq B(1-\delta^K)$. This inequality holds since (13) and (9) imply that $(\delta^K - \delta^T) V_i(\pi) > (\delta^K - \delta^T)\xi > B(1-\delta^K)$. ■

A.1.1 The strategy profile

We define the desired strategy profile $f \in F^p$ as follows. For any $k \in \mathbb{N}$, $0 \leq k \leq M$, let

$$H_1^k = \{h \in H : T^k(h) = (\pi^1, \ldots, \pi^k)\}, \tag{14}$$

$$H_2^k = \{(\pi^1, \ldots, \pi^k)\} \text{ if } k > 0 \text{ and } H_2^k = \{H_0\} \text{ if } k = 0, \text{ and} \tag{15}$$

$$H_3^k = \{h \in H : T^M(h) = ((\bar{m}; M-k), \pi^1, \ldots, \pi^k)\}. \tag{16}$$

We additionally define

$$H^k = \begin{cases} H_1^k & \text{if } k \geq p^1 \\ H_2^k \cup H_3^k & \text{if } k < p^1, \end{cases} \tag{17}$$

$H^E = \cup_{k=0}^{M} H^k$ and $H^P = H \setminus H^E$. Then f is defined by

$$f(h) = \begin{cases} \pi^{k+1} & \text{if } h \in H^k \text{ for some } 0 \leq k \leq M, \\ \bar{m} & \text{otherwise.} \end{cases} \tag{18}$$

Claim 2 *The strategy profile $f \in F^p$ is a well defined M-memory strategy.*

Proof. By the definition of H_i^k, $i = 1, 2, 3$, the following must hold: (i) If $h \in H_1^k \cap H_1^{k'}$ for some $k > k' \geq p^1$, then it must be that $k = k' + \alpha K$ for some $\alpha \in \mathbb{N}$, implying that $\pi^{k+1} = \pi^{k'+1}$. (ii) For any $k \geq p^1$ and $k' < p^1$, $H_1^k \cap H_2^{k'} = \emptyset$ and $H_1^k \cap H_3^{k'} = \emptyset$ (if the latter were not to hold, we would have $\pi^1 = \bar{m}$, a contradiction). (iii) For any $k, k' < p^1$, $k \neq k'$, $H_i^k \cap H_j^{k'} = \emptyset$ for any $i, j \in \{2, 3\}$. It then follows from (i)–(iii) that f is well-defined.

Finally, note that f is a M-memory strategy because its definition is such that $f(h)$ depends only on $T^M(h)$ for all $h \in H$. ■

A.1.2 Outcome paths induced by f and by one-shot deviations from f

The next two claims establish the continuation paths f induces after any history.

Claim 3 *If $h \in H^k$ for some $0 \leq k \leq M$, then $\pi(f|h) = (\pi^{k+1}, \pi^{k+2}, \ldots)$.*

Proof. We prove this in several steps.

Step 1: If $h \in H_1^k$ and $p^1 \leq k \leq M$, then $h \cdot f(h) \in H_1^{k'+1}$ for some k' such that $p^1 \leq k' \leq M$ and $k = \alpha K + k'$ for some $\alpha \in \mathbb{N}$. Suppose that $p^1 \leq k \leq M$ and $h \in H_1^k$. Then we must have that $T^k(h) = (\pi^1, \ldots, \pi^k)$ and $f(h) = \pi^{k+1}$. This implies that $T^{k+1}(h \cdot f(h)) = (\pi^1, \ldots, \pi^{k+1})$. If $k < M$, the claim of this step holds because $(h \cdot \pi^{k+1}) \in H_1^{k+1}$ and $p^1 \leq k + 1 \leq M$. If $k = M$, then $M \geq 2K$ implies that $T^M(h \cdot f(h)) = (\pi^2, \ldots, \pi^{k+1}) = (\pi^2, \ldots, \pi^K, \pi^1, \ldots \pi^{k-K+1})$ with $k - K + 1 = M - K + 1 > p^1$. Hence, the claim of this step holds because $h \cdot f(h) = h \cdot \pi^{k-K+1} \in H_1^{k-K+1}$ and $k - (k - K) = K$.

Step 2: If $h \in H_1^k$ and $p^1 \leq k \leq M$, then $\pi(f|h) = (\pi^{k+1}, \pi^{k+2}, \ldots)$. This follows by induction from Step 1 and by noting that $\pi^{k'+1} = \pi^{k+1}$ if $k = \alpha K + k'$ for some $\alpha \in \mathbb{N}$.

Step 3: If $h \in H_2^k \cup H_3^k$ and $0 \leq k < p^1$, then $\pi(f|h) = (\pi^{k+1}, \pi^{k+2}, \ldots)$. If $h \in H_2^k \cup H_3^k$ and $0 \leq k < p^1$, then by induction, f induces the outcome $(\pi^{k+1}, \ldots, \pi^{p^1})$ after h. But since $h \cdot (\pi^{k+1}, \ldots, \pi^{p^1}) \in H_1^{p^1}$, the claim of this step follows from Step 2. ■

It follows trivially from Claim 3 that $\pi(f) = (\pi^1, \pi^2, \ldots)$. Hence, f implements π.

Claim 4 *If $h \in H^P$ and $k = \max\{0 \leq k' \leq M : T^{k'}(h) = (\bar{m}; k')\}$, then $k < M$ and $\pi(f|h) = ((\bar{m}; M - k), \pi^1, \pi^2, \ldots)$.*

Proof. Fix any $h \in H^P$ and let k be as defined above.

Step 1: $k < M$. Otherwise, $k = M$ and $T^M(h) = (\bar{m}; M)$ producing a contradiction because then $h \in H_3^0 \subseteq H \setminus H^P$.

Step 2: If $h \cdot (\bar{m}; l-1) \in H^P$ for some $l \in \{1, \ldots, M-k-1\}$, then $h \cdot (\bar{m}; l) \in H^P$. Suppose not; then $h \cdot (\bar{m}; l-1) \in H^P$ and $h \cdot (\bar{m}; l) \in H^{k'}$ for some $0 \leq k' \leq M$. Since $a^1 \neq \bar{m}$ and for any $\tau \leq p^1$ we have that $(\pi^1, \ldots, \pi^\tau) = (a^1; \tau)$, it follows from $h \cdot (\bar{m}; l) \in H^{k'}$ that either $h \cdot (\bar{m}; l) \in H_1^{k'}$ and $k' \geq p^1$ or $T^M(h \cdot (\bar{m}; l)) = (\bar{m}; M)$. But the latter is not possible because we have by assumption $l < M - k$ (in fact, if $T^M(h \cdot (\bar{m}; l)) = (\bar{m}; M)$, then $T^{M-l}(h) = (\bar{m}; M-l)$ and so $k \geq M - l$); therefore, consider the former case. Then $T^{k'-1}(h \cdot (\bar{m}; l-1)) = (\pi^1, \ldots, \pi^{k'-1})$. Since $h \cdot (\bar{m}; l-1) \in H^P$, it must be that $k' - 1 < p^1$. Hence, $k' = p^1$, $k' - 1 = p^1 - 1 \geq 1$ and $\bar{m} = \pi^{k'-1} = a^1$; but this is a contradiction.

Step 3: $h \cdot (\bar{m}; l) \in H^P$ for all $l = 0, \ldots, M-k-1$. Since $h \in H^P$ and $f(h') = \bar{m}$ for all h'^P, this step follows by induction from the previous step.

Step 4: $\pi(f|h) = ((\bar{m}; M-k), \pi^1, \pi^2, \ldots)$. By the previous step, f results in $(\bar{m}; M-k)$ after h. Since $T^M(h \cdot (\bar{m}; M-k)) = (\bar{m}; M) \in H_3^0$, it then follows from Claim 3 that $\pi(f|h) = ((\bar{m}; M-k), \pi^1, \pi^2, \ldots)$. ■

The following three claims characterize the consequences of a single deviation by one player from f.

Claim 5 *If $h \in H^E$, $a_i \neq f_i(h)$ and $a_{-i} = f_{-i}(h)$ for some $i \in \{1, 2\}$, then $h \cdot a \in H^P$.*

Proof. Suppose not; then $h \in H^E$, $a_i \neq f_i(h)$, $a_{-i} = f_{-i}(h)$ for some $i \in \{1, 2\}$ and $h \cdot a \in H^k$ for some $0 \leq k \leq M$. There are three different cases to consider.

Case 1: $h \cdot a = (\pi^1, \ldots, \pi^k) \in H_2^k$ for some $k < p^1$. Then we must have $a = \pi^k$, $h \in H_2^{k-1}$ and $k - 1 < p^1$. But then $f(h) = \pi^k = a$; a contradiction.

Case 2: $h \cdot a \in H_1^k$ for some $k \geq p^1$. Then $T^k(h \cdot a) = (\pi^1, \ldots, \pi^k)$, $a = \pi^k$ and $T^{k-1}(h) = (\pi^1, \ldots, \pi^{k-1})$. If $k > p^1$, then $h \in H_1^{k-1}$ and $f(h) = \pi^k = a$; a contradiction. Thus, $k = p^1$, $a = \pi^k = a^1$ and $T^{p^1-1}(h) = (a^1; p^1 - 1)$. Also, by construction $p^1 - 1 \geq 1$. Therefore, it follows from the construction of π (a^1 is followed by $a^2 = \bar{m}$) and the definition of f that $f(h) = a^1$ or $f(h) = \bar{m}$. Thus, either $f(h) = a$ or $f_j(h) \neq a_j$ for all $j = 1, 2$. But both cases contradict our initial supposition that $a_i \neq f_i(h)$ and $a_{-i} = f_{-i}(h)$.

Case 3: $h \cdot a \in H_k^3$ for some $0 \leq k < p^1$. If $k = 0$, then $T^M(h \cdot a) = (\bar{m}; M)$, $a = \bar{m}$ and $T^M(h) = (a', (\bar{m}; M-1))$ for some $a' \in A$. But, since $h \in H^E$, it must also be that $a' = \bar{m}$. Thus, $T^M(h) = (\bar{m}; M)$ and $f(h) = a^1$. But this is a contradiction because it implies that $a_{-i} = \bar{m}_{-i} \neq a^1_{-i} = f_{-i}(h)$. Hence, it must be that $k > 0$. Then $a = a^1$ and

$T^M(h) = (a', (\bar{m}; M-k), (a^1; k-1))$ for some $a' \in A$. Since $k-1 < p^1$, $h \in H^E$ implies that $a' = \bar{m}$, and thus, $T^M(h) = ((\bar{m}; M-(k-1)), (a^1; k-1))$. But this is a contradiction because it implies that $f(h) = a^1 = a$. ■

Claim 6 *If $h \in H^E$ and $a_i \neq f_i(h)$ and $a_{-i} = f_{-i}(h)$ for some $i \in \{1,2\}$, then*

$$\pi(f|h \cdot a) = \begin{cases} ((\bar{m}; M), \pi^1, \pi^2, \ldots) & \text{if } a \neq \bar{m}, \\ ((\bar{m}; M-1), \pi^1, \pi^2, \ldots) & \text{if } a = \bar{m} \text{ and } T^1(h) \neq \bar{m}, \\ ((\bar{m}; M-p^2-1), \pi^1, \pi^2, \ldots) & \text{if } a = T^1(h) = \bar{m}. \end{cases} \tag{19}$$

