REPEATED GAMES WITH INCOMPLETE INFORMATION OVER PREDICTABLE SYSTEMS

EHUD LEHRER AND DIMITRY SHAIDERMAN

Abstract. Consider a stationary process $\xi = (\xi_n)_{n \geq 1}$ taking values in a finite state space. Each state is associated with a finite one-shot zero-sum game. We investigate the infinitely repeated zero-sum game with incomplete information on one side, in which the state of the game evolves according to ξ. Two players, named the observer and the adversary, play the following game, denoted $\Gamma(\xi)$. At the beginning of any stage n, only the observer is informed of the state ξ_n, and therefore he is the only one who knows the identity of the forthcoming one-shot game. Then, both players take actions, which become publicly known. The paper shows the existence of a uniform value in a new class of stationary processes: ergodic Kronecker systems. Techniques from ergodic theory, probability theory and game theory are involved in the description of the optimal strategies of the two players.

Date: January 30, 2022
Classification numbers: 91A20, 91A27, 60G10, 37A50
Keywords: repeated games; incomplete information on one side; stationary processes; Kronecker systems; uniform value; irrational rotation of the unit circle; odometers

The first author acknowledges the support of grants ISF 591/21 and DFG KA 5609/1-1. Andrei Iacob, the two anonymous referees and the associate editor of MOR are greatly acknowledged for their comments.

1. Introduction

Uniform Value and Incomplete Information on One Side - Background The study of repeated games with incomplete information was initiated by Aumann and Maschler [2] in the 1960's. They consider a dynamic model where at the beginning of the game one of finitely many states is chosen according to a lottery whose laws are known to both players. Player 1 (the maximizer, or the observer) is informed of the chosen state, while Player 2 (the minimizer, or the adversary) is not. Each state is associated with a finite zero-sum game and the specific game corresponding to the chosen state is played repeatedly either finitely or infinitely many times, with the actions realized at each stage becoming publicly known.

Aumann and Maschler showed that the uniform value of the infinitely repeated game exists. Moreover, they provided an explicit description of optimal strategies of the two players. In particular, Aumann and Maschler characterized how much of his extra information the observer should reveal to his adversary in order to guarantee the value.

Ever since this pioneering work of Aumann and Maschler, repeated games with incomplete information on one side remain a flourishing mathematical field [19]. An important development in the field concerns the study of situations in which, different from what happens in the game of Aumann and Maschler, the state evolves stochastically over time.

Renault (2006) dealt with a game in which the state evolves according to a Markov chain whose laws are known to both players. Similarly to the Aumann and Maschler game, only the observer is informed at every stage of the realized state, and thus learns which game is being played at that stage. Renault (2006) proved that the uniform value exists. Neyman (2008) provided an alternative proof of Renault's result by constructing a reduction of the Markovian model to that of Aumann and Maschler. Computing the uniform value in the model of Renault (2006), even in the case of two states, turned out to be a difficult problem (see Hörner, Rosenberg, Solan, Vieille (2010) and Bressaud and Quas (2017)).

In general, the uniform value need not exist in every zero-sum repeated game with incomplete information on one side. Sorin (1984) described a stochastic game over two states with absorbing payoffs, similar to the 'big match' of Blackwell and Ferguson [4], in which one of the states is chosen once and for all. He proved that in the situation of incomplete information in which only the observer is informed of the chosen state (and thereby knows his stage payoffs), the uniform value need not exist. Another instance exhibiting a similar phenomenon involves a controller in a stochastic game, that is, a player whose actions at any state determine the transition rule of the states. Rosenberg, Solan and Vieille (2004) described a stochastic game with incomplete information on one side, with the adversary being the controller, in which the uniform value need not exist.

A first general affirmative result regarding the existence of a uniform value in games with incomplete information on one side and controllers was obtained by Renault (2012) in the framework of stochastic games with signals. He showed that whenever the controller is also fully informed of the state at every stage and of the signals his adversary receives, the uniform value exists. Finally, Gensbittel, Oliu-Barton, and Venel (2014) described the notion of a more informed controller (compared to the fully informed controller studied by Renault (2012)). They generalized the results of Renault (2012) by showing that the uniform value of any stochastic game with signals exists whenever one of the players is a more informed controller.

The Key Role of Stationary Dynamics When the game is played only finitely many times, it has a value that depends on its duration. In the model of Aumann and Maschler as well as in that of Renault (2006), the uniform value of the infinite game is equal to the limit (as the duration tends to infinity) of the values of the finitely repeated games. The analysis of these values is carried out by using classical notions such as non-revealing strategies, and tools such as martingales of posteriors and recursive formulas. It culminates with a closed formula for the limit of the values of the finite games.

A connection between the formula related to the limit of the values of the finitely repeated games and the infinitely repeated game in the models of Aumann and Maschler (1960's) and Renault (2006) is obtained by applying the following strategic reasoning. Consider an adversary's strategy that prescribes him to consecutively play optimal strategies of finite-duration games, while ignoring any information collected at any time such an optimal strategy starts. In other words, the adversary is prescribed to 'patch' optimal strategies of the finitely repeated games. The fact that future states still follow the same Markovian laws (potentially with a different initial probability) implies that, using this 'patching' strategy, the adversary can guarantee the limit of the values of the finitely repeated games.

In order for these strategies to work, the optimal strategies used should remain optimal despite the fact that they are played at different stages rather than at the beginning of the game. In other words, the dynamics according to which the states evolve should be invariant under time translations. Equivalently, in probabilistic terms, the states should follow a stationary process.[1] This observation leads us to analyze repeated games based on general stationary processes.

The Research Question Consider a stationary process $\xi = (\xi_n)_{n\geq 1}$ that attains finitely many states. Similarly to Renault (2006), at any stage n, only the observer observes the realized state ξ_n (thus is informed of the one-shot zero-sum game to be played), and both players take actions that become publicly known. We denote the corresponding infinite game by $\Gamma(\xi)$ and the value of the n-stage game by $v_n(\xi)$.

Section 2 shows that the mere stationarity assumption implies that the patching strategies of the adversary guarantee that $\lim_{n\to\infty} v_n(\xi)$ exists. Moreover, the adversary can guarantee it in $\Gamma(\xi)$. At this point the following open problem arises naturally:

$$\text{Can the observer guarantee } \lim_{n\to\infty} v_n(\xi) \text{ in } \Gamma(\xi)?$$

[1] A Markov chain over finitely many states is stationary if and only if the initial distribution over the states of the chain is an invariant distribution for the transition matrix. However, any Markov chain becomes almost stationary (up to a period) after a while and therefore patching strategies can be tailored to guarantee the limit of the finite values.

An affirmative answer to this question would imply that the uniform value of $\Gamma(\xi)$ exists and equals to $\lim_{n\to\infty} v_n(\xi)$.

Can the observer guarantee $\lim_{n\to\infty} v_n(\xi)$ by employing similar patching strategies? The answer is negative. By playing optimally in some finite-duration games the observer might reveal important information about future steps of ξ that the adversary might use later. Therefore, playing patching strategies by the observer might lead to an inefficiency that would reduce his ability to guarantee the values of finite-duration games in the future.

Informational Aspects and Predictive Properties Another path towards answering the open problem posed above it to understand the tradeoff between exploiting the extra information the observer has in order to increase his payoff and the downside of loosing an edge by revealing valuable information. In other words, the task is to find the observer's optimal ways to use his private information. The private information of the observer up to stage n consists of the realization of $\xi_1, ..., \xi_n$. At this stage, the adversary is interested only in the conditional laws of future states ξ_ℓ, $\ell > n$, as they determine the games to be played in the future. This implies that the observer must take into account the predictive properties of ξ whenever he decides to reveal information. In the Aumann-Maschler model, for instance, the same game is played over and over again, and therefore knowing the past history of states enables the observer to fully predict the future. In the Markovian model, on the other hand, the system returns to an invariant distribution even if the informed player uses his extra knowledge and thereby (partially) reveals the identity of the realized state. Thus, knowing the past history of states might provide a partial predictive power for the short run, but almost nothing for the long run.

The prediction and inference abilities of an observer of a stationary process lie at the heart of many mathematical disciplines such as Information Theory, Ergodic Theory and Probability [12].

It is now evident that, based on the predictive properties of ξ, the observer should identify the information regarding the process ξ of interest to his adversary. The classic method that quantifies the dynamics of information revelation is the martingale of beliefs technique: the observer finds a martingale of beliefs that summmarizes the information which is of interest to his adversary, and then manipulates its evolution through his actions to his advantage. It turns out that this kind of martingale is not unique. Indeed, the proofs of Renault (2006) and Neyman (2008) show that in the Markovian model there exist at least two such martingales.

A Reformulation via Ergodic Theory A basic result from ergodic theory, establishes a correspondence between a stationary process that attains finitely many states and a finite partition of a measure-preserving system (see Subsection 3.1). The correspondence is achieved by considering the following auxiliary (stationary) process that arises from a finite partition of the measure-preserving system: at each time period, the process reveals the partition element containing the current position of the system. The research question may thus be reformulated in terms of measure-preserving systems. Indeed, as different finite partitions of a given measure-preserving system give rise to different stationary processes, one could first fix a measure preserving system, and then investigate the research question for each and every one of its finite partitions. Thus, an affirmative answer for a given measure-preserving system covers a wide range of stationary processes at once.

Main Results In this paper we explore the research question for a fundamental class of measure-preserving systems, called ergodic Kronecker systems (see Furstenberg (1981)). Examples of ergodic Kronecker systems include the irrational rotation of the unit circle and odometers. Our main result (see Theorem 1) shows that if ξ arises from a finite partition of an ergodic Kronecker system, then the uniform value of $\Gamma(\xi)$ exists and equals $\lim_{n\to\infty} v_n(\xi)$. In Theorem 2 we provide a formula for the uniform value, presenting it as a limit of concave envelopes of auxiliary functions. Based on this formula and a new game-theoretic result by Ashkenazi, Solan and Zseleva (2020), we prove a new limit theorem for ergodic Kronecker systems, Theorem 3.

Why ergodic Kronecker Systems? In relation to dynamic games, the interest in ergodic Kronecker systems comes from the fact that their finite partitions generate ergodic processes with perishing information. We expand on these two properties, emphasizing theirs relation to the Aumann and Maschler (1960's) and Renault (2006) models.

I. Ergodicity. The ergodicity property of processes generated by (partitions of) ergodic Kronecker systems reflects the fact they are constantly changing, as opposed to the motionless process considered by Aumann and Maschler (1960's). This property implies that any possible one-shot game is, almost surely, played infinitely often. Moreover, the order in which the possible one-shot games occur during the course of time is almost surely non-periodic, thus having a chaotic nature. Those features are also common to the ergodic Markov chains (i.e., irreducible and aperiodic) covered in Renault (2006).

II. Perishing Information. The processes generated by ergodic Kronecker systems are highly predictable. That is, after a long enough history of state realizations the observer learns to predict, with a high precision, which states will show up in future stages. It means that in such systems, the information of interest to both players lies in long histories of observations. Any exploitation of such information by the observer may have consequences for the entire future. This informational aspect is also shared by Aumann and Maschler's model and thus suggests that the observer should split the beliefs of the adversary over long histories, taking into account a certain underlying target function.

This informational aspect stands in polar opposite to the nature of ergodic Markov chains. Indeed, in the latter the information is generated anew in the sense that the distribution of the steps of the chain converges to the invariant distribution at a high rate, regardless of the initial probability of the chain. This allows the observer to exploit his extra information to his benefit for any finite duration he desires, without having to worry about future consequences of his use of private information. Indeed, if the observer refrains from using his information, by playing for a short while[2] a non-revealing strategy, the adversary posterior belief regarding the next step of the chain becomes

[2]This time depends only on the Markovian transition rule, and not on the strategy used by the observer.

very close to its' invariant distribution, and thus is unaffected by past information revealed by the observer.

III. Bridging the gap between the players. The problem of existence of a uniform value in ergodic Kronecker systems shares common aspects with the works of both Aumann and Maschler (1960's) and of Renault (2006). Ergodic Kronecker systems share similar dynamical aspects (such as ergodicity) with the latter, and others (predictability, perishing information) with the former. The main difficulty ergodic Kronecker systems pose, however, is to connect the target function used by the observer for splitting and the limit of the values of the n-stage games, which the adversary can guarantee. The difficulty to make this connection is rooted in the chaotic dynamics of the games under consideration: for the observer the question is how much information to use after observing a long history of realized states, while for the adversary it is how to exploit the information revealed by the observer. It turns out that such a connection can be made due to specific topological and stochastic properties that ergodic Kronecker systems posses, which both players have to utilize to their own benefit.

Key ingredients of the proof

I. Long-Term Predictability. Using classical tools from topological dynamics we identify the specific predictive property of the processes ξ that is used for the strategic analysis. Roughly speaking, this property, termed long-term predictability (see Definition 3), states that with a high probability one can, given a large enough number K of observations, predict the future correctly over a set of stages that has a high density. This property, being a distributional property of ξ, is thus commonly known among the observer and the adversary.

II. The Predictor Game. The long-term predictability property implies that one can approximate $\Gamma(\xi)$ with an auxiliary game resembling that of Aumann and Maschler. In this auxiliary game a state is defined as the set of states consisting of the history of realizations of $(\xi_1, ..., \xi_K)$ in the first K stages, that occur with positive probability. Each such state (again, it is a finite string of states of the original game) is associated with an infinite (deterministic) sequence of states, so that the long-term predictability property is satisfied with respect to those sequences of states. Such sequences can be thought of as predictions made by the observer for future states based on histories of length K. If we adopt this point of view, then this auxiliary game depicts a situation in which the observer alternates the dynamics of $\Gamma(\xi)$ by assuming that the state at time $K + \ell$ (in $\Gamma(\xi)$) is the one corresponding to the ℓ-th state in his respective prediction, rather then $\xi_{K+\ell}$. The game, called a predictor game, begins with a random choice of a state, selected according to the distribution induced by $(\xi_1, ..., \xi_K)$. Once a state is realized, the incomplete information game evolves deterministically: the one-shot games played at future stages are determined by the predictions associated with this state.

III. Splitting in the Predictor Game à la Aumann and Maschler. The predictor game differs from the classic Aumann-Maschler game in that the one-shot games played after stage K are not identical and do not occur in a periodic manner: they posses a chaotic nature. As in the Aumann-Maschler game, once a state is chosen, the entire future is determined. Moreover, as in the

Aumann-Maschler game, by using non-revealing strategies one obtains a 'value' function defined on the set of probability distributions over the set of states. It is important to note that the predictions satisfying the long-term predictability property after K stages are by no means unique, and that the 'value' function depends on the particular set of predictions (one for each history) chosen. The similarity between the predictor game and that of Aumann and Maschler is also reflected in that the observer can employ a splitting technique in the predictor game, to obtain the concave envelope of each such 'value' function. We show that the long-term predictability property suffices to ensure that the observer can guarantee the limit (as $K \to \infty$) of those concave envelopes in the original game $\Gamma(\xi)$.

IV. Martingale Analysis in the Predictor Game. In the predictor game the martingale of beliefs consists of the posterior probabilities the adversary assigns to the chosen state given the observer's actions. The analysis of this martingale provides upper bounds to the values of finitely repeated predictor games which depend on the predictions (i.e., the infinite sequence of states) the adversary associates to the histories of length K.[3] Based on the long-term predictability property, those upper bounds are then linked to the behavior of the $v_n(\xi)$'s.

V. The Main Difficulty. The main difficulty in the proof is to connect between the upper bounds on the $v_n(\xi)$'s, which take into account the predictions of the adversary, and the limit of the concave envelopes of the non-revealing 'value' functions which are based on the predictions used by the observer. In the models of Aumann and Maschler (1960's) and Renault (2006) the proper martingale and the game-theoretic analysis based on non-revealing strategies were sufficient to make this connection. Here, however, these are insufficient.

VI. Solution: Back to Ergodic Theory. In order to address the main difficulty, it is necessary to bridge between the predictions made by the two players. We do so by employing tools from topological dynamics and ergodic theory. The key finding (see Theorem 5) is that one can associate each point in the ergodic Kronecker system with predictions that follow histories $\xi_1, ..., \xi_K$ of length K, so that for a set of points of positive measure, the corresponding predictions satisfy the long-term predictability property. Consequently, in view of the ergodic properties of the system, the players can simultaneously select points from this set of positive measure for which the corresponding predictions satisfy essential regularity conditions, global and local, over the set of stages. By an appropriate selection of points we prove the desired connection between $\lim_{n\to\infty} v_n(\xi)$ and the limit of concave envelopes of the 'value' functions. Ergodic theory, probability theory and game theory are all involved in the description of the optimal use of information by the observer, which is one of the main insights of this paper.

Structure of the Paper The paper is organized as follows. In Section 2 we give a formal description of our model. Section 3 is devoted to a review of basic notions and results from Ergodic Theory and Dynamical Systems.

[3] As the adversary knows ξ, he in particular knows those infinite sequences of states whose association to K histories satisfies the long-term predictability property.

The main results of the paper are stated in Section 4. This section starts with a formal exposition to ergodic Kronecker systems and proceeds with the statements and discussion of the main results, given in Theorems 1, 2, and 3.

The proofs in the paper involve many technical arguments. Therefore, we urge the interested reader to begin with the sketch of the proof, provided in Section 5. After becoming familiar with the main ideas, tools, and difficulties, the transition to Section 6, the one that provides the detailed proofs, should be easier.

Finally, we have devoted two Appendices to the proofs of important technical steps involved in the proofs of our main results. Appendix A starts with a review of basic definitions and results from topological dynamics, and proceeds with all the ergodic-theory related proofs in the paper. Appendix B is devoted to game-theoretic proofs.

2. The Model

This section introduces the zero-sum repeated game with incomplete information on one-side based on a stationary process.

Definition 1. A *stationary process* ξ is a sequence of real-valued random variables $(\xi_n)_{n\geq 1}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that:

$$\mathbb{P}((\xi_1, \xi_2, ...) \in B) = \mathbb{P}((\xi_k, \xi_{k+1}, ...) \in B), \quad \forall k \geq 1, \forall B \in \mathcal{B}(\mathbb{R}^\infty).$$ [4]

We proceed to the description of the zero-sum repeated game with incomplete information on one side $\Gamma(\xi)$. This game is specified by a 5-tuple $\langle A, \xi, I, J, g\rangle$, where (i) A is a finite set of states, (ii) $\xi = (\xi)_{n\geq 1}$ is a stationary process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ whose steps take values in A, (iii) I and J are finite action sets of the observer and the adversary, respectively, and (iv) $g : A \times I \times J \to \mathbb{R}_+$ is a payoff function.

The game $\Gamma(\xi)$ proceeds as follows. A point $\omega \in \Omega$ is chosen at random according to $\mathbb{P}$. Then, at each stage $n \in \mathbb{N}$, the observer is informed of $\xi_n(\omega) \in A$ and the players choose actions $i_n \in I$ and $j_n \in J$, which are then publicly announced and become known to both players. The stage payoff to the observer equals $g(\xi_n(\omega), i_n, j_n)$. It is assumed that the description of the model is known to both players.

Since the players have perfect recall, by Aumann (1964) (e.g., [3]), we may restrict ourselves to behavioral strategies. A *behavioral strategy* σ of the observer is a sequence of stage strategies $(\sigma_n)_{n\geq 1}$, such that $\sigma_n : (A \times I \times J)^{n-1} \times A \to \Delta(I)$.[5] That is, at stage n, based on his information (i.e., $(\xi_1(\omega), i_1, j_1, ..., \xi_{n-1}(\omega), i_{n-1}, j_{n-1}, \xi_n(\omega))$), the observer chooses a mixed action. Similarly, a *behavioral strategy* τ of the adversary consists of sequences of stage strategies $(\tau_n)_{n\geq 1}$, such that $\tau_n : (I \times J)^{n-1} \to \Delta(J)$.

Denote by Σ (resp., $\mathcal{T}$) the set of behavioral strategies of the observer (resp. adversary). Consider the measurable space $\mathcal{O} = \Omega \times (I \times J)^{\mathbb{N}}$. The Ionescu-Tulcea extension theorem (see Neveu, 1970,

[4] For a topological space X we denote by $\mathcal{B}(X)$ the Borel σ-field on X and $\mathbb{R}^\infty$ is endowed with the product topology.

[5] For a finite set C we define $\Delta(C)$ to be the simplex of probability distributions on the elements of C.

Proposition V.1.1) implies that each pair $(\sigma,\tau)\in\Sigma\times\mathcal{T}$, together with the process $(\xi_n)_{n\geq 1}$ induce a probability measure $\mathbb{P}^{\xi}_{\sigma,\tau}$ on $\mathcal{O}$. Indeed, for every $A\in\mathcal{B}(\Omega)$ and every finite history of actions $h=(i_1,j_1,...,i_N,j_N)$, the measure $\mathbb{P}^{\xi}_{\sigma,\tau}$ is uniquely determined by the laws

$$\begin{aligned}\mathbb{P}^{\xi}_{\sigma,\tau}(A\times\{h\}) &= \int_A \Bigg(\prod_{n\leq N}\tau_n(i_1,j_1,...,i_{n-1},j_{n-1})[j_n] \\ &\times \prod_{n\leq N}\sigma_n(\xi_1(\omega),i_1,j_1,...,\xi_{n-1}(\omega),i_{n-1},j_{n-1},\xi_n(\omega))[i_n]\Bigg)\,d\mathbb{P}(\omega).\end{aligned}$$

We denote by $\Gamma_N(\xi)$ the N-stage game in which the payoff corresponding to the pair $(\sigma,\tau)\in\Sigma\times\mathcal{T}$ is the expected average payoff, i.e.,

$$\gamma_N(\sigma,\tau)=\mathbb{E}^{\xi}_{\sigma,\tau}\left(\frac{1}{N}\sum_{n=1}^{N}g(\xi_n,i_n,j_n)\right),$$

where $\mathbb{E}^{\xi}_{\sigma,\tau}$ is the expectation operator w.r.t. $\mathbb{P}^{\xi}_{\sigma,\tau}$. As $\Gamma_N(\xi)$ can be viewed as a finite game in extensive form, its value, denoted by $v_N(\xi)$, exists. The standard notion of value for the infinitely repeated game is that of the uniform value.

