Nonparametric Methods For Repeated Games

Christos A. Ioannou [*†]
Université Paris 1 Panthéon - Sorbonne

Laurent Mathevet [‡]
University of Arizona

Julian Romero [§]
University of Limassol

Huanren Zhang [¶]
University of Southern Denmark

This draft: August 3, 2026

Abstract

We develop two nonparametric, pattern-mining methods to investigate behavioral regularities and evolutions *within* infinitely-repeated games. Our first approach, the *action-convergence criterion*, draws on string-searching algorithms to identify recurring action profiles by minimizing mismatches over variable time horizons. Our second approach utilizes a modified *k-means clustering algorithm* to endogenously reveal distinct behavioral 'storylines' across the short, medium, and long run. We run experiments on a rich set of 2×2 repeated games with high discount factors ($\delta = 0.99, 0.995$), and apply our methods to the dataset. Despite the broad set of outcomes permitted by theory, long-run play is highly concentrated in our data, with behavior becoming markedly more stable over time. Efficiency and egalitarianism emerge as powerful learned attractors, while subjects build more complex dynamic arrangements as play progresses. Finally, by clustering entire trajectories of play, we characterize how behavior unfolds and organize the data into a small number of recurring, interpretable storylines.

[*]The paper has benefited greatly from the comments of Masaki Aoyaki, Syngjoo Choi, David J. Cooper, Luis Corchón, Guillaume Fréchette, Amanda Friedenberg, Nobuyuki Hanaki, Natalie Khalil, Yaroslav Rosokha and Doron Sonsino. We would also like to thank seminar audiences at the UC Irvine, European University Institute, University of Cyprus, Complutense University of Madrid, Institute of Social and Economic Research - Osaka University, Seoul National University, Waseda University, University of Limassol, Purdue University, Université Paris 1 Panthéon - Sorbonne, University of Arizona, and conference audiences at Barcelona GSE and SAET for helpful comments and discussions. This work has been funded by a French government subsidy managed by the Agence Nationale de la Recherche under the framework of the Investissements d'Avenir Programme (reference ANR-17-EURE-001). The usual disclaimer applies.

[†]Centre d'Économie de la Sorbonne. Email: christos.ioannou@univ-paris1.fr

[‡]Department of Economics, University of Arizona. Email: lmathevet@arizona.edu

[§]Department of Economics and Finance, University of Limassol. Email: julian.romero@uol.ac.cy

[¶]Department of Economics, University of Southern Denmark. Email: huanren@sam.sdu.dk

JEL: C72, C91, C81, C63, D83

Keywords: Pattern Mining, Nonparametric Methods, Experiments, Repeated Games

1 Introduction

One of the central findings from experimental work on repeated games is that the 'shadow of the future' is real: players often act today in ways that violate their immediate incentives to avoid future punishment. The canonical illustration is cooperation in the repeated Prisoner's Dilemma, which has remained a focal point of the literature since Dal Bó (2005).[1]

Although this literature has followed diverse trajectories, a common practice has been to rematch players across multiple repeated games,[2] assuming that by the time they play the later games, they are effectively in equilibrium from the outset. In this context, where within-game experience is bounded by the number of periods, and across-game experience lies outside the theory, researchers have typically deemphasized learning[3] and adopted discount factors only as high as needed to sustain the equilibrium outcomes of interest.

In this paper, we shift the focus to the evolution of behavior *within* long-horizon, repeated games. Our motivation is simple. Standard repeated-game theory is concerned with behavior within a single match and, for patient players, permits a vast

[1] Dal Bó (2005) finds that higher discount factors lead to higher cooperation rates. Dal Bó and Fréchette (2011) examine how cooperation evolves with experience and show that while being part of an equilibrium strategy may be necessary for cooperation to emerge, it is not sufficient. Cooperation is also found to be robust to changes in the monitoring structure in Aoyagi, Bhaskar, and Fréchette (2019) (see also Aoyagi and Fréchette (2009), Fudenberg, Rand, and Dreber (2012) and Kayaba, Matsushima, and Toyama (2016)). In a different vein, Proto, Rustichini, and Sofianos (2019, 2022) study the role of heterogeneity in intelligence, personality traits, and cognitive skills in shaping cooperation. Lambrecht et al. (2024) further explore how disclosing information about players' intelligence affects cooperative behavior. Finally, Gill and Rosokha (2024) elicit beliefs about others' strategies to study their role in sustaining cooperation.

[2] According to Dal Bó and Fréchette (2018), which survey decades of repeated-game experiments, the highest discount factor ever used was $\delta = 0.95$ with the norm being $\delta \leq 0.875$ (see their Table 3). But note that $\delta = 0.95$ induces 20 periods of game play in expectation, whereas $\delta \leq 0.875$ induces 8 or fewer periods.

[3] For example, Blonski, Ockenfels, and Spagnolo (2011) write: "... we are interested in cooperation as an equilibrium selection phenomenon per se. Therefore, we want to deemphasize learning, as (i) learning within a repeated game and (ii) learning by playing many repeated games" (p. 179).

multiplicity of (equilibrium) outcomes. This richness naturally raises the question of whether there are systematic regularities when interactions are sufficiently long.

We approach this question by developing nonparametric methods that study observed play directly, without imposing a model of behavior or attempting to recover underlying strategies. The starting point is that repeated-game data are naturally sequential, and thus amenable to temporal data mining. Guided by this observation, we draw on the frequent event-mining literature (Agrawal and Srikant (1995) and Mannila, Toivonen, and Verkamo (1997)), which centers on the concept of an *episode*: a finite, ordered sequence (of action profiles) that may begin at any of its elements. Episodes formalize the basic recurring patterns to search for.[4] Similar approaches have uncovered meaningful regularities in other scientific domains. In bioinformatics and molecular biology, for example, they are used to identify recurring motifs in DNA and protein sequences (see, e.g., Hertz and Stormo (1999) and Jumper et al. (2021)); in neuroscience, they help recover temporal patterns in the activity of groups of neurons, providing insight into how neural systems are organized (see Patnaik, Sastry, and Unnikrishnan (2008)); and in developmental biology, they have been applied to developmental timing data, where the extracted patterns can be used to compare species and even reconstruct evolutionary relationships based on gene expression and physical traits (see Bathoorn et al. (2010)).

Our perspective is model-light and discovery-oriented: it lets the data reveal regularities that theory does not instruct the analyst to look for ex ante. In this spirit, we introduce techniques designed to investigate sequences of action profiles – whether they converge, to which episodes, and how these episodes form over time.

Our first technique, which we call the *action-convergence criterion*, combines episode mining (Agrawal and Srikant (1995) and Mannila, Toivonen, and Verkamo (1997)) with a stricter local comparison rule inspired by exact string-search methods (see, e.g., Boyer and Moore (1977) and Knuth, Morris, and Pratt (1977)). The criterion evaluates how well an episode fits a sequence of observed play by performing 'character comparisons' over an interval of periods and counting mismatches (or errors). Given an experimental sequence, an episode is $(1-\phi)$-frequent over a time interval T if it matches the elements of the sequence over T at least $(1-\phi)$ of the time.[5] Frequent event discovery involves settling important methodological questions,

[4]Henceforth, we use the term 'episode' instead of 'pattern' to be consistent with the relevant literature.

[5]The term ϕ is defined as mismatches over time periods; therefore, $(1-\phi)$ is the (required)

such as where to restart the episode after an error, and what constitutes a match (or an error) in the first place. These points are addressed in Subsection 2.1.

Our second technique is a modified version of the k-means clustering algorithm (see, e.g., MacQueen (1967), Hartigan (1975), Hartigan and Wong (1979)). The k-means method partitions a dataset of n points in an m-dimensional space into k clusters so that points within a cluster are more similar to each other than to those in other clusters. Standard clustering methods, however, require numerical representations of the data in order to define distances between observations. We therefore begin by proposing a digitization procedure that maps sequences of action profiles into high-dimensional vectors. By applying discounted, uniform, and reverse-discounted weights, we capture behavioral shifts over the short, medium, and long run. This unsupervised procedure endogenously reveals distinct 'storylines' that characterize the progression of play.

We apply our methods to a rich class of 2×2 repeated games capturing a wide range of strategic interactions, including – but not limited to – the classic Prisoner's Dilemma. Each pair of players is matched to play a single repeated game with exceptionally high discount factors ($\delta = 0.99, 0.995$), corresponding to expected horizons of up to 200 periods. These long horizons allow complex play to unfold yet stabilize.

In the first set of results, based on the action-convergence criterion, we compare early and late behavior by focusing on the first and last 20 periods of each sequence of play. We define a *short-run outcome* as a 0.9-frequent episode over the initial 20 periods, and a *long-run outcome* as a 0.9-frequent episode over the final 20 periods. First, we establish a counterpoint to the theoretical abundance of equilibrium outcomes: long-run play is highly concentrated and the result of increasing stability over time, with convergence rising by 31% on average over the course of the interaction. Second, behavioral complexity – measured by the prevalence of multi-element episodes[6] – emerges gradually, with episodes of length two or more being three times more frequent in the long run than in the short run. Third, we identify efficiency and egalitarianism (defined as equal payoffs between the two players in a pair over the evaluation interval) as strong behavioral attractors, whose influence may increase as interactions mature.

weighted frequency of fit.

[6]The length of an episode is the number of action profiles (elements) it contains (see Subsection 2.1).

Our second set of results, based on the k-means clustering approach, provides a complementary perspective. While the action-convergence analysis focuses on local matching at the beginning or end of the game, the clustering method analyzes the entire trajectory of play and groups sequences according to the similarity of their full-game dynamics. This approach identifies broader behavioral trajectories that characterize how play unfolds over time. Importantly, despite the richness of possible paths, the data organize into a small number of recurring and interpretable storylines. First, in symmetric stage games where strategic considerations may conflict with efficiency, such as the Prisoner's Dilemma and Stag Hunt, behavioral trajectories are largely structured around early and persistent convergence to efficient coordination, alongside a smaller set of noisier or delayed paths. Second, in games where equilibrium play is efficient but delivers asymmetric payoffs, such as the Samaritan's Dilemma variants, storylines are more dynamic, with players transitioning across multiple episodes before (sometimes) settling into a long-run pattern. Third, across many games, trajectories exhibit a common pattern of midstream adjustment followed by stabilization.

The paper is structured as follows. Section 2 develops the two mining approaches, while Section 3 presents the experimental design and the games investigated. Section 4 reports the findings of each approach. Finally, Section 5 presents extensions to the action-convergence analysis, and then concludes.

2 Episode Mining in Repeated Games

Episode discovery aims to detect short[7] and ordered sequences of events that occur often enough in the data. The main technical questions are when to declare a match and where to assign an error when the match fails. For example, when testing the frequency of a word in a sentence, one could measure the frequency of words that are not an exact match to the target or the frequency of spelling discrepancies.[8] Which approach is preferable depends on the context.

[7] Short relative to the full sequence under consideration.

[8] For example, the word 'the' does not appear at all in 'this and that,' yet we could also argue that it is more than a 0% match, because 'th' makes up half of two of the three words. The latter reflects the intuition that 'the' seems closer to that sequence than to, say, 'English is a language.'

In our setting, the data consist of sequences of play generated by repeated games. We therefore begin by defining the repeated-game environment. A stage game is a finite two-player normal-form game defined by a tuple $(\mathcal{A}_1, \mathcal{A}_2, u_1, u_2)$, where $\mathcal{A}_i$ is player i's action set, $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ is the set of action profiles, and $u_i : \mathcal{A} \to \mathbb{R}$ is i's utility function. An infinitely repeated game (with perfect monitoring) is an infinite sequence of repetitions of a commonly known stage game at discrete periods $t = 1, 2, \ldots$, in which players choose actions simultaneously and the resulting action profile becomes common knowledge at the end of each period.

2.1 Preliminaries

Given a set of action profiles $\mathcal{A}$, an *episode* is a finite sequence of action profiles that is not obtained by repeating a shorter sequence. Formally, a (deterministic) episode of length $n < \infty$ is a sequence $p \in \mathcal{A}^n$, written as $p = (p^1 p^2 \ldots p^n)$, such that $p^t \in \mathcal{A}$ for all $t = 1, \ldots, n$ and there does not exist an integer $1 \leq \ell < n$ such that $p^t = p^{t+\ell}$ for all $t = 1, \ldots, n - \ell$ and $p^t = p^{t+\ell-n}$ for all $t = n - \ell + 1, \ldots, n$. Let $|p|$ denote the length of episode p. This definition avoids the redundancies caused by episodes contained in other episodes. For example, (aa) is excluded because it repeats (a) and $(abab)$ because it repeats (ab).