Proof. By Claim 5, $h \cdot a \in H^P$. Therefore, it follows from Claim 4 that $\pi(f|h \cdot a) = ((\bar{m}; M-k), \pi^1, \pi^2, \ldots)$, where $k = \max\{0 \leq k' \leq M : T^{k'}(h) = (\bar{m}; k')\}$. This means that $\pi(f|h \cdot a) = ((\bar{m}; M), \pi^1, \pi^2, \ldots)$ if $a \neq \bar{m}$ and $\pi(f|h \cdot a) = ((\bar{m}; M-1), \pi^1, \pi^2, \ldots)$ if $a = \bar{m}$ and $T^1(h) \neq \bar{m}$. Finally, consider the case $a = T^1(h) = \bar{m}$. Since $f_{-i}(h) = a_{-i} = \bar{m}_{-i} \neq a^1_{-i}$, we have $f(h) \neq a^1$. This rules out the possibility that $h \in H_2^{k'} \cup H_3^{k'}$ for some $k' < p^1$. Therefore, since $h \in H^E$, it must be that $T^{k'}(h) = (\pi^1, \ldots, \pi^{k'})$ for some $k' \geq p^1$. Also, $\pi^{k'} = T^1(h) = \bar{m}$ and $\pi^{k'+1} = f(h) \neq a = \bar{m}$; therefore, we must have $k' = p^1 + p^2$. But this implies that $k = p^2 + 1$. Hence, we have $\pi(f|h \cdot a) = ((\bar{m}; M-p^2-1), \pi^1, \pi^2, \ldots)$. ■

Claim 7 *If $h \in H^P$, $a_i \neq f_i(h)$ and $a_{-i} = f_{-i}(h)$ for some $i \in \{1,2\}$, then $h \cdot a \in H^P$ and $\pi(f|h \cdot a) = ((\bar{m}; M), \pi^1, \pi^2, \ldots)$.*

Proof. It follows from $h \in H^P$ that $f(h) = \bar{m}$. Thus, $a \neq \bar{m}$ and $a \neq a^1$. We will next prove that $h \cdot a \in H^P$ by showing that $h \cdot a \notin H^k$ for any $0 \leq k \leq M$: First, since $\pi^k = a^1$ for any $k < p^1$, $a \neq a^1$ implies that $h \cdot a \notin H_2^k$ for any $k < p^1$. Second, $h \cdot a \notin H_3^k$ for any $0 \leq k < p^1$ because otherwise $a = \bar{m}$ (if $k = 0$) or $a = a^1$ (if $k > 0$); a contradiction. And third, if $h \cdot a \in H_1^k$ for some $k \geq p^1$ then $\pi^k = a \neq \bar{m}$ and $\pi^k = a \neq a^1$. This implies that $k > p^1 + p^2$. Hence, $h \in H_1^{k-1}$ for some $k - 1 \geq p^1$; but this contradicts $h \in H^P$.

It follows from above that $h \cdot a \in H^P$. Since $a \neq \bar{m}$, it follows from Claim 4 that $\pi(f|h \cdot a) = ((\bar{m}; M), \pi^1, \pi^2, \ldots)$. ■

A.1.3 Incentive conditions

Claim 8 *The strategy profile $f \in F^p$ is SPE.*

Proof. We demonstrate this result by showing that one-shot deviations are not profitable at any history.

Fix any player i, any $h \in H$ and any strategy $g_i \in F_i$ that only differs from f_i at h; thus, $g_i(h) \neq f_i(h)$ and $g_i(h') = f_i(h')$ for all $h' \in H \setminus \{h\}$. We need to show that $U_i(f|h) \geq U_i(g_i, f_{-i}|h)$. To show this consider the two possible cases.

Case 1: $h \in H^k$ for some $0 \leq k \leq M$. In this case, by Claim 3 and Claim 6 respectively, $\pi(f|h) = (\pi^{k+1}, \pi^{k+2}, \ldots)$ and $\pi(g_i, f_{-i}|h) = ((a_i, \pi_{-i}^{k+1}), (\bar{m}; t), \pi^1, \pi^2, \ldots)$ for some $a_i \in A_i$ and $t \geq M - (p^2 + 1)$. Then we have

$$\begin{aligned} &U_i(f|h) - U_i(g_i, f_{-i}|h) = V_i^{k+1}(\pi) - [(1-\delta)u_i(a_i, \pi_{-i}^{k+1}) + \delta V_i((\bar{m}; t) \cdot \pi)] \geq \\ &V_i^{k+1}(\pi) - [(1-\delta)B + \delta^{2T+1} V_i(\pi)] \geq \delta^T (1 - \delta^{T+1}) V_i(\pi) - (1-\delta)B \end{aligned} \tag{20}$$

where the three inequalities in the above follow, respectively, from $u_i(a_i, \pi_{-i}^{k+1}) \leq B$, $u_i(\bar{m}) = 0$, $t \geq M - (p^2+1) \geq M - K \geq 2T$ (the inequality $M - K \geq 2T$ following from (8)) and Claim 1. By (13), we have $V_i(\pi) > v_i + \xi \geq \xi$. By (9), we have $\delta^T(1 - \delta^{T+1})\xi > (1-\delta)B$. Therefore, it follows from (20) that $U_i(f|h) - U_i(g_i, f_{-i}|h) > 0$.

Case 2: $h \in H^P$. In this case, by Claim 4 and Claim 7 respectively, $\pi(f|h) = ((\bar{m}; M - t), \pi^1, \pi^2, \ldots)$ for some $0 \leq t < M$ and $\pi(g_i, f_{-i}|h) = ((a_i, \bar{m}_{-i}), (\bar{m}; M), \pi^1, \pi^2, \ldots)$ for some $a_i \in A_i$. Since $u_i(\bar{m}) = 0$ and $u_i(a_i, \bar{m}_{-i}) \leq \max_{a_i' \in A_i} u_i(a_i', \bar{m}_{-i}) = v_i$, we must then have

$$\begin{aligned} &U_i(f|h) - U_i(g_i, f_{-i}|h) \geq \delta^{M-t} V_i(\pi) - [(1-\delta)v_i + \delta^{M+1} V_i(\pi)] \geq \\ &\delta^M V_i(\pi) - [(1-\delta)v_i + \delta^{M+1} V_i(\pi)] \geq (1-\delta)(\delta^M V_i(\pi) - v_i). \end{aligned} \tag{21}$$

By (13), we have that $V_i(\pi) > v_i + 2\xi$. By (9), we have $\delta^M > \frac{B}{B+\xi} \geq \frac{v_i}{v_i + \xi}$. Therefore, $\delta^M V_i(\pi) - v_i > 0$. But then, by (21), we have $U_i(f|h) - U_i(g_i, f_{-i}|h) > 0$. ■

A.2 More than 2-player case

In this subsection, for convenience, we normalize payoffs so that $v_i = 0$ for all $i \in N$.

Fix any $\varepsilon > 0$ and any $u \in \mathcal{U}^0$. Then, by Theorem 1 (Step 1) in **?**, for all $i \in N$ there exists $y^i \in \mathcal{U}^0$ satisfying the following property: for some $0 < \zeta' < \min_i y_i^i$, $y_i^i + \zeta' < u_i$ and $y_i^i + \zeta' < y_i^j$ for all $j \in N$ with $j \neq i$. Define $\xi > 0$ to be such that $2\xi < \varepsilon$ and $4\xi < \zeta'$ and $\zeta = \zeta' - 4\xi$.

Fix any s and s', both in A, such that $s_i \neq s_i'$ for all $i \in N$. For all $k \in \mathbb{N}$, let $\mathcal{V}_k$ be the set of $u' \in \text{co}(u(A))$ such that $u' = \sum_{a \in A} p^a u(a)/k$ for some $\{p^a\}_{a \in A}$ satisfying $p^a \in \mathbb{N}$ and $p^a \geq 2n+2$ for all $a \in A \setminus \{s', s\}$, $p^{s'} \geq 3$, $p^s \geq 4n+4$ and $\sum_{a \in A} p^a = k$. Using an analogous

argument to **?**, it follows that $\mathcal{V}_k$ converges to $\text{co}(u(A))$. Therefore, there must exist $K \in \mathbb{N}$ such that

$$\text{co}(u(A)) \subseteq \cup_{x \in \mathcal{V}_K} B_\xi(x). \tag{22}$$

For all $\hat{a} \in A$ and $j \in N$, let $D_j(\hat{a}) = \{a \in A : a_{-j} = \hat{a}_{-j}\}$ and $\bar{D}_j(\hat{a}) = D_j(\hat{a}) \backslash \{\hat{a}\}$. Define $D(\hat{a}) = \cup_{j \in N} D_j(\hat{a})$ and $\bar{D}(\hat{a}) = \cup_{j \in N} \bar{D}_j(\hat{a})$. Order all the actions in $A = \{a^1, \ldots, a^{|A|}\}$ as follows: $a^1 = s$, $a^2, \ldots, a^{|\bar{D}(s)|+1}$ are the different elements $\bar{D}(s)$, in any order, $a^{|\bar{D}(s)|+2} = s'$, $a^{|\bar{D}(s)|+3}, \ldots, a^{|\bar{D}(s)|+|\bar{D}(s')|+2}$ are the different elements of $\bar{D}(s')$, in any order, and all the remaining actions are then ordered arbitrarily.

To simplify notation, we also denote $y^0 = u$. For all $i \in \{0, \ldots, n\}$, let $x^i \in \mathcal{V}_K$ be such that $||x^i - y^i|| < \xi$ and $\{p^i_a\}_{a \in A}$ be such that $\frac{1}{K} \sum_{a \in A}^K p^i_a u_j(a) = x^i_j$, for all $j \in N$. For all $i \in \{0, \ldots, n\}$, define $\hat{\pi}^{(i)}$ as the repetition of the cycle

$$\left((s'; 2), (s; n+i+2), s', (a^1; p^{(i),1}), \ldots, (a^{|A|}; p^{(i),|A|})\right),$$

where $p^{(i),j} = p^i_{a^j} - 3$ if $a^j = s'$, $p^{(i),j} = p^i_{a^j} - (n+i+2)$ if $a^j = s$ and $p^{(i),j} = p^i_{a^j}$ otherwise. Note that the length of the cycle is K, i.e., $\sum_{j=1}^r p^{(i),j} + n + 5 + i = K$ for all $i \in \{0, \ldots, n\}$. In the construction below, $\hat{\pi}^{(i)}$ will be the equilibrium path when $i = 0$ (also sometimes denoted by $\pi^{(0)}$) and the reward path of player i when $i > 0$.

Let $T \in \mathbb{N}$ be such that

$$T > 2 \max\left\{(K+n+6)\frac{B}{\zeta}, K\right\}. \tag{23}$$

Also let $\pi^{(i)} = \left((s'; 2), (s; i+1), s', (m^i; T), \hat{\pi}^{(i)}\right)$ and

$$\pi^{(i)}(\theta, t) = \begin{cases} \left(\pi^{(i),t}, \ldots, \pi^{(i),i+4}, (m^i; (\theta+1)T), \hat{\pi}^{(i)}\right) & \text{if } t \leq i+4 \text{ and} \\ \left((m^i; (\theta+1)T), \hat{\pi}^{(i)}\right) & \text{if } t = i+5, \end{cases}$$

for any $i \in N$, $\theta \in \mathbb{N}_0$ and $t \in \{1, \ldots, i+5\}$. Define the size of the memory $M \in \mathbb{N}$ be such that

$$M \geq T(n+5) + (n+4) + (2n+5) + (n+4) = 4n + 13 + T(n+5).$$

Also, let $\delta^* \in (0,1)$ be such that $\delta \geq \delta^*$ implies

$$\min\left\{\frac{\delta^K - \delta^{n+6+T}}{2 - \delta^K - \delta^{n+6}}, \frac{\delta^{n+5+(n+4)T}(1-\delta^T)}{2(1-\delta^{n+5})}, \frac{\delta^{(n+5)(T+1)}}{2 - 2\delta^{(n+5)(T+1)} - \delta^{n+6}(1-\delta^T)}\right\} > \frac{B}{\zeta} \tag{24}$$

$$\sup_{x \in [-B,B]^K} \left| \frac{1-\delta}{1-\delta^K} \sum_{k=1}^K \delta^{k-1} x^k - \frac{1}{K} \sum_{k=1}^K x^k \right| < \xi. \tag{25}$$

Note that such $\delta^* \in (0,1)$ exists because the limit of the left hand side of (24) and (25) as $\delta \to 1$ are, respectively, $\min\{(T+n+6-K)/(K+n+6), T/2(n+5)\}$ and 0, and the former limit exceeds B/ζ by (23).