The scalar $v\in\mathbb{R}_+$ is called the *uniform value* of $\Gamma(\xi)$ if the following two conditions are satisfied:

1. The observer *can guarantee* v, i.e.,

$$\forall\varepsilon>0,\ \exists\,\sigma_\varepsilon\in\Sigma,\ \exists N_0\in\mathbb{N},\quad \text{s.t. } \gamma_N(\sigma_\varepsilon,\tau)\geq v-\varepsilon,\quad \forall N>N_0,\ \forall\tau\in\mathcal{T}.$$

2. The adversary *can guarantee* v, i.e.,

$$\forall\varepsilon>0,\ \exists\,\tau_\varepsilon\in\mathcal{T},\ \exists N_0\in\mathbb{N},\quad \text{s.t. } \gamma_N(\sigma,\tau_\varepsilon)\leq v+\varepsilon,\quad \forall N>N_0,\ \forall\sigma\in\Sigma.$$

If the uniform value of $\Gamma(\xi)$ exists, we denote it by $v(\xi)$. In that case, a strategy $\sigma\in\Sigma$ of the observer is said to be *optimal* if it satisfies

$$\forall\varepsilon>0,\ \exists N_0\in\mathbb{N},\quad \text{s.t. } \gamma_N(\sigma,\tau)\geq v(\xi)-\varepsilon,\quad \forall N>N_0,\ \forall\tau\in\mathcal{T},$$

whereas a strategy $\tau\in\mathcal{T}$ of the adversary is said to be an *optimal strategy* if it satisfies

$$\forall\varepsilon>0,\ \exists N_0\in\mathbb{N},\quad \text{s.t. } \gamma_N(\sigma,\tau)\leq v(\xi)+\varepsilon,\quad \forall N>N_0,\ \forall\sigma\in\Sigma.$$

Remark 1. *The assumption that the observer and his adversary know* $(\Omega,\mathcal{B}(\Omega),\mathbb{P})$ *can be relaxed to the one that both of them are only familiar with the distribution laws of* ξ. *Indeed, the uniform value takes into account only the finite duration payoffs* $\gamma_N(\sigma,\tau)$, $N\geq 1$, *and the latter are determined by the distribution laws of* ξ, *and not the pointwise laws* $\xi(\omega)$, $\omega\in\Omega$.

Remark 2. *The game* $\Gamma(\xi)$ *can be viewed as a repeated game with more informed observer (e.g., Subsection* 3.3. *of* [9]) *in which the state space is infinite. Indeed, consider the state space* $K=A^{\mathbb{N}}$ *of infinite sequences over the alphabet* A. *The evolution of the state over time is described by the deterministic motion* $k_{n+1}=T_{\leftarrow}k_n$, *where* $T_{\leftarrow}:A^{\mathbb{N}}\to A^{\mathbb{N}}$ *is the* backward shift operator *defined by* $T_{\leftarrow}(a_1,a_2,a_3,...):=(a_2,a_3,a_4,...)$. *The actions sets are* I,J *as before. The prior probability* π

is given by $\pi(B) = \mathbb{P}(\xi \in B)$ *for every* $B \in \mathcal{B}(A^{\mathbb{N}})$*. The private signal of the observer at stage* n*, denoted* s_n*, is independent of the actions played and equals* a_n *if the state* $(a_1, a_2, ..., a_n, ...) \in A^{\mathbb{N}}$ *was chosen initially according to* π*. Since the signals* $(s_n)_{n\geq 1}$*, defined on the Borel probability space* $(A^{\mathbb{N}}, \pi)$*, have the same distribution as* ξ *(e.g., p. 405 in [22]), Remark 1 implies that such a game is equivalent to* $\Gamma(\xi)$*.*

The following cornerstone observation inspired our research.

Observation 1. *The sequence* $(v_n(\xi))_{n\geq 1}$ *converges.*

Proof. We claim that the sequence $(nv_n(\xi))_{n\geq 1}$ is sub-additive, i.e., $(\ell+m)v_{\ell+m}(\xi) \leq \ell v_\ell(\xi)+mv_m(\xi)$ for all $\ell, m \in \mathbb{N}$. Indeed, consider the following strategy $\tau^* \in \mathcal{T}$ of the adversary: play an optimal strategy of $\Gamma_\ell(\xi)$ along the first ℓ stages, then, regardless of the past history of actions observed, play an optimal strategy of $\Gamma_m(\xi)$ in the following m stages; lastly, at every stage $n \geq \ell+m$, τ^* instructs the adversary to play the same arbitrary action $j_0 \in J$. The stationarity of ξ implies that under τ^* the payoff $\gamma_{\ell+m}(\sigma, \tau^*)$ in $\Gamma_{n\ell+m}(\xi)$ is less than $(\ell v_\ell(\xi)+mv_m(\xi))/(\ell+m)$ across all $\sigma \in \Sigma$. Therefore, by the minimax theorem applied to $\Gamma_{\ell+m}(\xi)$, we obtain $v_{\ell+m}(\xi) \leq (\ell v_\ell(\xi)+mv_m(\xi))/(\ell+m)$, proving the sub-additivity of $(nv_n(\xi))_{n\geq 1}$. In particular, we obtain that $\lim_{n\to\infty} v_n(\xi)$ exists.[6]

□

The patching strategy $\tau_p \in \mathcal{T}$ of the adversary is defined as follows: play an optimal strategy of $\Gamma_1(\xi)$, then regardless of the past history of actions observed, play an optimal strategy of $\Gamma_2(\xi)$, etc. Described differently, for every $n \in \mathbb{N}$, play an optimal strategy of $\Gamma_n(\xi)$ along the stages $n(n-1)/2+1, ..., n(n+1)/2$. Observation 1 is sufficient to deduce that:

Corollary 1. *For every* $\varepsilon > 0$ *there exists* $N_0 \in \mathbb{N}$ *such that*

$$\gamma_N(\sigma, \tau_p) \geq \lim_{n\to\infty} v_n(\xi) - \varepsilon, \quad \forall N > N_0, \forall \sigma \in \Sigma. \tag{2.1}$$

In particular, Corollary 1 implies that the adversary can guarantee $\lim_{n\to\infty} v_n(\xi)$ in $\Gamma(\xi)$. Let ξ be a stationary process taking values in a finite set A. It is natural to pose the following three open questions.

Main Questions. *Let* ξ *be a stationary process taking values in a finite set* A*.*

Question 1. *Can the observer guarantee* $\lim_{n\to\infty} v_n(\xi)$ *in* $\Gamma(\xi)$*?*

Question 2. *Does the uniform value of* $\Gamma(\xi)$ *exists?*

Question 3. *Is it possible that the uniform value exists and is strictly less then* $\lim_{n\to\infty} v_n(\xi)$*?*

In view of Observation 1, a positive answer to Question 1 implies that $\lim_{n\to\infty} v_n(\xi)$ is the uniform value of $\Gamma(\xi)$, giving a positive answer to Question 2. On the contrary, if the answer to Question 3 is positive, then an affirmative answer to Question 2 would in some cases (those processes ξ for which the answer to Question 3 is positive) imply a negative answer to Question 1.

[6] It is known that if $(a_n)_{n\geq 1}$ is sub-additive, then $\lim_{n\to\infty} \frac{a_n}{n}$ exists.

In this paper, we will address these problems by reformulating them in terms and notions from Ergodic Theory. In the next section we recall important facts related to measure-preserving systems, survey the relations between stationary process taking finitely many values and finite partitions of measure-preserving systems, and finally, review some basic definitions and results from Ergodic Theory and Topological Dynamics.

3. Preliminaries from Ergodic Theory and Dynamical Systems

3.1. **Measure-Preseving Systems and Stationary processes.** A *measure-preserving system* consists of a 4-tuple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ is an invertible, measurable, measure-preserving transformation, that is,

$$T^{-1}B \in \mathcal{B} \text{ and } \mu(T^{-1}B) = \mu(B) \ \ \forall B \in \mathcal{B},$$

where $T^{-1}B = \{x \in X : Tx \in B\}$.

We shall now describe how every finite partition of a measure-preserving system gives rise to a stationary process. We first need some notations. The *orbit of* $x \in X$ is defined to be the set $\{T^{(n-1)}x : n \in \mathbb{N}\}$. Consider now a finite measurable partition $\mathcal{P} = \{P_1, ..., P_k\}$ of X. The n'th step, $n \in \mathbb{N}$, of the *process generated by* $(\mathcal{P}, T)$ is described by the formula

$$\xi_n(x) = \sum_{a=1}^{k} (a-1) \mathbb{1}\{T^{-(n-1)} P_a\}(x) = \sum_{a=1}^{k} (a-1) \mathbb{1}\{P_a\}(T^{(n-1)}x), \quad x \in X. \tag{3.1}$$

Consider now an observer of this process. Prior to his first observation, the observer's belief is that a point $x \in X$ was chosen at random according to the law μ. By observing ξ_n he learns in which of the partition elements $\{P_1, ..., P_k\}$ the orbit of the chosen point $x \in X$ lies at time n. To simplify notation we write $\xi = (\mathcal{P}, T)$, where $\xi = (\xi_n)_{n \geq 1}$ is defined by (3.1). The following simple identity, which follows immediately from (3.1), will be repeatedly used throughout the paper:

$$\xi_{n+m}(x) = \xi_n(T^m x), \quad \forall n, m \in \mathbb{N}, \ \forall x \in X. \tag{3.2}$$

Since T is measure-preserving, the process $(\xi_n)_{n \geq 1}$ is stationary. Conversely, each stationary process $\xi = (\xi_n)_{n \geq 1}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ whose steps take value in a finite alphabet A can be regarded as being generated by a finite partition of a measure-preserving system. Indeed, consider $A^{\mathbb{N}}$ equipped with the product topology, and let $\mathcal{B}(A^{\mathbb{N}})$ be the Borel σ-field on $A^{\mathbb{N}}$. Let the probability measure μ_ξ be the push forward of $\mathbb{P}$ on $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}))$ by ξ, that is:

$$\mu_\xi(B) = \mathbb{P}((\xi_1, \xi_2, ...) \in B) \ \ \forall B \in \mathcal{B}(A^{\mathbb{N}}).$$

Together with the backward shift operator $T_\leftarrow$ defined in Remark 2, the system $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu_\xi, T_\leftarrow)$ is measure-preserving. The stationary process $\xi = (\xi_n)_{n \geq 1}$ is distributed as the stationary process generated by $(\mathcal{P}, T_\leftarrow)$, where $\mathcal{P} = \{\{[a]\}_{a \in A}\}$ and $[a] = \{(a_1, a_2, ...) \in A^{\mathbb{N}} : a_1 = a\}$.

To summarize the discussion, we may reformulate the main research questions in terms of measure-preserving systems and theirs finite partitions. For instance, Question 2 can now be reformulated as follows: Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, and let $\mathcal{P}$ be a finite measurable partition of X. Does the uniform value of $\Gamma(\xi)$ exists, where $\xi = (\mathcal{P}, T)$?

3.2. **Basic facts from Ergodic Theory.** The measure-preserving system $(X, \mathcal{B}, \mu, T)$ is said to be *ergodic* if for every $A, B \in \mathcal{B}$ such that $\mu(A) > 0$ and $\mu(B) > 0$ there exists an $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-n}B) > 0.$$

That is, if we look at T as a transformation describing the time evolution of the system, then in ergodic systems, regardless of which positive-measure set $A \in \mathcal{B}$ one starts with, for each positive measure set $B \in \mathcal{B}$ there exists a time $n \in \mathbb{N}$ at which a positive-measure set of points from A will visit B. An equivalent definition of ergodicity is the following: a set $A \in \mathcal{B}$ is called *invariant* if $T^{-1}A = A$; then, the measure-preserving system $(X, \mathcal{B}, \mu, T)$ is ergodic if and only if $\mu(A) \in \{0, 1\}$ for every invariant set A. We say that the stationary process $(\xi_n)_{n\geq 1}$ taking values in a finite alphabet A is *ergodic*, whenever the measure-preserving system $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu_\xi, T_{\leftarrow})$ is ergodic.

A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is said to be *uniquely ergodic* if μ is the only probability measure for which $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system; that is, if $(X, \mathcal{B}, \nu, T)$ is measure-preserving then $\nu = \mu$. By the Ergodic Decomposition Theorem (e.g., Theorem 6.2. in [6]), every uniquely ergodic system is ergodic.

A central tool from ergodic theory that will play a key role in the paper is Birkhoff's Pointwise Ergodic Theorem:

Theorem (Birkhoff's Pointwise Ergodic Theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system. Then for any $f \in L^1(X, \mathcal{B}, \mu)$ we have*

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{(n-1)}x) \to \int f d\mu \quad \textit{as} \quad N \to \infty \quad \mu\textit{-a.e. and in } L^1(X, \mathcal{B}, \mu). \tag{3.3}$$

Conversely, let us now describe how a relatively simple topological structure suffices to produce a measure-preserving system.

3.3. **Topological Dynamical Systems.** A *topological dynamical system* (X, T) consists of a compact metric space (X, d) and a homeomorphism $T : X \to X$ of X onto itself. The Krylov-Bogolyubov Theorem asserts that there exists a Borel probability measure μ that is supported on X such that $(X, \mathcal{B}(X), \mu, T)$ is a measure-preserving system. Two topological dynamical systems (X, T) and $(\mathcal{Y}, S)$ are said to be *topologically conjugate* if there exists a homeomorphism $\phi : X \to \mathcal{Y}$ such that $\phi \circ T = S \circ \phi$.

4. The Main Results

Definition 2. A topological dynamical system (X, T) is said to be a *Kronecker system* if it is topologically conjugate to a system of the form (G, S_g), where $(G, +)$ is a compact metrizable Abelian topological group and $S_g : h \mapsto g + h$ is the group rotation by $g \in G$.

Kronecker systems enjoy many special properties and are considered among the most fundamental systems studied in Ergodic Theory[7]. By Haar's Theorem, every Kronecker system (X, T) can be viewed as a measure-preserving system $(X, \mathcal{B}(X), T, \mu)$, where μ corresponds to the Haar measure. If such a system is ergodic, then it is well known that it must also be uniquely ergodic. In such a case we say that (X, T) is an *ergodic Kronecker system*. The following examples are all ergodic Kronecker systems.

Example 1. *The irrational rotation of the circle* $(\mathbb{T}, R_\alpha)$. We identify the unit circle $\mathbb{T}$ with the compact metrizable Abelian group[8] $(\mathbb{R}/\mathbb{Z}, +)$, and $R_\alpha : \mathbb{T} \to \mathbb{T}$, defined by $R_\alpha(x) := x + \alpha$, is the rotation by $\alpha \in \mathbb{T}$. The Haar measure μ is the Lebesgue measure on $\mathbb{T}$. It is well known that the Kronecker system $(\mathbb{T}, R_\alpha)$ is ergodic if and only if $\alpha \in \mathbb{T} \setminus \mathbb{Q}$, i.e., if α is irrational. We will now show the necessity of the latter, and for the sufficiency refer the reader to Proposition 2.16. on p. 26 in Einsiedler and Ward [6].

Assume that $\alpha = m/n + \mathbb{Z}$. Consider the interval (arc) $I = [0, \frac{1}{2n}) + \mathbb{Z} \subset \mathbb{T}$. Define

$$A = I \cup R_\alpha^{-1} I \cup \cdots \cup R_\alpha^{-(n-1)} I.$$

Since R_α^n is the identity mapping on $\mathbb{T}$, we see that A is invariant. Thus, as $\frac{1}{2n} = \mu(I) < \mu(A) < n\mu(I) = 1/2$, we deduce that $(\mathbb{T}, R_\alpha)$ is not ergodic whenever $\alpha \in \mathbb{Q}$.

Example 2. *Odometers* (O, σ). Let $\langle k_n : n \in \mathbb{N} \rangle$ be a sequence of integers greater then or equal to 2. Let $\mathbb{Z}_{k_n} = \mathbb{Z}/k_n\mathbb{Z}$ be equipped with the discrete topology. Consider the product space

$$O = \prod_{n=1}^{\infty} \mathbb{Z}_{k_n},$$

equipped with the product topology. Let $+_*$ be the addition and "carry to the right" binary operation on O. Formally, for any two element $a = (a_n)_{n \geq 1}$ and $b = (b_n)_{n \geq 1}$ in O we define the sequence of binary remainders $r = (r_n)_{n \geq 0}$ by the following recursive law:

$$r_n = \begin{cases} 0, & \text{if } n = 0 \\ \mathbb{1}\{a_n + b_n + r_{n-1} \geq k_n\}, & \text{if } n \geq 1 \end{cases}.$$

We now define for $n \geq 1$ the n'th coordinate, $(a +_* b)_n$, of $a +_* b$ to equal $(a_n + b_n + r_{n-1}) \bmod k_n$. To illustrate the formal definition, let us take $k_n = 3$ for every n, and the two elements $a =$

[7] Furstenberg's seminal multiple recurrence theorem, proved in 1977, states that in a measure-preserving system $(X, \mathcal{B}, \mu, T)$, for every $A \in \mathcal{B}$ with $\mu(A) > 0$ and every $k \geq 1$, there exists $n \in \mathbb{N}$ such that

$$\mu\left(A \cap T^{-n} A \cap T^{-2n} A \cap \cdots \cap T^{-kn} A\right) > 0.$$

Showing that the latter implies the famous Szemerédi theorem, gave rise to a branch of mathematics now known as Ergodic Ramsey Theory. The remarkable proof of Furstenberg was done by a sequence of reductions designed to lower at each step the complexity of the systems obtained along the sequence (see Chapter 7 in [6]). At the basis of this reductions stand ergodic Kronecker systems (see Proposition 7.12. on p. 189 in [6] for a proof of Furstenberg's multiple recurrence theorem for ergodic Kronecker systems).

[8] The binary addition operator $+$ on $\mathbb{T}$ is defined as follows; If $x = (a + \mathbb{Z}) \in \mathbb{T}$ and $y = (b + \mathbb{Z}) \in \mathbb{T}$, then $x + y := (a + b) + \mathbb{Z} \in \mathbb{T}$.

$(2,1,0,1,2,0,0,...)$ and $b=(2,1,2,2,2,0,0,...)$ in O. The reader may verify easily that $a +_* b = (1,0,0,1,2,1,0,0,...)$.

We have that $(O,+_*)$ is a compact metrizable Abelian topological group. Denote by $\bar{1}$ the group element $(1,0,0,...) \in O$. Let $\sigma : O \to O$ be the group rotation by $\bar{1}$, i.e., $\sigma : a \mapsto a +_* \bar{1}$ for every $a \in O$. The transformation σ obeys the following rules:

(i) $\sigma(k_1-1, k_2-1, ...) = (0,0,...)$ if each coordinate corresponding to $\mathbb{Z}_{k_n}$ equals $k_n - 1$.

(ii) $\sigma(k_1-1, k_2-1, ..., k_n-1, x_1, x_2, ...) = (0,...,0,x_1+1,x_2,...)$ when $x_1 \neq k_{n+1}-1$. Note that x_1+1 appears in the $(n+1)$'st coordinate of $(0,...,0,x_1+1,x_2,...)$.

The Haar measure μ on O is the product measure $\bigotimes_{n=1}^{\infty} \nu_i$, where ν_i is the probability measure on $\mathbb{Z}_{k_i}$ that assigns to each point a $1/k_i$ distribution mass. For more information and new results on odometers we refer the reader to Foreman and Weiss (2020).

Example 3. *Finite ergodic systems* $(\mathbb{Z}_n, \pi)$, where $\pi \in \mathbb{S}_n$ is a cyclic permutation on $\mathbb{Z}_n$, and the Haar measure μ is the uniform measure on $\mathbb{Z}_n$.

The main theorem of the paper gives positive answers to Questions 1 and 2 in the following case:

Theorem 1. *Let (X,T) be an ergodic Kronecker system. Then, for every finite Borel-measurable partition $\mathcal{P}$, the uniform value $v(\xi)$ of $\Gamma(\xi)$, where $\xi = (\mathcal{P}, T)$, exists and equals* $\lim_{n\to\infty} v_n(\xi)$.

Let us note first that the ergodicity assumption on the Kronecker system implies that the state changes over time and does not remain fixed as in Aumann and Maschler's model. Secondly, the states in the processes ξ considered in Theorem 1 do not evolve according to a non-deterministic[9] Markov chain[10]. Therefore, previous results in the literature on repeated games with incomplete information on one side do not imply the result stated in Theorem 1, nor follow from it.

The following corollary follows immediately from Theorem 1 and Corollary 1.

Corollary 2. *Let (X,T) be an ergodic Kronecker system. Then, for every finite Borel-measurable partition $\mathcal{P}$, the adversary has an optimal strategy in $\Gamma(\xi)$, where $\xi = (\mathcal{P}, T)$.*

Let us fix the finite Borel-measurable partition $\mathcal{P}$ of the ergodic Kronecker system (X,T) once and for all. Let $\xi = (\mathcal{P}, T)$ be the process it generates, and[11] $A = \{0, ..., |\mathcal{P}|-1\}$ be the set of states in $\Gamma(\xi)$. Our goal is now to describe a formula for $v(\xi)$. We need a few notations and definitions. For each $n \in \mathbb{N}$ we denote by $\gamma(n)$ the number of elements of positive μ-measure (Haar measure) in the joining of the transformed partitions $\bigvee_{\ell=1}^{n} T^{-(\ell-1)}\mathcal{P}$. Denote those elements by $P_n^1, ..., P_n^{\gamma(n)}$. By Eq. (3.1), for every $n \in \mathbb{N}$ and $1 \le r \le \gamma(n)$ there exists a sequence $(a_1^r, ..., a_n^r) \in A^n$ such that

$$P_n^r = \{x \in X : (\xi_1(x), ..., \xi_n(x)) = (a_1^r, ..., a_n^r)\}. \tag{4.1}$$

In words, each of the sets $P_n^1, ..., P_n^{\gamma(n)}$ contains all infinite histories that share the same positive-probability n-prefix of ξ. Set $\mathcal{S}_n = \{P_n^1, ..., P_n^{\gamma(n)}\}$ for every $n \in \mathbb{N}$. Hence, with μ-measure 1 the

[9] In a *deterministic* Markov chain the transition matrix consists only of 0's and 1's.