Given that an episode should repeat itself to be significant, as in $(p^1 p^2 p^1 p^2 \ldots)$, p^1 will appear both before and after p^2. Which element precedes the other within an episode is thus somewhat immaterial. What matters is that all elements in the episode repeat themselves in a specific order. For this reason, we work with equivalence classes of episodes; that is,

$$[p] := \big\{ q \in \mathcal{A}^{|p|} : \text{there exists some } 0 < \ell < n \text{ such that } p^t = q^{k+\ell} \text{ for all } 1 \leq t \leq |p| - \ell \text{ and } p^t = q^{k+\ell-n} \text{ for all } |p| - \ell + 1 \leq t \leq |p| \big\}. \quad (1)$$

An equivalence class contains all permutations of the same elements that preserve the order between them. For example, letting $a, b, c \in \mathcal{A}$, the equivalence class $[(abc)]$ is equal to $\{(abc), (bca), (cab)\}$. Each episode is contained in one and only one equivalence class. From here, we arbitrarily choose one member from each class and refer to it as the equivalence class. Let $\mathcal{P}(n)$ be the set of all (equivalence classes of)

episodes of length n or less.[9]

Our approach to declare a match (and to assign an error when the match fails) is based on the WINEPI algorithm of Mannila, Toivonen, and Verkamo (1997), which insists on exact matches rather than subsequences.[10] Given a finite sequence of action profiles $s \in \bigcup_{n=1}^{\infty} \mathcal{A}^n$, written $s = (s^1 s^2 \dots s^n)$, an episode $p \in \mathcal{P}(n)$ *matches* s in period t if there exists $q \in [p]$ such that $s^{t+k-1} = q^k$ for all $1 \leq k \leq \min\{|p|, |s|-t+1\}$; that is, a match is declared in period t if the episode perfectly matches the sequence in all periods from t to $t + |p| - 1$ (or until the end of the sequence if less than $|p|$ periods remain). If the episode does not match, then this is a mismatch or an error. Let

$$m(p, s, t) = \begin{cases} 1 & \text{if } p \text{ matches } s \text{ in period } t \\ 0 & \text{otherwise} \end{cases}$$

be the match function between p and s in t. In Figure 1, we illustrate our approach with a few examples. The match function is used in both the action-convergence criterion and the modified k-means clustering approach.

We conclude with two remarks on our methodological approach. When comparing an episode to an experimental sequence, we could record errors in the exact location in which they occur. Instead, we have chosen to declare an error in period t, if the episode does not match perfectly actual play from period t onwards. For example, when testing episode $(p^1 p^2 p^3)$, an error occurs in period t if $p^1 = s^t$, $p^2 = s^{t+1}$ but $p^3 \neq s^{t+2}$. This is a very stringent criterion for a match. An alternative, weaker criterion, would be to pinpoint errors where they occur. In the previous example, an error could be declared in period $t + 2$ but not in t or $t + 1$. However, this may favor longer episodes as they may be considered a good match even though they never appear *fully* in the data. We thus opted for the stringent criterion. Another important choice is to decide where to restart the episode after an error. In repeated games,

[9] In the application (i.e. Section 4), we will assume episodes of length 6 or less, totalling 964 episodes (see Figure A in the Supplemental Appendix for episodes of length less or equal to 4). The total number of equivalence classes for all episodes of length k is

$$n(k) = \frac{4^k - \sum_{j=1}^{k-1} j \times n(j) \times \mathbb{1}_{\{k \text{ is divisible by } j\}}}{k}.$$

[10] For example, (abc) is a subsequence of (adbc); thus, we could say that (abc) occurs within (adbc). However, we choose not to say that and insist, similar to Mannila, Toivonen, and Verkamo (1997), on exact matches rather than subsequences. In the above, (abc) clearly does not match (adb) or (dbc).

Figure 1: EXAMPLES

Notes: Each row contains an episode to test and numbers that indicate which element the episode should start from to guarantee a match with the sequence of game play. In case of a mismatch, a red X is displayed at the start of the testing interval.

when an episode is happening but a mistake occurs, players may try to re-implement the episode by trial and error. If they were following *abc* and, after a mistake, end up adopting *bca*, the difference seems immaterial. This motivates our aforementioned definition of equivalence classes, from which it follows that an episode is tested in each period by starting from *any element* that favors fitness to the data.

2.2 Action-Convergence Criterion

Our first method parallels the standard approach in the string-searching literature (Boyer and Moore (1977) and Knuth, Morris, and Pratt (1977)). We evaluate how well an episode fits a sequence of observed play by performing 'character comparisons' over an interval of periods and counting errors.

For any episode p, s and interval of periods T, let

$$\mathcal{X}(p, s, T) = \frac{\sum_{t \in T} (1 - m(p, s, t))}{|T|}$$

be the frequency of periods in T in which p mismatches sequence s. Furthermore, we define ϕ as the number of mismatches over the interval of periods T. Therefore, an episode p is said to be $(1-\phi)$-frequent in s over interval T, if $\mathcal{X}(p, s, T) = \min_{p' \in \mathcal{P}(n)} \mathcal{X}(p', s, T) \leq \phi$. This definition requires p to minimize the frequency of errors in $\mathcal{P}(n)$ and to induce at most a weighted frequency ϕ. Analogously, a sequence s $(1-\phi)$-converges to p (or is $(1-\phi)$-convergent) over T, if it admits a $(1-\phi)$-frequent episode over T. Otherwise, the sequence is $(1-\phi)$-divergent. Note that if many episodes minimize the frequency of errors, then we select the shortest episode.

In Figure 1, for example, assuming a maximum of two mismatches and the time interval consisting of periods 1 through 20 (i.e. $\phi = \frac{2}{20} = 0.10$), we can say that s $(1 - 0.10)$-converges to episode #2. Similarly, in the exposition of the results in Subsection 4.1, we will be looking at the first 20 periods of game play as well as the last 20 periods while allowing a maximum of two mismatches within each interval so that $\phi = 0.10$.

2.2.1 Action-Convergence Criterion and Finite Mixture Models

A natural benchmark for our approach is the finite-mixture methodology used in the repeated-games literature, most prominently by Dal Bó and Fréchette (2011). In that framework, repeated-game experimental data are represented as arising from a finite set of researcher-specified candidate strategies, and the estimation procedure recovers the population shares associated with each strategy. Individual behavior is then interpreted as reflecting (possibly with noise) one of these candidate rules. In one respect, this approach is more ambitious than ours. If one can credibly recover the strategies that generate behavior, then one also recovers the implied patterns of play. In this sense, structural strategy elicitation subsumes the description of realized paths.

This additional ambition, however, comes at the cost of stronger assumptions. Finite Mixture Models (FMM) fit a statistical model; consequently, they require

significant variability across both periods and games to achieve statistical identification.[11] In settings with high continuation probabilities, where the analyst may have access to a small number of very long interactions, the likelihood surface for an FMM can become prohibitively flat, rendering parameter estimation unstable or weakly identified. Furthermore, FMM require the researcher to specify ex ante a menu of candidate strategies and to estimate the population shares associated with them. Their success therefore hinges, not only on statistical fit, but also on whether the chosen strategy library is behaviorally relevant for the environment under study. In canonical settings, such as the repeated Prisoner's Dilemma, strategies such as Grim Trigger, Tit-for-Tat, or Win-Stay, Lose-Shift are natural candidates. But in less-studied environments, such as the Samaritan's Dilemma, or in datasets spanning many strategically distinct games (see Section 4.2.2), specifying ex ante a plausible and sufficiently rich strategy menu becomes considerably more demanding.

Our action-convergence criterion takes a deliberately more modest route, but can reveal regularities that theory does not instruct the analyst to look for ex ante. Rather than inferring unobserved strategies, it identifies recurrent episodes directly in observed play. This makes the method less structurally ambitious, but also far less dependent on game-specific prior knowledge. Given a dataset spanning multiple stage games, the same procedure can be applied uniformly without constructing a bespoke strategy library for each stage game. In this sense, the method is model-light and discovery-oriented; it is thus particularly well suited to (i) uncovering regularities in unfamiliar strategic settings where the risk of strategy-misspecification is high or (ii) comparing behavioral organization across games. This parsimony comes with a limitation: a recurrent episode need not correspond to a unique underlying strategy profile, and the method does not, by itself, deliver the off-path or counterfactual predictions that a correctly specified structural model can provide.

2.3 Modified k-Means Clustering Method

Our second method is a modified version of the celebrated clustering technique known as k-means clustering (MacQueen (1967)). This technique automatically partitions a dataset into k groups after first mapping data into numerical vectors. The conversion

[11] In this context, for instance, in Dal Bó and Fréchette (2011), subjects played from 33 to 77 games with discount factors $\delta = 0.5$ and $\delta = 0.75$.

of the sequences of game play into numerical vectors is explained next.

2.3.1 Digitization

Digitization is the step that makes clustering possible. Given that the k-means algorithm operates on numerical vectors and groups them by similarity, each observed sequence of play must first be mapped into a collection of numerical features. This representation determines how similarity between sequences is assessed. For instance, if we have a population of individuals and turn each person into a 2-dimensional vector (*weight*, *size*) after taking the relevant measurements, then the algorithm will group these individuals by similarity along the aforementioned physical characteristics. Obviously, the groups might change if we used (*IQ*, *salary*) instead.

Our objective is to measure similarity not only in terms of which episodes appear in a sequence, but also in terms of *when* they appear. Accordingly, we represent each sequence by the fit of every episode under three temporal weighting schemes that emphasize early, overall, and late play.

Let $\mathcal{P} = \{p_1, p_2, \ldots, p_\nu\} \subseteq \mathcal{P}(n)$ denote the indexed set of all canonical episodes of length less than or equal to n. Each element p_k uniquely identifies an equivalence class $[p_k]$ under the permutation-invariant ordering relation defined in Subsection 2.1.[12]

Fix a sequence $s \in \mathcal{A}^d$, where d denotes the number of periods of game play.[13] For any episode $p \in \mathcal{P}$ and any weighting parameter $\gamma > 0$, define

$$f(s, p, \gamma) = \frac{\sum_{t=1}^{|s|-|p|+1} \gamma^t m(p, s, t)}{\sum_{t=1}^{|s|-|p|+1} \gamma^t},$$

where $m(p, s, t)$ is the match indicator from Subsection 2.1. Thus, $f(s, p, \gamma)$ is the weighted frequency with which episode p matches sequence s. When $\gamma < 1$, earlier matches receive more weight; when $\gamma = 1$, all periods are weighted equally; and when $\gamma^{-1} > 1$, later matches receive more weight. We then define the attribute vector

[12] Recall that in Section 4, we will assume episodes of length 6 or less that total 964 episodes.

[13] In our application, the actual length of the repeated games with continuation probability 0.99 was 116 periods, whereas the length of the repeated games with continuation probability 0.995 was 188 periods (i.e. $d \in \{116, 188\}$). The actual draw for the lengths was done ex ante (see Subsection 3.2.1 for the details).

$$v(s) = \Big(f(s, p_1, \gamma), \ldots, f(s, p_\nu, \gamma),\ f(s, p_1, 1), \ldots, f(s, p_\nu, 1), \\ f(s, p_1, \gamma^{-1}), \ldots, f(s, p_\nu, \gamma^{-1}) \Big) \in \mathbb{R}^{3\nu},$$

with $\gamma \in (0, 1)$. The value of γ should be chosen to make time matter in the representation: values of γ close to 1 place nearly equal weight on all periods and collapse the early-, middle-, and late-weighted components into similar objects.[14] Importantly, γ is a feature of the procedure and is not directly related to the discount factor δ.

The first ν entries of $v(s)$ emphasize the beginning of the interaction, because matches occurring earlier in the sequence receive larger weights under γ. The next ν entries use equal weighting and therefore summarize the sequence as a whole. The final ν entries emphasize the end of the interaction, because later matches receive larger weights under $\gamma^{-1} > 1$.

In summary, digitization maps each of the N experimental sequences of play into a 3ν-dimensional vector that records how strongly each episode appears early, overall, and late. Two sequences are therefore close in the clustering stage when they exhibit similar episodic structure under these three temporal lenses.

Letting $\{s^1, \ldots, s^N\}$ be the dataset of sequences of play and $v^i = v(s^i)$, the digitized dataset is

$$V = \{v^1, \ldots, v^N\} \subseteq \mathbb{R}^{3\nu}.$$

2.3.2 k-Means Algorithm

We use the k-means clustering procedure to divide V into groups of sequences with similar temporal profiles, which we interpret as behavioral *storylines*. We first choose a number of clusters k, and then the algorithm determines (i) the k centers, $C = \{c^1, \ldots, c^k\} \subseteq \mathbb{R}^{3\nu}$, and (ii) an assignment function $a : V \to C$ that maps each element of V to one of the k centers. The objective is to find centers and an assignment of points to centers that minimize the Sum of Mean Squared Errors (SMSE) between

[14] In the exposition of the results in Subsection 4.2, we set $\gamma = 0.9$ because it creates a clear separation between early and late behavior, allowing the clustering procedure to detect temporal differences.

the data points and their corresponding centers; that is,

$$\min_{a} \quad \text{SMSE}\,(a|V) := \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{3\nu}\left(v_j^n - a_j(v^n)\right)^2.$$

Initially, the N vectors are assigned to k (randomly generated) clusters where each vector is assigned to the cluster that minimizes its MSE. From there, k new centers are calculated based on the vectors within each cluster. The calculation of the new center is simply the mean value of all of the vectors within the cluster. The N vectors are then (potentially) repositioned to the nearest center and, afterwards, a new center is again calculated. The procedure continues until the N vectors are stable within their cluster; that is, there is no more reshuffling of vectors across clusters. All in all, for every k, the clustering algorithm outputs the optimal assignment function (including the optimal centers) and the minimized SMSE. Clearly, the minimized SMSE decreases as k increases.