Fix any $\delta \geq \delta^*$. We will now demonstrate the result by constructing a (pure) M-memory SPE strategy profile $f \in F^p$ with $||U(f) - u|| < \varepsilon$.

Note that

$$||V(\hat{\pi}^{(0)}, \delta) - u|| \leq ||V(\hat{\pi}^{(0)}, \delta) - x^0|| + ||x^0 - u|| < 2\xi < \varepsilon,$$

where the second inequality follows from (25) and the definition of x^0 and the third from $\xi < \varepsilon/2$. So it suffices to show that there is a (pure) M-memory SPE $f \in F^p$ with $\pi(f) = \hat{\pi}^{(0)}$.

A.2.1 The strategy profile

For all $\tau \in \mathbb{N}$ and $d \in N$, define

$$\begin{aligned}\Sigma^{d,\tau} = \; & \big\{h \in H : h = (a^t)_{t=1}^{\tau} \text{ such that } a^t \in D_d(s) \text{ if } t = 3, \ldots, d+3 \\ & \text{and } a^t \in D_d(s') \text{ if } t = 1, 2, d+4\big\}\end{aligned}$$

and $\Sigma^{d,0} = \{H_0\}$ for all $d \in N$. Also, for all $\tau \geq d+4$ and all $h \in \Sigma^{d,\tau}$, let

$$\theta(h) = |\{t \in \{1, 2, d+4\} : a_d^t \neq s_d'\}| + |\{t \in \{3, \ldots, d+3\} : a_d^t \neq s_d\}|.$$

For all $d \in N$, define $\Gamma^{d,0} = \{H_0\}$ and, for all $\tau \in \mathbb{N}$,

$$\Gamma^{d,\tau} = \left\{h \in H : h = (a^t)_{t=1}^{\tau} \text{ and } a^t \in D_d(m^d) \text{ for all } 1 \leq t \leq \tau\right\}.$$

Define for all $k \in \{1,\ldots,M\}$, $i \in \{0,\ldots,n\}$, $d \in N$, and $\tau, r \in \mathbb{N}_0$ the following sets:[45]

$$
\begin{aligned}
H_{1,a}^{(i),k} = &\ \left\{h \in H : T^k(h) = \left(\hat{\pi}^{(i),1},\ldots,\hat{\pi}^{(i),k}\right)\right\},\\
H_{1,b}^{(i),k} = &\ \left\{h \in H : h = \left(\hat{\pi}^{(i),1},\ldots,\hat{\pi}^{(i),k}\right)\right\},\\
H_1^{(i),k} = &\ H_{1,a}^{(i),k} \cup H_{1,b}^{(i),k},\\
H_2^{k,d,\tau} = &\ \Big\{h \in H : T^k(h) = \bar{h}\cdot a\cdot \tilde{h} \text{ such that for some } k' \le k \text{ and } i \in \{0,\ldots,n\}\\
&(1)\quad \text{either } \bar{h} \in H_{1,a}^{(i),k'} \text{ with } k' \ge n+i+5 \text{ or } \bar{h} \in H_{1,b}^{(i),k'} \text{ with } \ell(h) = k,\\
&k' < n+5 \text{ and } i = 0, (2)\quad a \in \bar{D}_d(\hat{\pi}^{(i),k'+1}),\quad (3)\quad \tilde{h} \in \Sigma^{d,\tau} \text{ and}\\
&(4)\quad \text{if } T^{d+3}(\bar{h}\cdot a) = ((s';2),(s;d),a) \text{ and } a \in \bar{D}_d(s),\ \text{then } \ell(\tilde{h}) = 0\Big\},\\
H_3^{k,d} = &\ \Big\{h \in H : T^k(h) = \bar{h}\cdot\tilde{h} \text{ such that}\quad (1)\quad \bar{h} \in \Sigma^{d,d+4} \text{ and}\\
&(2)\quad \tilde{h} \in \Gamma^{d,l} \text{ for some } 0 \le l < (\theta(\bar{h})+1)T\Big\},\\
H_4^{k,d,r} = &\ \Big\{h \in H : T^k(h) = \bar{h}\cdot\hat{h}\cdot\tilde{h} \text{ such that } (1)\quad \bar{h} \in \Sigma^{d,d+4},\\
&(2)\quad \hat{h} \in \Gamma^{d,l} \text{ with } l = (\theta(\bar{h})+1)T \text{ and } (3)\quad \tilde{h} \in H_{1,b}^{(d),r}\Big\},\\
H_5^{k,d,\tau} = &\ \Big\{h \in H : T^k(h) = \bar{h}\cdot a\cdot\tilde{h} \text{ such that for some } k' \le k \text{ and } i \in N\\
&(1)\quad \text{either } \bar{h} \in H_3^{k',i}, a \in \bar{D}_d(m^i) \text{ and } d \ne i\\
&\text{or } \bar{h} \in H_4^{k',i,r} \text{ and } a \in \bar{D}_d(\hat{\pi}^{(i),r+1}) \text{ for some } r < n+i+5,\quad (2)\quad \tilde{h} \in \Sigma^{d,\tau} \text{ and}\\
&(3)\quad \text{if } T^{d+3}(\bar{h}\cdot a) = ((s';2),(s;d),a) \text{ and } a \in \bar{D}_d(s),\ \text{then } \ell(\tilde{h}) = 0\Big\}.
\end{aligned}
$$

We next define $H_{1,a} = \cup_{i=0}^{n}\left(\cup_{k=n+i+5}^{M} H_{1,a}^{(i),k}\right)$, $H_{1,b}^{(0),0} = \{H_0\}$, $H_{1,b} = \cup_{k=0}^{n+4} H_{1,b}^{(0),k}$, $H_1 = H_{1,a}\cup H_{1,b}$, $H_2 = \cup_{k=1}^{M}\left(\cup_{d\in N}\left(\cup_{\tau=0}^{d+3} H_2^{k,d,\tau}\right)\right)$, $H_3 = \cup_{k=1}^{M}\left(\cup_{d\in N} H_3^{k,d}\right)$, $H_4 = \cup_{k=1}^{M}\left(\cup_{d\in N}\left(\cup_{r=0}^{n+d+4} H_4^{k,d,r}\right)\right)$, and $H_5 = \cup_{k=1}^{M}\left(\cup_{d\in N}\left(\cup_{\tau=0}^{d+3} H_5^{k,d,\tau}\right)\right)$.

Let $\tilde{\Sigma}^{d,\tau} = \{h \in H : T^{\tau+1}(h) = a\cdot\tilde{h}, \tilde{h} \in \Sigma^{d,\tau} \text{ and } a \in \bar{D}_d(s)\cup\bar{D}_d(s')\}$ for all $d \in N$ and $\tau \in \mathbb{N}_0$. Define, for all $d \in N$ and $\tau \in \{0,\ldots,d+3\}$,

$$H_6^{d,\tau} = \left(H \setminus \cup_{l=1}^{5} H_l\right) \cap \tilde{\Sigma}^{d,\tau}. \tag{26}$$

Let $H_6 = \cup_{d\in N}\left(\cup_{\tau=0}^{d+3} H_6^{d,\tau}\right)$. Also, for all $t \in \{0,\ldots,n+4\}$, define

$$H_7^t = \{h \in H \setminus \cup_{l=1}^{6} H_l : T^t(h) \in H_{1,b}^{(0),t}\}. \tag{27}$$

[45]Note that, for some of these parameters, the sets below may be empty in some cases.

The strategy $f \in F^p$ is now defined as follows: For any $h \in H$,

$$f(h) = \begin{cases} \hat{\pi}^{(0),k+1} & \text{if } h \in H_{1,b}^{(0),k} \text{ for some } k \in \{0,\ldots,n+4\}, \\ \hat{\pi}^{(i),k+1} & \text{if } h \in H_{1,a}^{(i),k} \text{ for some } i \in \{0,\ldots,n\} \text{ and } k \in \{n+i+5,\ldots,M\}, \\ s & \text{if } h \in \left(\cup_{k=1}^{M} \cup_{d\in N} \cup_{\tau=2}^{d+2} (H_2^{k,d,\tau} \cup H_5^{k,d,\tau} \cup H_6^{d,\tau})\right) \cup \left(\cup_{t=2}^{n+3} H_7^t\right), \\ m^d & \text{if } h \in \cup_{k=1}^{M} H_3^{k,d} \text{ for some } d \in N, \\ \hat{\pi}^{(d),r+1} & \text{if } h \in \cup_{k=1}^{M} H_4^{k,d,r} \text{ for some } d \in N \text{ and } r \in \{0,\ldots,n+d+4\}, \\ s' & \text{otherwise.} \end{cases} \tag{28}$$

Clearly, f is M-memory. In the supplementary materials, we show that f is well-defined.

To enable us to show that f is a SPE and $\pi(f) = \hat{\pi}^{(0)}$, we first need to describe the outcome paths induced by strategy profile f and those obtained by one-shot deviations from f. Claims 9-19 describe these outcome paths while their proofs are presented in the supplementary materials, Section A.3.

Claim 9 *If $h \in H_{1,a}^{(i),k}$ for some $i \in \{0,\ldots,n\}$ and $k \in \{n+i+5,\ldots,M\}$, then $\pi(f|h) = (\hat{\pi}^{(i),k+1}, \hat{\pi}^{(i),k+2}, \ldots)$.*

Claim 10 *If $h \in H_{1,b}^{(0),k}$ for some $k \in \{0,\ldots,n+4\}$, then $\pi(f|h) = (\pi^{(0),k+1}, \pi^{(0),k+2}, \ldots)$.*

Claim 11 *If $h \in H_4^{k,d,r}$ for some $k \in \{1,\ldots,M\}$, $d \in N$ and $r \in \{0,\ldots,n+d+4\}$, then $\pi(f|h) = (\hat{\pi}^{(d),r+1}, \hat{\pi}^{(d),r+2}, \ldots)$.*

Claim 12 *If $h \in H_3^{k,d}$ for some $k \in \{1,\ldots,M\}$ and $d \in N$, then $\pi(f|h) = ((m^d; (\theta+1)T - k + d + 4), \hat{\pi}^{(d),1}, \hat{\pi}^{(d),2}, \ldots)$ where $\theta = \theta(T^k(h))$.*

Claim 13 *Let $h \in H_2^{k,d,\tau} \cup H_5^{k,d,\tau} \cup H_6^{d,\tau}$ for some $k \in \{1,\ldots,M\}$, $d \in N$ and $\tau \in \{0,\ldots,d+3\}$, and $\bar{a} \in D_d(f(h))$. Then there exists $\bar{\theta}(\bar{a}) \in \{0,\ldots,d+4\}$ and $t(\bar{a}) \in \{1,\ldots,d+5\}$ such that $\pi(f|h \cdot \bar{a}) = \pi^{(d)}(\bar{\theta}(\bar{a}), t(\bar{a}))$ and $\bar{\theta}(f(h)) < \bar{\theta}(\bar{a})$ for any $\bar{a} \in \bar{D}_d(f(h))$.*

Claim 14 *Let $h \in H_7^k$ for some $k \in \{0,\ldots,n+4\}$. If $h \notin \cup_{k=2}^{n+3} H_7^k$ and $T^{d+3}(h) \in \Sigma^{d,d+3}$ for some $d \in N$, then $\pi(f|h) = \left(s', (m^d; (\theta+1)T), \hat{\pi}^{(d),1}, \ldots\right)$ where $\theta = \theta(T^{d+3}(h) \cdot s')$. Otherwise, $\pi(f|h) = \left(\hat{\pi}^{(i),k'+1}, \hat{\pi}^{(i),k'+2}, \ldots\right)$ for some $i \in \{0,\ldots,n\}$ and $k' \in \{0,\ldots,n+i+4\}$.*

Claim 15 *Let $h \in H_1$ and $\bar{a} \in \bar{D}_d(f(h))$ for some $d \in N$. Then $\pi(f|h \cdot \bar{a}) = f(h \cdot \bar{a}) \cdot \pi^d(\theta, t)$ for some $\theta \in \{0,\ldots,d+4\}$ and $t \in \{1,\ldots,d+5\}$.*

Claim 16 *Let* $h \in H_2^{k,d',\tau'} \cup H_5^{k,d',\tau'} \cup H_6^{d',\tau'}$ *for some* $k \in \{1,\ldots,M\}$, $d' \in N$ *and* $\tau' \in \{0,\ldots,d'+3\}$. *Let* $d \neq d'$ *and* $\bar{a} \in \bar{D}_d(f(h))$. *Then, for some* $\theta \in \{0,\ldots,d+4\}$ *and* $t \in \{1,\ldots,d+5\}$, *either* $\pi(f|h\cdot\bar{a}) = \pi^{(d)}(\theta, d+5)$ *or* $\pi(f|h\cdot\bar{a}) = f(h\cdot\bar{a})\cdot\pi^{(d)}(\theta,t)$.