[10] This is due to the notion of entropy (e.g., Chapter V in Smorodinsky (1971)). Non-deterministic Markov chains over finite state spaces have positive entropy, whereas ergodic Kronecker systems have zero entropy.

[11] For every set B we denote by $|B|$ the cardinality of B.

observer learns in which element of $\mathcal{S}_n$ the chosen point $x \in X$ lies, upon observing the first n outcomes of ξ.[12]. Next, for every $n \in \mathbb{N}$ define $\pi_n \in \Delta(\mathcal{S}_n)$ by

$$\pi_n(P_n^r) = \mu(P_n^r), \quad \forall r = 1, ..., \gamma(n). \tag{4.2}$$

Also, as each $q \in \Delta(\mathcal{S}_n)$ can be viewed as a probability vector in $\mathbb{R}^{\gamma(n)}$, it is convenient to denote the r'th coordinate of q by q^r, so that $q^r = q(P_n^r)$.

Denote by $\mathcal{W}_n = A^{\gamma(n)}$ the set of all words of length $\gamma(n)$ over the alphabet A. For every $q \in \Delta(\mathcal{S}_n)$ and $w \in \mathcal{W}_n$, define the one-shot game $G(q, w)$ by

$$G(q, w) = \sum_{r=1}^{\gamma(n)} q^r g(w^r, \cdot, \cdot), \tag{4.3}$$

where w^r is the r'th letter of w. That is, $G(q, w)$ is a convex combination of the games $g(w^r, \cdot, \cdot)$, $r = 1, ..., \gamma(n)$, where the weight of a particular $g(w^r, \cdot, \cdot)$ is q^r. We define $u : \Delta(\mathcal{S}_n) \times \mathcal{W}_n \to \mathbb{R}_+$ by[13]

$$u(q, w) = \operatorname{val} G(q, w). \tag{4.4}$$

Also, for every bounded real-valued function h, defined on a set D, we denote $\|h\|_\infty = \sup\{|h(x)| : x \in D\}$. Finally, for any function $f : E \to \mathbb{R}$, where $E \subseteq \mathbb{R}^d$ is a closed convex set, define the function $\operatorname{Cav} f$ by

$$(\operatorname{Cav} f)(y) = \inf\{g(y) \,:\, g : E \to \mathbb{R} \text{ concave}, \, g \geq f\}, \qquad \forall y \in E.$$

The formula for the uniform value is provided in the following theorem.

Theorem 2. *There exists a triangular array* $t = (t_n^r)_{r=1\;,n=1}^{\gamma(n),\,\infty}$, *consisting of positive integers, such that*

(a) $\mu\left(T^{-t_n^1} P_n^1 \cap \cdots \cap T^{-t_n^{\gamma(n)}} P_n^{\gamma(n)}\right) > 0$ *for every* $n \in \mathbb{N}$.

(b) *Define* $\Phi_n : \Delta(\mathcal{S}_n) \to \mathbb{R}_+$ *by*

$$\Phi_n(q) = \mathbb{E}\left[u\left(q, (\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}})\right)\right],$$

where $\mathbb{E}$ *is the expectation operator w.r.t.* μ. *In words,* $\Phi_n(q)$ *is the expected value of* $u(q, w)$, *where* $w \in \mathcal{W}_n$ *is the random word* $(\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}})$. *Then,* $v(\xi) = \lim_{n\to\infty} (\operatorname{Cav} \Phi_n)(\pi_n)$.

Item (a) in Theorem 2 guarantees the existence of sequences such that in each row n, the times $(t_n^r)_{r=1}^{\gamma(n)}$, satisfy a 'visit time property'. That is, along those times, a set of points of positive μ-measure should visit all of the 'information sets' available to the observer (with μ-measure 1) at

[12]In light of Remark 1, even if the observer and his adversary are assumed to know only the distribution of ξ, they are indifferent about the underlying probability space $(\Omega, \mathcal{B}, \mathbb{P})$ on which ξ is defined. Thus, both of them can choose this underlying space to be the probability space induced by the ergodic Kronecker system (X, T, μ), and assume that ξ is given by (3.1)

[13]We omit the parameter n from the mappings G and u as it will be unambiguous, given the context and based on the relevant dimensions of the q's and w's.

time n. In view of Eq. (4.1), this means that the process ξ evaluated at those points, would print the history $(a_1^1, ..., a_n^1)$ starting from the $t_n^1 + 1$'st step $\xi_{t_n^1+1}$, the history $(a_1^2, ..., a_n^2)$ starting from the $t_n^2 + 1$'st step $\xi_{t_n^2+1}$, and so on, so that generally the history $(a_1^r, ..., a_n^r)$ would be printed by ξ starting from the $t_n^r + 1$'st step $\xi_{t_n^r+1}$, where $r = 1, 2, ..., \gamma(n)$.

As for item (b), note that it describes the uniform value $v(\xi)$ of $\Gamma(\xi)$ in terms of the primitives of the model. The payoff functions $g(a, \cdot, \cdot)$, $a \in A$, leave their mark through the function u (see (4.4)). The process ξ has a double role in the formula. First in the random word $(\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}}) \in \mathcal{W}_n$ appearing in the second argument of u in the definition of Φ_n. Second, in the probability vector π_n, which describes the distribution of the n-prefix $(\xi_1, ..., \xi_n)$ of ξ.

Another phenomenon worth mentioning is that in some cases, the uniform value depends on the distribution laws of infinitely many steps of the process $(\xi_n)_{n\geq 1}$. Indeed, as the sets $P_n^1, ..., P_n^{\gamma(n)}$ are pairwise disjoint, it follows that $(t_n^r)_{r=1}^{\gamma(n)}$ are pairwise disjoint times as well. Therefore, such a phenomenon holds if $\gamma(n) \to \infty$ as $n \to \infty$. For instance, in the irrational rotation of the unit circle $(\mathbb{T}, R_\alpha)$, $g(n) \to \infty$ as $n \to \infty$ for every non-degenerate[14] partition $\mathcal{P}$.

It turns out that this formula coupled with a new result of Ashkenazi-Golan, Solan, and Zseleva (2020) and some additional game-theoretic arguments, suffices to deduce a new underlying dynamical property of ergodic Kronecker systems, in the case where ξ is generated by a binary partition. This property concerns the asymptotic behavior of the random word $(\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}})$, which, as we saw, has a central role in the formula for the uniform value.

Theorem 3. *Let (X, T) be an ergodic Kronecker system and let $\mathcal{Q}$ be a finite partition of X with $|\mathcal{Q}| = 2$ (i.e., $\mathcal{Q}$ is a partition of X into two subsets). Denote by $\{Q_n^1, ..., Q_n^{\alpha(\mathcal{Q},n)}\}$ the elements of positive μ-measure in the joining of the transformed partitions $\bigvee_{\ell=1}^n T^{-(\ell-1)}\mathcal{Q}$. Then, there exists a triangular array $t = (t_n^r)_{r=1,\, n=1}^{\alpha(\mathcal{Q},n),\, \infty}$ of positive integers such that*

(a) $\mu\left(T^{-t_n^1} Q_n^1 \cap \cdots \cap T^{-t_n^{\alpha(\mathcal{Q},n)}} Q_n^{\alpha(\mathcal{Q},n)}\right) > 0$ *for every* $n \in \mathbb{N}$.

(b) $\mu(Q_n^1)\eta_{t_n^1} + \cdots + \mu(Q_n^{\alpha(\mathcal{Q},n)})\eta_{t_n^{\alpha(\mathcal{Q},n)}} \to \mathbb{E}\,\eta_1$ *in* $L^1(X, \mathcal{B}(X), \mu)$,

where $\eta = (\mathcal{Q}, T)$, and $\mathbb{E}$ is the expectation operator w.r.t. μ.

First, note that the theorem is stated in terms of the partition $\mathcal{Q}$ and not $\mathcal{P}$, as the partition $\mathcal{P}$ was fixed to simplify notation in future analysis. As a consequence, the quantity $\gamma(n)$ used before was strictly adapted to the fixed partition $\mathcal{P}$. For a general partition $\mathcal{Q}$, the number of elements of positive μ-measure in the partition $\bigvee_{\ell=1}^n T^{-(\ell-1)}\mathcal{Q}$ does depend on $\mathcal{Q}$, which justifies the notation $\alpha(\mathcal{Q}, n)$.

Next, let us try and elaborate on the result of Theorem 3. Formally, Theorem 3 states that if one takes the weighted average of the steps of η at the times given by the n'th row of the array t w.r.t. the weights $(\mu(Q_n^1), ..., \mu(Q_n^{\alpha(\mathcal{Q},n)}))$, then for sufficiently large values of n, such an average is close to the expected value of the first step of the process, η_1. One can also describe the result in terms of the notions of Theorem 2. In that case, Theorem 3 implies that whenever $|\mathcal{P}| = 2$, if one takes

[14] i.e., contains an element with μ-measure different from 0 and 1.

the weighted average of the letters of the word $(\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}}) \in \mathcal{W}_n$ w.r.t. the probability vector π_n (describing the distribution of $(\xi_1, ..., \xi_n)$), then for sufficiently large n, this average will be close to the expected letter of the process ξ, being $\mathbb{E}\,\xi_1 = \mathbb{E}\,\xi_2 = \cdots = \mathbb{E}\,\xi_\ell = \cdots$. Thus, the limit law given in Theorem 3 can be thought of combining two asymptotic behaviors associated with the process η. The first, is the distribution of the n-prefix $(\eta_1, ..., \eta_n)$ (described by the $\mu(Q_n^l)$'s, $l = 1, ..., \alpha(\mathcal{Q}, n)$) and second, the random variable $(\eta_{t_n^1}, ..., \eta_{t_n^{\alpha(\mathcal{Q},n)}})$.

5. A Sketch of the Proof

The dynamics. The proofs of Theorems 1 and 2 begin by identifying the information the observer can extract from his first K observations of ξ. This information is valuable, as it enables the observer to predict future outcomes. To describe it, we resort to two basic definitions. First, the *lower density* of a set $E \subseteq \mathbb{N}$ is defined as

$$\liminf_{n\to\infty} \frac{|E \cap \{1, ..., n\}|}{n}.$$

The second definition is related to a key probabilistic property of the process ξ, termed long-term predictability.

Definition 3. We say that the process ξ is *long-term predictable* if for every $\varepsilon > 0$ there exists a number K, such that with μ-measure of at least $1 - \varepsilon$, upon observing $\xi_1(x), ..., \xi_K(x)$, the observer is able to correctly predict $(\xi_n(x))_{n>K}$ over a set of stages $Z \subset \mathbb{N}$ whose lower density is greater then or equal to $1 - \varepsilon$.

Assume that for some large value of K, the observer decides to assign to each history $h = (a_1, ..., a_K)$ a prediction $y^h = (y_1^h, y_2^h, ..., y_n^h, ...)$, where $y_n^h \in A$ for every history h and $n \in \mathbb{N}$. The role of these infinite sequences is to predict the entire future of outcomes. In other words, after the history h, the observer predicts that at time $K + n$ the outcome will be y_n^h. As the payoffs in the game take into account only histories of positive measure, using (4.1) we may denote for those histories y^h by y^r, $r = 1, ..., \gamma(K)$, with the interpretation that y^r is the set of predictions assigned to the history corresponding to P_K^r. Whenever the set of predictions $Y = \{y^r\}_{r=1}^{\gamma(K)}$ satisfies the long-term predictability property with precision $\varepsilon > 0$ we say that Y is a $(K, 1-\varepsilon)$-long-term predictor for ξ (see Subsection 6.1 for exact details).

The predictor game. The long-term predictability property enables one to approximate $\Gamma(\xi)$ with an auxiliary repeated game whose dynamics follow those governed by some $(K, 1-\varepsilon)$-long-term predictor Y for ξ. Such a game begins with a random choice of an element P_K^r from $\mathcal{S}_K$ according to the probability distribution π_K (see (4.2)). Equivalently, π_K selects the sequence of predictions y^r from Y. This choice is known to the observer, as he obtains this information in $\Gamma(\xi)$ at the beginning of the K'th stage. The adversary, on the other hand, is assumed to know only the law of the lottery π_K[15]. Then, at every stage $n \geq 1$, the zero-sum one-shot game $g(y_n^r, \cdot, \cdot) : I \times J \to \mathbb{R}_+$ is

[15]This assumption is natural, as the adversary's knowledge of the distribution of ξ in $\Gamma(\xi)$ implies that he knowns π_K in $\Gamma(\xi)$.

being played. As in $\Gamma(\xi)$, the actions i_n, j_n played by the observer and his adversary, respectively, at stage n are publicly announced and become common knowledge among the two. The payoff to the observer after N stages equals $\frac{1}{N}\sum_{n=1}^{N} g(y_n^r, i_n, j_n)$.

For every $(K, 1-\varepsilon)$-long-term predictor $Y = \{y^r\}_{r=1}^{\gamma(K)}$ we denote the corresponding game by $\widehat{\Gamma}(Y)$. Also, we denote by $\widehat{v}_n(Y)$ the value of the finite n-stage game corresponding to $\widehat{\Gamma}(Y)$. Those auxiliary games, which all depend on the predictions made by the observer, are called predictor games (see Subsection 6.2).

In a given predictor game, there are only finitely many information sets $P_K^1, ..., P_K^{\gamma(K)}$, and each uniquely determines all of the future states. The game thus obtained resembles that of Aumann and Maschler ([2]). It stands to reason that the optimal use of information by the observer should, as in the Aumann-Maschler game, be carried out via a splitting scheme. However, there is a major difference between the current model and that of Aumann and Maschler.

In Aumann and Maschler's game, the information set coincides with the state of the game, which remains fixed for ever. Here, in contrast, when the information set P_K^r is chosen, the sequence of future states follows $(y_n^r)_{n\geq 1}$. This sequence typically exhibits an erratic behavior, in particular it is neither constant nor periodic. This makes it difficult to identify which target function should the observer split in order to obtain its concave envelope. Moreover, such an optimal split clearly depends on the choice of the $(K, 1-\varepsilon)$-long-term predictor Y.

Non-revealing strategies and splitting. We proceed with the non-revealing strategy of the observer in the predictor game $\widehat{\Gamma}(Y)$. When playing this strategy, the observer completely ignores his additional private information. Specifically, at each stage $n \in \mathbb{N}$ the observer plays an optimal action in the one-shot zero-sum game $G(\pi_K, w_n)$ (see (4.3)), where $w_n = (y_n^1, ..., y_n^{\gamma(K)})$ will be called the *n'th word induced by Y*.

Based on long-term predictability, Proposition 1 shows that whenever Y is a $(K, 1-\varepsilon)$-long-term predictor, there exists $\sigma^Y \in \Sigma$ such that for any duration N sufficiently large,

$$\gamma_N(\sigma^Y, \tau) \geq \liminf_{L\to\infty} \frac{1}{L}\sum_{\ell=1}^{L} u(\pi_K, w_\ell) - 2\varepsilon - 3\|g\|_\infty \varepsilon, \quad \forall \tau \in \mathcal{T}. \tag{5.1}$$

The strategy $\sigma^Y \in \Sigma$ is parallel to the non-revealing strategy in $\widehat{\Gamma}(Y)$. We now define two mappings. For $q \in \Delta(\mathcal{S}_K)$ (which now can be thought of as a distribution over the predictions $\{y^r\}_{r=1}^{\gamma(K)}$) let

$$I_K[Y](q) = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} u(q, w_n) \tag{5.2}$$

and

$$I_K^{m,L}[Y](q) = \frac{1}{L}\sum_{\ell=1}^{L} u(q, w_{mL+\ell}), \quad m \geq 0, \quad L \in \mathbb{N}. \tag{5.3}$$

Note that the dependence of these mappings on Y is through the words $\{w_n\}_{n\geq 1}$ it induces. We need to keep Y in the notation because we later introduce another predictor and compare between the two. Note also that the mapping $I_K^{m,L}[Y]$ is determined by the words in the m'th L-block (i.e., $\{w_{mL+1}, ..., w_{(m+1)L}\}$).

Subsection 6.3 describes a standard 'splitting' strategy σ_ε^Y with respect to the mapping $I_K[Y]$, and Proposition 2 shows that, whenever Y is a $(K, 1-\varepsilon)$-long-term predictor, then for every N sufficiently large,

$$\gamma_N(\sigma_\varepsilon^Y, \tau) \geq (\text{Cav}\, I_K[Y])\,(\pi_K) - 2\varepsilon - 3\|g\|_\infty \varepsilon, \quad \forall \tau \in \mathcal{T}. \tag{5.4}$$

In words, for a large number of observations K, the splitting strategy of the observer can guarantee almost $(\text{Cav}\, I_K[Y])\,(\pi_K)$.

The adversary side. We now switch our attention to the adversary. To show that the uniform value exists, we describe a strategy of the adversary that guarantees a payoff close to $(\text{Cav}\, I_K[Y])\,(\pi_K)$ for large values of K. By Observation 1, the adversary can guarantee $\lim_{n\to\infty} v_n(\xi)$. Thus, in order to prove Theorem 1 it is sufficient to show that for large values of n and K, $v_n(\xi)$ is close to $(\text{Cav}\, I_K[Y])\,(\pi_K)$.

It seems, at a first glance, that the adversary can analyze $v_n(\xi)$ using the predictor game $\widehat{\Gamma}(Y)$. However, by the definition of the uniform value, the adversary should come up with a strategy independent of that of the observer. In particular, it should be independent of the choice of the observer's $(K, 1-\varepsilon)$-long-term predictor Y and so of the number of observations K on which the observer decides to base his predictions.[16]

The adversary can construct his own $(K', 1-\varepsilon)$-long-term predictor $Y' = \{y'^r\}_{r=1}^{\gamma(K')}$. Let us elaborate on this somewhat confusing statement. The adversary himself might not have his own prediction y'^r on the future states, as he does not observe the outcomes of the first K' steps of ξ. Nevertheless, the existence of an $(K', 1-\varepsilon)$-long-term predictor of ξ is guaranteed based on the distribution of ξ, which is known to the adversary. That is, the adversary can construct a predictor $Y' = \{y'^r\}_{r=1}^{\gamma(K')}$ (that ascribes the history of outcomes that corresponds to $P_{K'}^r$ the prediction y'^r) which is a $(K', 1-\varepsilon)$-long-term predictor for ξ. Proposition 4 shows that in order to get a tight upper bound on $v_n(\xi)$, it suffices to analyze $\widehat{v}_n(Y')$, for large values of n. Thus, the adversary might as well play as if the underlying process follows Y', rather than ξ.

The analysis of $\widehat{\Gamma}(Y')$ is made possible by the fact that $\widehat{\Gamma}(Y')$ is a repeated game with incomplete information on one side over a finitely many information sets. To this game we apply the classical technique known as 'martingales of posteriors' (see Subsection 6.4). This analysis culminates with Corollary 3 which states that for every $L \in \mathbb{N}$,

$$v_{ML}(\xi) \leq \frac{1}{M} \sum_{m=0}^{M-1} (\text{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) + 6\|g\|_\infty \varepsilon, \tag{5.5}$$

whenever M is large enough. In words, since ξ is stationary the adversary can guarantee $v_{ML}(\xi)$ for all $M, L \in \mathbb{N}$. Thus, roughly speaking, we have shown that for large values of K' and M the

[16]Note that the existence of a $(K, 1-\varepsilon)$-long-term predictor implies the existence of $(K+\ell, 1-\varepsilon)$-long term predictor, for every $\ell \in \mathbb{N}$. Thus, a precision level of $1-\varepsilon$ can be achieved for infinitely many sets of observations.

adversary can guarantee almost $\frac{1}{M}\sum_{m=0}^{M-1}(\text{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'})$. Hence, Eqs. (5.4) and (5.5) suggest that the main question regarding the existence of the uniform value is reduced to the following problem.

Do there exist two predictors Y and Y', which are, respectively, $(K, 1-\varepsilon)$ and $(K', 1-\varepsilon)$-long-term predictors, and a positive integer L, such that

$$\left|(\text{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) - (\text{Cav}\, I_K[Y])\,(\pi_K)\right| < 3\varepsilon, \quad \forall m = 0, ..., M-1? \tag{5.6}$$

In order to find such predictors one should know what kind of predictors there are at the disposal of the observer and the adversary, how many of them exist, and whether or not they posses useful properties. While so far we have applied well-known game-theoretic techniques, addressing these issues requires new dynamical methods.

Satisfactory long-term predictors and where to find them. It turns out that the unique topological structure of the ergodic Kronecker system (X, T), together with the properties of the Haar measure μ, can be employed in order to obtain a 'positive-measure' set of long-term predictors.

Let $t = (t_n^r)_{r=1\ ,n=1}^{\gamma(n),\infty}$ be a triangular array of times (positive integers). Based on this array, for every $n \in \mathbb{N}$ and $x \in X$, define

$$y^{n,r}(x) = \xi_{n+t_n^r+1}(x), \xi_{n+t_n^r+2}(x), ..., \xi_{n+t_n^r+n}(x), ...;$$

$y^{n,r}(x)$ is the sequence of future realizations of states that starts at time $n + t_n^r + 1$ and corresponds to the point x. Let $Y^n(x) = \{y^{n,r}(x)\}_{r=1}^{\gamma(n)}$ be the predictor determined by the predictions $y^{n,r}(x)$, $r = 1, ..., \gamma(n)$. The key step in the proof is described in Theorem 5. It states that there exists a triangular array of times $t = (t_n^r)_{r=1\ ,n=1}^{\gamma(n),\infty}$, satisfying item (a) of Theorem 2, for which $Y^n(x) = \{y^{n,r}(x)\}_{r=1}^{\gamma(n)}$ has the following key property: for every $\varepsilon > 0$ there exists a required number of observation K_ε such that for any $K \geq K_\varepsilon$ there exists a set of points $x \in X$ of positive measure for which $Y^K(x)$ is a $(K, 1-\varepsilon)$-long-term predictor.

In words, both the observer and his adversary can choose a long-term predictor from a relatively large pool of predictors, namely one that has a positive Haar measure. The latter will play a crucial role in finding Y and Y' (see Lemmas 2, 3, and 4, which lead to the choice of points (long-term predictors) $x, x' \in X$ that determine Y, Y', respectively).