A drawback of the k-means clustering approach is that the number of clusters k has to be declared ex ante; that is, the number of clusters k does not emerge endogenously. In an attempt to endogenize k, we adopt a cost-benefit logic using the Kneedle approach (Satopaa et al. (2011)). Plotting SMSE against the number of clusters, the Kneedle algorithm identifies the optimal cluster count by pinpointing the point of greatest slope change, where additional clusters provide diminishing improvements. At that number of clusters, we terminate the procedure and look at the storylines that emerge from each cluster.[15]

2.3.3 k-Means Clustering and Finite Mixture Models

As discussed in Section 2.2.1, the finite-mixture methodology is naturally geared toward strategy elicitation: given an ex ante library of candidate strategies, it asks which strategies account for the data and in what proportions. Our clustering procedure, by contrast, is geared toward trajectory discovery: it asks whether entire sequences of play can be organized into a small number of empirical storylines with similar temporal profiles.

[15] This automated 'elbow method' enhances objectivity thus simplifying the selection of the ideal number of clusters.

This difference matters because the modified k-means method operates on full sequences rather than on local episodes. Each observed sequence is first digitized into a multi-temporal attribute vector, and sequences are then grouped by similarity in their beginning-, middle-, and end-of-game behavior. By defining similarity over the discounted, equal, and reverse-discounted components of these vectors, the clustering algorithm explicitly accounts for the temporal evolution of play – distinguishing, for example, between pairs that cooperate from the outset and those that learn to coordinate only after a period of initial conflict.

As in Subsection 2.2.1, the gain from this more agnostic approach is portability and discovery. The same clustering procedure can be applied across games without constructing a bespoke strategy menu for each environment. The cost is that the resulting clusters are empirical summaries rather than structural primitives: they do not, by themselves, identify named strategies or generate off-path or counterfactual predictions. Moreover, the procedure still depends on design choices, including the digitization of sequences, the similarity metric, and the number of clusters. The Kneedle method helps discipline the last of these choices, but it does not eliminate researcher judgment.

3 Experimental Design

3.1 Stage Games

We conducted a large-scale experiment (with 1510 participants) that covered sixteen 2×2 stage games. The experiments were conducted in two phases. The first phase included the ten strict ordinal games shown in Figure 2. The second phase added another six games introduced to probe behavioral patterns raised in the first phase; these are displayed in Figure 3.

A strict ordinal game is one in which each u_i is injective so that players strictly rank all four outcomes. Focusing on this class allows us to define a well-structured space of strategic environments that can be systematically classified. This classification, in turn, enables the selection of a representative set of games that spans a wide range of qualitatively-distinct interactions. To begin, we follow the approach of Randolph Bruns (2015), which builds on the topological framework introduced by

Figure 2: Strict Ordinal Stage Games

	A	B
A	3,3	1,4
B	4,1	2,2

(a) Prisoner's Dilemma (PD)

	A	B
A	1,1	4,3
B	3,4	2,2

(b) Battle of the Sexes (BO)

	A	B
A	3,3	0,2
B	2,0	1,1

(c) Stag Hunt (SH)

	A	B
A	3,3	1,4
B	4,1	0,0

(d) Chicken (CH)

	A	B
A	2,2	1,1
B	4,3	3,4

(e) Samaritan's Dilemma (SD)

	A	B
A	2,4	4,2
B	1,1	3,3

(f) Samson (SA)

	A	B
A	1,4	3,2
B	2,3	4,1

(g) Total Conflict (TC)

	A	B
A	3,3	1,2
B	2,1	0,0

(h) Common Interest (CI)

	A	B
A	1,4	4,1
B	3,2	2,3

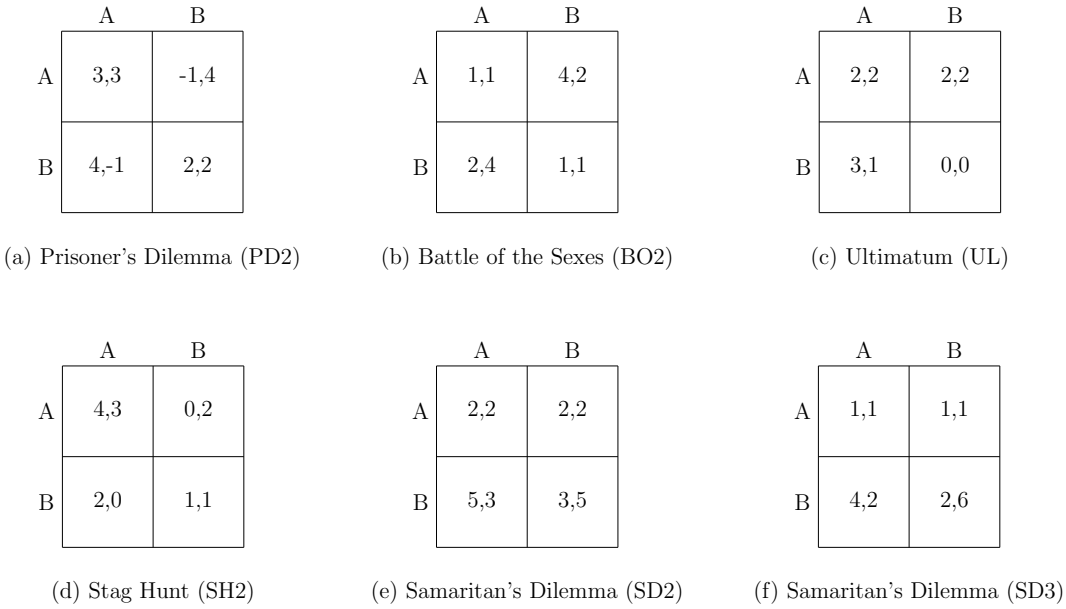
(i) Unique Mixed (UM)

	A	B
A	3,4	4,2
B	1,1	2,3

(j) Complex (CO)

Notes: Figure 2 displays the payoff matrices of the 2×2 strict ordinal stage games.

Figure 3: Other Stage Games



(a) Prisoner's Dilemma (PD2) (b) Battle of the Sexes (BO2) (c) Ultimatum (UL)

(d) Stag Hunt (SH2) (e) Samaritan's Dilemma (SD2) (f) Samaritan's Dilemma (SD3)

Notes: Figure 3 displays the payoff matrices of an additional set of 2×2 stage games.

Robinson and Goforth (2005). This framework organizes 2×2 ordinal games along three key dimensions:

1. **Dominance structure**: A game exhibits **two-sided** dominance if both players have a strictly dominant action, **one-sided** dominance if only one does, and **zero-sided** dominance if neither does.

2. **Number of pure-strategy Nash equilibria**: Games with zero-sided dominance may have either zero or two pure-strategy Nash equilibria, while those with one- or two-sided dominance have exactly one.

3. **Equilibrium payoff family**: We classify each pure-strategy Nash equilibrium (NE) based on the ordinal ranks of the payoffs it yields for both players. Let $u_{i,n}$ denote the payoff received by player i from their n^{th} most preferred outcome, with $n = 1, 2, 3, 4$. A NE is *Pareto-dominant* if it gives both players their most preferred payoff, $u_{\cdot,1}$; *biased* if it gives one player $u_{i,1}$ and the other $u_{j,2}$, thereby favoring one side; *unfair* if it gives one player $u_{i,1}$ and the other $u_{j,3}$, resulting in

a highly asymmetric outcome; *sad* if it gives both players intermediate payoffs ($u_{i,2}$ or $u_{i,3}$); and *trapped* if it gives both players the lowest payoff, $u_{\cdot,3}$. These five types capture the efficiency and asymmetry of the equilibrium.

Among the ten canonical games, four exhibit zero-sided dominance (BO, SH, CH, UM), three have two-sided dominance (PD, TC, CI), and three have one-sided dominance (SD, SA, CO). In games with zero-sided dominance, there can be either zero or two pure-strategy Nash equilibria. UM has none; BO, SH, and CH each have two. All three are symmetric games, and so are their equilibria. In BO, both Nash equilibria fall into the biased family; in CH, they are unfair; and in SH, one is Pareto-dominant and the other is trapped.

Games with one- or two-sided dominance each yield exactly one pure-strategy NE. In the two-sided group, the NE in PD is trapped, in TC it is sad, and in CI it is Pareto-dominant. In the one-sided group, the NE is biased in SD and CO, and unfair in SA. To further distinguish the one-sided dominance games, we introduce a fourth classification dimension: the **length of the shortest episode** that is both egalitarian and Pareto-efficient. This dimension captures the minimal coordination required to reach a fair and efficient outcome. In SA, the shortest such episode has length 1; in SD, length 2; and in CO, length 3. These differences allow us to test whether players are able to coordinate on efficient outcomes when doing so requires greater behavioral complexity.

The first phase of the experiments raised several follow-up questions about the roles of dominance, payoff asymmetries, and coordination complexity in shaping behavior across games. To address these questions, we introduced a second phase with six additional games displayed in Figure 3.

3.2 Experiments

Given the large sample size required to gather data for all games, we decided to run the experiments on Amazon's Mechanical Turk (henceforth, MTurk), where the subject pool is sufficiently large. We used two continuation probabilities on the MTurk platform, $\delta = 0.99$ and $\delta = 0.995$. The discount factor $\delta = 0.99$ was implemented in all games; however, in three games (in particular, PD2, SH2 and SD3), the discount

factor $\delta = 0.995$ was also implemented.[16] Table 1 displays the number of subjects who played each game on the platform under the respective continuation probability.

Table 1: Characteristics of the Games

Game	Acronym	# of Subjects ($\delta = 0.99$)	# of Subjects ($\delta = 0.995$)
Panel A			
Prisoner's Dilemma	PD	78	
Battle of the Sexes	BO	84	
Stag Hunt	SH	82	
Chicken	CH	76	
Samaritan's Dilemma	SD	76	
Samson	SA	84	
Total Conflict	TC	78	
Common Interest	CI	52	
Unique Mixed	UM	80	
Complex	CO	80	
Panel B			
Prisoner's Dilemma	PD2(L)	80	80
Battle of the Sexes	BO2	100	
Ultimatum	UL	76	
Stag Hunt	SH2(L)	76	88
Samaritan's Dilemma	SD2	82	
Samaritan's Dilemma	SD3(L)	72	86

Notes: Table 1 displays the number of subjects who played the corresponding game under the respective continuation probability. A data point consists of a pair of subjects (hence the numbers of data points are half of the above numbers). The acronym includes 'L' (for *longer* interaction) when it refers to $\delta = 0.995$, otherwise it refers to $\delta = 0.99$.

[16] These three games are arguably more complicated than the other ones. Subjects perhaps needed more time to gain experience, thus we granted them with another 100 periods (in expectation) of game play.

3.2.1 MTurk Experiments

The online experiments were conducted on the MTurk platform. MTurk is an online labor market that consists of workers and requesters. Requesters post human intelligence tasks (HITs) that are completed by the workers in exchange for payment. In the social sciences, HITs take the form of experiments conducted by researchers with workers taking up the role of subjects. Despite MTurk workers being significantly more heterogeneous than university subject pools, there exists a growing body of literature supporting the efficacy of MTurk (see, e.g., Horton, Rand, and Zeckhauser (2011), Amir, Rand, and Gal (2012), Arechar, Kraft-Todd, and Rand (2017), Arechar, Gächter, and Molleman (2018)).

The MTurk experiments were conducted in three waves. The first wave took place from May to August, 2021; the second wave lasted from June to October, 2023; the third and final wave took place from February to March, 2024. One of the challenges we faced was the need to have a large number of *reliable* workers to pair up in the experiment. Similarly to Goodman, Cryder, and Cheema (2013) and Hauser and Schwarz (2016), we imposed a restriction (through the *qualification* HITs) that only US[17] workers who had completed at least 100 HITs with a HIT approval rate of, at least, 90%[18] could sign up for an experimental session. In addition, workers were also asked to indicate their available time slots. The HIT for each experimental session was released about 30 minutes before the scheduled time, and only workers who indicated their availability at that time slot could participate.

Upon accepting the HIT, participants were provided with the url of the experiment. To make sure everybody started around the same time, while on the portal, participants were asked to read general information about the experiment up until the scheduled starting time. Participants could participate in only one experimental session. Furthermore, participants who started an experimental session were subsequently disqualified from accepting another HIT of the experiment.

The participants were first asked to provide informed consent. After that, they went through the experimental instructions and had to complete a quiz to confirm their understanding of the game. The instructions made it clear that participants

[17]This requirement was crucial to minimize the chance of including individuals who were not proficient in English (the language used in the experimental instructions).

[18]With the completion of a HIT, the requester approves the participation of the worker. Therefore, what we required was that at least 90% of each participant's previous submissions were approved by the requesters.

needed to answer all quiz questions correctly within a time-frame to proceed to the game-play stage.[19] After successfully completing the quiz, participants entered the (virtual) waiting room where matching occurred. Once two participants were successfully matched, a game from Figure 2/3 was randomly assigned to them and the pair began their repeated interaction with the respective values denoted in *experimental francs*. Participants played a single infinitely-repeated game with perfect monitoring. In the experiment, the two actions of the stage games 'A' and 'B' in Figures 2 and 3 were labeled as 'Y' and 'W.' Additionally, the labelling of 'Y' and 'W' was randomized across repeated games so that, for instance, (Y, Y) was the NE in some Samaritan's Dilemmas and (W, W) was the NE in others. The experiments were programmed and conducted using oTree (Chen, Martin, and Wickens (2016)).