Claim 17 *Let* $h \in H_3^{k,d'}$ *for some* $k \in \{1,\ldots,M\}$ *and* $d' \in N$ *and let* $\bar{a} \in \bar{D}_d(f(h))$ *for some* $d \in N$. *If* $d = d'$, *then* $\pi(f|h\cdot\bar{a}) = \big((m^d; (\theta+1)T - [k+1-(d+4)]), \hat{\pi}^{(d),1}, \ldots\big)$ *where* $\theta = \theta(T^k(h))$. *If* $d \neq d'$, *then* $\pi(f|h\cdot\bar{a}) = f(h\cdot\bar{a})\cdot\pi^{(d)}(\theta,t)$ *for some* $\theta \in \{0,\ldots,d+4\}$ *and* $t \in \{1,\ldots,d+5\}$.

Claim 18 *Let* $h \in H_4^{k,d',r}$ *for some* $k \in \{1,\ldots,M\}$, $d' \in N$ *and* $r \in \{0,\ldots,n+d'+4\}$, *and let* $\bar{a} \in \bar{D}_d(f(h))$ *for some* $d \in N$. *Then* $\pi(f|h\cdot\bar{a}) = f(h\cdot\bar{a})\cdot\pi^{(d)}(\theta,t)$ *for some* $\theta \in \{0,\ldots,d+4\}$ *and* $t \in \{1,\ldots,d+5\}$.

Claim 19 *Let* $h \in H_7^k$ *for some* $k \in \{0,\ldots,n+4\}$, *and* $\bar{a} \in \bar{D}_d(f(h))$ *for some* $d \in N$. *If* $h \notin \cup_{k=2}^{n+3} H_7^k$ *and* $T^{d+3}(h) \in \Sigma^{d,d+3}$, *then* $\pi(f|h\cdot\bar{a}) = \big((m^d; (\theta+1)T), \hat{\pi}^{(d),1}, \ldots\big)$ *where* $\theta = \theta(T^{d+3}(h)\cdot s') + 1$. *Otherwise,* $\pi(f|h\cdot\bar{a}) = f(h\cdot\bar{a})\cdot\pi^{(d)}(\theta,t)$ *for some* $\theta \in \{0,\ldots,d+4\}$ *and* $t \in \{1,\ldots,d+5\}$.

A.2.2 f is subgame perfect and $\pi(f) = \hat{\pi}^{(0)}$

Claims 9 and 10 establish that the strategy profile $f \in F^p$ induces $\hat{\pi}^{(0)}$, hence, $\pi(f) = \hat{\pi}^{(0)}$.

Before showing f is subgame perfect, it needs to be pointed out that in Claim A.1 in the supplementary material we show that the payoffs of different paths $(\hat{\pi}^{(0)},\ldots,\hat{\pi}^{(n)})$ satisfy the following inequalities: for all $i \in \{0,\ldots,n\}$ and $d, d' \in N$ with $d \neq d'$:

$$-B(1-\delta^{(n+5)(T+1)}) + \delta^{(n+5)(T+1)} V_d(\hat{\pi}^{(d')}) > B(1-\delta^{n+6}) + \delta^{n+6+T} V_d(\hat{\pi}^{(d)}), \quad (29)$$

$$-B(1-\delta^K) + \delta^K V_d(\hat{\pi}^{(i)}) > (1-\delta^{n+6})B + \delta^{n+6+T} V_d(\hat{\pi}^{(d)}), \quad (30)$$

$$-(1-\delta^{n+5})B + \delta^{n+5+(n+4)T} V_d(\hat{\pi}^{(d)}) > (1-\delta^{n+5})B + \delta^{(n+5)(T+1)} V_d(\hat{\pi}^{(d)}). \quad (31)$$

To show that f is subgame perfect we need to establish the following for all $h \in H$:

$$V_d(\pi(f|h)) \geq (1-\delta)u_d(\bar{a}) + \delta V_d(\pi(f|h\cdot\bar{a})) \text{ for all } d \in N \text{ and } \bar{a} \in \bar{D}_d(f(h)). \quad (32)$$

Case 1: $h \in H_1 \cup H_4$. In this case, by Claims 9, 10 and 11, $\pi(f|h) = (\hat{\pi}^{(i),k}, \ldots)$ for some $i \in \{0,\ldots,n\}$ and $k \leq M$. Also, by Claims 15 and 18, $\pi(f|h\cdot\bar{a}) = f(h\cdot\bar{a})\cdot\pi^{(d)}(\theta,t)$ for

some $\theta \in \{0, \ldots, d+4\}$ and $t \in \{1, \ldots, d+5\}$. Therefore, the left-hand side of (32) must be greater or equal to $-B(1-\delta^{K-k+1}) + \delta^{K-k+1} V_d(\hat{\pi}^{(i)}) \geq -B(1-\delta^K) + \delta^K V_d(\hat{\pi}^{(i)})$ and the right-hand side of (32) is less than or equal to $(1-\delta^{d+7-t})B + \delta^{d+7-t+(\theta+1)T} V_d(\hat{\pi}^{(d)}) \leq (1-\delta^{d+6})B + \delta^{d+6+T} V_d(\hat{\pi}^{(d)})$. Thus, by (30), (32) must hold.

Case 2: $h \in H_3^{k,d'}$ for some $k \in \{1, \ldots, M\}$ and $d' \in N$. Claim 12 implies that $\pi(f|h) = ((m^{d'}; (\theta+1)T - [k-(d'+4)], \hat{\pi}^{(d'),1}, \ldots)$, where $\theta = \theta(T^k(h))$. Claim 17 implies that $\pi(f|h \cdot \bar{a}) = \big((m^d; (\theta+1)T - [k+1-(d+4)], \hat{\pi}^{(d),1}, \ldots\big)$ if $d = d'$ and $\pi(f|h \cdot \bar{a}) = f(h \cdot \bar{a}) \cdot \pi^{(d)}(\theta, t)$ for some $\theta \in \{0, \ldots, d+4\}$ and $t \in \{1, \ldots, d+5\}$ if $d \neq d'$. Clearly, the deviation is not profitable if $d = d'$. When $d \neq d'$, the left-hand side of (32) must be greater or equal to $(1-\delta^{(\theta+1)T-k+d'+4}) u_d(m^{d'}) + \delta^{(\theta+1)T-k+d'+4} V_d(\hat{\pi}^{(d')}) \geq -(1-\delta^{(n+5)T})B + \delta^{(n+5)T} V_d(\hat{\pi}^{(d')})$ and the right-hand side of (32) is less than or equal to $(1-\delta^{d+6})B + \delta^{d+6+T} V_d(\hat{\pi}^{(d)})$. Thus, by (29), (32) must hold.

Case 3: $h \in H_2^{k,d,\tau} \cup H_5^{k,d,\tau} \cup H_6^{d,\tau}$ for some $k \in \{1, \ldots, M\}$ and $\tau \in \{0, \ldots, d+3\}$. Claim 13 implies that $\pi(f|h) = f(h) \cdot \pi^{(d)}(\theta, t)$ and $\pi(f|h \cdot \bar{a}) = \pi^{(d)}(\theta', t')$ for some $t, t' \in \{1, \ldots, d+5\}$ and $\theta, \theta' \in \{0, \ldots, d+4\}$ such that $\theta < \theta'$. Therefore, the left-hand side of (32) must be greater or equal to $-(1-\delta^{d+6-t})B + \delta^{d+6-t+(\theta+1)T} V_d(\hat{\pi}^{(d)}) \geq -(1-\delta^{d+5})B + \delta^{d+5+(\theta+1)T} V_d(\hat{\pi}^{(d)})$ and the right-hand side of (32) is less than or equal to $(1-\delta^{d+6-t'})B + \delta^{d+6-t'+(\theta'+1)T} V_d(\hat{\pi}^{(d)}) \leq (1-\delta^{d+5})B + \delta^{d+5+(\theta'+1)T} V_d(\hat{\pi}^{(d)})$. Thus, by (31) and $\theta < \theta'$, (32) must hold.

Case 4: $h \in H_2^{k,d',\tau} \cup H_5^{k,d',\tau} \cup H_6^{d',\tau}$ for some $k \in \{1, \ldots, M\}$ and $d' \in N$ and $\tau \in \{0, \ldots, d'+3\}$ such that $d' \neq d$. Claim 13 implies that $\pi(f|h) = f(h) \cdot \pi^{(d')}(\theta', t')$ for some $t' \in \{1, \ldots, d'+5\}$ and $\theta' \in \{0, \ldots, d'+4\}$. Also, Claim 16 implies that, for some $t \in \{1, \ldots, d+5\}$ and $\theta \in \{0, \ldots, d+4\}$, $\pi(f|h \cdot \bar{a}) = \pi^{(d)}(\theta, d+5)$ or $\pi(f|h \cdot \bar{a}) = f(h \cdot \bar{a}) \cdot \pi^{(d)}(\theta, t)$. Therefore, the left-hand side of (32) must be greater or equal to $-(1-\delta^{d'+6-t'+(\theta'+1)T})B + \delta^{d'+6-t'+(\theta'+1)T} V_d(\hat{\pi}^{(d')}) \geq -(1-\delta^{(n+5)(T+1)})B + \delta^{(n+5)(T+1)} V_d(\hat{\pi}^{(d')})$. The right-hand side of (32) is less than or equal to $(1-\delta)B + \delta^{T+1} V_d(\hat{\pi}^{(d)}) \leq (1-\delta^{n+6})B + \delta^{n+6+T} V_d(\hat{\pi}^{(d)})$ if $\pi(f|h \cdot \bar{a}) = \pi^{(d)}(\theta, d+5)$ and is less than or equal to $(1-\delta^{d+6})B + \delta^{d+6+(\theta+1)T} V_d(\hat{\pi}^{(d)}) \leq (1-\delta^{n+6})B + \delta^{n+6+T} V_d(\hat{\pi}^{(d)})$ if $\pi(f|h \cdot \bar{a}) = f(h \cdot \bar{a}) \cdot \pi^{(d)}(\theta, t)$. Thus, by (29), (32) must hold.

Case 5: $h \in H_7^k$ for some $k \in \{0, \ldots, n+4\}$ and $h \notin \cup_{k=2}^{n+3} H_7^k$ and $T^{d+3}(h) \in \Sigma^{d,d+3}$. Claims 14 and 19 imply that we must have $\pi(f|h) = \big(s'^d; (\theta+1)T), \hat{\pi}^{(d),1}, \ldots\big)$ and $\pi(f|h \cdot \bar{a}) = \big((m^d; (\theta+2)T), \hat{\pi}^{(d),1}, \ldots\big)$ for some $\theta \in \{0, \ldots, d+3\}$. Therefore, the left-hand side of (32)

must be greater or equal to $-(1-\delta)B + \delta^{(\theta+1)T+1} V_d(\hat{\pi}^{(d)})$ and the right-hand side of (32) is less than or equal to $(1-\delta)B + \delta^{(\theta+2)T+1} V_d(\hat{\pi}^{(d)})$. Thus, by (31), (32) must hold.