The existence of the two predictors. We begin with the observer. Let the observer choose $K \geq K_\varepsilon$ such that

$$\left|(\text{Cav}\, \Phi_K)(\pi_K) - \limsup_{n\to\infty}(\text{Cav}\, \Phi_n)(\pi_n)\right| \leq \varepsilon.$$

We argue in Subsection 6.6 that the observer can choose $x \in X$ such that (i) $Y^K(x)$ is a $(K, 1-\varepsilon)$-long-term predictor, and (ii) $I_K[Y^K(x)] = \Phi_K$. By (ii) and the choice of K,

$$\left|\,(\text{Cav}\, I_K[Y])\,(\pi_K) - \limsup_{n\to\infty}(\text{Cav}\, \Phi_n)(\pi_n)\right| < \varepsilon. \tag{5.7}$$

Now let us look at the adversary. The adversary will choose $K' \geq K_\varepsilon$ such that

$$\left|(\mathrm{Cav}\,\Phi_{K'})(\pi_{K'}) - \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n)\right| < \varepsilon.$$

Lemmas 3 and 4 describe how to find a sufficiently large M such that for each $\beta > 0$ the adversary can choose x' and L such that for the $(K', 1-\varepsilon)$-long-term predictor $Y' = Y^{K'}(x')$ the following two properties hold:

(i) $v_{ML}(\xi) \leq \frac{1}{M}\sum_{m=0}^{M-1}(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) + 6\|g\|_\infty\varepsilon$, and

(ii) $\left|(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) - (\mathrm{Cav}\,\Phi_{K'})\,(\pi_{K'})\right| \leq |W_{K'}|\|g\|_\infty\beta, \quad \forall m = 0, ..., M-1.$

The combination of (i) and (ii), together with the choice of K', implies that

$$v_{ML}(\xi) \leq \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) + 6\|g\|_\infty\varepsilon + \varepsilon + |W_{K'}|\|g\|_\infty\beta. \tag{5.8}$$

We next note that the adversary cannot guarantee less than what the observer can guarantee. Thus, by applying (5.4) to $Y = Y^K(x)$ and using the fact that the adversary can guarantee $v_{ML}(\xi)$, we obtain, with the help of Eqs. (5.6) and (5.8), that

$$\limsup_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) - 2\varepsilon - 2\|g\|_\infty\varepsilon \leq \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) + 6\|g\|_\infty\varepsilon + \varepsilon + |W_{K'}|\|g\|_\infty\beta. \tag{5.9}$$

Since $\varepsilon > 0$ and $\beta > 0$ are arbitrary, we see that

$$\liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) = \limsup_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n). \tag{5.10}$$

The latter is sufficient to deduce that for the choices $Y = Y^K(x)$, $Y' = Y^{K'}(x')$ and L, we have

$$\left|(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) - (\mathrm{Cav}\, I_K[Y])\,(\pi_K)\right| < |W_{K'}|\|g\|_\infty\beta + 2\varepsilon. \tag{5.11}$$

Thus, by taking β sufficiently small, we obtain an affirmative answer to the question posed in (5.6). Moreover, it follows that the observer can guarantee $\lim_{n\to\infty} v_n(\xi)$ and that the uniform value $v(\xi)$ must be equal to $\lim_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n)$, as stated in Theorem 2.

6. Detailed Proofs

We divide the outline of the proofs into multiple steps. We start with the notion of long-term predictors.

6.1. Long-Term Predictors.

The analysis of the predictive properties of the stationary processes considered in the paper has led us to the following new definition:

Definition 4. Let $Z = (Z_n)_{n\geq 1}$ and $Y = (Y_n)_{n\geq 1}$ be two sequences of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Y is said to be a $(K, 1-\varepsilon)$-*long-term predictor* (LT-predictor) for Z if

(i) $Y_n \in \mathcal{F}_K$ for all $n \in \mathbb{N}$, where $\mathcal{F}_K := \sigma(Z_1, ..., Z_K)$.

(ii) $\mathbb{P}\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^n \mathbb{1}\{Z_\ell = Y_\ell\} \geq 1-\varepsilon\right) \geq 1-\varepsilon.$

That is, with probability of at least $1-\varepsilon$, after observing the first K steps of Z we can predict the future steps of Z correctly at times whose lower density is at least $1-\varepsilon$. In that case we say that Z *admits a* $(K, 1-\varepsilon)$-LT-predictor.

Theorem 4. *For every $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that ξ admits a $(K, 1-\varepsilon)$-LT-predictor.*

The proof of Theorem 4 is given in Appendix A.

Each $(K, 1-\varepsilon)$-LT-predictor Y for ξ induces a sequence of words $\{w_n\}_{n\geq 1}$ in $\mathcal{W}_K$. Indeed, denote the unique value Y_{n+K} attains on the event P_K^r by y_n^r, $r = 1, ..., \gamma(K)$ (all the steps of Y are measurable w.r.t. $(\xi_1, ..., \xi_K)$). The *n'th word induced by* Y is set to be $w_n = (y_n^1, ..., y_n^{\gamma(K)})$.

6.2. **The Predictor Game $\widehat{\Gamma}(Y)$.** The goal of this subsection is to introduce the *predictor game* $\widehat{\Gamma}(Y)$ associated with a $(K, 1-\varepsilon)$-LT-predictor Y for ξ. This game will also be a zero-sum repeated game with incomplete information on one side.

The description of $\widehat{\Gamma}(Y)$ goes as follows. As in $\Gamma(\xi)$, the actions available to the observer and his adversary are I and J, respectively. At the start of the game an element $P_K^r \in \mathcal{S}_K$ is chosen at random according to the probability vector π_K defined by Eq. (4.2). The observer is informed of the chosen element, whereas his adversary only knows the law of π_K. At each stage $n \in \mathbb{N}$ the observer and his adversary are instructed to choose actions. The pair of chosen actions $(i_n, j_n) \in I \times J$ is publicly announced and becomes common knowledge among the two. The n'th stage payoff to the observer is defined to be $g(y_n^r, i_n, j_n)$. The payoff evaluation along the first N stages of $\widehat{\Gamma}(Y)$ is the average per-stage payoff, i.e., $\frac{1}{N}\sum_{n=1}^{N} g(y_n^r, i_n, j_n)$.

The information setup in $\widehat{\Gamma}(Y)$ is strongly reminiscent to that considered in the original Aumann and Maschler model (e.g., Chapter 3 in [2]). Indeed, upon learning the chosen element $P_K^r \in \mathcal{S}_K$, the observer learns the state of the game at any given stage $n \in \mathbb{N}$. The difference of course lies in the fact that, unlike in the Aumann and Maschler model, this state is not fixed across all future stages, but rather follows the dynamics of the **deterministic** sequence $(y_n^r)_{n\geq 1}$. As we shall see in the paper, the dynamics of the latter can be chosen to be of the form $(\xi_n(x_r))_{n\geq 1}$ for some $x_r \in X$. In ergodic Kronecker systems, such as in the irrational rotation of the circle, such dynamics are extremely non-periodic. Nevertheless, the **only information** of interest to the adversary remains which $P_K^r \in \mathcal{S}_K$ was chosen. Thus, as in the Aumann-Maschler model, it should not be a surprise that the true 'state' space for the adversary is the belief space $\Delta(\mathcal{S}_K)$. The dynamics in this 'state' space are fully controlled by the observer's actions.

6.3. **Non-Revealing and Splitting Strategies of the Observer.** Let Y be some $(K, 1-\varepsilon)$-LT-predictor for ξ. Consider the following non-revealing strategy σ^Y of the observer:

- At stages $1, ..., K$, play the same pure action $i_0 \in I$.
- At each stage $K+n$, $n \in \mathbb{N}$, play an optimal mixed action in the one-shot zero-sum game $G(\pi_n, w_n)$, where w_n is the n'th word induced by Y.

The simple idea behind this strategy is that the observer essentially waits for K steps, and then plays an optimal non-revealing mixed action in each of the stages in the game $\widehat{\Gamma}(Y)$.

Proposition 1. *Let $Y=(Y_n)_{n\geq 1}$ be a $(K,1-\varepsilon)$-*LT*-predictor for ξ. Then, there exists $M_Y \in \mathbb{N}$ such that*

$$\gamma_N(\sigma^Y,\tau) \geq \liminf_{L\to\infty} \frac{1}{L}\sum_{\ell=1}^{L} u(q,w_\ell) - 2\varepsilon - 3\|g\|_\infty \varepsilon, \quad \forall N > M_Y, \forall \tau \in \mathcal{T}. \tag{6.1}$$

The proof of Proposition 1 is quite technical and can be found in Appendix B.

Let us proceed by defining the function $I_K[Y] : \Delta(\mathcal{S}_K) \to \mathbb{R}_+$ by

$$I_K[Y](q) = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} u(q,w_n), \tag{6.2}$$

where we note that the dependence on the $(K,1-\varepsilon)$-LT-predictor Y is through the sequence of words $\{w_n\}_{n\geq 1}$ induced by it. Carathéodory's Theorem ensures that there exist non-negative weights $(\alpha_\ell)_{\ell=1}^{\gamma(K)+1}$ and distributions $(q_\ell)_{\ell=1}^{\gamma(K)+1} \in \Delta(\mathcal{S}_K)$ such that

(1) $\sum_{\ell=1}^{\gamma(K)+1} \alpha_\ell = 1$.
(2) $\sum_{\ell=1}^{\gamma(K)+1} \alpha_\ell q_\ell = \pi_K$.
(3) $(\mathrm{Cav}\, I_K[Y])(\pi_K) \leq \sum_{\ell=1}^{\gamma(K)+1} \alpha_\ell\, I_K[Y](q_\ell) + \varepsilon$.

Consider the lottery $Z : X \to \{1,...,\gamma(K)\}$ whose distribution satisfies

$$\mu\left(Z=\ell \,|\, P_K^r\right) = \frac{\alpha_\ell q_\ell^r}{\pi_K^r}, \tag{6.3}$$

for every $\ell \in \{1,...,\gamma(K)+1\}$ and $r \in \{1,...,\gamma(K)\}$. Standard computations show that for every $\ell \in \{1,...,\gamma(K)+1\}$ we have (i) $\mu(Z=\ell)=\alpha_\ell$ as well as (ii) $\mu(P_K^r \,|\, Z=\ell) = q_\ell^r$ for $r=1,...,\gamma(K)$. Define the strategy σ_ε^Y of the observer as follows:

- At stages $1,...,K$, play the same pure action $i_0 \in I$.
- 'Splitting': observe $(\xi_1,...,\xi_K)$ and learn which of the elements of the partition that generates $\mathcal{F}_K$ was realized. If the realized partition element is some $P_K^r \in \mathcal{S}_K$ (note that this happens μ-a.s.), perform the lottery Z conditional on P_K^r, i.e., according to Eq. (6.3). Otherwise, play the arbitrary pure action $i_0 \in I$ in any future stage.
- Conditional on the event $\{Z=\ell\}$, play at each stage $K+n$, $n\geq 1$, an optimal mixed action in the one-shot game $G(q_\ell,w_n)$, where w_n is the n'th word induced by Y.

Proposition 2. *Let $Y=(Y_n)_{n\geq 1}$ be a $(K,1-\varepsilon)$-*LT*-predictor for ξ. Then, there exists $N_Y \in \mathbb{N}$ such that*

$$\gamma_N(\sigma_\varepsilon^Y,\tau) \geq (\mathrm{Cav}\, I_K[Y])(\pi_K) - 2\varepsilon - 3\|g\|_\infty \varepsilon, \quad \forall N > N_Y, \forall \tau \in \mathcal{T}. \tag{6.4}$$

As the proof of Proposition 2 is quite technical, we relegate it to Appendix B as well.

We now switch our focus to the strategic behavior of the adversary. Our starting point is the following.

6.4. **Martingales of Posteriors for the Adversary.** The goal of this subsection is to show how martingales of posteriors, a common tool in the field of repeated games with incomplete information on one side, apply to $\widehat{\Gamma}(Y')$, where Y' is some $(K', 1-\varepsilon)$-LT-predictor for ξ that the adversary decides to construct. Denote by $\{w'_n\}_{n\geq 1}$ the sequence of words induced by Y'.

The description of $\widehat{\Gamma}(Y')$ implies that the set of behavior strategies of the observer in $\widehat{\Gamma}(Y')$ can be represented as

$$\widehat{\Sigma} = \left\{ \begin{array}{c} \sigma \quad = \quad (\sigma_n)_{n\geq 1}\colon \quad \sigma_n \quad = \quad \{\sigma_n^s\}_{s\in\mathcal{S}_{K'}}, \quad \text{such that} \\ \sigma_n^s : (I\times J)^{n-1} \to \Delta(I) \end{array} \right\}. \tag{6.5}$$

The space of behavior strategies of the adversary in $\widehat{\Gamma}(Y')$ is the same as in $\Gamma(\xi)$, i.e., $\mathcal{T}$. Thus, the space of finite histories $\widehat{H}$ in $\widehat{\Gamma}(Y')$ can be described as $\widehat{H} = \bigcup_{n\in\mathbb{N}} \mathcal{S}_{K'} \times (I\times J)^n$. Consider the product space $\widehat{\mathcal{H}} = \mathcal{S}_{K'} \times (I\times J)^{\mathbb{N}}$, equipped with the product topology. The cylinder sets in $\widehat{\mathcal{H}}$, induced by the elements of $\widehat{H}$, form a basis for the product topology on $\widehat{\mathcal{H}}$. By the Ionescu-Tulcea Extension Theorem (see Neveu, 1970, Proposition V.1.1), each pair $(\sigma,\tau)\in\widehat{\Sigma}\times\mathcal{T}$, together with $\pi_{K'}$ induce a Borel probability measure on $\widehat{\mathcal{H}}$, denoted $\mathbb{P}^{K'}_{\sigma,\tau}$, which is determined by the laws

$$\begin{aligned} \mathbb{P}^{K'}_{\sigma,\tau}(h) = \pi_{K'}[s] \quad &\times \prod_{n\leq N} \sigma_n^s(i_1,j_1,...,i_{n-1},j_{n-1})[i_n] \\ &\times \prod_{n\leq N} \tau_n(i_1,j_1,...,i_{n-1},j_{n-1})[j_n]. \end{aligned}$$

The sequence $h = (s, i_1, j_1, ..., i_N, j_N) \in \widehat{H}$ is identified with the cylinder it induces on $\widehat{\mathcal{H}}$. Next, for each $n\in\mathbb{N}$ we let $\widehat{\mathcal{H}}^2_n$ denote the σ-field generated by the elements of $(I\times J)^{n-1}$ on $\widehat{\mathcal{H}}$. The *martingale of posteriors* $p = (p_n)_{n\geq 1}$ induced by (σ,τ) is a sequence of random variables such that $p_n \in \Delta(\mathcal{S}_{K'})$ is given by

$$p_n^r = \mathbb{P}^{K'}_{\sigma,\tau}(P^r_{K'} \,|\, \widehat{\mathcal{H}}^2_n), \quad \forall r = 1,...,\gamma(K'),\ \forall n\in\mathbb{N}. \tag{6.6}$$

The random sequence $(p_n^r)_{n\geq 1}$ corresponds to the sequence of posterior probabilities the adversary ascribes for $P^r_{K'}\in\mathcal{S}_{K'}$, being the chosen state, given that he knows the strategy σ of the observer. Note that $p_1 = \pi_{K'}$. It is well known that the distribution of the martingale of posteriors p is determined solely by $\pi_{K'}$ and σ. Thus, if the adversary knows the strategy $\sigma\in\widehat{\Sigma}$ of the observer, he can compute p_n at any stage n, based on the history of actions played up to that stage, $(i_1, j_1, ..., i_{n-1}, j_{n-1})$. Consider now the following behavior strategy τ_* of the adversary. At stage n compute p_n, and play an optimal mixed action in the one-shot game $G(p_n, w'_n)$ (e.g., (4.3)), where w'_n is the n'th word induced by Y'. By following the footsteps of the proofs of Lemmas V.2.3-2.4 and Proposition V.2.7 in Mertens et. al. (2014), we obtain[17]:

$$\mathbb{E}^{K'}_{\sigma,\tau_*}\left(g(Y'_{K'+n}, i_n, j_n) \,\middle|\, \widehat{\mathcal{H}}^2_n\right) \leq u\left(p_n, w'_n\right) + \|g\|_\infty\, \mathbb{E}^{K'}_{\sigma,\tau_*}\left(\|p_{n+1}-p_n\|_1 \,\middle|\, \widehat{\mathcal{H}}^2_n\right), \tag{6.7}$$

where $\mathbb{E}^{K'}_{\sigma,\tau_*}$ is the expectation operator w.r.t. $\mathbb{P}^{K'}_{\sigma,\tau_*}$, and $\|p_{n+1}-p_n\|_1 = \sum_{r=1}^{\gamma(K')} |p^r_{n+1} - p^r_n|$. The meticulous reader will notice that $Y'_{K'+n}$ is not measurable w.r.t. the σ-field generated by the

[17] This footsteps are made available due to the special structure of $\widehat{\Sigma}$.

elements of $\mathcal{S}_{K'}$, i.e., $P_1^{K'}, ..., P_{K'}^{\gamma(K')}$. As a result, one cannot take the expectation on the left-hand side of Eq. (6.7). To overcome this difficulty, we identify $Y'_{K'+n}$ with the random variable $\widehat{Y}'_{K'+n}$ defined on the probability space $(\mathcal{S}_{K'}, \pi_{K'})$, whose values on each $P_{K'}^r \in \mathcal{S}_{K'}$ agree with the values of $Y'_{K'+n}$ on $P_{K'}^r$. By taking the expectation in Eq. (6.7) and averaging over $n = 1, ..., N$, we get that the expected payoff along the first N stages of $\widehat{\Gamma}(Y')$ under (σ, τ_*), denoted $\widehat{\gamma}_N(\sigma, \tau_*)$, satisfies

$$\widehat{\gamma}_N(\sigma,\tau_*) \leq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{\sigma,\tau_*}^{K'} u\left(p_n, w'_n\right) + \frac{\|g\|_\infty}{N}\sum_{n=1}^{N}\mathbb{E}_{\sigma,\tau_*}^{K'}\|p_{n+1}-p_n\|_1 \tag{6.8}$$

Neyman (2013) proved the sharp upper bound

$$\sum_{n=1}^{N}\mathbb{E}_{\sigma,\tau_*}^{K'}\|p_{n+1}-p_n\|_1 \leq \sqrt{2N\,H(\pi_{K'})}\,, \tag{6.9}$$

where $H(\pi_{K'}) = -\sum_{r=1}^{\gamma(K')}\pi_{K'}^r \log(\pi_{K'}^r)$ is Shannon's entropy function, which together with Eq. (6.8) implies that

$$\widehat{\gamma}_N(\sigma,\tau_*) \leq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{\sigma,\tau_*}^{K'} u\left(p_n, w'_n\right) + \frac{\|g\|_\infty\sqrt{2\,H(\pi_{K'})}}{\sqrt{N}}. \tag{6.10}$$

We now proceed with the following technical result.

Proposition 3. *For any two positive integers M, L, it holds that*

$$\frac{1}{ML}\sum_{n=1}^{ML}\mathbb{E}_{\sigma,\tau_*}^{K'} u\left(p_n, w'_n\right) \leq \frac{1}{M}\sum_{m=0}^{M-1}\mathbb{E}_{\sigma,\tau_*}^{K'}\left(\frac{1}{L}\sum_{\ell=1}^{L} u\left(p_{mL+1}, w'_{mL+\ell}\right)\right) + \frac{\|g\|_\infty\gamma(K')}{\sqrt{M}}. \tag{6.11}$$

Proposition 3 implies that for large values of M, there is an arbitrary small gain in payoff (from the adversary's perspective) if one fixes the belief p_{mL+1} throughout the m'th block of length L, i.e., along stages $mL+1, ..., (m+1)L$, for all $m = 0, ..., M-1$. The proof of Proposition 3 can be found in Appendix B. Let us define, for every $m \in \mathbb{N}$ and $L \in \mathbb{N}$, the function $I_{K'}^{m,L}[Y'] : \Delta(\mathcal{S}_{K'}) \to \mathbb{R}_+$ by

$$I_{K'}^{m,L}[Y'](q) := \frac{1}{L}\sum_{\ell=1}^{L} u\left(q, w'_{mL+\ell}\right).$$

Note that the dependence on the $(K', 1-\varepsilon)$-LT-predictor Y' is through the sequence $\{w'_n\}_{n\geq 1}$ of words induced by it. Jensen's inequality, in conjunction with the martingale of posteriors property of $(p_n)_{n\geq 1}$, implies that

$$\begin{aligned} \mathbb{E}_{\sigma,\tau_*}^{K'}\left[I_{K'}^{m,L}[Y'](p_{mL+1})\right] &\leq \mathbb{E}_{\sigma,\tau_*}^{K'}\left[(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(p_{mL+1})\right] \\ &\leq (\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\mathbb{E}_{\sigma,\tau_*}^{K'}[p_{mL+1}]) \\ &= (\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}). \end{aligned} \tag{6.12}$$

Thus, by combining Eqs. (6.10), (6.11) with the upper bound given in Eq. (6.12) we obtain

$$\widehat{\gamma}_{ML}(\sigma,\tau_*) \leq \frac{1}{M}\sum_{m=0}^{M-1}(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) + \frac{\|g\|_\infty \gamma(K')}{\sqrt{M}} + \frac{\|g\|_\infty \sqrt{2\,H(\pi_{K'})}}{\sqrt{ML}}. \tag{6.13}$$

Hence, by the minimax theorem, for every $M, L \in \mathbb{N}$

$$\widehat{v}_{ML}(Y') \leq \frac{1}{M}\sum_{m=0}^{M-1}(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) + \frac{\|g\|_\infty \gamma(K')}{\sqrt{M}} + \frac{\|g\|_\infty \sqrt{2\,H(\pi_{K'})}}{\sqrt{ML}}. \tag{6.14}$$

The next proposition, whose proof is relegated to Appendix B, relates the values $v_n(\xi)$ to the values $\widehat{v}_n(Y')$, where Y' is some $(K', 1-\varepsilon)$-LT-predictor for ξ.