Participants who failed to reach a decision within one minute were marked as *dropouts*.[20] The dropout rate in our experiments was less than 5%. After a participant dropped out, the experiment would end for both individuals in the pair. Individuals who dropped out were not paid, while partners of dropouts were paid based on the earnings they had accumulated until the last interaction plus the participation fee. This information was common knowledge. Data points involving a dropout were excluded from the analysis.

We used high continuation probabilities, $\delta = 0.99$ and $\delta = 0.995$, to induce long time series between the same two participants. Furthermore, the termination rule was common knowledge. The actual draw for the lengths of the repeated games was done *ex ante* and hard coded into the program. We had one draw for all repeated games with continuation probability 0.99, and one draw for all repeated games with continuation probability 0.995. The actual length of the repeated games with continuation probability 0.99 was 116, whereas the length of the repeated games with

[19]Participants that gave five wrong answers were immediately dismissed from the experiment. Though we allowed participants that gave four wrong answers to proceed to the game-play stage, we only included in the dataset participants that gave three or fewer wrong answers. Ex ante, the screening choice is arbitrary. However, ex post, we chose to exclude participants from the dataset that gave four wrong answers after we observed that many of them chose actions that seemed to reflect a wrong interpretation of the instructions. Our decision is backed up by Smith (1994) who points out that we do not learn about the relevant behavior when instructions are misunderstood (p. 129). Having said this, including in the dataset the 38 participants that gave four wrong answers would not affect the qualitative results of our analysis.

[20]On average, the per-period decision time of the participants was less than 2 seconds; the median was about one second. Consequently, participants had a sufficient amount of time to reach their decision comfortably before being marked as dropouts.

continuation probability 0.995 was 188. We fixed the length of the repeated games to make consistent comparisons across games without introducing variations in learning/fatigue, which could confound our findings. Each player's payoff in the repeated game consisted of the sum of the experimental francs accumulated from each round. The exchange rate was common knowledge: 100 experimental francs for $1.20.

After playing the repeated game, participants were asked to complete a survey and to provide demographic information. For the $\delta = 0.99$ sessions, the 1256 participants on average took 26 minutes to finish the experiment; for the $\delta = 0.995$ sessions, the 254 participants on average took 35 minutes to finish. Overall, the average hourly payment was $16.12, which is a generous payment for standard MTurk experiments, but comparable to the average hourly payment in a typical laboratory experiment. The experimental instructions, interface, and the quiz questions of the experiments are provided in the Supplemental Appendix.

4 Findings

We organize our results by the methodological approach, first presenting the findings based on the action-convergence criterion, followed by the storylines identified through the modified k-means clustering procedure. In the analysis that follows, we will be using $n = 6$; that is, focusing on episodes of length 6 or less, totalling 964 episodes.

4.1 Action-Convergence Criterion

In repeated-game experiments, researchers are often interested in behavior after subjects have gained experience as well as in comparing it with behavior early in the experiment. This motivates the following definitions:

Definition 1 *A long-run outcome is an episode that is* 90%*-frequent over* $T_{End} = \{$*last 20 periods*$\}$*;*

Definition 2 *A short-run outcome is an episode that is* 90%*-frequent over* $T_{Beg} = \{$*first 20 periods*$\}$*.*

Table 2 reports the short-run and long-run outcomes extracted from the experimental data.[21] Convergence is not expected to occur within the same number of periods across all games. Since our analysis focuses on the beginning and end of play rather than the speed of convergence, recall that we granted an additional 100 periods (in expectation) to the three more complex games – PD2, SH2 and SD3 – by considering, in addition to $\delta = 0.99$, the higher discount factor of $\delta = 0.995$.

The main results here will speak about the concentration of long-run outcomes in the data (Result 1), the growth of frequent episodes over time (Result 2) and the increased complexity of frequent episodes over time (Result 3).

Result 1 (Concentration) *The empirical distributions of long-run play exhibit entropy reductions ranging from 18% to 83% relative to the uniform benchmark, with the exceptions of SD2 (approximately -3%) and UM (approximately 2%).*

This result provides an empirical counterpoint to the theoretical abundance of outcomes in repeated games. The Folk Theorem states that, with sufficiently patient players, a broad set of feasible and individually rational payoff vectors can be supported in a Subgame Perfect Nash equilibrium – and under other equilibrium refinements. This multiplicity extends beyond payoff vectors to the on-path sequences of action profiles that can arise in equilibrium. In the data, however, long-run play is highly concentrated. Thus, the theoretical abundance of supportable outcomes provides limited guidance about what is most likely to happen in these games.

Result 1 does not contradict the Folk Theorem. With a sufficiently large number of experiments and observations, one could, in principle, observe a broader set of long-run outcomes consistent with the theory. Even so, the theorem provides no reason to expect these outcomes to occur with comparable frequency. More fundamentally, classical infinitely repeated-game theory aims to identify what can be supported in equilibrium rather than how frequently different equilibrium outcomes should be observed. This distinction underscores our point: while theory permits many outcomes, empirical play tends to concentrate on a small subset. Understanding which episodes emerge – and how frequently – therefore motivates complementing the standard framework with additional predictive structure.

[21] An additional episode, $((B,A),(B,A),(B,A),(B,B))$, was identified in TC with one short-run outcome that persisted until the end. This episode appears to reflect spurious, rather than meaningful, convergence and is therefore omitted from the table and the discussion that follows.

Table 2: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	22 → 32	3 → 25	34 → 38	18 → 28	8 → 16	6 → 17	5 → 24	21 → 25	→ 7	10 → 20	13 → 37	20 → 36	3 → 30	20 → 30	26 → 33	25 → 39	19 → 20	9 → 22	13 → 26
A A / A A	20 → 22 (-1)	→ 4	34 → 35 (-1)*	18 → 26 (-1)		3 → 6 (-1)*	→ 1	21 → 25*	→ 2	9 → 17 (-1)*	10 → 17 (-1)	17 → 24 (-2)	1 → 1 (-1)	19 → 21 (-2)	26 → 29 (-2)*	25 → 31 (-2)*			
A A / A B		→ 4*			→ 1								1 → 8*	→ 5*			→ 1	2 → 2 (-2)	→ 3
B B / A A		3 → 12*		→ 1*	1 → (-1)		5 → 21*		→ 3				→ 5*	1 → 4*				→ 1	→ 1
B B / B A	2 → 10*		→ 3*	→ 1	5 → 13 (-2)*	2 → 11			→ 1		3 → 20*	3 → 12 (-2)*	→ 1		→ 4*	→ 8*	11 → 9 (-4)*	4 → 11 (-1)*	12 → 17 (-3)*
A A / A B						1 → (-1)			→ 1										
A B / A A													→ 1						
A B / A B							→ 2								*	*			
A B / B A		→ 5*		*									1 → 14*	*					
B B / A B					2 → 2 (-1)												8 → 10 (-1)	→ 4	1 → 2 (-1)
A A A / A A B										1 → 3									
B B B / A A B																		3 → 4 (-1)	→ 3
Divergent	17 → 7	39 → 17	7 → 3	20 → 10	30 → 22	36 → 25	34 → 15	5 → 1	40 → 33	30 → 20	27 → 3	20 → 4	47 → 20	18 → 8	12 → 5	19 → 5	22 → 21	27 → 14	30 → 17
Entropy	3.0 → 1.9	5.3 → 3.7	1.1 → 0.9	3.3 → 2.0	4.9 → 3.9	5.2 → 4.1	5.0 → 2.9	1.2 → 0.2	5.3 → 5.2	4.6 → 3.5	4.4 → 1.4	3.5 → 1.5	5.6 → 3.9	3.1 → 2.3	2.0 → 1.3	2.8 → 1.4	3.8 → 3.9	4.8 → 3.6	4.4 → 3.5

* Stage Nash Pareto Efficient Equal Payoff

Notes: The columns indicate the games and the rows indicate the episodes. The number that precedes an arrow is that of short-run outcomes and the number an arrow points to is that of long-run outcomes. The number of convergent pairs in the beginning that did not converge to the same episode at the end is displayed under the arrows. In the symmetric games BO, CH and BO2 the episodes (A, B) and (B, A) are interchangeable (in BO2 also the episodes (A, A) and (B, B)); therefore, the values could have been placed in either row. Green fonts flag an increase in convergent sequences, whereas red fonts flag a decrease in convergence. Pareto efficient episodes are displayed with a green background, whereas egalitarian episodes are displayed with a green border. The ⋆ denotes a NE episode.

Result 2 (Convergence) *Frequent episodes increase over time:*

(i) 36% (275/755) of observed sequences converge within the first 20 periods, compared to 67% (505/755) in the last 20 periods – a difference of 31 percentage points, indicating substantially greater convergence toward the end of play.

(ii) Among sequences that do not converge within the first 20 periods, 54% (257/480) eventually converge to a long-run outcome (57% (250/440) excluding the UM game).[22]

Table 2 shows that once convergence occurs, it is often permanent, which is not obvious given that, when reached early, many periods of play still remain. Figures 4 and 5 break down these trends by game. In Figure 4, most games (except UM, SA, SD, and SD2) have total convergence rates at or above 0.5, with some (PD2, SH, CI) approaching 1. For sequences that were initially divergent, late convergence rates (blue bars in Figure 5) are particularly high in CI, SH2L, PD2, PD2L and BO2 – ranging from 60% in BO2 (28/47) to 93% in PD2 (25/27). In contrast, the share of sequences that remained divergent (orange bars) is much smaller in these games, highlighting how rarely initial divergence persists to the end.

Figure 6 tracks the percentage of sequences containing a 90%-frequent episode over rolling 20-period windows $T_0 = [0, 20]$, $T_1 = [1, 21]$, $T_2 = [2, 22]$, $\ldots$, $T_n = [|s|-20, |s|]$, revealing three distinct convergence trajectories: (i) relatively flat convergence (SH, SD2), (ii) steadily increasing convergence (BO, TC, BO2, UL, SH2L, SD3, SD3L), and (iii) an initial jump followed by a plateau (PD, CI, UM, SH2, CO, CH, PD2, PD2L, SD, SA).

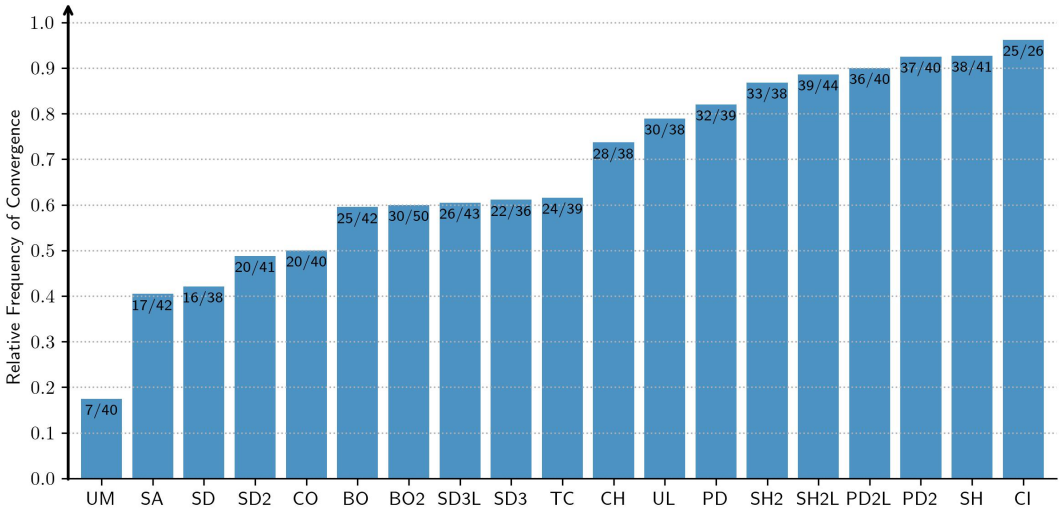
For the next result, we refer to a long episode as an episode of length 2 or more.

Result 3 (More complexity over time) *There are three times as many long episodes among long-run outcomes than among short-run outcomes.*

Thus, about 67% (34/51) of long episodes formed after the first 20 periods, indicating that this phenomenon often takes time to emerge. The fact that long episodes represent only 10% (51/505) of all long-run outcomes may reflect the nature of our setups. If, as proposed in Mathevet (2018)'s axioms, subjects are willing to adopt more complex arrangements only when such arrangements benefit at least one of them, then

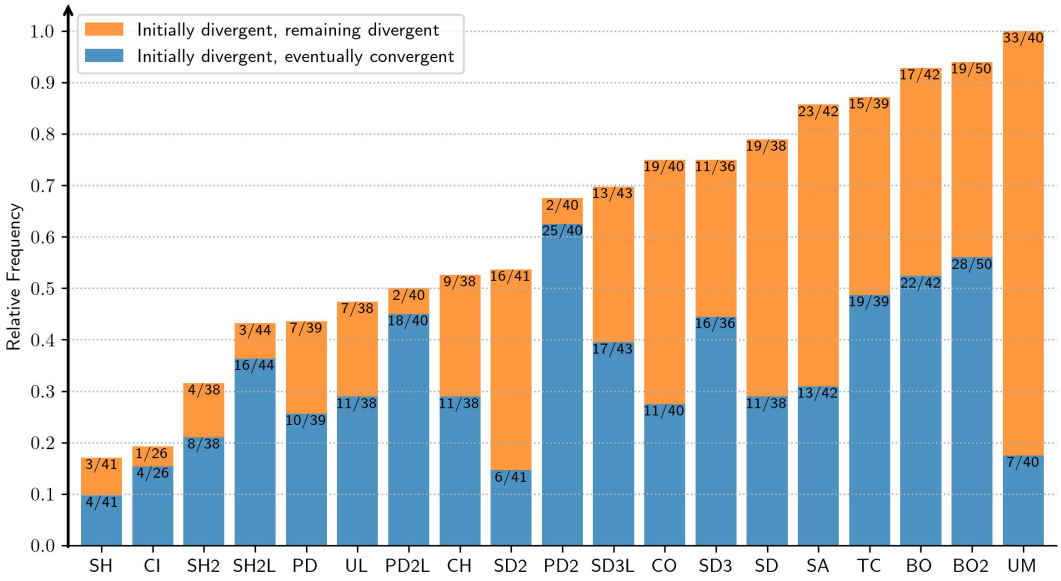
[22] In the UM game, deterministic notions of convergence are not well-suited and are therefore excluded from this analysis.