Case 6: $h \in H_7^k$ for some $k \in \{0, \ldots, n+4\}$ and $h \notin \cup_{k=2}^{n+3} H_7^k$ and $T^{d'+3}(h) \in \Sigma^{d',d'+3}$ for some $d' \neq d$. Claims 14 and 19 imply that $\pi(f|h) = \left(s'^{d'}; (\theta'+1)T), \hat{\pi}^{(d'),1}, \ldots\right)$ for some $\theta' \in \{0, \ldots, d'+4\}$ and $\pi(f|h \cdot \bar{a}) = f(h \cdot \bar{a}) \cdot \pi^{(d)}(\theta, t)$ for some $\theta \in \{0, \ldots, d+4\}$ and $t \in \{1, \ldots, d+5\}$. Therefore, the left-hand side of (32) must be greater or equal to $-(1-\delta^{1+(\theta'+1)T})B + \delta^{1+(\theta'+1)T} V_d(\hat{\pi}^{(d')}) \geq -(1-\delta^{(n+5)(T+1)})B + \delta^{(n+5)(T+1)} V_d(\hat{\pi}^{(d')})$ and the right-hand side of (32) is less than or equal to $(1-\delta^{d+6})B + \delta^{d+6+(\theta+1)T} V_d(\hat{\pi}^{(d)}) \leq (1-\delta^{n+6})B + \delta^{n+6+T} V_d(\hat{\pi}^{(d)})$. Thus, by (29), (32) must hold.

Case 7: $h \in H_7^k$ and either $h \in \cup_{k=2}^{n+3} H_7^k$ or $T^{d'+3}(h) \notin \Sigma^{d',d'+3}$ for all $d' \in N$. Claim 14 implies that $\pi(f|h) = (\hat{\pi}^{(i),k}, \ldots)$ for some $i \in \{0, \ldots, n\}$ and $k \leq M$. Also, by Claim 19, $\pi(f|h \cdot \bar{a}) = f(h \cdot \bar{a}) \cdot \pi^{(d)}(\theta, t)$ for some $\theta \in \{0, \ldots, d+4\}$ and $t \in \{1, \ldots, d+5\}$. Therefore, by an identical argument as in Case 1, (32) must hold.

B Proof of Theorem 2 (mixed Folk Theorem)

We normalize payoffs so that the mixed strategy minmax payoff $\tilde{v}_i = 0$ for all $i \in N$.

B.1 Some preliminary results

First, we establish two lemmas needed in the construction of our equilibrium strategies.

By the full-dimensionality assumption, for each $i \in N$, there exist action profiles $\bar{a}^{(i)} \in A$ such that $u_i(m^i) < \max_{a_i \in A_i} u_i(a_i, \bar{a}_{-i}^{(i)}) = u_i(\bar{a}^{(i)})$. As before, $B = \max_{i \in N, a \in A} |u_i(a)|$. Let $\gamma \in \mathbb{R}$ be such that

$$\gamma(u_i(\bar{a}^{(i)}) - u_i(m^i)) > 2B \text{ for all } i \in N. \tag{33}$$

As before, fix any s and s', both in A, such that $s_i \neq s_i'$ for all $i \in N$. Then for all

$d \in N, Q \in \mathbb{N}, \hat{h} = (\hat{a}^1, \ldots, \hat{a}^t) \in H, \delta \in (0,1), i \in N$ and $\eta > 0$ let

$$\hat{\Sigma}^{d,\tau}(Q) = \big\{h \in H : h = (a^t)_{t=1}^{\tau} \text{ such that } a^t \in D_d(s) \text{ if } t = 2, \ldots, d+Q+1$$
$$\text{and } a^t \in D_d(s') \text{ if } t = 1\big\},$$
$$S(\hat{h}, Q) = \{1\} \cup \Big\{k \in \{Q+2, \ldots, t\} : T^{Q+1}(\hat{a}^1, \ldots, \hat{a}^{k-1}) \in \hat{\Sigma}^{l,(Q+1)}(Q) \text{ for some } l \in N\Big\},$$
$$\lambda(a, \hat{h}, Q, \delta) = \sum_{k \notin S(\hat{h},Q)} \delta^{k-1} 1_a(\hat{a}^k),$$
$$\lambda_i(a_{-i}, \hat{h}, Q, \delta) = \sum_{b_i \in A_i} \lambda((b_i, a_{-i}), \hat{h}, Q, \delta),$$
$$\Phi_i^d(\hat{h}, Q, \delta) = \frac{1-\delta}{1-\delta^t} \sum_{a \in A} |\lambda(a, \hat{h}, Q, \delta) - \lambda_i(a_{-i}, \hat{h}, Q, \delta)\mu_i^d(a_i)|,$$
$$\alpha_i^d(\hat{h}, Q, \delta, \eta) = \begin{cases} 1 & \text{if } \Phi_i^d(\hat{h}, Q, \delta) < \eta, \\ 0 & \text{otherwise.} \end{cases}$$

We can then state our first lemma.

Lemma 1 *For every $\zeta > 0$ and $0 < \varepsilon_1 < 1$, there exist $Q \in \mathbb{N}$ with $Q > \gamma$ and $\eta > 0$ such that for every $d \in N$ and $\delta \in (0,1)$ and $t \in \mathbb{N}$ and $\hat{h} = (\hat{a}^1, \ldots, \hat{a}^t) \in H_t$ such that $\delta^t \geq \zeta$ and $\alpha_i^d(\hat{h}, \delta, Q, \eta) = 1$ for all $i \neq d$, we have*

$$\frac{1-\delta}{1-\delta^t} \sum_{k=1}^{t} \delta^{k-1} u_d(\hat{a}^k) < \varepsilon_1.$$

The proof of the above lemma has some similarity to that of Lemma 1 in **?**. See the supplementary materials for the proof of Lemma 1.

Next fix $i, d \in N$ with $i \neq d$ and let $\tilde{\mu}_i^d \in F_i^m$ be player i's strategy consisting of playing μ_i^d each period independently of the history. Recall that for any strategy $f_{-i} \in F_{-i}^m$ for the remaining players and $t \in \mathbb{N}$, $P_{(\tilde{\mu}_i^d, f_{-i}),t}$ denotes the probability measure on H_t induced by $(\tilde{\mu}_i^d, f_{-i})$. We next establish our second lemma.

Lemma 2 *For all $c \in (0,1)$, $Q \in \mathbb{N}$, $\eta > 0$ and $\varepsilon_2 > 0$, there exists a function $t(\cdot) : (0,1) \to \mathbb{N}$ and $\bar{\delta} \in (0,1)$ such that, for all $\delta \geq \bar{\delta}$, $|\delta^{t(\delta)} - c| < \varepsilon_2$ and*

$$P_{(\tilde{\mu}_i^d, f_{-i}), t(\delta)} \Big(\{\hat{h} \in H_{t(\delta)} : \Phi_i^d(\hat{h}, Q, \delta) \geq \eta\}\Big) < \varepsilon_2$$

for all $i, d \in N$ with $i \neq d$ and for any $f_{-i} \in F_{-i}^m$. Furthermore, we have that $\lim_{\delta \to 1} \delta^{t(\delta)} = c$.

Henceforth, we shall omit the dependence of $\hat{\Sigma}^d(.), S(.), \lambda(.), \lambda_i(.), \Phi_i^d(.), \alpha_i^d(.)$ on $\hat{h}, Q, \delta$ or η whenever the meaning is clear.

Proof. Let $c \in (0,1)$, $Q \in \mathbb{N}$, $\eta > 0$ and $\varepsilon_2 > 0$ be given. For each $\delta \in (0,1)$, let $t(\delta)$ be the highest integer $t \in \mathbb{N}$ such that $\delta^t \geq c$. Then $|\delta^{t(\delta)} - c| < (1-\delta)/\delta$ and therefore $\lim_{\delta \to 1} \delta^{t(\delta)} = c$. Hence, there exists $\delta_1 \in (0,1)$ such that $|\delta^{t(\delta)} - c| < \varepsilon_2$ for all $\delta \geq \delta_1$.

For each $\delta \in (0,1)$, let

$$b(\delta) = \frac{(1-\delta)(1+\delta^{t(\delta)})}{(1+\delta)(1-\delta^{t(\delta)})}.$$

Also, let $\kappa = \eta/|A|$. Since $\lim_{\delta \to 1} b(\delta) = 0$, there exists $\bar{\delta} \in [\delta_1, 1)$ such that $e^{-\frac{2\kappa^2}{b(\delta)}} < \varepsilon_2/|A|$ for all $\delta > \bar{\delta}$.

Next, note that for all $i, d \in N$ with $i \neq d$ and $f_{-i} \in F_{-i}^m$

$$P_{(\tilde{\mu}_i^d, f_{-i}), t(\delta)} \left(\left\{ \hat{h} \in H_{t(\delta)} : \sum_{a \in A} \left| \frac{1-\delta}{1-\delta^{t(\delta)}} (\lambda(a) - \lambda_i(a_{-i}) \mu_i^d(a_i)) \right| \geq \eta \right\} \right) \leq$$
$$\sum_a P_{(\tilde{\mu}_i^d, f_{-i}), t(\delta)} \left(\left\{ \hat{h} \in H_{t(\delta)} : \left| \frac{1-\delta}{1-\delta^{t(\delta)}} (\lambda(a) - \lambda_i(a_{-i}) \mu_i^d(a_i)) \right| \geq \kappa \right\} \right).$$

Since $e^{-\frac{2\kappa^2}{b(\delta)}} < \varepsilon_2/|A|$ for all $\delta \geq \bar{\delta}$, to demonstrate the result, it suffices to show that for all $\delta \geq \bar{\delta}$

$$P_{(\tilde{\mu}_i^d, f_{-i}), t(\delta)} \left(\left\{ \hat{h} \in H_{t(\delta)} : \left| \frac{1-\delta}{1-\delta^{t(\delta)}} (\lambda(a) - \lambda(a_{-i}) \mu_i^d(a_i)) \right| \geq \kappa \right\} \right) \leq e^{-\frac{2\kappa^2}{b(\delta)}}, \tag{34}$$

for every $i, d \in N$ with $i \neq d$ and $a \in A$ and $f_{-i} \in F_{-i}^m$. In the rest of the proof we shall fix $i, d \in N$ with $i \neq d$ and $a \in A$ and $f_{-i} \in F_{-i}^m$ and $\delta \geq \bar{\delta}$, and show that (34) holds.

To simplify the notation, when the meaning is clear, we shall denote $t(\delta)$ by t.

For any $\hat{h} = (\hat{a}^1, \ldots, \hat{a}^t) \in H_t$ and any $1 \leq k \leq t$, let $X_k(\hat{h}) = 1_{\{\hat{a}_i^k = a_i\}}$, $Y_k(\hat{h}) = 1_{\{\hat{a}_{-i}^k = a_{-i}\}}$,

$$W_k(\hat{h}) = \begin{cases} 1_{\{T^{Q+1}(\hat{a}^1, \ldots, \hat{a}^{k-1}) \notin \cup_{l \in N} \hat{\Sigma}^{l, Q+1}\}} & \text{if } k > 1, \\ 0 & \text{if } k = 1, \end{cases}$$

and $Z_k(\hat{h}) = (X_k(\hat{h}), Y_k(\hat{h}), W_k(\hat{h}))$. Again when the meaning is clear, we shall denote $X_k(\hat{h}), Y_k(\hat{h}), W_k(\hat{h})$ and $Z_k(\hat{h})$, by X_k, Y_k, W_k and Z_k, respectively.