Proposition 4. *Let Y' be a $(K', 1-\varepsilon)$-LT-predictor for ξ. Then, there exists $N_{Y'} \in \mathbb{N}$ such that $v_n(\xi) \leq \widehat{v}_n(Y') + 5\|g\|_\infty\varepsilon$ for every $n \geq N_{Y'}$.*

Combining the upper bound on $\widehat{v}_{ML}(Y')$ given in Eq. (6.14) with Proposition 4 we get the following corollary.

Corollary 3. *Let Y' be a $(K', 1-\varepsilon)$-predictor for ξ. Then, there exists $R_{Y',\varepsilon} \in \mathbb{N}$ such that*

$$v_{ML}(\xi) \leq \frac{1}{M}\sum_{m=0}^{M-1}(\mathrm{Cav}\, I_{K'}^{m,L}[Y'])(\pi_{K'}) + 6\|g\|_\infty\varepsilon, \tag{6.15}$$

for every $M \geq R_{Y',\varepsilon}$ and $L \in \mathbb{N}$.

Proof of Corollary 3. In view of Proposition 4 and Eq. (6.14), it suffices to take $R_{Y',\varepsilon} \geq N_{Y'}$, where $N_{Y'}$ is described in Proposition 4 to satisfy

$$\frac{\|g\|_\infty \gamma(K')}{\sqrt{R_{Y',\varepsilon}}} + \frac{\|g\|_\infty \sqrt{2\,H(\pi_{K'})}}{\sqrt{R_{Y',\varepsilon} L}} \leq \|g\|_\infty \varepsilon.$$

□

6.5. **Long-term predictors of a special form.**

Theorem 5. *There exists a triangular-array $t = (t_n^r)_{r=1\ ,n=1}^{\gamma(n),\,\infty}$ of positive integers, such that*

(a) $\mu\left(T^{-t_n^1}P_n^1 \cap \ldots \cap T^{-t_n^{\gamma(n)}}P_n^{\gamma(n)}\right) > 0$ *for every $n \in \mathbb{N}$.*

(b) *Define for each $x \in X$ a sequence $Y^n(x) = (Y_\ell^n(x))_{\ell\geq 1}$ of $\mathcal{F}_n$-measurable random variables by*

$$Y_\ell^n(x) := \sum_{r=1}^{\gamma(n)} \xi_\ell(T^{t_n^r}x)\mathbb{1}\{P_n^r\} = \sum_{r=1}^{\gamma(n)} \xi_{\ell+t_n^r}(x)\mathbb{1}\{P_n^r\}, \quad \forall \ell \in \mathbb{N}, \tag{6.16}$$

and define the event

$$C_{K,\varepsilon} = \left\{x \in X : Y^K(x) \text{ is an } (K, 1-\varepsilon)\text{-LT-predictor for } \xi\right\}. \tag{6.17}$$

Then, for every $\varepsilon > 0$ there exists $K_\varepsilon \in \mathbb{N}$ such that $\mu(C_{K,\varepsilon}) > 0$ for every $K \geq K_\varepsilon$.

The proof of Theorem 5 is based on tools and techniques from Topological Dynamics and classic Ergodic Theory. Let us give a very brief introduction to some of these tools (a formal exposition can be found in Subsection A.1 of Appendix A).

The field of topological dynamics comes into play in several directions. First, it is well known (see Theorem A.1) that T must be an *isometry* of X, i.e., T preserves distances in the metric space X. Second, the set of visit times $N(x,U) = \{n \in \mathbb{N} : T^n x \in U\}$ in any open set $U \subseteq X$ of the orbit of any $x \in X$ has bounded gaps, or in other words is *syndetic*.

The connection between the topological dynamics of T on X and the Haar measure μ is achieved via the basic notion of a generic point $x \in X$. The frequency of times, $\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{A\}(T^n x)$, in which the orbit of a generic point x visits a set A, converges (as $N \to \infty$) to the Haar measure of A, in the case where the topological boundary of A has vanishing Haar measure. The key point is that any $x \in X$ in an ergodic Kronecker system must be generic. The proof of this theorem, given in Appendix A, is a significant part of the mathematical novelty of this paper.

Prior to showing why Theorem 5 solves our dynamical difficulties, we provide some intuition as to why the special form of LT-predictors described in Theorem 5 exists. First, for a given $n \in \mathbb{N}$ the existence of $(t_n^r)_{r=1}^{\gamma(n)}$ satisfying item (a) in Theorem 5 is guaranteed by the assumed ergodicity of the Kronecker system (X,T) (note that the sets P_n^r are of positive μ-measure). Second, in view of Theorem 4, for most $r = 1, ..., \gamma(K)$, the (deterministic) sequence of outcomes $(y_n^r)_{n\geq 1}$ predicts well most of the outcomes of $(\xi_n)_{n\geq 1}$ on P_K^r . An a posteriori transitivity argument thus implies that the sequences of outcomes $(\xi_n(x))_{n\geq 1}$ on P_K^r are 'close' (agree most of the times on average) on most points inside $x \in P_K^r$. Thus, the remaining goal is to choose a set of points $x \in X$ of positive μ-measure such that $T^{t_K^r} x \in P_K^r$ are elements of this set of most 'close' points, across most $r = 1, ..., \gamma(K)$.

We now introduce some key definitions, which play a central role in establishing a formula for the uniform value. For every $K \in \mathbb{N}$ and every $w \in \mathcal{W}_K$ define the event

$$R_K^w := \Big\{x \in X : \Big(\xi_{K+t_1^K+1}(x), ..., \xi_{K+t_{\gamma(K)+1}}(x)\Big) = w\Big\}. \tag{6.18}$$

Define the process $\mathbb{I}_K^w = (\mathbb{I}_n^{K,w})_{n\geq 1}$ by the rule

$$\mathbb{I}_n^{K,w}(x) := \mathbb{1}\{T^{-(n-1)} R_K^w\}(x), \quad \forall x \in X,\ \forall n \in \mathbb{N}. \tag{6.19}$$

It follows from the definition given in Eq. (6.16) that $\mathbb{I}_n^{K,w}(x)$ indicates whether or not the n'th word induced by $Y^K(x)$ equals w. Thus, $\frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_n^{K,w}(x)$ measures the frequency of occurrences of the word w along the first N words induced by $Y^K(x)$.

6.6. **The predictions of interest to the observer.** The observer begins by fixing $K \geq K_\varepsilon$ for which,

$$\Big|(\mathrm{Cav}\,\Phi_K)(\pi_K) - \limsup_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n)\Big| \leq \varepsilon. \tag{6.20}$$

Next, define the set

$$O_K = \bigcap_{w \in \mathcal{W}_K} \left\{ x \in X \, : \, \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_n^{K,w}(x) \to \mu(R_K^w) \right\}. \tag{6.21}$$

In words, O_K consists of those points $x \in X$ for which the limit frequency of occurrences of every word $w \in \mathcal{W}_K$ along the entire sequence of words induced by $Y^K(x)$ equals $\mu(R_K^w)$.

Lemma 1. *$I_K[Y^K(x)] = \Phi_K$ for any $x \in O_K$.*

Proof of Lemma 1. Since $\sum_{w \in \mathcal{W}_K} \mathbb{I}_n^{K,w}(x) = 1$ for every $n \in \mathbb{N}$, and $\{\mathbb{I}_n^{K,w}(x)\}_w$ are all binary, we have

$$\begin{aligned} I_K[Y^K(x)](q) &= \liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{w \in \mathcal{W}_K} \mathbb{I}_n^{K,w}(x) u(q,w) \\ &= \liminf_{N \to \infty} \sum_{w \in \mathcal{W}_K} \left[\frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_n^{K,w}(x) \right] u(q,w). \end{aligned} \tag{6.22}$$

Since $x \in O_K$, we may move to the limit in Eq. (6.22) to obtain

$$I_K[Y^K(x)](q) = \sum_{w \in \mathcal{W}_K} \mu(R_K^w) u(q,w) = \Phi_K(q). \tag{6.23}$$

$\square$

Lemma 2. *The observer can guarantee $\limsup_{n \to \infty} (\mathrm{Cav}\, \Phi_n)(\pi_n)$.*

Proof of Lemma 2. Since (X, T, μ) is ergodic and $|\mathcal{W}_K| < \infty$, Birkhoff's Pointwise Ergodic Theorem implies that O_K is a finite intersection of probability 1 events, so that $\mu(O_K) = 1$. Since $K \geq K_\varepsilon$, Theorem 5 implies that

$$\mu(C_{K,\varepsilon} \cap O_K) > 0. \tag{6.24}$$

Hence, the observer can choose some $x \in C_{K,\varepsilon} \cap O_K$. Since $Y^K(x)$ is a $(K, 1-\varepsilon)$-LT-predictor, we can apply Proposition 2 (for $Y = Y^K(x)$) to obtain that

$$\begin{aligned} \gamma_N(\sigma_\varepsilon^{Y^K(x)}, \tau) &\geq (\mathrm{Cav}\, I_K[Y^K(x)])(\pi_K) - 2\varepsilon - 3\|g\|_\infty \varepsilon \\ &= (\mathrm{Cav}\, \Phi_K)(p_K) - \varepsilon - 2\|g\|_\infty \varepsilon \\ &\geq \limsup_{n \to \infty} (\mathrm{Cav}\, \Phi_n)(\pi_n) - 2\varepsilon - 2\|g\|_\infty \varepsilon, \quad \forall N > N_{Y^K(x)}, \forall \tau \in \mathcal{T}, \end{aligned} \tag{6.25}$$

where the equality is due to Lemma 1, and the second inequality follows from the choice of K. $\square$

6.7. **The predictions of interest to the adversary.** The adversary begins by fixing $K' \geq K_\varepsilon$ such that

$$\left|(\text{Cav}\,\Phi_{K'})(\pi_{K'}) - \liminf_{n\to\infty}(\text{Cav}\,\Phi_n)(\pi_n)\right| < \varepsilon. \tag{6.26}$$

We will need the following measurable version of Corollary 3.

Lemma 3. *There exist a positive integer $M_{K',\varepsilon} \in \mathbb{N}$ and an event $G_{K',\varepsilon} \subseteq C_{K',\varepsilon}$ with $\mu(G_{K',\varepsilon}) > 0$, such that if $x \in G_{K',\varepsilon}$ then*

$$v_{ML}(\xi) \leq \frac{1}{M}\sum_{m=0}^{M-1}(\text{Cav}\,I_{K'}^{m,L}[Y^{K'}(x)])(\pi_{K'}) + 6\|g\|_\infty\varepsilon, \tag{6.27}$$

for every $M \geq M_{K',\varepsilon}$ and $L \in \mathbb{N}$.

This proposition is of great importance for us, as the cutoff $M_{K',\varepsilon}$ is uniform across all $(K', 1-\varepsilon)$-LT-predictors $\{Y^{K'}(x)\}_{x\in G_{K',\varepsilon}}$, unlike the cutoff $R_{Y',\varepsilon}$ given in Corollary 3, whose value depend on a specific $(K', 1-\varepsilon)$-LT-predictor Y' the adversary decides to use. The proof of Lemma 3 relies heavily on the proof of Proposition 4 and will be given in Appendix B.

In order to relate $(\text{Cav}\,I_{K'}^{m,L}[Y^{K'}(x')])(\pi_{K'})$, $m = 0, ..., M-1$, to $(\text{Cav}\,\Phi_{K'})(\pi_{K'})$ for some $x' \in G_{K',\varepsilon}$, the adversary desires that at each block $mL+1, ..., (m+1)L$ of words induced by $Y^{K'}(x')$ the frequency of every word $w \in \mathcal{W}_{K'}$ is 'close' to $\mu(R^w_{K'})$. For this purpose we define for each $L \in \mathbb{N}$ and $m \in \mathbb{N}\cup\{0\}$ the event

$$D_{L,m}(\beta) := \bigcap_{w\in\mathcal{W}_{K'}}\left\{x \in X \,:\, \left|\frac{1}{L}\sum_{\ell=1}^{L}\mathbb{I}^{K',w}_{mL+\ell}(x) - \mu(R^w_{K'})\right| < \beta\right\}.$$

Note that $\frac{1}{L}\sum_{\ell=1}^{L}\mathbb{I}^{K',w}_{mL+\ell}(x)$ is equal to the frequency of w along the m'th L-block of words induced by $Y^{K'}(x)$. Thus, the event $D_{L,m}(\beta)$ collects those $x \in X$ for which the frequency of each word $w \in \mathcal{W}_{K'}$ along the m'th L-block of words induced by $Y^{K'}(x)$ is β-close to $\mu(R^w_{K'})$.

The following lemma will help the adversary choose the predictor $Y^{K'}(x)$ he is interested in.

Lemma 4. *For every positive integer M and every $\beta > 0$, there exists a positive integer $L = L(M,\beta)$ such that*

$$\mu\left(G_{K',\varepsilon} \cap \left(\bigcap_{m=0}^{M-1} D_{L,m}(\beta)\right)\right) > 0. \tag{6.28}$$

The proof of Lemma 4, which mostly relies on Birkhoff's Pointwise Ergodic Theorem, is deferred to Appendix B.

Let us fix $M \geq M_{K',\varepsilon}$. Lemma 4 implies that, for a given $\beta > 0$, the adversary can choose $L \in \mathbb{N}$ and $x' \in G^b_{K',\varepsilon} \cap (\cap_{m=0}^{M-1} D_{L,m}(\beta))$ so that $Y' = Y^{K'}(x')$ satisfies the following two key properties:

(i) $v_{ML}(\xi) \leq \frac{1}{M}\sum_{m=0}^{M-1}(\text{Cav}\,I_{K'}^{m,L}[Y'])(\pi_{K'}) + 6\|g\|_\infty\varepsilon$ (since $x \in G_{K',\varepsilon}$).

(ii) For every $m = 0, ..., M-1$ the frequency of each word $w \in \mathcal{W}_{K'}$ along the m'th L-block of words induced by Y' is β-close to $\mu(R^w_{K'})$ (since $x \in \bigcap_{m=0}^{M-1} D_{L,m}(\beta)$).

The following simple claim, whose proof appears in Appendix B, is important for later analysis.

Claim 1. *Let $f, h : E \to \mathbb{R}$ be two continuous functions, where E is a closed convex subset of $\mathbb{R}^d$, $d \in \mathbb{N}$. Then*

$$\|f - h\|_\infty < \rho \implies \|\mathrm{Cav}\, f - \mathrm{Cav}\, h\|_\infty < \rho. \tag{6.29}$$

The approximation of $\mathrm{Cav}\, I^{m,L}_{K'}[Y']$ to $(\mathrm{Cav}\,\Phi_{K'})(\pi_{K'})$ is achieved as follows. First, observe that the definitions imply the identity

$$I^{m,L}_{K'}[Y'](q) = \sum_{w\in\mathcal{W}_{K'}} \left[\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}^{K',w}_{mL+\ell}(x')\right] u\left(q, w\right), \quad \forall q \in \Delta(\mathcal{S}_{K'}). \tag{6.30}$$

Second, since $x \in \bigcap_{m=0}^{M-1} D_{L,m}(\beta)$ (where $L = L(M, \beta)$ is described in Lemma 4), we have for every $m = 0, ..., M-1$ that

$$\begin{aligned} \|I^{m,L}_{K'}[Y'] - \Phi_{K'}\|_\infty &= \left\| \sum_{w\in\mathcal{W}_{K'}} \left(\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}^{K',w}_{mL+\ell}(x') - \mu(R^w_{K'})\right) u\left(\cdot, w\right) \right\|_\infty \\ &\le \sum_{w\in\mathcal{W}_{K'}} \left|\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}^{K',w}_{mL+\ell}(x') - \mu(R^w_{K'})\right| \|u\left(\cdot, w\right)\|_\infty \le |\mathcal{W}_{K'}|\|g\|_\infty\beta. \end{aligned} \tag{6.31}$$

Third, by Claim 1,

$$\|\mathrm{Cav}\, I^{m,L}_{K'}[Y'] - \mathrm{Cav}\,\Phi_{K'}\|_\infty \le |\mathcal{W}_{K'}|\|g\|_\infty\beta, \tag{6.32}$$

which in conjunction with property (i) of Y' yields

$$\begin{aligned} v_{ML}(\xi) &\le \frac{1}{M}\sum_{m=0}^{M-1}\Big[(\mathrm{Cav}\,\Phi_{K'})(\pi_{K'}) + |\mathcal{W}_{K'}|\|g\|_\infty\beta\Big] + 6\|g\|_\infty\varepsilon \\ &\le \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) + \varepsilon + |\mathcal{W}_{K'}|\|g\|_\infty\beta + 6\|g\|_\infty\varepsilon, \end{aligned} \tag{6.33}$$

where the last inequality is due to the choice of K'. As $M \ge M_{K',\varepsilon}$ is arbitrary, we can let $M \to \infty$, and since $v_n(\xi)$ converges as $n \to \infty$ (see Observation 1) we see that

$$\lim_{n\to\infty} v_n(\xi) \le \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) + \varepsilon + |\mathcal{W}_{K'}|\|g\|_\infty\beta + 6\|g\|_\infty\varepsilon. \tag{6.34}$$

Since $\varepsilon > 0$ and $\beta > 0$ were arbitrary all along, we conclude that

Corollary 4. $\lim_{n\to\infty} v_n(\xi) \le \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n)$.

Let us now prove Theorems 1 and 2 simultaneously.

Proof of Theorems 1 and 2. By Lemma 2 and Corollary 4,

$$\limsup_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n) \le \liminf_{n\to\infty}(\mathrm{Cav}\,\Phi_n)(\pi_n). \tag{6.35}$$

Hence, $v(\xi)$ exists and equals $\lim_{n\to\infty}(\operatorname{Cav}\Phi_n)(\pi_n)$, which proves Theorem 2. Next, by Observation 1, Lemma 2 and Corollary 4,

$$\lim_{n\to\infty} v_n(\xi) \le \liminf_{n\to\infty}(\operatorname{Cav}\Phi_n)(\pi_n) \le \limsup_{n\to\infty}(\operatorname{Cav}\Phi_n)(\pi_n) \le \lim_{n\to\infty} v_n(\xi), \tag{6.36}$$

which proves Theorem 1. □

6.8. **Proof of the result on visiting times to Kronecker systems.** Let us prove now Theorem 3.

Proof of Theorem 3. We first note that it suffices to show that

$$\mu(P_n^1)\xi_{t_n^1} + \cdots + \mu(P_n^{\gamma(n)})\xi_{t_n^{\gamma(n)}} \to \mathbb{E}\,\xi_1 \quad \text{in } L^1(X, \mathcal{B}(X), \mu), \tag{6.37}$$

for the triangular array $t = (t_n^r)_{r=1,\,n=1}^{\gamma(n),\,\infty}$ described in Theorem 2, whenever $|\mathcal{P}| = 2$. As $\mathcal{P}$ is an arbitrary partition we do not lose generality. We find such a reduction useful, as it significantly helps simplify notation.

Define $w : [0,1] \to \mathbb{R}_+$ by

$$w(s) = \operatorname{val}\left((1-s)g(0,\cdot,\cdot) + sg(1,\cdot,\cdot)\right).$$

That is, $w(s)$ is the value of the zero-sum game in which the payoff for the actions pair $(i,j) \in I \times J$ equals $(1-s)g(0,i,j) + sg(1,i,j)$. Also, for each $n \in \mathbb{N}$, define $\Psi_n : \Delta(\mathcal{S}_n) \to \mathbb{R}$ by

$$\Psi_n(q) := \mathbb{E}\left[w(\,\langle q, (\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}})\rangle)\right], \tag{6.38}$$

where $\mathbb{E}$ is the expectation operator w.r.t. μ, $\langle\cdot,\cdot\rangle$ is the scalar product on $\mathbb{R}^{\gamma(n)}$, and $q \in \Delta(\mathcal{S}_n)$ is viewed as a probability vector in $\mathbb{R}^{\gamma(n)}$. Since we assume that $|\mathcal{P}| = 2$, the process ξ has binary steps (see (3.1)). Consequently, $\Psi_n = \Phi_n$, which implies $\operatorname{Cav}\Phi_n = \operatorname{Cav}\Psi_n$ for every $n \in \mathbb{N}$. By Theorem 2, $\lim_{n\to\infty}(\operatorname{Cav}\Psi_n)(\pi_n)$ exists and is equal to the uniform value of $\Gamma(\xi)$.

We now state the following key proposition.

Proposition 5. *Assume that w is a concave function. Then*

$$\lim_{n\to\infty} \Psi_n(\pi_n) = w(\mathbb{E}\,\xi_1). \tag{6.39}$$

The proof can be found in Appendix B. We now proceed with the definition of a piecewise rational function.

Definition 5. A function $h : [0,1] \to \mathbb{R}$ is said to be *piecewise rational*, if there exist a positive integer N and points in the unit interval $0 = s_0 < s_1 < s_2 < \cdots < s_N = 1$ such that

$$h = \sum_{n=1}^{N-1} \mathbb{1}\{[s_{n-1}, s_n)\}\frac{P_n}{Q_n} + \mathbb{1}\{[s_{N-1}, 1]\}\frac{P_N}{Q_N}, \tag{6.40}$$

where $(P_n)_{n=1}^N$ and $(Q_n)_{n=1}^N$ are polynomials.