Figure 4: Total Convergence



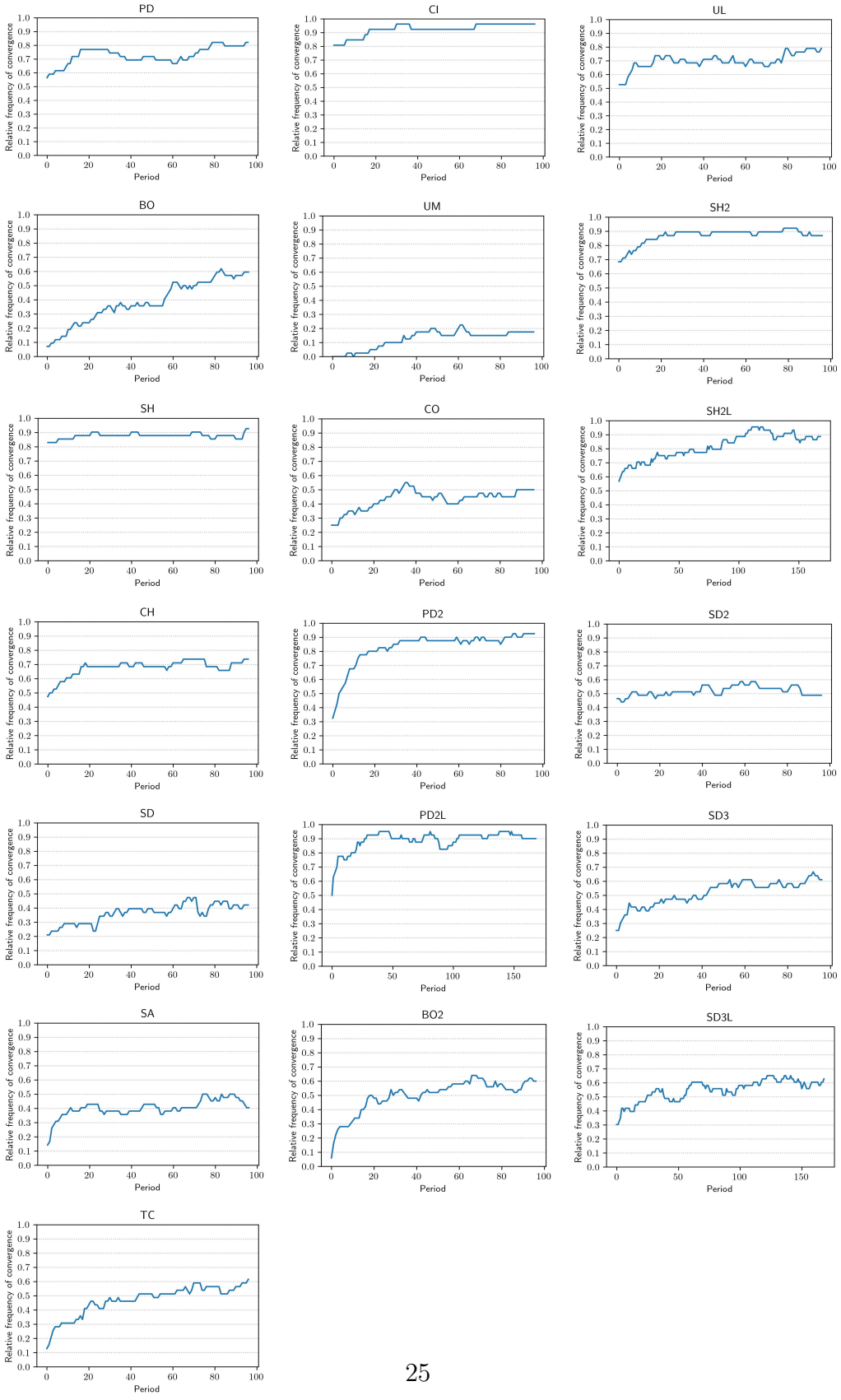
Notes: Figure 4 displays the relative frequencies of total convergence.

Figure 5: Late Convergence



Notes: Figure 5 displays the relative frequencies of long-run outcomes that emerged from divergent sequences (in blue), and the relative frequencies of divergent sequences that emerged from divergent ones (in orange).

Figure 6: Stabilization



such arrangements would (and indeed do) arise primarily in BO, BO2, SD2, SD3 and SD3L. Hence, only a few of our setups create conditions under which subjects might want to implement complex arrangements.

Discussion

Results 2 and 3 reveal a form of failure of recursive play in the data. To see this, define

$$\mathcal{S}_t^{20} = \{\alpha_t = (a_t, \ldots, a_{t+19}) : \exists\, \sigma \in \text{SPNE such that } \sigma \text{ induces } \alpha_t\}\,.$$

In a standard repeated game, subgame perfection implies that $\mathcal{S}_t^{20} = \mathcal{S}_{t'}^{20}$ for all t, t'. Thus, the SPNE correspondence offers the same set of 20-period continuation paths – simple or complex – at every date. This is an implication of recursiveness at the level of possibility. Results 2 and 3 concern frequency instead. Let $Q_t \in \Delta(\mathcal{A}^{20})$ denote the empirical distribution, across pairs, of 20-period paths beginning at t. Let $\mathcal{C} \subseteq \mathcal{A}^{20}$ denote the set of paths satisfying the action-convergence criterion, and let $\mathcal{C}_{\geq 2} \subseteq \mathcal{C}$ denote those whose selected episode has length at least two. We find

$$Q_{\text{Beg}}(\mathcal{C}) = \frac{275}{755} = 0.36, \qquad Q_{\text{End}}(\mathcal{C}) = \frac{505}{755} = 0.67,$$

while

$$Q_{\text{Beg}}(\mathcal{C}_{\geq 2}) = \frac{17}{755}, \qquad Q_{\text{End}}(\mathcal{C}_{\geq 2}) = \frac{51}{755}.$$

Therefore, elapsed experience changes not only the likelihood of settling into a recurring episode, but also the frequency of multi-profile arrangements. Complex episodes remain a minority of long-run outcomes, but 34 of the 51 observed at the end emerged only after the first 20 periods. These findings do not require the support of play to change. Noisy or complex paths observed early may also occur late, and a sufficiently large sample might reveal equally rich supports at both dates. What changes sharply is the weight placed on different paths. Over time, play becomes more stable, while some pairs also construct more elaborate recurring arrangements. SPNE can accommodate both developments because strategies may condition on the entire history. But this flexibility cuts both ways: the SPNE correspondence does not privilege dispersion followed by stability over stability followed by dispersion, or simple play followed by complex coordination over the reverse. The data do. At the level of possibility, a complex arrangement is no less supportable at the beginning than

at the end; what SPNE does not explain is why its empirical prevalence builds with experience. These evolutions are consistent either with subjects passing through an initial out-of-equilibrium learning phase before coordinating in the intended repeated game, or with equilibrium from the outset in a richer dynamic game in which beliefs, reputations, or conventions evolve with experience (as in, e.g., Fudenberg and Levine (1989)).

4.1.1 Pareto Efficiency

We next compare short-run and long-run outcomes with respect to Pareto efficiency and Pareto improvement.

Result 4 (Long-run efficiency) *There are 64% more efficient long-run outcomes than efficient short-run ones (264 vs. 433).*

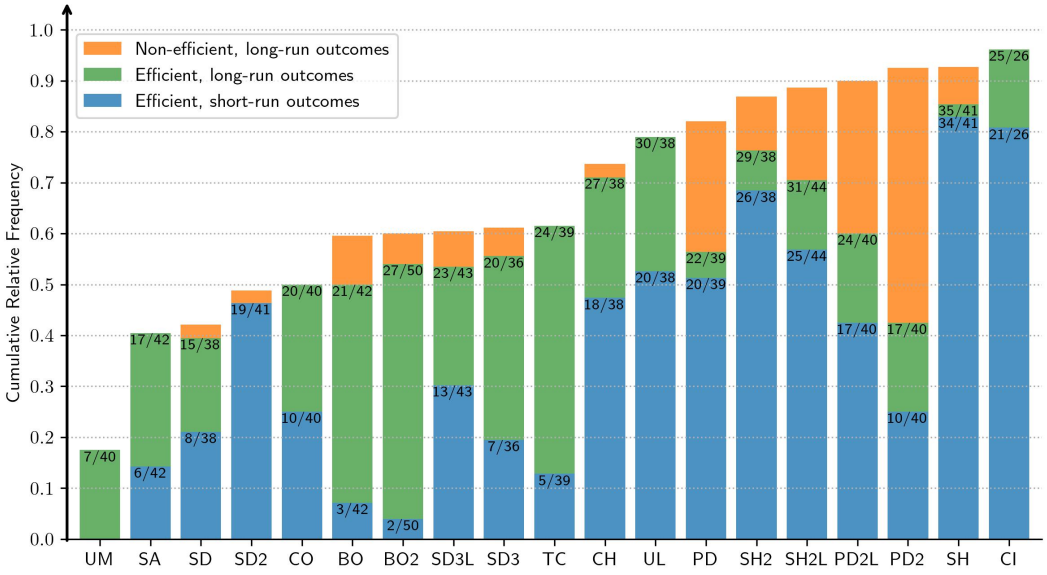
Time allows additional efficient agreements to emerge (433 in the long run vs. 264 in the short run). However, a higher share of short-run outcomes are efficient: about 96% (264/275) of short-run outcomes are efficient, compared to roughly 86% (433/505) of long-run outcomes. Thus, when convergence occurs quickly, it is almost always towards an efficient outcome.

Figure 7 illustrates substantial variations across games. At one extreme, SH and CI exhibit widespread early convergence to the efficient (static NE) outcome. At the other, BO2, BO and TC display no early convergence, but eventually reach an efficient episode. Intuitively, the speed of convergence appears related to the complexity of the efficient norms: if efficiency does not require alternation, subjects typically converge quickly to the efficient action profile.

Even when efficiency is not achieved immediately, subjects often generate additional value over time. Specifically, for sequences that diverged in the first 20 periods but converged in the last 20, we computed the difference in average payoffs between the early and late periods. We refer to this as 'value creation.' Figure 8 presents the value creation across all games.

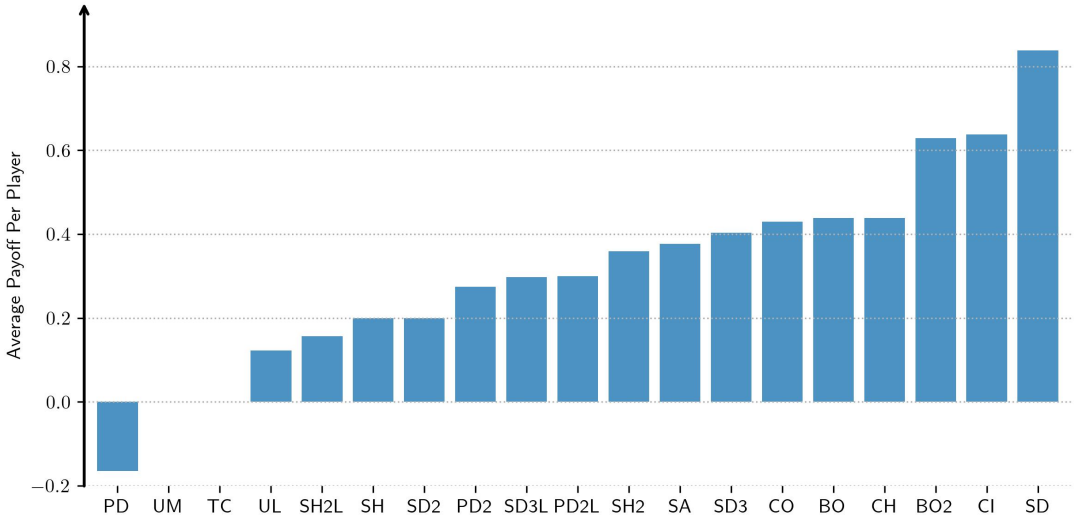
Result 5 (Long-run value creation) *In all games except PD, UM and TC there is a strictly positive average Pareto improvement between the last and the first 20-period payoffs for sequences that failed to converge in the first 20 periods, but converged to a long-run outcome.*

Figure 7: Pareto Efficiency



Notes: Figure 7 displays the relative frequencies of short-run outcomes that are efficient (in blue), the cumulative relative frequencies of long-run outcomes that are efficient (in green), and the relative frequencies of non-efficient, long-run outcomes (in orange).