Denote $\mu_i^d(a_i)$ by p and let

$$f(Z_1, \ldots, Z_t) = \frac{1-\delta}{1-\delta^t} \sum_{k=1}^t \delta^{k-1} (X_k - p) Y_k W_k.$$

Then the RHS of the last equation is equal to

$$\begin{aligned}\frac{1-\delta}{1-\delta^t}\sum_{k=1}^{t}\delta^{k-1}(1_{a_i}(\hat{a}_i^k)-p)1_{a_{-i}}(\hat{a}_{-i}^k)W_k &= \frac{1-\delta}{1-\delta^t}\sum_{k=1}^{t}\delta^{k-1}\left(1_a(\hat{a}^k)-1_{a_{-i}}(\hat{a}_{-i}^k)\mu_i^d(a_i)\right)W_k \\ &= \frac{1-\delta}{1-\delta^t}(\lambda(a)-\lambda(a_{-i})\mu_i^d(a_i)).\end{aligned}$$

Hence, (34) is equivalent to the following condition

$$P_{(\tilde{\mu}_i^d,f_{-i})}\left(\{\hat{h}\in H_t:|f(Z_1,\ldots,Z_t)|\geq\kappa\}\right)\leq e^{-\frac{2\kappa^2}{b(\delta)}} \tag{35}$$

To show this, assume that $(\tilde{\mu}_i^d, f_{-i})$ is chosen and, for any $\mathbf{z}=(z_1,\ldots,z_t)\in\{0,1\}^{3t}$, $k\in\{1,\ldots,t\}$ and $z=(x,y,w)\in\{0,1\}^3$, let $B_k(\mathbf{z})=\{Z_l=z_l \text{ for all } l=1,\ldots,k-1\}$ and

$$g_k(z,\mathbf{z})=E(f(Z_1,\ldots,Z_t)|B_k(\mathbf{z}),Z_k=z)-E(f(Z_1,\ldots,Z_t)|B_k(\mathbf{z})).$$

Also, let $\text{ran}_k(\mathbf{z})=\sup\{|g_k(z,\mathbf{z})-g_k(z',\mathbf{z})|:z,z'\in\{0,1\}^3\}$, $R^2(\mathbf{z})=\sum_{k=1}^t(\text{ran}_k(\mathbf{z}))^2$ and $\hat{r}^2=\sup_{\mathbf{z}\in\{0,1\}^{3t}}R^2(\mathbf{z})$.

Then, by Theorem 3.7 in **?**,

$$P_{(\tilde{\mu}_i^d,f_{-i})}\left(\{\hat{h}\in H_t:|f(Z_1,\ldots,Z_t)|-E(f(Z_1,\ldots,Z_t))\geq\kappa\}\right)\leq e^{-\frac{2\kappa^2}{\hat{r}^2}}. \tag{36}$$

Therefore, to complete the proof of this lemma we need to show that $E(f(Z_1,\ldots,Z_t))=0$ and $\hat{r}^2=b(\delta)$. We establish these in the next three claims.

Claim 20 *$E(f(Z_1,\ldots,Z_t))=0$.*

Proof. We have that

$$\begin{aligned}E(f(Z_1,\ldots,Z_t)) &= \frac{1-\delta}{1-\delta^t}\sum_{k=1}^{t}\delta^{k-1}E\left((X_k-p)Y_kW_k\right) \\ &= \frac{1-\delta}{1-\delta^t}\sum_{k=1}^{t}\delta^{k-1}E\left(X_k-p\right)E(Y_kW_k)=0\end{aligned}$$

where the second equality above follows from X_k and (Y_k,W_k) being independent and the third equality from $E(X_k)=p$ for all $1\leq k\leq t$. ■

Claim 21 *For each $\mathbf{z}=(z_1,\ldots,z_t)\in\{0,1\}^{3t}$ and $k\in\{1,\ldots,t\}$ and $z=(x,y,w)\in\{0,1\}^3$,*

$$g_k(z,\mathbf{z})=\frac{1-\delta}{1-\delta^t}\delta^{k-1}(x-p)yw.$$

Proof. Fix any $\mathbf{z} = (z_1, \ldots, z_t) \in \{0,1\}^{3t}$ and $k \in \{1, \ldots, t\}$ and $z = (x, y, w) \in \{0,1\}^3$, and denote $B_k(\mathbf{z})$ by B_k for simplicity in the rest of this proof. Note that (i) for each $k' \geq k$, $X_{k'}$ and $(Y_{k'}, W_{k'})$ are independent given B_k and $E(X_{k'}|B_k) = p$, and (ii) for each $k' > k$, $X_{k'}$ and $(Y_{k'}, W_{k'})$ are independent given $B_k \cap \{Z_k = z\}$ and $E(X_{k'}|B_k, Z_k = z) = p$. So

$$\begin{aligned} E(f(Z_1, \ldots, Z_t)|B_k) &= \frac{1-\delta}{1-\delta^t}\left(\sum_{l=1}^{k-1}\delta^{l-1}(x_l - p)y_l w_l + \sum_{l=k}^{t}\delta^{l-1}E((X_l - p)Y_l W_l|B_k)\right) \\ &= \frac{1-\delta}{1-\delta^t}\sum_{l=1}^{k-1}\delta^{l-1}(x_l - p)y_l w_l \end{aligned}$$

and

$$\begin{aligned} E(f(Z_1, \ldots, Z_t)|B_k, Z_k = z_k) &= \frac{1-\delta}{1-\delta^t}\sum_{l=1}^{k}\delta^{l-1}(x_l - p)y_l w_l \\ &+ \frac{1-\delta}{1-\delta^t}\sum_{l=k+1}^{t}\delta^{l-1}E((X_l - p)Y_l W_l|B_k, Z_k = z_k) \\ &= \frac{1-\delta}{1-\delta^t}\sum_{l=1}^{k}\delta^{l-1}(x_l - p)y_l w_l. \end{aligned}$$

Hence, the result follows. ■

Claim 22 $\hat{r}^2 = b(\delta)$.

Proof. We start by establishing that

$$\mathrm{ran}_k(\mathbf{z}) = \frac{1-\delta}{1-\delta^t}\delta^{k-1}$$

for all $k \in \{1, \ldots, t\}$ and $\mathbf{z} = (z_1, \ldots, z_t \in \{0,1\}^{3(t)}$. Indeed, it follows by Claim 21 that for all z and $z' \in \{0,1\}^3$

$$|g_k(z, \mathbf{z}) - g_k(z', \mathbf{z})| = \frac{1-\delta}{1-\delta^t}\delta^{k-1}|(x-p)yw - (x'-p)y'w'|.$$

Note that $(x-p)yw \in \{-p, 0, 1-p\}$ and similarly for $(x'-p)y'w'$. Hence, $|(x-p)yw - (x'-p)y'w'| \leq 1 - p - (-p) = 1$, thus, $\mathrm{ran}(z_1, \ldots, z_{k-1}) = \frac{1-\delta}{1-\delta^t}\delta^{k-1}$.

Therefore, for all $\mathbf{z} = (z_1, \ldots, z_t) \in \{0,1\}^{3t}$,

$$R^2(\mathbf{z}) = \left(\frac{1-\delta}{1-\delta^t}\right)^2 \sum_{k=1}^{t}\delta^{2(k-1)},$$

hence,

$$\hat{r}^2 = \left(\frac{1-\delta}{1-\delta^t}\right)^2 \sum_{k=1}^{t} \delta^{2(k-1)} = \frac{(1-\delta)^2}{(1-\delta^t)^2}\frac{1-\delta^{2t}}{1-\delta^2} = \frac{1-\delta}{1+\delta}\frac{1+\delta^t}{1-\delta^t} = b(\delta).$$

■

Claims 20 and 22 complete the proof of the lemma. ■

B.2 The strategy profile

This section describes the equilibrium strategy profile employed in the proof.

B.2.1 Parametrization

Since $\tilde{\mathcal{U}}$ is compact and $\tilde{\mathcal{U}}$ equals the closure of $\tilde{\mathcal{U}}^0$, to establish Theorem 2, it suffices to show that, for all $\varepsilon > 0$ and $u \in \tilde{\mathcal{U}}^0$, there exists $\delta^* \in (0,1)$ such that, for all $\delta \geq \delta^*$, there are $M \in \mathbb{N}$ and a M-memory SPE f of $G_m^\infty(\delta)$ with $\|U(f,\delta) - u\| < \varepsilon$.

Fix any $\varepsilon > 0$ and any $u \in \tilde{\mathcal{U}}^0$. Since G is full-dimensional, we may assume that $u \in \text{int}(\tilde{\mathcal{U}}^0)$. Let $u' \in \text{int}(\tilde{\mathcal{U}}^0)$ such that $u' < u$, and $\rho > 0$ be such that (i) $u'_i + \rho < u_i$ for all $i \in N$ and (ii) $||\hat{u} - u'|| \leq \rho$ implies $\hat{u} \in \tilde{\mathcal{U}}^0$. Also, fix any

$$\varepsilon_1 \in (0, \min_{d \in N} u'_d). \tag{37}$$

Let $Q \in \mathbb{N}$ and $\eta > 0$ be as in Lemma 1, corresponding to $\zeta = 1/2$ and to ε_1 defined by (37). Let $\varepsilon_2 > 0$ be such that

$$\varepsilon_2 < 1 - \zeta, \text{ and} \tag{38}$$

$$\bar{\varepsilon} < \min_{d \in N} u'_d, \tag{39}$$

where $\bar{\varepsilon}$ is defined by

$$\bar{\varepsilon} = (1-2\varepsilon_2)^n \varepsilon_1 + (1-(1-2\varepsilon_2)^n) B.$$

Let $c \in (0,1)$ be such that

$$c > \zeta + \varepsilon_2, \tag{40}$$

$$c\rho\varepsilon_2 > (1-c)2B, \text{ and} \tag{41}$$

$$c(1-2\varepsilon_2)\rho > (1-c)(B+\bar{\varepsilon}). \tag{42}$$

By (38), it is clear that such c exists. Let $\bar{\delta}$ be as in Lemma 2, corresponding to Q, η, ε_2 and c as defined above.

Define $\xi > 0$ to be such $2\xi < \varepsilon$ and

$$(1-c)(\min_d u'_d - \bar{\varepsilon}) > (1+c)2\xi, \tag{43}$$

$$c(\rho\varepsilon_2 - 4\xi) > (1-c)2B, \tag{44}$$

$$c((1-2\varepsilon_2)\rho - 4\xi) > (1-c)(B+\bar{\varepsilon}). \tag{45}$$

Such $\xi > 0$ exists due to (39), (41) and (42), respectively.

For all $i = 1, \ldots, n$ and $\beta \in \mathbb{R}^n$, let $u^i(\beta)$ be defined by $u^i_i(\beta) = u'_i$ and $u^i_j(\beta) = u'_j + \beta_j\rho$. Furthermore, define

$$W_i = \{u^i(\beta) : \beta_j \in \{0,1\} \text{ for all } j \in N \setminus \{i\}\}.$$

Due to our choice of ρ (recall that (ii) above states that $||\hat{u} - u'|| \leq \rho$ implies $\hat{u} \in \tilde{\mathcal{U}}^0$), $W_i \subseteq \tilde{\mathcal{U}}^0$. Define $\hat{W} = \cup_{i=1}^n W_i$. Since $\hat{W}$ is finite, order $\hat{W} = \{\hat{u}^1, \ldots, \hat{u}^{\bar{\omega}}\}$, where $\bar{\omega} = |\hat{W}|$. For notational convenience, let $\hat{u}^0 = u$ and $W = \hat{W} \cup \{\hat{u}^0\}$.