Recently, Ashkenazi-Golan, Solan, and Zseleva (2020) proved the following result:

Theorem 5. *A function $h : [0,1] \to \mathbb{R}$ is a continuous piecewise rational function iff there exist two matrices A and B of equal dimensions such that:*

$$h(s) = \mathrm{val}\,(sA + (1-s)B)\,. \tag{6.41}$$

Consider the sequence of concave polynomials $w_k : [0,1] \to [-1,0]$ defined by $w_k(s) := -s^k$ for every $k \in \mathbb{N}$. Using Theorem 5 one can select for each k matrices $G^k(0)$ and $G^k(1)$ such that

$$w_k(s) = \mathrm{val}((1-s)G^k(0) + sG^k(1))$$

for all $s \in [0,1]$. By considering the repeated game $\Gamma_\infty(\xi)$ with payoff function $g(a,i,j) = G^k(a)[i,j]$, we deduce from Proposition 5 that

$$\lim_{n\to\infty} \mathbb{E}\left[\,(\,\langle \pi_n, (\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}})\rangle\,)^k\right] = \mathbb{E}[(\xi_1)^k],\ \forall k \in \mathbb{N}. \tag{6.42}$$

Since the constant random variable $\mathbb{E}(\xi_1)$ satisfies the Carleman condition, i.e.,

$$\sum_{n=1}^{\infty} \left(\mathbb{E}(\xi_1)^{2n}\right)^{-\frac{1}{2n}} = \sum_{n=1}^{\infty} \frac{1}{\mathbb{E}(\xi_1)} = +\infty, \tag{6.43}$$

it is uniquely determined by its moments. The method of moments [16] implies that $\langle \pi_n, (\xi_{t_n^1}, ..., \xi_{t_n^{\gamma(n)}})\rangle$ converges in distribution to $\mathbb{E}\,\xi_1$. As the latter is a constant random variable, one obtains convergence in probability. Hence, as ξ has binary steps, convergence in $L^1(X, \mathcal{B}(X), \mu)$ holds as well, thus proving (6.37). □

Acknowledgements. The authors wish to thank an anonymous associate editor and two anonymous referees for their valuable comments, which greatly contributed to the improvement of the paper. The authors also wish to thank Andrei Iacob, Eilon Solan, Barak Weiss, and Benjamin Weiss for their contribution to this work.

References

[1] Ashkenazi-Golan, G., Solan E., and Zseleva, A. (2020), Characterizing the value functions of polynomial games. *Operations Research Letters*, 48: 93–95

[2] Aumann, R.J. and Maschler, M. (1995), *Repeated games with incomplete information.* with the collaboration of R. Stearns, Cambridge, MA: MIT Press.

[3] Aumann, R. J. (1964), *Mixed and behavior strategies in infinite extensive games.* In: Advances in Game Theory pp. 627–650, Princeton Univ. Press, Princeton, N.J.

[4] Blackwell, D. and Ferguson, T. S. (1968), The big match. *Ann. Math. Statist.* 39: 159–163.

[5] Bressaud, X. and Quas, A. (2017), Dynamical analysis of a repeated game with incomplete information. *Mathematics of Operations Research*, 42, 4: 1085–1105.

[6] Einsiedler, M. and Ward, T. (2011) *Ergodic theory with a view towards number theory.* Graduate Texts in Mathematics, 259. Springer-Verlag, London.

[7] Forman, M. and Weiss, B. (2020), Odometer based systems. *manuscript.*

[8] Furstenberg, H. (1981). *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, Princeton, N.J.

[9] Gensbittel, F., Oliu-Barton, M. and Venel, X. (2014), Existence of the uniform value in zero-sum repeated games with a more informed controller. *J. Dyn. Games*, 1,3: 411–445.
[10] Hörner, J., Rosenberg, D., Solan, E., and Vieille, N. (2010), On a Markov game with one-sided incomplete information. *Operations Research*, 58:1107–1115.
[11] Mertens, J.F., Sorin, S. and Zamir, S. (2015), *Repeated games.* Cambridge: Cambridge University Press.
[12] Morvai, G. and Weiss, B. (2007), On sequential estimation and prediction for discrete time series. *Stochastics and Dynamics*, 7, 4: 417–437.
[13] Neveu, J. (1970), *Bases mathématiques du calcul des probabilités.* Masson et Cie.
[14] Neyman, A. (2008), Existence of optimal strategies in Markov games with incomplete information. *Int. J. Game Theory*, 37: 581–596.
[15] Neyman, A. (2013), The maximal variation of martingales of probabilities and repeated games with incomplete information. *J. Theor. Probab.*, 26: 557–567.
[16] Prokhorov, A.V., Moments, method of (in probability theory). In: M. Hazewinkel (ed.). *Encyclopedia of Mathematics.*
[17] Renault, J. (2006), The value of Markov chain games with lack of information on one side. *Mathematics of Operations Research*, 31, 3: 433–648.
[18] Renault, J. (2012), The value of repeated games with an informed controller. *Mathematics of Operations Research*, 37, 1: 154–179.
[19] Renault, J. (2018), *Repeated games with incomplete information.* In: Meyers R. (eds) Encyclopedia of Complexity and Systems Science. Springer, Berlin, Heidelberg.
[20] Rockafellar, R. T. (1970), *Convex analysis.* Princeton University Press.
[21] Rosenberg, D., Solan, E. and Vieille, N. (2004), Stochastic games with a single controller and incomplete information. *SIAM J. Control Optim.*, 43, 1: 86–110.
[22] Shiryaev, A.N. (1996), *Probability.* Translated from the first (1980) Russian edition by R. P. Boas. Second edition. Graduate Texts in Mathematics, 95. Springer-Verlag, New York.
[23] Smorodinsky, M. (1971), *Ergodic theory, entropy.* Lecture Notes in Mathematics, 214, Springer-Verlag Berlin-Heidelberg-New York.
[24] Sorin, S. (1984), Big Match with lack of information on one side (part I). *Int. J. Game Theory*, 13: 201–255.

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail address: `lehrer@post.tau.ac.il`

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail address: `dima.shaiderman@gmail.com`

Appendix A. Long-Term Predictors of a Special Form

A.1. **Tools from Topological Dynamics.** The goal of this section is to introduce some basic notions and review fundamental theorems and facts from topological dynamics which will play a key role in the proofs of Theorems 4 and 5. The items in this short list together with their proofs are scattered throughout Chapters 1 and 3 in Furstenberg (1981).

Theorem A.1. *Any Kronecker system* $(\mathcal{Y}, S)$ *is isometric, i.e., it admits a compatible metric* $\tilde{d}$ *on* $\mathcal{Y}$ *with respect to which* S *is an isometry, i.e.,* $\tilde{d}(Sy, Sy') = \tilde{d}(y, y')$ *for all* $y, y' \in \mathcal{Y}$.

A topological dynamical system $(\mathcal{Y}, S)$ is said to be *transitive* if $(S^n y)_{n\in\mathbb{Z}}$ is a dense subset of $\mathcal{Y}$ for every $y \in \mathcal{Y}$. A subset $I \subset \mathbb{N}$ is *syndetic* it there is a constant L such that

$$[m, m+L] \cap I \neq \emptyset, \quad \forall m \in \mathbb{N},$$

that is, the gaps in I are of lengths bounded from above by L. The number L is called the *syndeticity constant* of I. The interplay between these properties is summarized in the following fundamental theorem:

Theorem A.2. *The following assertions are equivalent:*

1. $(\mathcal{Y}, S)$ *is transitive.*
2. *For every* $y \in \mathcal{Y}$ *and every non-empty open set* $U \subseteq \mathcal{Y}$*, the set* $N(y, U) := \{n \in \mathbb{N} : S^n y \in U\}$ *is syndetic.*

It is well known that ergodic Kronecker systems are transitive. We conclude with a useful result on generic points. Let $(\mathcal{Y}, S)$ be a topological dynamical system. A point $y \in \mathcal{Y}$ is said to be *generic* for a Borel measure μ on $\mathcal{Y}$ if

$$\frac{1}{N}\sum_{n=1}^{N} f(S^n y) \to \int f d\mu, \quad \forall f \in C(\mathcal{Y}),$$

where we recall that $C(\mathcal{Y})$ stands for the space of real-valued continuous functions with domain $\mathcal{Y}$. It is well known that in a uniquely ergodic topological dynamical system $(\mathcal{Y}, S)$ every point $y \in \mathcal{Y}$ is generic with respect to the (unique) invariant measure μ. The following fact regarding generic points will be crucial for our proofs:

Fact A.1. *If* y *is generic for* μ *and* $A \in \mathcal{B}(\mathcal{Y})$ *satisfies* $\mu(\partial A) = 0$ *(*∂A *denotes the boundary of the set* A*), then* $\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{A\}(S^n x) \to \mu(A)$ *as* $N \to \infty$*.*

A.2. **Proofs.** Recall that in Section 4 we fixed an ergodic Kronecker system (X, T) and a finite partition $\mathcal{P}$ of X. We also denoted $\xi = (\mathcal{P}, T)$. The set of states is thus $A = \{0, ..., |\mathcal{P}| - 1\}$. Our starting point is the following proposition.

Proposition A.1. *There exists a countable basis* $\mathscr{B}$ *for the topology on* X*, such that for every finite partition* $\mathcal{Q}$ *measurable w.r.t.* $\mathcal{A}(\mathscr{B})$ *(the algebra of sets generated by the elements of* $\mathscr{B}$*) and every* $\varepsilon > 0$ *there exist a positive integer* N *and a number* $\delta > 0$ *such that for every* $x, y \in X$ *we have*

$$d(x, y) < \delta \implies \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\eta_\ell(x) \neq \eta_\ell(y)\} < \varepsilon, \quad \forall n \geq N,$$

where $\eta = (\mathcal{Q}, T)$*.*

Proof of Proposition A.1. We first need to exhibit a countable basis $\mathscr{B}$ for which the conclusion of Proposition A.1 is true. To do so, we recall that, by the definition of a Kronecker system (see Definition 2), there exists a homeomorphism $\phi : X \to G$ such that $\phi \circ T = S_g \circ \phi$, where (G, S_g) is a topological dynamical system, with $(G,+)$ a compact metrizable Abelian topological group and $S_g : h \mapsto g+h$ is the group rotation by $g \in G$.

By Theorem A.1, there exists a compatible metric d_g on G such that S_g is an isometry of d_g. As the Haar measure μ on G is invariant under left rotations and S_g is an isometry, one can choose a decreasing sequence $r_m \searrow 0$ such that $\mu(\partial B_{d_g}(ng, r_m)) = 0$ for all $n \in \mathbb{Z}$ and $m \in \mathbb{N}$ (here ng is the element of G obtained by adding g (resp. $-g$) to itself n times if $n \geq 0$ (resp. $n < 0$), and $B_{d_g}(ng, r_m)$ denotes the ball of radius r_m around ng w.r.t. the metric d_g). Since (G, S_g) is transitive, the set $\{ng : n \in \mathbb{Z}\}$ is dense in G. Hence, the family $\mathscr{B}_g = \{B_{d_g}(ng, r_m) : n \in \mathbb{Z}, m \in \mathbb{N}\}$ forms a countable basis of G. Since ϕ is a homeomorphism, we may define the countable basis $\mathscr{B}$ by

$$\mathscr{B} = \phi^{-1}\mathscr{B}_g := \{\phi^{-1} B_{d_g}(ng, r_m) : n \in \mathbb{Z}, m \in \mathbb{N}\}. \tag{A.1}$$

Fix a finite partition $\mathcal{Q} = \{Q_1, ..., Q_k\}$ measurable w.r.t. $\mathcal{A}(\mathscr{B}_g)$ and fix $\varepsilon > 0$. For each $a = 1, ..., k$, let $R_a = \phi\, Q_k = \{\phi(x) : x \in Q_a\}$. Thus, the finite partition $\mathcal{R} = \{R_1, ..., R_k\}$ is measurable w.r.t. $\mathcal{A}(\mathscr{B}_g)$. Consider the process $\zeta = (\mathcal{R}, S_g)$. Then, for every $x \in X$ and $\ell \in \mathbb{N}$,

$$\begin{aligned} \eta_\ell(x) &= \sum_{a=1}^{|\mathcal{Q}|} (a-1)\mathbb{1}\{T^{-(\ell-1)}Q_a\}(\phi^{-1}(\phi(x))) \\ &= \sum_{a=1}^{|\mathcal{Q}|} (a-1)\mathbb{1}\{\phi \circ T^{-(\ell-1)}Q_a\}(\phi(x)) \\ &= \sum_{a=1}^{|\mathcal{Q}|} (a-1)\mathbb{1}\{S_g^{-(\ell-1)} \circ \phi\ Q_a\}(\phi(x)) \\ &= \sum_{a=1}^{|\mathcal{R}|} (a-1)\mathbb{1}\{(S_g^{-(\ell-1)} R_a)\}(\phi(x)) \\ &= \zeta_\ell(\phi(x)). \end{aligned} \tag{A.2}$$

Eq. (A.2) implies that, for every $n \geq 1$,

$$\frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\eta_\ell(x) \neq \eta_\ell(y)\} = \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\zeta_\ell(\phi(x)) \neq \zeta_\ell(\phi(y))\}. \tag{A.3}$$

We now claim that to prove Proposition A.1 it suffices to find an $N \in \mathbb{N}$ and $\delta' > 0$ such that for every $h, f \in G$ it holds that

$$d_g(h,f) < \delta' \implies \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\zeta_\ell(h) \neq \zeta_\ell(f)\} < \varepsilon, \quad \forall n \geq N. \tag{A.4}$$

To see this, we note that since (X, d) is compact, ϕ is uniformly continuous, and thus there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $d_g(\phi(x), \phi(y)) < \delta'$ for all $x, y \in X$. Then, Eq. (A.3) would guarantee that the conclusion of Proposition A.1 holds.

Thus, it remains to prove the reduction described in (A.4). We begin by introducing the open sets

$$U_{a,\alpha} = \{h \in G \, : \, d(h, \partial R_a) < \alpha\}, \qquad \forall a = 1, ..., |\mathcal{R}|. \tag{A.5}$$

Since ∂R_a is closed, $U_{a,\alpha} \searrow \partial R_a$ as $\alpha \searrow 0$, for all a. Moreover, since $\mu(\partial R_a) = 0$ for all a (recall the definition of $\mathscr{B}_g$), we can choose α_0 such that (i) $\sum_a \mu(U_{a,\alpha_0}) < \varepsilon/2$ and (ii) $\mu\left(\partial U_{a,\alpha_0}\right) = 0$ for all a. Set $\delta' := \alpha_0/4$, and let $\{h_j\}_{j=1}^{\mathcal{J}}$ be a δ'-net in G. Since (G, S_g) is uniquely ergodic, h_j is generic for every $j = 1, ..., \mathcal{J}$. Thus, as

$$\partial\left(\bigcup_{a=1}^{|\mathcal{R}|} U_{a,\alpha_0}\right) \subseteq \bigcup_{a=1}^{|\mathcal{R}|} \partial U_{a,\alpha_0} \implies \mu\left(\partial\left(\bigcup_{a=1}^{|\mathcal{R}|} U_{a,\alpha_0}\right)\right) = 0, \tag{A.6}$$

Fact A.1 ensures the existence of a positive integer N such that for every $n \geq N$ it holds that

$$\left|\frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\bigcup_{a=1}^{|\mathcal{R}|} U_{a,\alpha_0}\}(S_g^{\ell-1}h_j) - \mu\left(\bigcup_{a=1}^{|\mathcal{R}|} U_{a,\alpha_0}\right)\right| < \frac{\varepsilon}{2}, \quad \forall j = 1, ..., \mathcal{J}. \tag{A.7}$$

If $S_g^{\ell-1}h_j \notin \bigcup_a U_{a,\alpha_0}$, the ball $B_{d_g}(S_g^{\ell-1}h_j, \alpha_0/2)$ must lie entirely inside one of the elements of the partition $\mathcal{R}$. Hence, by exploiting the fact that S_g is an isometry of d_g once more, we see that, for every $h, f \in G$ such that $d_g(h, f) < \delta'$,

$$\begin{aligned} \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\zeta_\ell(h) \neq \zeta_\ell(f)\} &\leq \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\bigcup_{a=1}^{|\mathcal{R}|} U_{a,\alpha_0}\}(S_g^{\ell-1}\hat{h}) \\ &\leq \frac{\varepsilon}{2} + \mu\left(\bigcup_{a=1}^{|\mathcal{R}|} U_{a,\alpha_0}\right) \leq \varepsilon, \quad \forall n \geq N, \end{aligned} \tag{A.8}$$

where $\hat{h} \in \{h_j\}_{j=1}^{\mathcal{J}}$ satisfies $d_g(\hat{h}, h) < \delta'$. This proves our reduction and thus completes the proof of Proposition A.1. □

Let us fix a countable basis $\mathscr{B}$ for the topology on X satisfying the conclusion of Proposition A.1. We need a number of additional definitions and notations. For each $n \in \mathbb{N}$, we define $\rho_n : A^{\mathbb{N}} \times A^{\mathbb{N}} \to [0, 1]$ by

$$\rho_n(\alpha, \beta) := \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}\{\alpha_\ell \neq \beta_\ell\}.$$

As the functions $(\rho_n)_{n\geq 1}$ are measurable w.r.t. the product σ-field $\mathcal{B}(A^{\mathbb{N}}) \otimes \mathcal{B}(A^{\mathbb{N}})$, so is the function $\rho : A^{\mathbb{N}} \times A^{\mathbb{N}} \to [0, 1]$ defined by

$$\rho(\alpha, \beta) := \limsup_{n\to\infty} \rho_n(\alpha, \beta).$$

Lemma A.1. *Let $\mathcal{Q}$ be a partition measurable w.r.t. $\mathcal{A}(\mathscr{B})$, $|\mathcal{Q}| = |\mathcal{P}|$, and let $\eta = (\eta_n)_{n\geq 1}$ be the process generated by $(\mathcal{Q}, T)$. For every $\varepsilon > 0$ there exists a positive integer $N(\eta, \varepsilon)$ such that for every $n \geq N(\eta, \varepsilon)$ and all $x, y \in X$,*

$$\rho_n(\eta(x), \eta(y)) < \varepsilon \implies \rho(\eta(x), \eta(y)) \leq 2\varepsilon. \tag{A.9}$$

Proof of Lemma A.1. First, by Theorem A.1, we may assume that T is an isometry of the metric d on X. By Proposition A.1, there exist $\delta > 0$ and $N \geq 1$ such that for all $x, x' \in X$ we have

$$d(x, x') < \delta \implies \rho_n(\eta(x), \eta(x')) < \varepsilon/4, \quad \forall n \geq N. \tag{A.10}$$

Since (X, T) is transitive, by Theorem A.2 for every $x \in X$ there exists a positive integer L_x which is the syndeticity constant of $N(x, B_d(x, \delta))$. Since T is an isometry, $N(x, B_d(x, \delta)) = N(y, B_d(y, \delta))$ for all $x, y \in X$. Thus $L_x = L_y$ for all $x, y \in X$. To simplify notation, we write $L = L_x$ for some $x \in X$. Choose $N(\eta, \varepsilon) > N$ such that $L < (\varepsilon/2)N(\eta, \varepsilon)$. Take $n \geq N(\eta, \varepsilon)$. We have $\rho_n(\eta(x), \eta(y)) < \varepsilon$. Thereafter, we have a waiting time of at most L till the orbit of $T^n x$ returns to $B_d(x, \delta)$. Since T is an isometry, whenever $T^n x$ returns to $B_d(x, \delta)$, $T^n y$ returns to $B_d(y, \delta)$. Let $m = \min\{0 \leq \ell < L : T^{n+\ell} x \in B_d(x, \delta)\}$. Combining Eq. (A.10) with the fact that $\rho_n(\eta(x), \eta(y)) < \varepsilon$, we see that $\rho_n(\eta(T^{n+m} x), \eta(T^{n+m} y)) < 3\varepsilon/2$. We continue the same line of logic iteratively, using the fact that the waiting time it takes $T^{2n+m} x$ to return to a small neighborhood of x is at most $L < (\varepsilon/2)N(\eta, \varepsilon)$, and thereafter we once again get a $3\varepsilon/2$ proximity of the corresponding orbits in the ρ_n metric, and so on and so forth. The result follows since the waiting times, which are at most $L < (\varepsilon/2)N(\eta, \varepsilon)$, are always negligible compared to the duration $n > N(\eta, \varepsilon)$ of proximity of orbits in the ρ_n metric. ☐

Lemma A.2. *For every $\varepsilon > 0$, there exists a partition $\mathcal{Q}^\varepsilon$, $|\mathcal{Q}^\varepsilon| = |\mathcal{P}|$, measurable w.r.t. $\mathcal{A}(\mathscr{B})$, such that the process $\eta^\varepsilon = (\eta_n^\varepsilon)_{n\geq 1}$ generated by $(\mathcal{Q}^\varepsilon, T)$ satisfies*

$$\mu\left(\{x \in X \,:\, \rho(\xi(x), \eta^\varepsilon(x)) < \varepsilon\}\right) = 1. \tag{A.11}$$

Proof of Lemma A.2. Fix $\varepsilon > 0$. First, we recall that the Haar measure is outer-regular, that is,

$$\mu(B) = \inf\{\mu(U) : U \text{ open},\ B \subseteq U\}, \quad \forall B \in \mathcal{B}(X). \tag{A.12}$$

Thus, as $\mathscr{B}$ is a countable basis, we can approximate[18] the partition $\mathcal{P}$ by a partition $\mathcal{Q}^\varepsilon = \{Q_1^\varepsilon, ..., Q_{|\mathcal{P}|}^\varepsilon\}$ measurable w.r.t. $\mathcal{A}(\mathscr{B})$, so that $\mu(P_a \triangle Q_a^\varepsilon) \leq \varepsilon/|\mathcal{P}|$[19] for every $a = 1, ..., |\mathcal{P}|$. Let η^ε

[18]This claim can be proved using induction on the cardinality of $\mathcal{P}$.