Figure 8: Average Payoff Creation in the Long Run



Notes: Figure 8 displays the average difference in the average payoffs between the first and the last 20 periods for the pairs that diverged in the first 20 periods and converged in the last 20 periods.

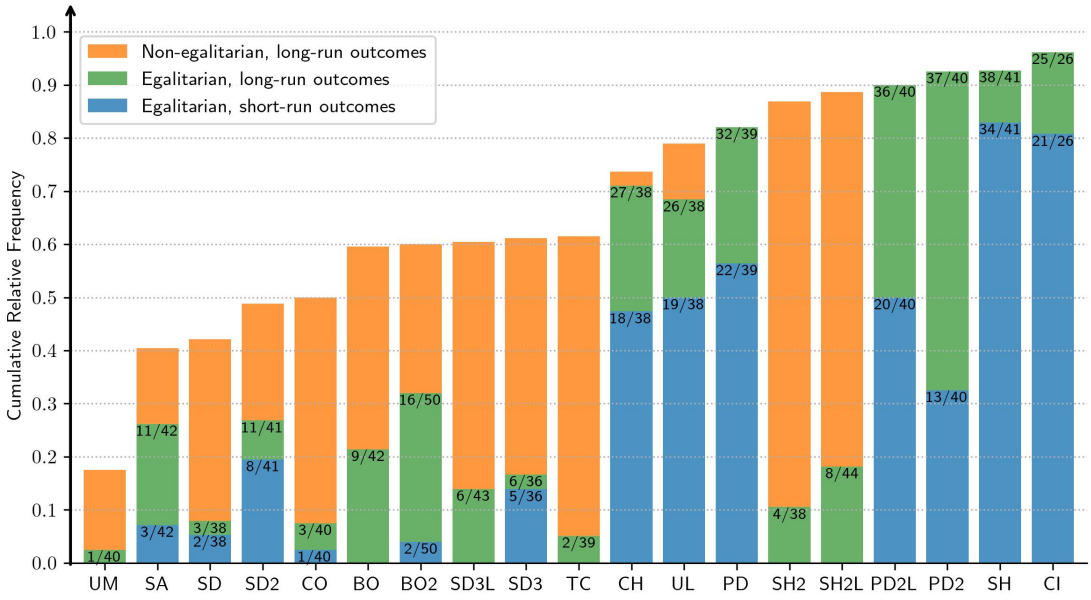
4.1.2 Egalitarianism

We also compare short-run and long-run outcomes with respect to fairness and movements toward equality.

Result 6 (Long-run egalitarianism) *There are 80% more egalitarian long-run outcomes than egalitarian short-run ones (167 vs. 301).*

Time also allows additional egalitarian agreements to emerge. However, unlike efficiency, the share of outcomes that are egalitarian remains nearly unchanged: about 61% (167/275) of short-run outcomes are egalitarian compared to roughly 60% (301/505) of long-run outcomes. Thus, when convergence is quick, it is just as likely to lead to egalitarian payoffs as when it occurs only in the long run.

Figure 9: Egalitarianism



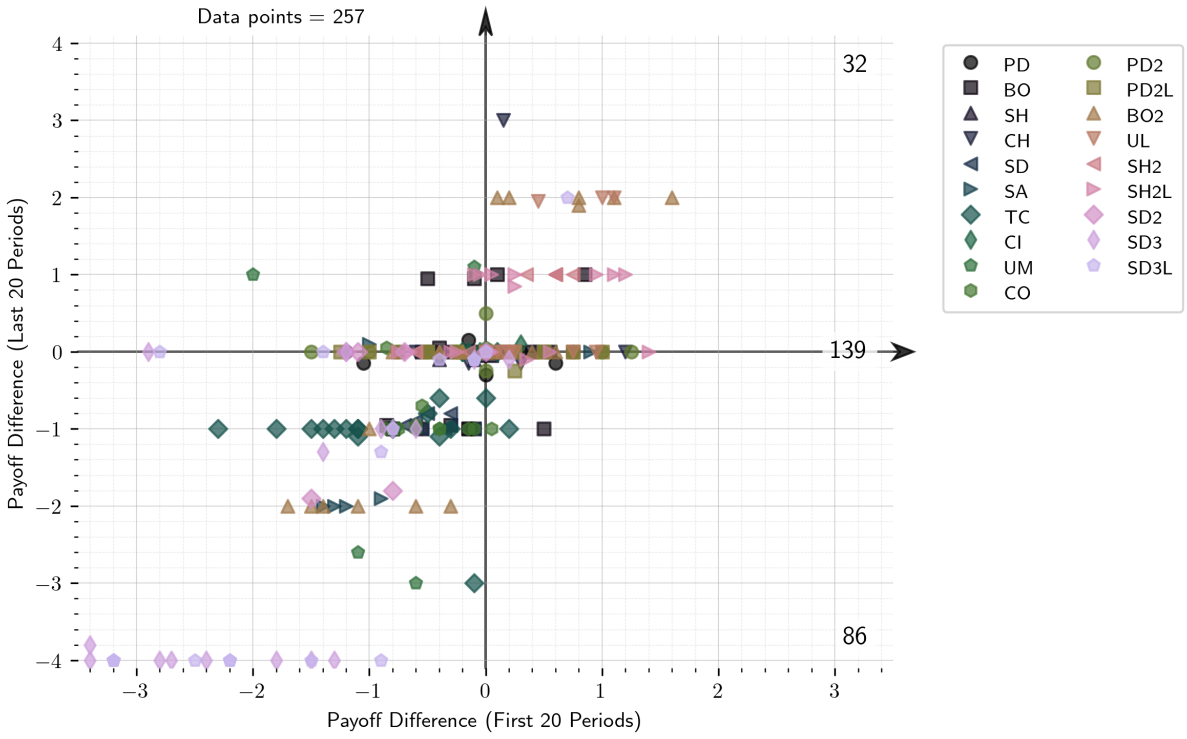
Notes: Figure 9 displays the relative frequencies of short-run outcomes that are egalitarian (in blue), the cumulative relative frequencies of long-run outcomes that are egaliatrian (in green), and the relative frequencies of non-egalitarian, long-run outcomes (in orange).

Figure 9 illustrates large variations across games. At one extreme, SH and CI exhibit widespread early convergence to static Nash equilibria, which yield fair pay-

offs. At the other extreme, BO2 shows no early convergence, but eventually reaches alternating behavior that produces equal payoffs.

We also examine whether outcomes move towards equality over time. Among the 257 sequences that diverged in the first 20 periods but converged in the last 20 periods, players' average payoffs shift closer to equality. As shown in Figure 10, in 139 pairs the two players earn the same average payoff in the last 20 periods, which was not the case initially.

Figure 10: Towards Equality in the Long Run



Notes: Figure 10 displays the difference between the players' average payoffs in the first 20 periods and in the last 20 periods. For example, if the average payoff profile is $(2.5, 3.5)$ in the first 20 periods and $(3, 3)$ in the last 20 periods, a game symbol is displayed at coordinates $(2.5-3.5, 3-3) = (-1, 0)$. There are 139 long-run outcomes where, on average, the players earn the same.

Result 7 (Long-run movement towards equality) *For sequences that failed to converge in the first 20 periods but converged in the last 20 periods, players' average payoffs move towards equality – with some qualifications in games that require long episodes to attain the efficient and egalitarian payoff.*

4.2 Modified k-Means Clustering Approach

We now apply the clustering approach from Section 2.3 to the entire dataset. Our modified k-means algorithm groups sequences of action profiles into a small number of representative storylines. Within a game (Section 4.2.1), each storyline captures not only the outcome at which play stabilizes, but also the path by which it emerges. Across games (Section 4.2.2), the same objects can be compared under a common digitization and clustering procedure. This approach therefore reduces numerous, complex behavioral trajectories to a handful of interpretable ones that can be compared both within and across environments.

4.2.1 Illustration: Prisoner's Dilemma

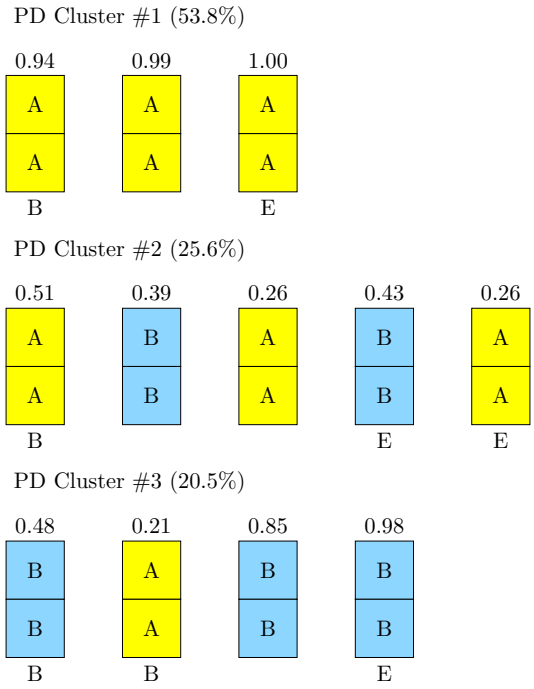
We begin with the standard Prisoner's Dilemma (PD) to illustrate how the k-means clustering approach works. Figure 11 displays the resulting clusters for PD. Recall that action A denotes 'cooperation' and action B denotes 'defection.' Each cluster is annotated with the percentage of sequences it contains, and with indicators of whether an episode was observed: at the beginning of play (with the letter 'B') based on discounted weighting, end of play (with the letter 'E') based on reverse-discounted weighting, or during intermediate periods (without a letter) based on equal weighting.[23] These annotations make it possible to trace storylines – not just endpoints.

In the baseline PD, the k-means procedure identifies three clusters:

1. **Sustained cooperation** – pairs quickly stabilize on (A, A) and maintain it throughout (see cluster #1).

[23] In the symmetric payoff matrices (i.e. PD, BO, SH, CH, CI, PD2, BO2), swapping the row player and the column player leaves the strategic environment unchanged. The same underlying dyadic behavior (for instance, when one player chooses A while the other chooses B) may appear in two orientations purely due to the arbitrary and random assignment of players to the row/column roles. The clusters are constructed from *ordered* action pairs (e.g. (A, B) versus (B, A)) – this labelling artifact can mechanically split a single behavioral class across two clusters, thereby misrepresenting its true prevalence. To ensure our findings are invariant to player assignment, we augment the dataset for symmetric games by adding a player-swapped duplicate for each observed sequence, where the players' actions are exchanged in every period. In the Figures of the symmetric games, we thus display both (flipped) clusters with, as expected, the same proportions given that they reflect equivalent behaviors (for an illustration of this, see BO cluster #2 and #3 in Figure 16). In the Supplemental Appendix, we report results based on the original, non-augmented dataset to illustrate how arbitrary role assignment can confound the emergent cluster structure.

Figure 11: k-MEANS IN THE PD GAME



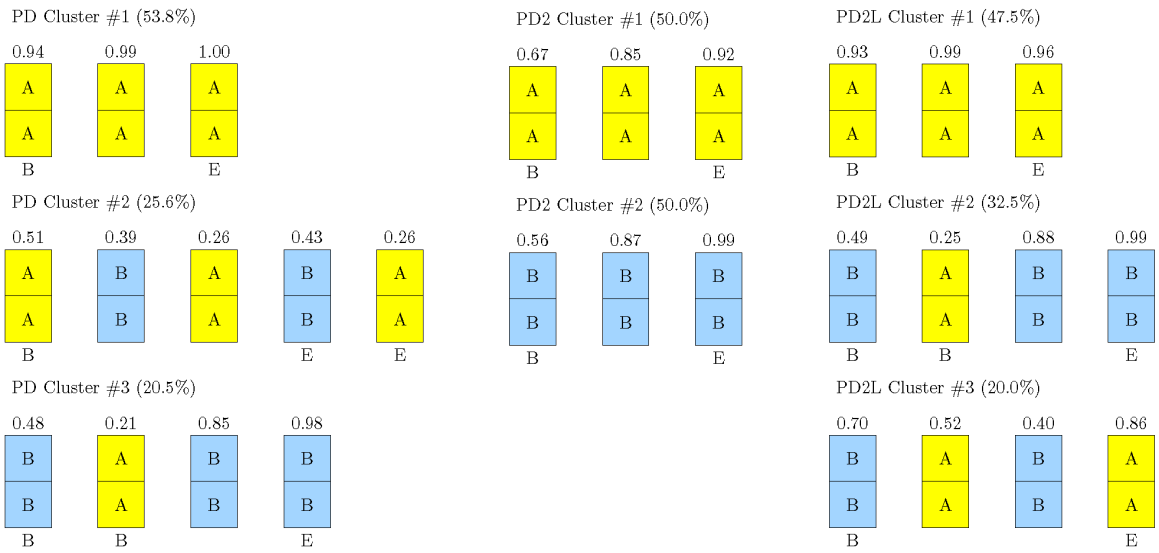
Notes: Figure 11 displays features of the clusters identified via the modified k-means approach in the PD game. The percentage in the brackets next to each cluster indicates the proportion of sequences that belong in the specific cluster. We provide, on top of the episodes, the discounted weighted frequency of matching in the beginning of game play (indicated with the letter B), during game play (without any letter), and at the end of game play (indicated with the letter E). Figures 12, 14-20 follow the same structure. Figures 14-20 are provided in the Supplemental Appendix.