For all $k \in \mathbb{N}$, let $\mathcal{V}_k$ be the set of $u'' \in \text{co}(u(A))$ such that $u'' = \sum_{a\in A} p^a u(a)/k$ for some $\{p^a\}_{a\in A}$ satisfying $p^a \in \mathbb{N}$ and $p^a \geq n + \bar{\omega} + Q + 2$ for all $a \in A \setminus \{s', s\}$, $p^{s'} \geq 3$, $p^s \geq 2(n + \bar{\omega} + Q + 2)$ and $\sum_{a\in A} p^a = k$. Using an analogous argument to **?**, it follows that $\mathcal{V}_k$ converges to $\text{co}(u(A))$. Therefore, let $K \in \mathbb{N}$ such that

$$K > n \text{ and } \text{co}(u(A)) \subseteq \cup_{x\in\mathcal{V}_K} B_\xi(x). \tag{46}$$

As before, order all the actions in $A = \{a^1, \ldots, a^{|A|}\}$ as follows: $a^1 = s$, $a^2, \ldots, a^{|\bar{D}(s)|+1}$ are the different elements $\bar{D}(s)$, in any order, $a^{|\bar{D}(s)|+2} = s'$, $a^{|\bar{D}(s)|+3}, \ldots, a^{|\bar{D}(s)|+|\bar{D}(s')|+2}$ are the different elements of $\bar{D}(s')$, in any order, and all the remaining actions are then ordered arbitrarily.

For all $\omega \in \{0, \ldots, \bar{\omega}\}$, let $x^\omega \in \mathcal{V}_K$ be such that

$$||x^\omega - \hat{u}^\omega|| < \xi \tag{47}$$

and $\{p^\omega_a\}_{a\in A}$ be such that $\frac{1}{K}\sum_{a\in A}^K p^\omega_a u_j(a) = x^\omega_j$ for all $j \in N$. For all $\omega \in \{0, \ldots, \bar{\omega}\}$, define $\hat{\pi}^{(\omega)}$ as the repetition of the cycle

$$\left((s';2), (s; n+\omega+2+Q), s', (a^1; p^{(\omega),1}), \ldots, (a^{|A|}; p^{(\omega),|A|})\right),$$

where $p^{(\omega),j} = p^{\omega}_{a^j} - 3$ if $a^j = s'$, $p^{(\omega),j} = p^{\omega}_{a^j} - (n+\omega+2+Q)$ if $a^j = s$ and $p^{(\omega),j} = p^{\omega}_{a^j}$ otherwise. Note that the length of the cycle is K, i.e. $\sum_{j=1}^{|A|} p^{(\omega),j} + n + 5 + \omega + Q = K$ for all $\omega \in \{0, \ldots, \bar{\omega}\}$. In the construction below, $\hat{\pi}^{(\omega)}$ will be the equilibrium path when $\omega = 0$ (also sometimes denoted by $\pi^{(0)}$) and a "reward path" when $\omega > 0$.

Let $\delta^* \in [\bar{\delta}, 1)$ be such that for all $\delta \geq \delta^*$, letting $t(\delta)$ be as in Lemma 2,

$$\sup_{x \in [-B,B]^K} \left| \frac{1-\delta}{1-\delta^K} \sum_{k=1}^{K} \delta^{k-1} x^k - \frac{1}{K} \sum_{k=1}^{K} x^k \right| < \xi, \tag{48}$$

$$|\delta^{t(\delta)} - c| < \varepsilon_2, \tag{49}$$

$$t(\delta) > n + Q + 4 + K, \tag{50}$$

$$\delta^{t(\delta)}(\rho\varepsilon_2 - 4\xi) > (1 - \delta^{t(\delta)})2B, \tag{51}$$

$$\delta^{(\gamma+2)(n+Q+4)} \frac{(1-\delta^{\gamma})(u_d(\bar{a}^{(d)}) - u_d(m^d))}{(1-\delta)2B} > 1, \tag{52}$$

$$\begin{aligned} -(1-\delta^K)B + \delta^K(u'_d - 2\xi) > (1 - \delta^{(\gamma+2)(n+Q+4)+1})B + \delta^{(\gamma+2)(n+Q+4)+1}(1-\delta^{t(\delta)})\bar{\varepsilon} \\ + \delta^{(\gamma+2)(n+Q+4)+1+t(\delta)}(u'_d + 2\xi) \text{ for all } d \in N, \end{aligned} \tag{53}$$

and

$$\begin{aligned} &-(1 - \delta^{(\gamma+2)(n+Q+4)+t(\delta)})B + \delta^{(\gamma+2)(n+Q+4)+t(\delta)}(u'_d + (1-2\varepsilon_2)\rho - 2\xi) > \\ &(1 - \delta^{(\gamma+2)(n+Q+4)+1})B + \delta^{(\gamma+2)(n+Q+4)+1}(1-\delta^{t(\delta)})\bar{\varepsilon} + \delta^{(\gamma+2)(n+Q+4)+1+t(\delta)}(u'_d + 2\xi) \\ &\text{for all } d \in N. \end{aligned} \tag{54}$$

Note that such $\delta^* \in (0,1)$ exists because (i) by Lemma 2, (49) holds for all $\delta > \bar{\delta}$, (ii) the limits of the left hand side of (48), (50), (51), (52), (53) and (54), as $\delta \to 1$ are, respectively, 0, $+\infty$, $c(\rho\varepsilon_2 - 4\xi)$, $\frac{\gamma(u_d(\bar{a}^{(d)}) - u_d(m^d))}{2B}$ (and the latter is strictly above 1 due to (33)), $(u'_d - 2\xi)$ and $-(1-c)B + c(u'_d + (1-2\varepsilon_2)\rho - 2\xi)$, (iii) the limits of the right hand side of (51), (53) and (54), as $\delta \to 1$ are, respectively, $(1-c)2B$, $(1-c)\bar{\varepsilon} + c(u'_d + 2\xi)$, $(1-c)\bar{\varepsilon} + c(u'_d + 2\xi)$ and (iv) conditions (43)–(45) hold.

Fix any $\delta \geq \delta^*$ and set $T = t(\delta)$. Note that by Lemma 2 and (40), $\delta^T > \zeta$. Thus, by Lemma 1, for every $d \in N$ and $\hat{h} = (\hat{a}^1, \ldots, \hat{a}^T) \in H_T$ such that $\alpha_i^d(\hat{h}, \delta, Q, \eta) = 1$ for all $i \neq d$, we have

$$\frac{1-\delta}{1-\delta^T} \sum_{k=1}^{T} \delta^{k-1} u_d(\hat{a}^k) < \varepsilon_1. \tag{55}$$

Define the size of the memory $M \in \mathbb{N}$ as follows

$$M \geq (\gamma + 3)(n + Q + 4) + T + n + \bar{\omega} + Q + 4. \tag{56}$$

We next define M-memory SPE strategy profile f with $||U(f) - u|| < \varepsilon$. We first start with the mixed actions to be played during the punishment phases.

B.2.2 The minmax and the payoffs of reward phases

Let $C : W \to \mathbb{R}^n$ be defined by setting $C(\hat{u}^\omega) = V(\hat{\pi}^{(\omega)})$ for all $0 \leq \omega \leq \bar{\omega}$. For any $d \in N$ and $\hat{h} = (\hat{a}^1, \ldots, \hat{a}^T) \in H_T$ we define the reward payoff after a punishment phase by

$$w(d, \hat{h}) = C(u^d(\alpha^d(\hat{h}))),$$

where $\alpha^d(\hat{h}) = (\alpha_1^d(\hat{h}), \ldots, \alpha_n^d(\hat{h}))$. Also, let $\omega(d, \hat{h})$ be such that $\hat{u}^{\omega(d,\hat{h})} = u^d(\alpha^d(\hat{h}))$.

For all $d \in N$ let $f^d : \cup_{t=0}^{T-1} H_t \to \Delta$ and $V^d : \cup_{t=0}^{T-1} H_t \to \mathbb{R}^n$ be such that the following property holds: For all $0 \leq t \leq T-1$, $\hat{h} \in H_t$ and $i \in N$:

(a) If $\ell(\hat{h}) = T - 1$ and $T^{Q+1}(\hat{h}) \in \cup_{l \in L} \hat{\Sigma}^{l,Q+1}$, then $f_{-d}^d(\hat{h}) = s'_{-d}$, $f_d^d(\hat{h})$ solves

$$\max_{a_d \in A_d} [(1-\delta) u_d(a_d, s'_{-d}) + \delta w_d(d, \hat{h} \cdot (a_d, s'_{-d}))]$$

and $V_i^d(\hat{h}) = (1-\delta) u_i(f^d(\hat{h})) + \delta w_i(d, \hat{h} \cdot f^d(\hat{h}))$.

(b) If $\ell(\hat{h}) = T - 1$ and $T^{Q+1}(\hat{h}) \notin \cup_{l \in N} \hat{\Sigma}^{l,Q+1}$, then $f_i^d(\hat{h})$ solves

$$\max_{\sigma_i \in \Delta_i} [(1-\delta) u_i(\sigma_i, f_{-i}^d(\hat{h})) + \delta \sum_{a \in A} (\sigma_i, f^d(\hat{h}))[a] w_i(d, \hat{h} \cdot a)]$$

and $V_i^d(\hat{h}) = (1-\delta) u_i(f^d(\hat{h})) + \delta \sum_{a \in A} f^d(\hat{h})[a] w_i(d, \hat{h} \cdot a)$.

(c) If either $\hat{h} = H_0$ or $\ell(\hat{h}) < T - 1$ and $T^{Q+1}(\hat{h}) \in \cup_{l \in N} \hat{\Sigma}^{l,Q+1}$, then $f_{-d}^d(\hat{h}) = s'_{-d}$, $f_d^d(\hat{h})$ solves

$$\max_{a_d \in A_d} [(1-\delta) u_d(a_d, s'_{-d}) + \delta V_d^{d,\hat{a}}(\hat{h} \cdot (a_d, s'_{-d}))]$$

and $V_i^d(\hat{h}) = (1-\delta) u_i(f^d(\hat{h})) + \delta V_i^d(\hat{h} \cdot f^d(\hat{h}))$.

(d) If $\ell(\hat{h}) < T - 1$ and $T^{Q+1}(\hat{h}) \notin \cup_{l \in N} \hat{\Sigma}^{l,Q+1}$, then $f_i^d(\hat{h})$ solves

$$\max_{\sigma_i \in \Delta_i} [(1-\delta) u_i(\sigma_i, f_{-i}^d(\hat{h})) + \delta \sum_{a \in A} (\sigma_i, f^d(\hat{h}))[a] V_i^d(\hat{h} \cdot a)]$$

and $V_i^d(\hat{h}) = (1-\delta) u_i(f^d(\hat{h})) + \delta \sum_{a \in A} f^d(\hat{h})[a] V_i^d(\hat{h} \cdot a)$.

The existence of f^d and V^d can be established using, for each fixed $(d, \hat{a})$, backwards inductions and Nash's existence theorem.

B.2.3 The complete profile

For all $d \in N$, let $s_d^* \in A_d$ be a static best-reply to s'_{-d}, i.e. $u_d(s_d^*, s'_{-d}) \geq u_d(a_d, s'_{-d})$ for all $a_d \in A_d$. For all $\tau \in \mathbb{N}$ and $d \in N$ define

$$\Sigma^{d,\tau} = \Big\{ h \in H : h = (a^t)_{t=1}^{\tau} \text{ such that } a^t \in D_d(s) \text{ if } t = 3, \ldots, d+Q+3$$
$$\text{and } a^t \in D_d(s') \text{ if } t = 1, 2, d+Q+4 \Big\}$$

with $\Sigma^{d,0} = \{H_0\}$.