[19]We denote by $\triangle$ the symmetric difference operation.

be the process generated by $(\mathcal{Q}^\varepsilon, T)$. By Birkhoff's Pointwise Ergodic Theorem,

$$
\begin{aligned}
\frac{1}{n}\sum_{\ell=1}^{n}\mathbb{1}\{\xi_\ell \neq \eta^\varepsilon_\ell\} &\leq \frac{1}{n}\sum_{\ell=1}^{n}\sum_{a=1}^{|\mathcal{P}|}\mathbb{1}\{T^{-(\ell-1)}P_a \triangle T^{-(\ell-1)}Q^\varepsilon_a\} \\
&= \sum_{a=1}^{|\mathcal{P}|}\frac{1}{n}\sum_{\ell=1}^{n}\mathbb{1}\{T^{-(\ell-1)}(P_a \triangle Q^\varepsilon_a)\} \\
&\to \sum_{a=1}^{|\mathcal{P}|}\mu(P_a \triangle Q^\varepsilon_a) \leq \varepsilon, \quad \mu\text{-a.e.,}
\end{aligned}
\tag{A.13}
$$

which proves (A.11). □

Proof of Theorems and 4 and 5. For each $k \in \mathbb{N}$, Lemma A.2 ensures the existence of a partition $\mathcal{Q}^k$, $|\mathcal{Q}^k| = |\mathcal{P}|$, measurable w.r.t. $\mathcal{A}(\mathscr{B})$, such that the event $O_k = \{x \in X : \rho(\xi(x), \eta^k(x)) \leq \frac{1}{16k}\}$, where $\eta^k = (\mathcal{Q}^k, T)$, satisfies $\mu(O_k) = 1$. Moreover, there exists a positive integer $M(k)$ such that

$$
\mu\left(\left\{x \in X \,:\, \rho_n(\xi(x), \eta^k(x)) \leq \frac{1}{8k},\ \ \forall n \geq M(k)\right\}\right) \geq 1 - \frac{1}{k}. \tag{A.14}
$$

For $n \in \mathbb{N}$, and $r = 1, ..., \gamma(n)$, define

$$
E^r_{n,k} := P^r_n \cap \left\{x \in X \,:\, \rho_n(\xi(x), \eta^k(x)) \leq \frac{1}{8k}\right\} \cap O_k.
$$

Denote $J^1_{n,k} := \{r \in \{1, ..., \gamma(n)\} : \mu(E^r_{n,k}) > 0\}$ and $J^2_{n,k} := \{1, ..., \gamma(n)\} \setminus J^1_{n,k}$. Since T is ergodic, for every $n, k \in \mathbb{N}$, there exist positive integers $t^1_{n,k}, t^2_{n,k}, ..., t^{\gamma(n)}_{n,k}$ such that

$$
A_{n,k} := \bigcap_{r \in J^1_{n,k}} T^{-t^r_{n,k}} E^r_{n,k} \cap \bigcap_{r \in J^2_{n,k}} T^{-t^r_{n,k}} P^r_n
$$

satisfies $\mu(A_{n,k}) > 0$. We now define for each pair $k, n \in \mathbb{N}$ and $x \in X$ the sequence $Y^{n,k}(x) = (Y^{n,k}_\ell(x))_{\ell \geq 1}$ of random variables, measurable w.r.t. $\mathcal{F}_n = \sigma(\xi_1, ..., \xi_n)$, by

$$
Y^{n,k}_\ell(x) := \sum_{r=1}^{\gamma(n)} \xi_\ell(T^{t^r_{n,k}}x)\mathbb{1}\{P^r_n\} = \sum_{r=1}^{\gamma(n)} \xi_{\ell + t^r_{n,k}}(x)\mathbb{1}\{P^r_n\}, \quad \forall \ell \in \mathbb{N}.
$$

The following proposition describes an event of positive μ-probability associated with a family of LT-predictors. Note that Theorem 4 follows immediately from it.

Proposition A.2. *For each pair $n, k \in \mathbb{N}$ define*

$$
B_{n,k} := \left\{x \in X : Y^{n,k}(x) \text{ is an } (n, 1 - k^{-1})\text{-LT-predictor for } \xi\right\}.
$$

Then $\mu(B_{n,k}) > 0$ for all $k \in \mathbb{N}$ and $n \geq \max\{M(k), N(\eta^k, \frac{1}{4k})\}$, where $M(k)$ is described in Eq. (A.14) *and $N(\eta^k, \frac{1}{4k})$ is described in the statement of* Lemma A.1.

Proof of Proposition A.2. Fix $n \geq \max\{M(k), N(\eta^k, \frac{1}{4k})\}$. We start by showing that the events $B_{n,k}$ are measurable for all $k, n \in \mathbb{N}$. For each $r = 1, ..., \gamma(n)$ we define the function $e^r_{n,k} : X \to [0,1]$ by

$$e^r_{n,k}(x) = \int_{P^r_n} \mathbb{1}\{\rho(\xi(x), \xi(y)) \leq k^{-1}\}\, d\mu(y). \tag{A.15}$$

By Tonelli's theorem, $e^r_{n,k}$ is measurable for every r. Now let us introduce the measurable functions $h^r_{n,k} = e^r_{n,k} \circ T^{t^r_{n,k}}$. As $Y^{n,k}_\ell(x) \in \mathcal{F}_n$ for every $\ell \in \mathbb{N}$ and $x \in X$, the definition of $Y^{n,k}(x)$ implies that

$$Y^{n,k}(x) \text{ is an } (n, 1-k^{-1})\text{-LT-predictor for } \xi \quad \Longleftrightarrow \quad \sum_{r=1}^{\gamma(n)} h^r_{n,k}(x) \geq 1 - k^{-1}. \tag{A.16}$$

Thus we have established that $B_{n,k}$ is measurable.

Next, consider $E^r_{n,k}$ for some $r \in \{1, ..., \gamma(n)\}$. Since $(\xi_1, ..., \xi_n)$ are constant on $E^r_{n,k}$, we have that $\rho_n(\xi(x), \xi(y)) = 0$ for all $x, y \in E^r_{n,k}$. Thus, for all $x, y \in E^r_{n,k}$,

$$\rho_n(\eta^k(x), \eta^k(y)) \leq \rho_n(\eta^k(x), \xi(x)) + \rho_n(\xi(x), \xi(y)) + \rho_n(\xi(y), \eta^k(y)) \leq \frac{1}{4k}. \tag{A.17}$$

Since $\mathcal{Q}^k$ is measurable w.r.t. $\mathcal{A}(\mathscr{B})$ and $n \geq N(\eta^k, \frac{1}{4k})$, Lemma A.1 coupled with Eq. (A.17) implies that $\rho(\eta^k(x), \eta^k(y)) \leq \frac{1}{2k}$ for all $x, y \in E^r_{n,k}$. Since $E^r_{n,k} \subseteq O_k$, we get that $\rho(\xi(x), \xi(y)) \leq \frac{1}{k}$ for all $x, y \in E^r_{n,k}$, which in conjunction with the fact that $T^{t^r_{n,k}} A_{n,k} \subseteq E^r_{n,k}$ for every $r \in J^1_{n,k}$ implies that, for every $x \in A_{n,k}$,

$$\int_{E^r_{n,k}} \mathbb{1}\{\rho(\xi(T^{t^r_{n,k}} x), \xi(y)) \leq k^{-1}\}\, d\mu(y) = \mu(E^r_{n,k}), \quad \forall r \in J^1_{n,k}. \tag{A.18}$$

Consequently, for every $x \in A_{n,k}$,

$$\begin{aligned} \sum_{r=1}^{\gamma(n)} h^r_{n,k}(x) &\geq \sum_{r \in J^1_{n,k}} h^r_{n,k}(x) \\ &\geq \sum_{r \in J^1_{n,k}} \int_{E^r_{n,k}} \mathbb{1}\{\rho(\xi(T^{t^r_{n,k}} x), \xi(y)) \leq k^{-1}\}\, d\mu(y) \\ &= \sum_{r \in J^1_{n,k}} \mu(E^r_{n,k}) = \mu\left(\{y \in X \,:\, \rho_n(\xi(y), \eta^k(y)) \leq \frac{1}{8k}\}\right) \geq 1 - \frac{1}{k}, \end{aligned} \tag{A.19}$$

where the equality follows from Eq. (A.18) while the last inequality follows from the fact that $n \geq M(k)$ (see Eq. (A.14)). Using the criterion (A.16), we see that $Y^{n,k}(x)$ is an $(n, 1-k^{-1})$-LT-predictor for ξ for all $x \in A_{n,k}$, i.e., $A_{n,k} \subseteq B_{n,k}$. Since $\mu(A_{n,k}) > 0$, the proof is complete. $\square$

To get the triangular array t described in the statement of Theorem 5 we shall perform a "diagonalization" procedure on the array consisting of $t^r_{n,k}$, $r = 1, ..., \gamma(n)$, $n, k \in \mathbb{N}$. Roughly speaking, the purpose is to reduce the dimension of the family $\{Y^{n,k}(x)\}_{n,k,x}$ so as to obtain a new family $\{Y^n(x)\}_{n,x}$, such that the precision of the long-term predictors in this family tends to 1 as $n \to \infty$. We proceed as follows. For every $k \in \mathbb{N}$, set $a_k := \max\{M(k), N(\eta^k, \frac{1}{4k})\}$. Without loss of generality we can assume that the sequence $(a_k)_k$ is strictly increasing. For each $n \geq a_1$, and $r = 1, ..., \gamma(n)$, set $t^r_n := t^r_{n,k}$ whenever $a_k \leq n < a_{k+1}$. For $n < a_1$, let $(t^r_n)_{r=1}^{\gamma(n)}$ be an arbitrary sequence satisfying item (a) of Theorem 5. Consider the triangular array $t := (t^r_n)_{r=1,\, n=1}^{\gamma(n),\infty}$. We note that this array satisfies item (a) in Theorem 5. As in the statement of Theorem 5, define for each $x \in X$ a sequence $Y^n(x) = (Y^n_\ell(x))_{\ell \geq 1}$ of $\mathcal{F}_n$-measurable random variables by

$$Y^n_\ell(x) := \sum_{r=1}^{\gamma(n)} \xi_\ell(T^{t^r_n} x) \mathbb{1}\{P^r_n\} = \sum_{r=1}^{\gamma(n)} \xi_{\ell + t^r_n}(x) \mathbb{1}\{P^r_n\}, \quad \forall \ell \in \mathbb{N}.$$

Also, in the statement of Theorem 5, we defined for every $K \in \mathbb{N}$ and $\varepsilon > 0$ the event

$$C_{K,\varepsilon} = \Big\{ x \in X : Y^K(x) \text{ is an } (K, 1-\varepsilon)\text{-LT-predictor for } \xi \Big\}.$$

Fix $\varepsilon > 0$ and take $k_\varepsilon \in \mathbb{N}$ with $k_\varepsilon^{-1} < \varepsilon$. Set $K_\varepsilon = a_{k_\varepsilon}$. Fix some $K \geq K_\varepsilon$ and let $k \geq k_\varepsilon$ be such that $a_k \leq K < a_{k+1}$. Since $Y^K(x) = Y^{K,k}(x)$ and $k^{-1} < \varepsilon$, we have $B_{K,k} \subseteq C_{K,\varepsilon}$. Since $a_k \leq K$, Proposition A.2 implies that $\mu(B_{K,k}) > 0$. Hence $\mu(C_{K,\varepsilon}) > 0$. The proof of Theorem 5 is complete, as $K \geq K_\varepsilon$ was arbitrary. □

Appendix B. Game-Theoretic Proofs

The following two claims will be necessary for proving several of the propositions stated in Section 6.

Claim B.1. *Let $(Y_n)_{n \geq 1}$ be a sequence of A-valued random variables on X. Then, for every pair $(\sigma, \tau) \in \Sigma \times \mathcal{T}$,*

$$\left| \gamma_N(\sigma, \tau) - \mathbb{E}^\xi_{\sigma,\tau} \left(\frac{1}{N} \sum_{n=1}^N g(Y_n, i_n, j_n) \right) \right| \leq \frac{\|g\|_\infty}{N} \sum_{n=1}^N \mu\left(\xi_n \neq Y_n\right), \tag{B.1}$$

Proof of Claim B.1. For every $n = 1, ..., N$ we have

$$\begin{aligned} \left| \mathbb{E}^\xi_{\sigma,\tau} g(\xi_n, i_n, j_n) - \mathbb{E}^\xi_{\sigma,\tau} g(Y_n, i_n, j_n) \right| &\leq \mathbb{E}^\xi_{\sigma,\tau} \left| g(\xi_n, i_n, j_n) - g(Y_n, i_n, j_n) \right| \\ &\leq \|g\|_\infty \, \mathbb{E}^\xi_{\sigma,\tau} \mathbb{1}\{\xi_n \neq Y_n\} = \|g\|_\infty \mu\left(\xi_n \neq Y_n\right). \end{aligned} \tag{B.2}$$

where we used the fact that the marginal of $\mathbb{P}^\xi_{\sigma,\tau}$ on X is μ, for every σ and τ. We complete the proof by applying the triangle inequality to the left-hand side in Eq. (B.1) and plugging in the upper bound given in Eq. (B.2). □

Claim B.2. *If $Y = (Y_n)_{n\geq 1}$ is a $(K, 1-\varepsilon)$-LT-predictor for ξ then*

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\left(\xi_n \neq Y_n\right) < 2\varepsilon. \tag{B.3}$$

Proof of Claim B.2. By the definition of a $(K, 1-\varepsilon)$-LT-predictor for ξ,

$$\mu\left(\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{\xi_n = Y_n\} \geq 1-\varepsilon\right) \geq 1-\varepsilon.$$

Applying Markov's inequality first, and then Fatou's inequality, we obtain

$$(1-\varepsilon)^2 \leq \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\left(\xi_n = Y_n\right) = 1 - \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\left(\xi_n \neq Y_n\right).$$

The proof follows immediately. □

We are now in position to prove Propositions 1 and 2. As Proposition 1 is a special case of Proposition 2 whenever $(\operatorname{Cav} I_K[Y])(\pi_K) = I_K[Y](\pi_K)$, it suffices to prove the latter.

Proof of Proposition 2. Combining Claims B.1 and B.2 for the $(K, 1-\varepsilon)$-LT-predictor Y, we see that there exists $N_Y^1 > K$ such that

$$\gamma_N(\sigma_\varepsilon^Y, \tau) \geq \mathbb{E}_{\sigma,\tau}^{\xi}\left(\frac{1}{N}\sum_{n=K+1}^{N} g(Y_n, i_n, j_n)\right) - 3\|g\|_\infty \varepsilon, \quad \forall N > N_Y^1, \forall \tau \in \mathcal{T}. \tag{B.4}$$

By conditioning on the outcome of the lottery Z we obtain for every $N \geq K+1$ and every $\tau \in \mathcal{T}$,

$$\begin{aligned}
\mathbb{E}_{\sigma_\varepsilon^Y,\tau}^{\xi}\left(\frac{1}{N}\sum_{n=K+1}^{N} g(Y_n, i_n, j_n)\right) &\geq \sum_{\ell=1}^{\gamma(K)+1} \mu(Z=\ell)\left[\mathbb{E}_{\sigma_\varepsilon^Y,\tau}^{\xi}\left(\frac{1}{N}\sum_{n=K+1}^{N} g(Y_n, i_n, j_n)\,\Big|\, Z=\ell\right)\right] \\
&= \sum_{\ell=1}^{\gamma(K)+1} \mu(Z=\ell)\left[\frac{1}{N}\sum_{n=K+1}^{N} \mathbb{E}_{\sigma_\varepsilon^Y,\tau}^{\xi}\left(g(Y_n, i_n, j_n)\,\Big|\, Z=\ell\right)\right] \\
&\geq \sum_{\ell=1}^{\gamma(K)+1} \mu(Z=\ell)\left[\frac{1}{N}\sum_{n=1}^{N-K} u\left(q_\ell, w_n\right)\right], \qquad \text{(B.5)}
\end{aligned}$$

where the last inequality follows from the definition of σ_ε^Y. Next, by the definition of the function $I_K[Y]$, we may choose $N_Y^2 > K$ so that for every $N > N_Y^2$ it holds that

$$\frac{1}{N}\sum_{n=1}^{N-K} u\left(q_\ell, w_n\right) \geq I_K[Y](q_\ell) - \varepsilon, \quad \forall \ell \in \{1, ..., \gamma(K)+1\}. \tag{B.6}$$

Set $N_Y = \max\{N_Y^1, N_Y^2\}$. Combining Eqs. (B.4), (B.5), and (B.6) we see that

$$\begin{aligned}\gamma_N(\sigma_\varepsilon^Y, \tau) &\geq \sum_{\ell=1}^{\gamma(K)+1} \mu(Z=\ell)\Big[I_K[Y](q_\ell) - \varepsilon\Big] - 3\|g\|_\infty \varepsilon \\ &\geq (\operatorname{Cav} I_K[Y])(\pi_K) - 2\varepsilon - 3\|g\|_\infty\varepsilon, \quad \forall N > N_Y, \forall \tau \in \mathcal{T},\end{aligned} \tag{B.7}$$

where the second inequality follows from the properties of the lottery Z. This completes the proof. $\square$

We break now the order of proofs, and continue with the proof of Proposition 4, as its proof heavily relies on Claims B.1 and B.2 as well.

Proof of Proposition 4. By Claims B.1 and B.2, there exists $N_{Y'}^1$ such that for every pair $(\sigma, \tau) \in \Sigma \times \mathcal{T}$ it holds that

$$\gamma_N(\sigma,\tau) \leq \mathbb{E}_{\sigma,\tau}^{\xi}\left(\frac{1}{N}\sum_{n=1}^{N} g(Y_n', i_n, j_n)\right) + 3\|g\|_\infty\varepsilon, \quad \forall N \geq N_{Y'}^1. \tag{B.8}$$

Let us denote by $G[Y_n']$ the random $|I| \times |J|$ zero-sum matrix game corresponding to $g(Y_n', \cdot, \cdot)$. Thus, if we identify σ_n and τ_n as $1 \times |I|$ and $|J| \times 1$ matrices respectively, we obtain:

$$\mathbb{E}_{\sigma,\tau}^{\xi}\left(g(Y_n', i_n, j_n)\right) = \mathbb{E}_{\sigma,\tau}^{\xi}\left(\sigma_n G[Y_n'] \tau_n\right), \quad \forall n \in \mathbb{N}. \tag{B.9}$$

Consider the subspace of behavioral strategies $\mathcal{T}_{K'}$ of the adversary, defined by

$$\mathcal{T}_{K'} = \{\tau = (\tau_n)_{n\geq 1} \in \mathcal{T} : \tau_{K'+n} : (I \times J)^{n-1} \to \Delta(J), \quad \forall n \geq 1\}. \tag{B.10}$$

A behavioral strategy $\widehat{\tau} \in \mathcal{T}_{K'}$ can be thought of as one in which the adversary losses his memory after the K''th stage, but thereafter remembers all moves from the $(K'+1)$'st stage onward. For every $n \geq 1$ denote by $\mathcal{G}_n$ the σ-field generated by $(\xi_1, ..., \xi_{K'})$ and the random variables $i_{K'+1}, j_{K'+1}, ..., i_{K'+n-1}, j_{K'+n-1}$ (which stand for the pairs of actions played along stages $K'+1, ..., K'+n-1$) on the Borel space $X \times (I \times J)^{\mathbb{N}}$. By the definition of $\mathcal{T}_{K'}$ and the fact that Y' is $\mathcal{F}_{K'}$-measurable, for every pair $(\sigma, \tau) \in \Sigma \times \mathcal{T}_{K'}$ it holds that

$$\mathbb{E}_{\sigma,\tau}^{\xi}\left(\sigma_{K'+n} G[Y_{K'+n}] \tau_{K'+n}\right) = \mathbb{E}_{\sigma,\tau}^{\xi}\left(\mathbb{E}_{\sigma,\tau}^{\xi}\left(\sigma_{K'+n} \,|\, \mathcal{G}_n\right) G[Y_{K'+n}]\tau_{K'+n}\right), \quad \forall n \geq 1. \tag{B.11}$$

Next, for every $\sigma = (\sigma_n)_{n\geq 1} \in \Sigma$ and $\tau \in \mathcal{T}_{K'}$ define $\sigma^\tau = (\sigma_n^\tau)_{n \geq 1}$, so that $\sigma_n^\tau = \{\sigma_n^{\tau,s}\}_{s \in \mathcal{S}_{K'}}$, where $\sigma_n^{\tau,s} : (I \times J)^{n-1} \to \Delta(J)$, by

$$\sigma_n^{\tau,s}(i_1^{'}, j_1^{'}, ..., i_{n-1}^{'}, j_{n-1}^{'}) = \mathbb{E}_{\sigma,\tau}^{\xi}\Big(\sigma_{K'+n} \,|\, s, i_{K'+1} = i_1^{'}, j_{K'+1} = j_1^{'}, ... \\ , i_{K'+n-1} = i_{n-1}^{'}, j_{K'+n-1} = j_{n-1}^{'}\Big). \tag{B.12}$$

It is evident that $\sigma^\tau = (\sigma_n^\tau)_{n\geq 1} \in \widehat{\Sigma}$ for every $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$. The definition of σ^τ, coupled with Eq. (B.11) and the definition of $\mathcal{T}_{K'}$, shows that the key relation

$$\mathbb{E}_{\sigma,\tau}^{\xi}\left(\sigma_{K'+n} G[Y_{K'+n}] \tau_{K'+n}\right) = \mathbb{E}_{\sigma^\tau, \widehat{\tau}}^{K'}\left(\sigma_n^\tau G[Y_{K'+n}] \widehat{\tau}_n\right) \tag{B.13}$$

holds for every pair $(\sigma,\tau)\in\Sigma\times\mathcal{T}_{K'}$ and $n\geq 1$, where we set $\widehat{\tau}_n=\tau_{K'+n}$ for every $n\geq 1$, and $\widehat{\tau}=(\widehat{\tau}_n)_{n\geq 1}$. We can rewrite Eq. (B.13), using Eq. (B.9), as

$$\mathbb{E}^{\xi}_{\sigma,\tau}\left(g(Y_{K'+n},i_{K'+n},j_{K'+n})\right)=\mathbb{E}^{K'}_{\sigma^\tau,\widehat{\tau}}\left(g(Y_{K'+n},i_n,j_n)\right). \tag{B.14}$$

Summing over $n=1,...,N$ for $N\geq N_1$, and using Eq. (B.8), we conclude that for every $\sigma\in\Sigma$ and $\tau\in\mathcal{T}_{K'}$,

$$\begin{aligned}\gamma_{K'+N}(\sigma,\tau)&\leq\widehat{\gamma}_N(\sigma^\tau,\widehat{\tau})+3\|g\|_\infty\varepsilon\\&\quad+\frac{1}{N+K'}\sum_{n=1}^{K'}\mathbb{E}^{\xi}_{\sigma,\tau}\left(g(Y_n,i_n,j_n)\right)\leq\widehat{\gamma}_N(\sigma^\tau,\widehat{\tau})+\|g\|_\infty\left(\frac{K'}{N+K'}+3\varepsilon\right).\end{aligned} \tag{B.15}$$