2. **Attempted cooperation** – early cooperative attempts either hold briefly or, more often, unravel and play ends near (B, B) (see cluster #2).

3. **Sustained defection** – players defect almost immediately and never return to cooperation (see cluster #3).

This approach provides a vocabulary – clusters – that captures the heterogeneity of play without losing interpretability. In the repeated Prisoner's Dilemma, a small set of distinct storylines coexist. Similar analysis can be carried out in each game to identify respective storylines based on their clusters. In the context of this PD

Figure 12: k-means in the PD, PD2 and PD2L Games



Notes: Figure 12 displays features of the clusters identified via the modified k-means approach in the PD, PD2 and PD2L games. The clusters in the PD game are reproduced from Figure 11.

illustration, we can easily extend the analysis to include PD2 and PD2L (see Figure 12). In PD2, where the incentive to play B is strengthened, half of the sequences converge to cooperation and half to defection, producing a sharp division. PD2L shares commonalities with both PD and PD2: cluster #1 is nearly identical to the cooperative clusters in the other two games, and cluster #2 is nearly identical to PD cluster #3, where players end up defecting despite some initial attempt to cooperate. Importantly, in moving from PD2 to PD2L, a third storyline emerges: players defect initially but later shift to (A, A), thereby achieving a late recovery of cooperation. Our findings in the family of PD games are summarized in the following result.

Result 8 (Recurring storylines in the repeated PDs) *In the PD, PD2, and PD2L games, the modified k-means procedure identifies three main storylines:*

(1) early stabilization at cooperation (A, A)*;*

(2) early stabilization at defection (B, B)*;*

(3) late stabilization after a different start.

Across the repeated PDs, the relative weight of these storylines varies systematically with the temptation to defect and the time horizon.

The PD family therefore already exhibits two of the three cross-game families emphasized in the next Subsection: early stabilization at a singleton profile and late stabilization after a different start. What the broader comparison adds is a third family, stabilization at complex recurring episodes involving multiple profiles.

4.2.2 Cross-Game Regularities

The value of clustering extends beyond describing storylines within a given game. Across distinct setups, it yields a data-driven taxonomy of repeated-game dynamics. The key point is that, in our dataset, cross-game differences rarely take the form of entirely new behavioral patterns. Rather, outside the mixed benchmark UM, they often reflect a re-weighting of a small set of recurring storyline families.

Result 9 (Recurring storylines across games) *Outside the mixed benchmark UM, the modified k-means procedure identifies three main storyline families:*

(1) early stabilization at an efficient or egalitarian profile;

(2) late stabilization at an efficient or egalitarian profile after a different start;

(3) stabilization at a complex episode, typically egalitarian and sometimes efficient.

Across our games, the relative weight of these families varies systematically with the alignment between efficiency and egalitarianism, the time horizon, and the complexity of efficient play.

Table 3 formalizes this empirical taxonomy by mapping each game to its corresponding dynamic trajectory family. A cluster is assigned to family (1) or (2) when its dominant ending episode (under E) is a singleton efficient or egalitarian profile with within-cluster frequency at least 0.6.[24] Family (1) applies when that same singleton is also the dominant beginning episode (under B). Family (2) applies when the dominant beginning episode differs from the dominant ending episode. Given that

[24]The 0.6 threshold is chosen deliberately to reflect a robust empirical supermajority, ensuring that the designated singleton or complex episode is the unambiguously dominant feature of the cluster's behavior. This cutoff disciplines the taxonomy by balancing Type I errors (misclassifying transient patterns as stable storylines) and Type II errors (over-filtering genuine behavioral variation).

beginnings are typically noisier than endings, no parallel 0.6 cutoff is imposed on B; what matters for family (2) is that the dominant episode at the beginning differs from the dominant episode at the end. Family (3) applies when the dominant ending episode is a non-singleton recurring episode involving at least two distinct action profiles, again with within-cluster frequency at least 0.6.

Table 3: Empirical Taxonomy of Dynamic Storyline Families

Storyline Family	**Structural Alignment & Dynamics**	**Games**
Family 1: Early Stabilization	Immediate coordination on a static target. Dominant beginning and ending episodes perfectly align on a single efficient or egalitarian profile.	SH, CI, CH, PD, PD2, PD2L, SH2, SH2L
Family 2: Late Redirection	Midstream dynamic path adjustments. The dominant ending profile differs from the initial noisy trajectory, capturing late recoveries or risk retreats.	PD2L, SH2L, UL, SD3L
Family 3: Complex Intertemporal	Stabilization on a multi-period alternating episode (length $\geq$ 2) that systematically resolves payoff asymmetries or conflicts over fairness.	BO, BO2, SD2, SD3, CO, TC

Notes: Table 3 displays the empirical taxonomy (Result 9) of the 18 non-UM setups.

At the extensive margin, this parsimonious taxonomy still covers a large share of the data. Equally weighting the 18 non-UM setups, the three families account on average for 81.9% of observed sequences. Family (1) is the baseline static case and carries most of that mass, averaging 70.7% across games, while families (2) and (3) add 5.4 and 5.8 percentage points, respectively. At the intensive margin, the fit sharpens over time. Restricting attention to clusters assigned to the three families and weighting each cluster by its size, the dominant episode has an average fit of 70% at the beginning, 83.8% during play, and 92% at the end. Combining these with the 81.9% extensive coverage yields overall fit of 57.3%, 68.7%, and 75.3%, respectively.

We next analyze the specific economic and structural determinants – namely, the alignment between efficiency and egalitarianism, the interaction horizon, and complexity – that systematically govern the empirical distribution of these storylines across our games.

Alignment between efficiency and egalitarianism. This dimension determines which target profiles families (1) and (2) point towards. When one profile is both efficient and egalitarian, family (1) typically dominates. In SH, cluster #1 (85.4%) stabilizes immediately at (A, A), and in CI, clusters #1–3 together account for 96.2% and all end at (A, A). CH is only slightly looser: clusters #1 and #2 account for 81.6%, and both end at (A, A). The PD family already shows a subtler version of the same logic: both clean singleton targets are egalitarian, but only one is efficient, so family (1) splits between (A, A) and (B, B) rather than concentrating on a unique endpoint. When efficiency and egalitarianism diverge more sharply, the split becomes clearer. In SH2, cluster #1 (84.2%) is family (1) at the efficient but unequal profile (A, A), while cluster #2 (15.8%) is family (1) at the egalitarian but inefficient profile (B, B). SH2L preserves the same two endpoints, but adds late redirection between them: cluster #2 (22.7%) ends at (B, B) after an (A, A)-heavy start, and cluster #3 (9.1%) ends at (A, A) after a (B, B)-heavy start. UL shows the same tension in a richer environment: cluster #1 (52.6%) is family (1) at the equal split (A, A), cluster #2 (18.4%) is family (2) and ends at the alternative equal split (A, B) after an (A, A) start, and cluster #4 (10.5%) is family (1) at the unequal profile (B, A). SA likewise splits the clean fitted mass between the egalitarian profile (B, B) and the unfair one-shot Nash profile (A, A), while leaving the larger mixed cluster #1 outside of the taxonomy.

Time horizon. Longer horizons matter mainly because they populate family (2); that is, late stabilization after a different start. In PD2, both clusters are family-(1) clusters: cluster #1 (50%) stabilizes early at (A, A), and cluster #2 (50%) at (B, B). In PD2L, the same two family-(1) endpoints remain, but cluster #3 (20%) becomes family (2): it begins most often at (B, B) with frequency 0.7 and ends at (A, A) with frequency 0.86. The SH2/SH2L comparison is analogous but richer. SH2 consists entirely of family-(1) clusters ending at (A, A) and (B, B). SH2L retains those same endpoints, but now adds late redirection in both directions: cluster #2 ends at (B, B) with frequency 0.94 after an (A, A)-heavy start, and cluster #3 ends at (A, A) with

frequency 1 after a (B, B)-heavy start. UL cluster #2 and SD3L cluster #3 show why family (2) should not be read as recovery only from an initially inefficient path. UL cluster #2 ends at the equal split (A, B) with frequency 0.88 after an (A, A) start, and SD3L cluster #3 ends at (A, B) with frequency 0.8 after a (B, B)-heavy start. Longer horizons therefore do not merely prolong play; they create room for late redirection towards a static efficient or egalitarian target.

Complexity of efficient play. When efficient play is simple and static, the fitted mass lies almost entirely in families (1) and (2). This is the case in PD, SH, CI, CH, PD2, PD2L, SH2, and SH2L, where the clean storyline clusters end at singleton profiles. As efficient play becomes intertemporal, family (3) becomes empirically important. In BO, cluster #4 (14.3%) stabilizes at the alternating episode $(A, B), (B, A)$, and in BO2 the same complex family becomes the largest clean cluster, with cluster #1 carrying 30% and ending at that alternating episode with frequency 0.99. In SD2, cluster #2 (31.7%) stabilizes at the recurring episode $(B, A), (B, B)$ with ending frequency 0.9. In SD3, cluster #4 (13.9%) stabilizes at a longer episode built around (B, A) and (B, B) with ending frequency 0.93. In CO, cluster #4 (10%) stabilizes at a recurring episode involving (A, A) and (A, B) with ending frequency 0.86. In TC, cluster #4 (5.1%) stabilizes at the two-period episode $(A, A), (B, B)$ with ending frequency 1. These complex episodes are typically intertemporally egalitarian. In BO, BO2, SD2, and TC they are also efficient on average; in SD3 and CO they remain egalitarian on average but are not fully efficient relative to the best static profile. Complexity therefore shifts mass away from immediate singleton stabilization and toward sharply organized intertemporal episodes.

Across our games, variation in dynamic behavior primarily reflects a re-weighting of a small number of recurring storyline families rather than the emergence of new patterns. The resulting taxonomy induces a classification of games by the similarity of their dynamic outcomes. In particular, games with similar structural features tend to generate similar storyline compositions, while games that differ along these dimensions exhibit systematically different dynamics. This suggests that observed behavior in one game could inform predictions in others within the same class.

5 Extensions and Concluding Remarks

Before concluding, we briefly summarize three supplementary exercises based on the action-convergence criterion (developed fully in the Supplemental Appendix). Section D uses the additional games in Figure 3 to examine how incentives, equilibrium salience, payoff tradeoffs, and complexity shape long-run play. Section E compares the identified episodes with an independent-play benchmark, while Section F studies sensitivity to the evaluation window and mismatch tolerance.

Comparative Evidence Across Related Games

The additional comparisons sharpen several patterns from the main analysis. In the Prisoner's Dilemma family, strengthening the incentive to choose the inefficient dominant action reduces the incidence of efficient long-run outcomes, while a longer expected horizon partially offsets this effect. Alternation between the two static Nash equilibria is more common in BO2 than in BO, where those equilibria are less salient. In UL, 26 of the 30 long-run outcomes yield equal payoffs, although most do so at the payoff-equivalent profile (A, A) rather than at the equilibrium involving a weakly-dominated action. In SH2 and SH2L, where efficiency and equality point to different singleton outcomes, long-run play predominantly selects the efficient but unequal profile. Finally, the Samaritan's Dilemma variants show that multi-profile episodes can reconcile efficiency and equality, but such outcomes are less common when their implementation requires a longer episode. This last comparison is consistent with, although it does not separately identify, a role for implementation complexity.

Independent-Play Benchmark

Section E asks whether the episodes identified by the action-convergence criterion could plausibly arise if action profiles were drawn independently across periods according to their empirical frequencies. For singleton episodes, the relevant probabilities are computed directly from the binomial distribution. For multi-profile episodes, we use an exact Dynamic Programming algorithm that accounts for overlapping comparisons, mismatches at different positions, and realignment following a mismatch. Conditional on the identified episodes, the resulting benchmark p-values are below 0.001. The exercise therefore shows that independent period-by-period play is an implausible account of the reported regularities. It does not, by itself, distinguish

intentional coordination from learning, persistence, or other sources of serial dependence.[25]

Sensitivity

The baseline analysis permits two mismatches within the first and last 20 periods of play. Section F varies the number of permitted mismatches between one and three and the evaluation window between 15 and 25 periods. Exact episode assignments and counts vary at the margin, but the central qualitative patterns remain: convergence is substantially more prevalent late in the interaction than early, and multi-profile episodes are identified more often at the end of play.

Concluding Remarks

Repeated games provide a natural model of relationships sustained by future interaction. Most experimental evidence, however, comes from subjects playing a succession of relatively short matches. We instead follow the same pair through a single, unusually long interaction. The paper develops two nonparametric approaches to one central question: how does behavior evolve within such a relationship? The action-convergence criterion identifies recurring episodes and when they emerge, while the modified k-means method organizes complete paths of play into recurring trajectory families. Both methods are model-light, but they recover different dimensions of behavioral organization.