Define for all $k \in \{1, \ldots, M\}$, $\omega \in \{0, \ldots, \bar{\omega}\}$, $d \in N$ and $\tau, r \in \mathbb{N}_0$ the following sets:[46]

$$H_{1,a}^{(\omega),k} = \left\{ h \in H : T^k(h) = \left(\hat{\pi}^{(\omega),1}, \ldots, \hat{\pi}^{(\omega),k}\right) \right\},$$
$$H_{1,b}^{(\omega),k} = \left\{ h \in H : h = \left(\hat{\pi}^{(\omega),1}, \ldots, \hat{\pi}^{(\omega),k}\right) \right\},$$
$$H_1^{(\omega),k} = H_{1,a}^{(\omega),k} \cup H_{1,b}^{(\omega),k},$$
$$H_2^{k,d,\tau} = \Big\{ h \in H : T^k(h) = \bar{h} \cdot a \cdot \tilde{h} \text{ such that for some } k' \leq k \text{ and } \omega \in \{0, \ldots, \bar{\omega}\}$$
$$(1) \quad \text{either } \bar{h} \in H_{1,a}^{(\omega),k'} \text{ with } k' \geq n + \omega + Q + 5 \text{ or } \bar{h} \in H_{1,b}^{(\omega),k'} \text{ with } \ell(h) = k,$$
$$k' < n + Q + 5 \text{ and } \omega = 0, (2) \quad a \in \bar{D}_d(\hat{\pi}^{(\omega),k'+1}), \quad (3) \quad \tilde{h} \in \Sigma^{d,\tau} \text{ and}$$
$$(4) \quad \text{if } T^{d+Q+3}(\bar{h} \cdot a) = ((s'; 2), (s; d+Q), a) \text{ and } a \in \bar{D}_d(s), \text{ then } \ell(\tilde{h}) = 0 \Big\},$$
$$\hat{H}_{3,a}^{k,d} = \Big\{ h \in H : T^k(h) = \bar{h} \cdot \tilde{h} \text{ such that } (1) \quad \bar{h} \in \Sigma^{d,d+Q+4} \text{ and}$$
$$(2) \text{ if } \ell(\tilde{h}) > 0 \text{ where } \bar{h} = (\bar{a}^1, \ldots, \bar{a}^{d+Q+4}) \text{ and } \tilde{h} = (\tilde{a}^1, \ldots, \tilde{a}^{\ell(\tilde{h})}) \text{ and}$$
$$(l-1)(\gamma+1) + 1 \leq t \leq l(\gamma+1) \text{ for some } l \in \{1, \ldots, d+Q+4\},$$
$$\text{then } \tilde{a}_{-d}^t = \begin{cases} s'_{-d} & \text{if } t = l(\gamma+1) \\ \bar{a}_{-d}^{(d)} & \text{if } \bar{a}^l \in \{s', s\} \text{ and } t \neq l(\gamma+1) \quad \} \\ m_{-d}^d & \text{if } \bar{a}^l \notin \{s', s\} \text{ and } t \neq l(\gamma+1). \end{cases}$$
$$H_{3,a}^{k,d} = \begin{cases} \hat{H}_{3,a}^{k,d} & \text{if } k < (\gamma+2)(d+Q+4), \\ \emptyset & \text{otherwise,} \end{cases}$$
$$\hat{H}_{3,b}^{k,d} = \Big\{ h \in H : T^k(h) = \bar{h} \cdot \tilde{h} \text{ with } (1) \quad \bar{h} \in \hat{H}_{3,a}^{(\gamma+2)(d+Q+4),d},$$
$$(2) \text{ if } \tilde{h} = (\tilde{a}^1, \ldots, \tilde{a}^{\ell(\tilde{h})}) \text{ and } l(\tilde{h}) > 0, \text{ then } \tilde{a}_{-d}^1 = s'_{-d}, \text{ and}$$
$$(3) \text{ if } (\tilde{a}^\tau, \ldots, \tilde{a}^{\tau+Q}) \in \hat{\Sigma}^{l,(Q+1)} \text{ for some } \tau < \ell(\tilde{h}) - Q \text{ and } l \in N, \text{ then } \tilde{a}_{-d}^{\tau+Q+1} = s'_{-d} \Big\},$$

[46] Note that, for some of these parameters, the sets below may be empty.

$$H_{3,b}^{k,d} = \begin{cases} \hat{H}_{3,b}^{k,d} & \text{if } k < (\gamma+2)(d+Q+4)+T, \\ \emptyset & \text{otherwise,} \end{cases}$$

$H_4^{k,\omega,r} = \Big\{ h \in H : T^k(h) = \bar{h} \cdot \tilde{h}$ such that for some $d \in N$,

(1) $\ell(\bar{h}) = (\gamma+2)(d+Q+4)+T$, (2) $\bar{h} \in \hat{H}_{3,b}^{\ell(\bar{h}),d}$, (3) $\tilde{h} \in H_{1,b}^{\omega,r}$, and

(4) $\omega = \omega(d, T^T(\bar{h}))\Big\}$,

$H_5^{k,d,\tau} = \Big\{ h \in H : T^k(h) = \bar{h} \cdot a \cdot \tilde{h}$ such that for some $k' \le k$ and $d' \in N$

(1) $\bar{h}$ and a satisfy one the following conditions :

(1.a) $d' \neq d$ and $(l-1)(\gamma+1)+1 \le k' - (d'+Q+4)+1 \le l(\gamma+1)$

for some $1 \le l \le d'+Q+4$ and $\bar{h} \in H_{3,a}^{k',d'}$ and

$$a \in \begin{cases} \bar{D}_d(s_{d'}^*, s'_{-d'}) & \text{if } k'-(d'+Q+4)+1 = l(\gamma+1), \\ \bar{D}_d(\bar{a}^{(d')}) & \text{if } k'-(d'+Q+4)+1 \neq l(\gamma+1) \text{ and } \bar{a}^l \in \{s', s\} \\ \bar{D}_d(m^{d'}) & \text{if } k'-(d'+Q+4)+1 \neq l(\gamma+1) \text{ and } \bar{a}^l \notin \{s', s\}, \end{cases}$$

(1.b) $d \neq d'$ and $\bar{h} \in H_{3,b}^{k',d'}$ and if either $k' = (\gamma+2)(d+Q+4)$

or $k' > (\gamma+2)(d+Q+4)$ and $T^{Q+1}(\bar{h}) \in \cup_{l \in N} \hat{\Sigma}^{l,Q+1}$ for some l, then $a_d \neq s'_d$,

(1.c) $\bar{h} \in H_4^{k',\omega,r}$ for some $\omega = 1, \ldots, \bar{\omega}$ and $a \in \bar{D}_d(\hat{\pi}^{(\omega),r+1})$

for some $r < n+\omega+Q+5$,

(2) $\tilde{h} \in \Sigma^{d,\tau}$ and

(3) if $T^{d+Q+3}(\bar{h} \cdot a) = ((s';2),(s;d+Q),a)$ and $a \in \bar{D}_d(s)$, then $\ell(\tilde{h}) = 0\Big\}$.

We next define $H_{1,a} = \cup_{\omega=0}^{\bar{\omega}} \left(\cup_{k=n+\omega+Q+5}^{M} H_{1,a}^{(\omega),k} \right)$, $H_{1,b}^{(0),0} = \{H_0\}$, $H_{1,b} = \cup_{k=0}^{n+Q+4} H_{1,b}^{(0),k}$, $H_1 = H_{1,a} \cup H_{1,b}$, $H_2 = \cup_{k=1}^{M} \left(\cup_{d \in N} \left(\cup_{\tau=0}^{d+Q+3} H_2^{k,d,\tau} \right) \right)$, $H_{3,a} = \cup_{k=1}^{M} \left(\cup_{d \in N} H_{3,a}^{k,d} \right)$, $H_{3,b} = \cup_{k=1}^{M} \left(\cup_{d \in N} H_{3,b}^{k,d} \right)$, $H_3 = H_{3,a} \cup H_{3,b}$, $H_4 = \cup_{k=1}^{M} \left(\cup_{\omega=1}^{\bar{\omega}} \left(\cup_{r=0}^{n+\omega+Q+4} H_4^{k,\omega,r} \right) \right)$, and $H_5 = \cup_{k=1}^{M} \left(\cup_{d \in N} \left(\cup_{\tau=0}^{d+Q+3} H_5^{k,d,\tau} \right) \right)$.

Let $\tilde{\Sigma}^{d,\tau} = \{h \in H : T^{\tau+1}(h) = a \cdot \tilde{h}, \tilde{h} \in \Sigma^{d,\tau}$ and $a \in \bar{D}_d(s) \cup \bar{D}_d(s')\}$ for all $d \in N$ and $\tau \in \mathbb{N}_0$. Define, for all $d \in N$ and $\tau \in \{0, \ldots, d+Q+3\}$,

$$H_6^{d,\tau} = \left(H \setminus \cup_{l=1}^{5} H_l \right) \cap \tilde{\Sigma}^{d,\tau}.$$

Let $H_6 = \cup_{d \in N} \left(\cup_{\tau=0}^{d+Q+3} H_6^{d,\tau} \right)$. Also, for all $t \in \{0, \ldots, n+Q+4\}$, define

$$H_7^t = \{h \in H \setminus \cup_{l=1}^{6} H_l : T^t(h) \in H_{1,b}^{(0),t}\}.$$

The strategy f is now defined as follows. Let $h \in H$.

If $h \in H_{1,b}^{(0),k}$ for some $k \in \{0, \ldots, n+Q+4\}$, then $f(h) = \hat{\pi}^{(0),k+1}$.

If $h \in H_{1,a}^{(\omega),k}$ for some $\omega \in \{0, \ldots, \bar{\omega}\}$ and $k \in \{n+\omega+Q+5, \ldots, M\}$, then $f(h) = \hat{\pi}^{(\omega),k+1}$.

If $h \in \left(\cup_{k=1}^{M} \cup_{d \in N} \cup_{\tau=2}^{d+Q+2} (H_2^{k,d,\tau} \cup H_5^{k,d,\tau} \cup H_6^{d,\tau})\right) \cup \left(\cup_{t=2}^{n+Q+3} H_7^t\right)$, then $f(h) = s$.

If $h \in H_{3,a}^{k,d}$ for some $k \in \{1, \ldots, M\}$ and $d \in N$ and $k - (d+Q+4) = l(\gamma+1) - 1$ for some $1 \leq l \leq d+Q+4$, then $f(h) = (s_d^*, s'_{-d})$.

If $h \in H_{3,a}^{k,d}$ for some $k \in \{1, \ldots, M\}$ and $d \in N$ and $(l-1)(\gamma+1) \leq k - (d+Q+4) < l(\gamma+1) - 1$ for some $1 \leq l \leq d+Q+4$ and $T^k(h) = (a^1, \ldots, a^k)$, then $f(h) = \bar{a}^{(d)}$ if $a^l \in \{s', s\}$ and $f(h) = m^d$ if $a^l \notin \{s', s\}$.

If $h \in H_{3,b}^{k,d}$ for some $k \in \{1, \ldots, M\}$ and $d \in N$ then $f(h) = f^d(T^{k-(\gamma+2)(d+Q+4)}(h))$.

If $h \in H_4^{k,\omega,r}$ for some $k \in \{1, \ldots, M\}$ and $r \in \{0, \ldots, n+\omega+Q+4\}$ and $\omega \in \{1, \ldots, \bar{\omega}\}$, then $f(h) = \hat{\pi}^{(\omega),r+1}$.

Otherwise, $f(h) = s'$.

By construction, clearly, f has M-memory.

Lemma 3 *f is a well-defined SPE strategy profile and $\pi(f) = \hat{\pi}^{(0)}$.*

The proof of Lemma 3 is presented in the supplementary materials.

To complete the proof of Theorem 2 we need to show that $\|U(f) - u\| < \varepsilon$. But this follows from $\pi(f) = \hat{\pi}^{(0)}$ in Lemma 3 and

$$||V(\hat{\pi}^{(0)}, \delta) - u|| \leq ||V(\hat{\pi}^{(0)}, \delta) - x^0|| + ||x^0 - u|| < 2\xi < \varepsilon$$

(the second inequality follows from (47) and (48) and the third from $\xi < \varepsilon/2$).