Thus, for every $\tau\in\mathcal{T}_{K'}$ it holds that

$$\begin{aligned}\max_{\sigma\in\Sigma}\gamma_{K'+N}(\sigma,\tau)&\leq\max_{\sigma\in\Sigma}\widehat{\gamma}_N(\sigma^\tau,\widehat{\tau})+\|g\|_\infty\left(\frac{K'}{N+K'}+3\varepsilon\right)\\&\leq\max_{\widehat{\sigma}\in\widehat{\Sigma}}\widehat{\gamma}_N(\widehat{\sigma},\widehat{\tau})+\|g\|_\infty\left(\frac{K'}{N+K'}+3\varepsilon\right),\end{aligned} \tag{B.16}$$

which in turn implies that

$$\min_{\tau\in\mathcal{T}_{K'}}\max_{\sigma\in\Sigma}\gamma_{K'+N}(\sigma,\tau)\leq\min_{\tau\in\mathcal{T}_{K'}}\max_{\widehat{\sigma}\in\widehat{\Sigma}}\widehat{\gamma}_N(\widehat{\sigma},\widehat{\tau})+\|g\|_\infty\left(\frac{K'}{N+K'}+3\varepsilon\right). \tag{B.17}$$

On the one hand, the definition of $\mathcal{T}_{K'}$ implies that

$$\min_{\tau\in\mathcal{T}_{K'}}\max_{\widehat{\sigma}\in\widehat{\Sigma}}\widehat{\gamma}_N(\widehat{\sigma},\widehat{\tau})=\min_{\tau\in\mathcal{T}}\max_{\widehat{\sigma}\in\widehat{\Sigma}}\widehat{\gamma}_N(\widehat{\sigma},\tau)=\widehat{v}_N(Y'),\quad\forall N\in\mathbb{N}. \tag{B.18}$$

On the other hand, the stationarity of ξ and the definition of $\mathcal{T}_{K'}$ imply that

$$\min_{\tau\in\mathcal{T}_{K'}}\max_{\sigma\in\Sigma}\gamma_{K'+N}(\sigma,\tau)=\frac{K'v_{K'}(\xi)+Nv_N(\xi)}{K'+N},\quad\forall N\in\mathbb{N}. \tag{B.19}$$

Combining Eqs. (B.17), (B.18), and (B.19) and performing some simple algebraic manipulations yields the inequality

$$v_N(\xi)\leq\widehat{v}_N(Y')+\|g\|_\infty\left(\frac{K'}{N+K'}+3\varepsilon\right)+\left(\frac{K'}{K'+N}\right)v_N(\xi), \tag{B.20}$$

which holds for every $N\geq N_1$. Since $v_N(\xi)\leq\|g\|_\infty$, we complete the proof by setting $N_{Y'}=\max\{N^1_{Y'},\lceil\frac{K'}{\varepsilon}\rceil-K'\}$. □

Proof of Lemma 3. For every $a\in\mathbb{N}$ define the set

$$F^a_{K',\varepsilon}=\left\{x\in X:\frac{1}{N}\sum_{n=1}^{N}\mu(\xi_n\neq Y^{K'}(x))<3\varepsilon,\quad\forall N\geq a\right\}. \tag{B.21}$$

We claim that $F^a_{K',\varepsilon}\in\mathcal{B}(X)$ for every $a\in\mathbb{N}$, that is that the $F^a_{K',\varepsilon}$'s are measurable. Indeed, recall that in Appendix A we defined for each $n\in\mathbb{N}$ the mapping $\rho_n:A^{\mathbb{N}}\times A^{\mathbb{N}}\to[0,1]$ by $\rho_n(\alpha,\beta):=\frac{1}{n}\sum_{\ell=1}^{n}\mathbb{1}\{\alpha_\ell\neq\beta_\ell\}$. This mapping is measurable w.r.t. the product σ-field $\mathcal{B}(A^{\mathbb{N}})\otimes\mathcal{B}(A^{\mathbb{N}})$. Since, by

Tonelli's Theorem, the mapping $x \mapsto \int_{P^r_{K'}} \rho_N(\xi(y), \xi(x))\, d\mu(y)$ is measurable, and since $T : X \to X$ is measurable, we have

$$\text{(B.22)}\quad \left\{ x \in X : \frac{1}{N} \sum_{n=1}^{N} \mu(\xi_n \neq Y^{K'}(x)) < 3\varepsilon \right\}$$
$$= \left\{ x \in X : \sum_{r=1}^{\gamma(K')} \int_{P^r_{K'}} \rho_N(\xi(y), \xi(T^{t^r_{K'}} x))\, d\mu(y) < 3\varepsilon \right\} \in \mathcal{B}(X).$$

The measurability of $F^a_{K',\varepsilon}$ now follows because

$$\text{(B.23)}\qquad F^a_{K',\varepsilon} = \bigcap_{N \geq a} \left\{ x \in X : \frac{1}{N} \sum_{n=1}^{N} \mu(\xi_n \neq Y^{K'}(x)) < 3\varepsilon \right\} \in \mathcal{B}(X).$$

By following the steps in the proof of Proposition 4, we see that if $x \in F^a_{K',\varepsilon}$ for some $a > \lceil \frac{K'}{\varepsilon} \rceil - K'$, then

$$\text{(B.24)}\qquad v_n(\xi) \leq \widehat{v}_n(Y^{K'}(x)) + 5\|g\|_\infty \varepsilon, \quad \forall n \geq a.$$

By Claim B.2 and Theorem 5, $C_{K',\varepsilon} \subseteq \bigcup_{a \in \mathbb{N}} F^a_{K',\varepsilon}$. Since $\{F^a_{K',\varepsilon}\}_{a \in \mathbb{N}}$ is an increasing sequence of events, there must exist $b = b(K', \varepsilon) > \lceil \frac{K'}{\varepsilon} \rceil - K'$ such that $\mu(C_{K',\varepsilon} \cap F^a_{K',\varepsilon}) > 0$ for every $a > b$. Moreover, for $x \in C_{K',\varepsilon}$ we may apply the bound given in Eq. (3) to the predictor $Y^{K'}(x)$ and deduce the existence of a number $M_{K',\varepsilon} > b$ such that

$$\text{(B.25)}\qquad \widehat{v}_{ML}(Y^{K'}(x)) \leq \frac{1}{M} \sum_{m=0}^{M-1} (\mathrm{Cav}\, I^{m,L}_{K'}[Y^{K'}(x)])(\pi_{K'}) + \|g\|_\infty \varepsilon,$$

for every $M \geq M_{K',\varepsilon}$ and $L \in \mathbb{N}$. Set $G_{K',\varepsilon} = C_{K',\varepsilon} \cap F^{M_{K',\varepsilon}}_{K',\varepsilon}$. Since $M_{K',\varepsilon} > b$, we have $\mu(G_{K',\varepsilon}) > 0$. By Eqs. (B.24) and (B.25), for every $x \in G_{K',\varepsilon}$ we have that

$$\text{(B.26)}\qquad v_{ML}(\xi) \leq \frac{1}{M} \sum_{m=0}^{M-1} (\mathrm{Cav}\, I^{m,L}_{K'}[Y^{K'}(x)])(\pi_{K'}) + 6\|g\|_\infty \varepsilon,$$

for every $M \geq M_{K',\varepsilon}$ and $L \in \mathbb{N}$. This completes the proof. □

We move on to the proof of Proposition 3.

Proof of Proposition 3. We start by recalling the following basic fact regarding the value of zero-sum matrix games:

Fact B.1. *For every two zero-sum matrix games A and B of equal dimensions we have*

$$|\mathrm{val}(A) - \mathrm{val}(B)| \leq \|A - B\|_\infty.$$

Claim B.3. *For every $p, q \in \Delta(\mathcal{S}_{K'})$ and every $w \in \mathcal{W}_{K'}$,*

$$|u\,(p, w) - u\,(q, w)\,| \leq \|g\|_\infty \,\|p - q\|_1.$$

Proof of Claim B.3. The definition of u together with Fact B.1 imply that

$$|u(p,w) - u(q,w)| \le \|G(p,w) - G(q,w)\|_\infty. \tag{B.27}$$

On the other hand, by definition (see (4.3)),

$$\|G(p,w) - G(q,w)\|_\infty = \max_{i\in I,\, j\in J} \left| \sum_{r=1}^{\gamma(K')} (p^r - q^r) g(w^r, i, j) \right| \le \|g\|_\infty \|p-q\|_1, \tag{B.28}$$

which in combination with Eq. (B.27) proves the claim. □

We return to the proof of Proposition 3. For each $m = 0, ..., M-1$ denote

$$I_m = \left| \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}^{K'}_{\sigma,\tau_*} u\left(p_{mL+\ell}, w'_{mL+\ell}\right) - \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}^{K'}_{\sigma,\tau_*} u\left(p_{mL+1}, w'_{mL+\ell}\right) \right|. \tag{B.29}$$

By the triangle inequality and Claim B.3,

$$I_m \le \frac{1}{L} \sum_{\ell=1}^{L} \|g\|_\infty \mathbb{E}^{K'}_{\sigma,\tau_*} \|p_{mL+\ell} - p_{mL+1}\|_1, \quad \forall m = 0, ..., M-1. \tag{B.30}$$

Changing the order of summation and then using the Hölder inequality for vector-valued functions (with $p = q = 2$) we see that, for every $m = 0, ..., M-1$,

$$\begin{aligned} \frac{1}{L} \sum_{\ell=1}^{L} \|g\|_\infty \mathbb{E}^{K'}_{\sigma,\tau_*} \|p_{mL+\ell} - p_{mL+1}\|_1 &= \frac{\|g\|_\infty}{L} \sum_{r=1}^{\gamma(K')} \mathbb{E}^{K'}_{\sigma,\tau_*} \left(\sum_{\ell=1}^{L} |p^r_{mL+\ell} - p^r_{mL+1}| \right) \\ &\le \frac{\|g\|_\infty}{L} \sum_{r=1}^{\gamma(K')} \sqrt{\mathbb{E}^{K'}_{\sigma,\tau_*} \left(\sum_{\ell=1}^{L} 1^2 \right)} \sqrt{\mathbb{E}^{K'}_{\sigma,\tau_*} \left(\sum_{\ell=1}^{L} (p^r_{mL+\ell} - p^r_{mL+1})^2 \right)} \\ &= \|g\|_\infty \sum_{r=1}^{\gamma(K')} \sqrt{\frac{1}{L} \mathbb{E}^{K'}_{\sigma,\tau_*} \left(\sum_{\ell=1}^{L} (p^r_{mL+\ell} - p^r_{mL+1})^2 \right)}. \end{aligned} \tag{B.31}$$

Utilizing first the martingale property of $(p^r_n)_{n\ge1}$, and second, the sub-martingale property of $((p^r_n)^2)_{n\ge1}$, we obtain

$$\begin{aligned} \frac{1}{L} \mathbb{E}^{K'}_{\sigma,\tau_*} \left(\sum_{\ell=1}^{L} (p^r_{mL+\ell} - p^r_{mL+1})^2 \right) &= \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+\ell})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+1})^2] \\ &\le \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+L})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+1})^2]. \end{aligned} \tag{B.32}$$

Combining the above upper bound with Eqs. (B.30) and (B.31), we see that, for every $m = 0, ..., M-1$,

$$I_m \le \|g\|_\infty \sum_{r=1}^{\gamma(K')} \sqrt{\mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{(m+1)L})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+1})^2]}. \tag{B.33}$$

Summing over $m = 0, ..., M-1$, changing the order of summation, and utilizing the concavity of the square root function we obtain

$$\begin{aligned}
\frac{I_0 + \cdots + I_{M-1}}{M} &\leq \frac{1}{M}\sum_{m=0}^{M-1}\|g\|_\infty \sum_{r=1}^{\gamma(K')}\sqrt{\mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{(m+1)L})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+1})^2]} \\
&= \|g\|_\infty \sum_{r=1}^{\gamma(K')}\frac{1}{M}\sum_{m=0}^{M-1}\sqrt{\mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{(m+1)L})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+1})^2]} \\
&\leq \|g\|_\infty \sum_{r=1}^{\gamma(K')}\sqrt{\frac{1}{M}\sum_{m=0}^{M-1}\left(\mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{(m+1)L})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{mL+1})^2]\right)} \\
&\leq \|g\|_\infty \sum_{r=1}^{\gamma(K')}\sqrt{\frac{1}{M}\left(\mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{ML})^2] - \mathbb{E}^{K'}_{\sigma,\tau_*}[(p^r_{1})^2]\right)}, \qquad \text{(B.34)}
\end{aligned}$$

where the last inequality requires uses the sub-martingale property of $((p^r_n)^2)_{n\geq 1}$. As the latter sequence of random variables is bounded from above by 1 for every $r = 1, ..., \gamma(K')$, the triangle inequality and Eq. (B.34) imply that

$$\text{(B.35)} \quad \left|\frac{1}{ML}\sum_{n=1}^{ML}\mathbb{E}^{K'}_{\sigma,\tau_*}u\left(p_n, w'_n\right) - \frac{1}{M}\sum_{m=0}^{M-1}\left(\frac{1}{L}\sum_{\ell=1}^{L}\mathbb{E}^{K'}_{\sigma,\tau_*}u\left(p_{mL+1}, w'_{mL+\ell}\right)\right)\right|$$

$$\leq \frac{I_0 + ... + I_{M-1}}{M} \leq \|g\|_\infty \sum_{r=1}^{\gamma(K')}\sqrt{\frac{1}{M}} \leq \frac{\|g\|_\infty \gamma(K')}{\sqrt{M}},$$

as required. □

Proof of Claim 1. Fix $x \in E$. Carathéodory's theorem (e.g., Rockafellar (1970), Corollary 17.1.5 p. 157) ensures the existence of convex weights $(\alpha_i)_{i=1}^d$ and points $(x_i)_{i=1}^d$ such that $(\mathrm{Cav}\, f)(x) = \sum_{i=1}^d \alpha_i f(x_i)$. Thus, using the concavity of $\mathrm{Cav}\, h$ we obtain

$$\text{(B.36)} \qquad (\mathrm{Cav}\, h)(x) \geq \sum_{i=1}^{d}\alpha_i h(x_i) \geq \sum_{i=1}^{d}\alpha_i f(x_i) - \rho = (\mathrm{Cav}\, f)(x) - \rho.$$

As the opposite inequality follows from symmetry, the proof is complete. □

Proof of Lemma 4. Fix $M \in \mathbb{N}$ and $\beta > 0$. Set

$$\alpha = \frac{\mu(G_{K',\varepsilon})}{2M|\mathcal{W}_{K'}|}.$$

Note that $\alpha > 0$ since $K' \geq K'_\varepsilon$. Birkhoff's Pointwise Ergodic Theorem implies that for every $w \in \mathcal{W}_{K'}$ we can choose $L_w = L_w(M, \beta)$ such that for every $L \geq L_w$ it holds that

$$\text{(B.37)} \qquad \mu\left(\left\{x \in X \,:\, \left|\frac{1}{L}\sum_{\ell=1}^{L}\mathbb{I}^{K',w}_\ell(x) - \mu(R^w_{K'})\right| < \beta\right\}\right) \geq 1 - \alpha.$$

Since the sequence of random variables $(\mathbb{I}_n^{K',w})_{n\geq 1}$ is stationary for every $w \in \mathcal{W}_{K'}$, it follows that

$$\mu\left(\left\{x \in X \,:\, \left|\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}_{mL+\ell}^{K',w}(x) - \mu(R_{K'}^w)\right| < \beta\right\}\right) \geq 1-\alpha, \tag{B.38}$$

for every $m \geq 1$, $w \in \mathcal{W}_{K'}$ and $L \geq L_w$. Consequently,

$$\mu(D_{L_*,m}(\beta)) \geq 1 - |\mathcal{W}_{K'}|\alpha, \quad \forall m \geq 1, \tag{B.39}$$

where $L_* = \max_{w\in\mathcal{W}_{K'}} L_w$. The proof now follows because

$$\begin{aligned} \mu\left(\bigcap_{m=0}^{M-1} D_{L_*,m}\right) &\geq 1 - \sum_{m=0}^{M-1} |\mathcal{W}_{K'}|\alpha \\ &= 1 - M|\mathcal{W}_{K'}|\alpha \\ &= 1 - \frac{\mu(G_{K',\varepsilon})}{2}, \end{aligned} \tag{B.40}$$

as the latter implies that Eq. (6.28) holds with $L = L_*$. □

Proof of Proposition 5. Since w is concave and the scalar product and expectation operators are linear, it follows that Ψ_n is concave as well. Thus, as $\Psi_n = \Phi_n$, we have the string of equalities

$$\operatorname{Cav}\Phi_n = \operatorname{Cav}\Psi_n = \Psi_n = \Phi_n. \tag{B.41}$$

Hence, by Theorem 2, $v(\xi) = \lim_{n\to\infty}\Psi_n(\pi_n)$. Thus, to show that (6.39) holds, it suffices to show that $v(\xi) = w(\mathbb{E}\,\xi_1)$. Assume that the observer ignores his private information and plays at every stage $n \in \mathbb{N}$ his optimal mixed action in the zero-sum one-shot game

$$(1-q_n)g(0,\cdot,\cdot) + q_n g(1,\cdot,\cdot),$$

where $q_n = \mu(\xi_n = 1)$. Since ξ is a binary stationary process, $q_n = \mu(\xi_1 = 1) = \mathbb{E}\,\xi_1$ for every $n \geq 1$. In this manner the observer guarantees himself a payoff of $w(\mathbb{E}(\xi_1))$ in $\Gamma_\infty(\xi)$. Thus we have shown that

$$w(\mathbb{E}\,\xi_1) \leq v(\xi). \tag{B.42}$$

To show that the opposite inequality holds as well, we return to the analysis carried out in Subsection 6.4. In particular, let us return to Eq. (6.10):

$$\begin{aligned} \widehat{\gamma}_N(\sigma,\tau_*) &\leq \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}_{\sigma,\tau_*}^{K'} u\left(p_n, w_n'\right) + \frac{\|g\|_\infty\sqrt{2\,H(\pi_{K'})}}{\sqrt{N}} \\ &= \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}_{\sigma,\tau_*}^{K'} w\left(\langle p_n, w_n'\rangle\right) + \frac{\|g\|_\infty\sqrt{2\,H(\pi_{K'})}}{\sqrt{N}} \\ &\leq \frac{1}{N}\sum_{n=1}^{N} w\left(\langle \pi_{K'}, w_n'\rangle\right) + \frac{\|g\|_\infty\sqrt{2\,H(\pi_{K'})}}{\sqrt{N}}, \end{aligned} \tag{B.43}$$

where the equality follows from the definitions of u, w and the fact that $A = \{0,1\}$, while the second inequality is obtained by applying Jensen's inequality to the concave mapping $q \mapsto w(\langle q, w'_n\rangle)$. Next, it holds that

$$\langle \pi_{K'}, w'_n\rangle = \mathbb{E}\, Y'_{K'+n}, \qquad \forall n \in \mathbb{N}. \tag{B.44}$$

Combining the Eqs. (B.43) and (B.43), and using the von Neumann Minmax Theorem we obtain that

$$\widehat{v}_N(Y') \leq \frac{1}{N}\sum_{n=1}^{N} w\left(\mathbb{E}\, Y'_{K'+n}\right) + \frac{\|g\|_\infty \sqrt{2\, H(\pi_{K'})}}{\sqrt{N}}. \tag{B.45}$$

By Fact B.1 and the fact that $\mathbb{E}\,\xi_1 = \mathbb{E}\,\xi_n$ for all $n \geq 1$, we see that, for every $n \geq 1$,

$$\begin{aligned} |w\left(\mathbb{E}\, Y'_{K'+n}\right) - w\left(\mathbb{E}\,\xi_1\right)| &\leq \|g\|_\infty \|(1-\mathbb{E}\, Y'_{K'+n}, \mathbb{E}\, Y'_{K'+n}) - (1-\mathbb{E}\,\xi_{K'+n}, \mathbb{E}\,\xi_{K'+n})\|_1 \\ &\leq 2\,\|g\|_\infty\, \mathbb{E}|Y'_{K'+n} - \xi_{K'+n}| = 2\,\|g\|_\infty\, \mu(Y'_{K'+n} \neq \xi_{K'+n}), \end{aligned} \tag{B.46}$$

where the equality follows from the relation $|Y'_{K'+n} - \xi_{K'+n}| = \mathbb{1}\{Y'_{K'+n} \neq \xi_{K'+n}\}$. Now Eqs. (B.46) and (B.45) imply that

$$\widehat{v}_N(Y') \leq w\left(\mathbb{E}\,\xi_1\right) + 2\,\|g\|_\infty \frac{1}{N}\sum_{n=1}^{N} \mu(Y'_{K'+n} \neq \xi_{K'+n}) + \frac{\|g\|_\infty \sqrt{2\, H(\pi_{K'})}}{\sqrt{N}}. \tag{B.47}$$

Applying Proposition 4 together with Claim B.2, we obtain

$$\begin{aligned} \lim_{N\to\infty} v_N(\xi) &\leq \limsup_{N\to\infty} \widehat{v}_N(Y') + 5\|g\|_\infty \varepsilon \\ &\leq w\left(\mathbb{E}\,\xi_1\right) + \limsup_{N\to\infty} 2\,\|g\|_\infty \frac{1}{N}\sum_{n=1}^{N} \mu(Y'_{K'+n} \neq \xi_{K'+n}) + 5\,\|g\|_\infty \varepsilon \\ &\leq w\left(\mathbb{E}\,\xi_1\right) + 4\,\|g\|_\infty \varepsilon + 5\,\|g\|_\infty \varepsilon. \end{aligned} \tag{B.48}$$

As ε was arbitrary all along, we have that $\lim_{n\to\infty} v_n(\xi) \leq w\left(\mathbb{E}\,\xi_1\right)$, and thus by Theorem 1 we obtain,

$$v(\xi) = \lim_{n\to\infty} v_n(\xi) \leq w\left(\mathbb{E}\,\xi_1\right), \tag{B.49}$$

which together with Eq. (B.42) shows that $v(\xi) = w\left(\mathbb{E}\,\xi_1\right)$, and thus completes the proof. □