The action-convergence analysis turns the theoretical multiplicity of repeated-game outcomes into an empirical selection question. Long-run play places substantial weight on a relatively small set of episodes, recurring play becomes more prevalent with experience, and multi-profile arrangements become more common as interactions mature. Efficiency and equality help organize the outcomes that emerge, although their influence depends on incentives and on the complexity required to implement them. These findings also highlight a distinction between the formal game and the relationship that develops within it. The actions, payoffs, and continuation probability remain fixed, but the relationship accumulates a history. SPNE can accommodate the resulting history dependence, yet it does not privilege the observed movement from dispersion towards stability, or from simpler towards more organized play, over

[25] Given that the candidate episodes and the null frequencies are estimated from the same data, these calculations are best interpreted as conditional benchmark comparisons rather than as tests that adjust for the episode-mining step.

the reverse transitions.

The clustering analysis makes a distinct contribution by studying complete trajectories rather than selected early and late windows. Across games, the paths of play organize into a small number of recurring families: early stabilization, late redirection, and stabilization at complex intertemporal episodes. Much of the variation across strategic environments reflects changes in the relative prevalence of these families rather than the appearance of wholly new forms of behavior. The resulting taxonomy captures not only where subjects end up, but also how they get there, and provides a common basis for comparing dynamic play across games. Its contribution is descriptive; establishing whether the taxonomy predicts behavior in new games is a separate question requiring out-of-sample evidence.

Together, the two methods characterize both the episodes that acquire empirical weight and the trajectories through which they emerge. They do not identify a unique underlying strategy or deliver structural counterfactuals. Instead, they uncover regularities that theories of learning, reputation, and equilibrium selection should seek to explain. Understanding long repeated interactions requires studying not only which paths can be supported, but also how experience redistributes behavior across them. The formal game repeats; the relationship evolves.

Beyond describing behavioral trajectories, our framework suggests a route towards prediction. If the taxonomy proves stable out of sample, it could be used to forecast play in related games and to compare institutional designs before costly laboratory or field implementation. Sequential pattern recognition could thus complement experiments by helping researchers screen and refine mechanisms before testing them in practice.

References

Agrawal, Rakesh and Ramakrishnan Srikant. 1995. "Mining Sequential Patterns." In *Proceedings of the* 11^{th} *International Conference on Data Engineering*. 3–14.

Amir, Ofra, David G. Rand, and Ya'akov Kobi Gal. 2012. "Economic Games on the Internet: The Effect of $1 Stakes." *PLoS ONE* 7 (2):e31461.

Aoyagi, Masaki, V. Bhaskar, and Guillaume R. Fréchette. 2019. "The Impact of Monitoring in Infinitely Repeated Games: Perfect, Public, and Private." *American Economic Journal: Microeconomics* 11 (1):1–43.

Aoyagi, Masaki and Guillaume R. Fréchette. 2009. "Collusion as Public Monitoring Becomes Noisy: Experimental Evidence." *Journal of Economic Theory* 144 (3):1135–65.

Arechar, Antonio A, Simon Gächter, and Lucas Molleman. 2018. "Conducting Interactive Experiments Online." *Experimental Economics* 21 (1):99–131.

Arechar, Antonio A., Gordon T. Kraft-Todd, and David G. Rand. 2017. "Turking Overtime: How Participant Characteristics and Behavior Vary Over Time and Day on Amazon Mechanical Turk." *Journal of the Economic Science Association* 3 (1):1–11.

Bathoorn, Rob, Monique Welten, Michael K. Richardson, Arno Siebes, and Fons J. Verbeek. 2010. "Frequent Episode Mining to Support Pattern Analysis in Developmental Biology." In *Pattern Recognition in Bioinformatics (PRIB 2010)*, *Lecture Notes in Computer Science*, vol. 6282, edited by Paulo G. Ferreira and André C. P. L. F. Carvalho. Springer, 220–231.

Blonski, Matthias, Peter Ockenfels, and Giancarlo Spagnolo. 2011. "Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence." *American Economic Journal: Microeconomics* 3 (3):164–92.

Boyer, Robert S. and J. Strother Moore. 1977. "A Fast String Searching Algorithm." *Communications of the ACM* 10 (20):762–72.

Chen, Daniel L., Schonger Martin, and Chris Wickens. 2016. "oTree – An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9:88–97.

Dal Bó, Pedro. 2005. "Cooperation Under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *American Economic Review* 95 (5):1591–604.

Dal Bó, Pedro and Guillaume R. Fréchette. 2011. "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." *American Economic Review* 101 (1):411–29.

———. 2018. "On the Determinants of Cooperation in Infinitely Repeated Games: A Survey." *Journal of Economic Literature* 56 (1):60–114.

Fudenberg, Drew and David Levine. 1989. "Reputation and Equilibrium Selection in Games with a Patient Player." *Econometrica* 57 (4):759–78.

Fudenberg, Drew, David G. Rand, and Anna Dreber. 2012. "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World." *American Economic Review* 102 (2):720–49.

Gill, David and Yaroslav Rosokha. 2024. "Beliefs, Learning, and Personality in the Indefinitely Repeated Prisoner's Dilemma." *American Economic Journal: Microeconomics* 16 (3):259–83.

Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26 (3):213–24.

Hartigan, John A. 1975. *Clustering Algorithms*. New York: Wiley.

Hartigan, John A. and Manchek A. Wong. 1979. "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society* 28 (1):100–8.

Hauser, David J. and Norbert Schwarz. 2016. "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants." *Behavior Research Methods* 48:400–7.

Hertz, Gerald Z. and Gary D. Stormo. 1999. "Identifying DNA and Protein Patterns with Statistically Significant Alignments of Multiple Sequences." *Bioinformatics* 15 (7):563–77.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. "The Online Laboratory: Conducting Experiments in a Real Labor Market." *Experimental Economics* 14 (3):399–425.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596 (7873):583–9.

Kayaba, Yutaka, Hitoshi Matsushima, and Tomohisa Toyama. 2016. "Accuracy and Retaliation in Repeated Games with Imperfect Private Monitoring: Experiments and Theory." Mimeo.

Knuth, Donald, James H. Morris, and Vaughan Pratt. 1977. "Fast Pattern Matching in Strings." *SIAM Journal on Computing* 2 (6):323–50.

Lambrecht, Marco, Eugenio Proto, Aldo Rustichini, and Andis Sofianos. 2024. "Intelligence Disclosure and Cooperation in Repeated Interactions." *American Economic Journal: Microeconomics* 16 (3):199–231.

Low-Kam, Cécile, Chedy Raïssi, Mehdi Kaytoue, and Jian Pei. 2013. "Mining Statistically Significant Sequential Patterns." In *2013 IEEE* 13^{th} *International Conference on Data Mining*. Dallas, TX, 488–97.

MacQueen, James B. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*. Berkeley, CA: University of California Press, 281–97.

Mannila, Heikki, Hannu Toivonen, and Inkeri Verkamo. 1997. "Discovery of Frequent Episodes in Event Sequences." *Data Mining and Knowledge Discovery* 1:259–89.

Mathevet, Laurent. 2018. "An Axiomatization of Plays in Repeated Games." *Games and Economic Behavior* 110:19–31.

Patnaik, Debprakash, P. S. Sastry, and K. P. Unnikrishnan. 2008. "Inferring Neuronal Network Connectivity from Spike Data: A Temporal Data Mining Approach." *Scientific Programming* 16 (1):49–77.

Proto, Eugenio, Aldo Rustichini, and Andis Sofianos. 2019. "Intelligence, Personality and Gains from Cooperation in Repeated Interactions." *Journal of Political Economy* 127 (3):1351–90.

———. 2022. "Intelligence, Errors and Cooperation in Repeated Interactions." *The Review of Economic Studies* 89 (5):2723–67.

Randolph Bruns, Bryan. 2015. "Names for Games: Locating 2x2 Games." *Games* 6:495–520.

Robinson, David and David Goforth. 2005. *The Topology of the 2x2 Games: A New Periodic Table*. Routledge, London.

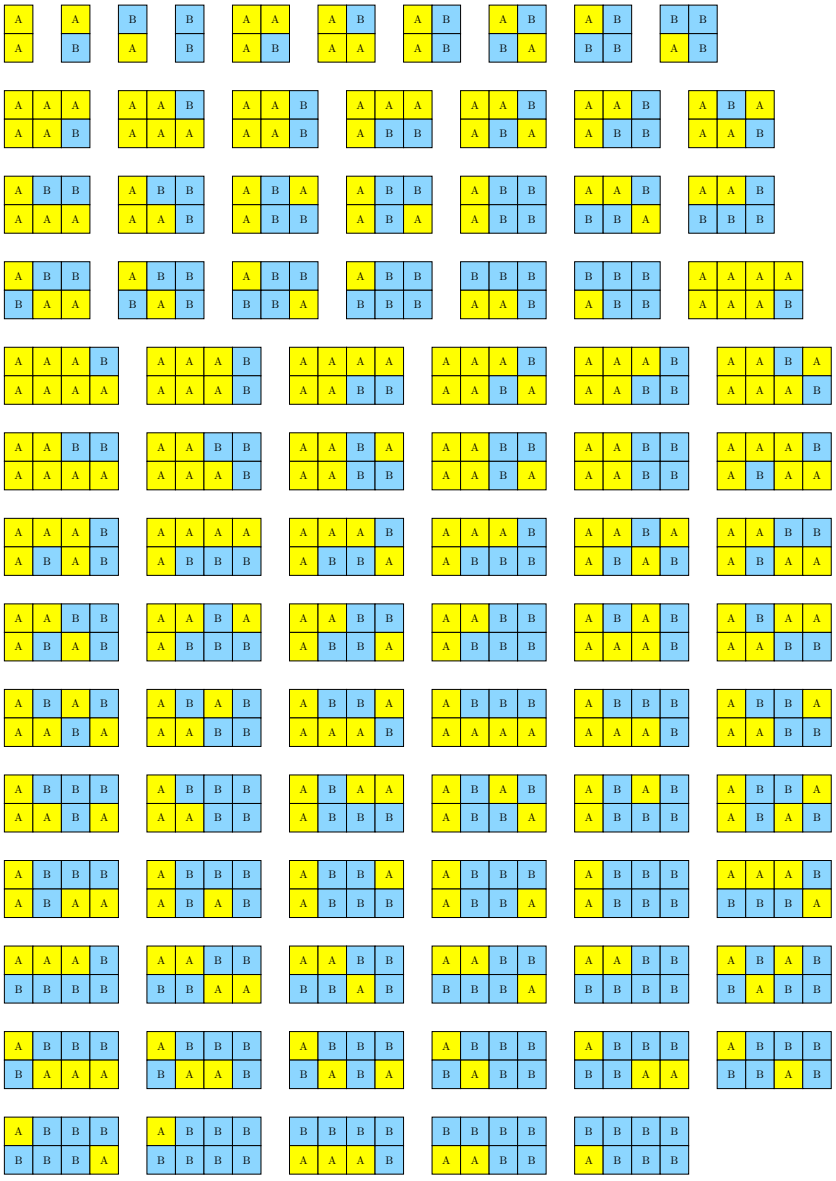
Satopaa, Ville, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior." 31^{st} International Conference on Distributed Computing Systems Workshops. 166–71.

Smith, Vernon L. 1994. "Economics in the Laboratory." *Journal of Economic Perspectives* 8 (1):113–31.

SUPPLEMENTAL APPENDIX

A Episodes

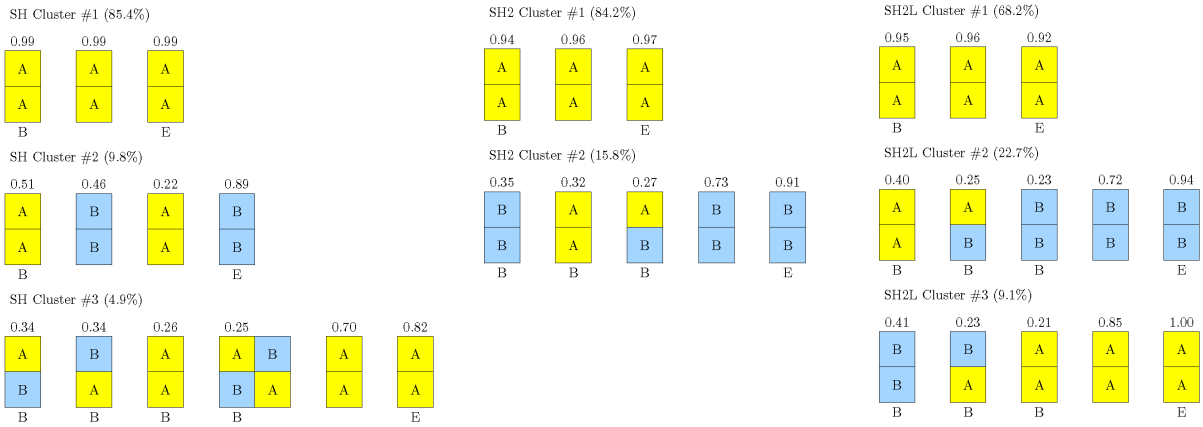
Figure 13: Episodes of Length ≤ 4



Notes: In the analysis, we focussed on episodes of length 6 or less, totalling 964 episodes. Episodes of length 5 and 6 are not displayed due to space constraints.

B k-Means Clustering in Other Games

Figure 14: k-means in the SH, SH2 and SH2L Games

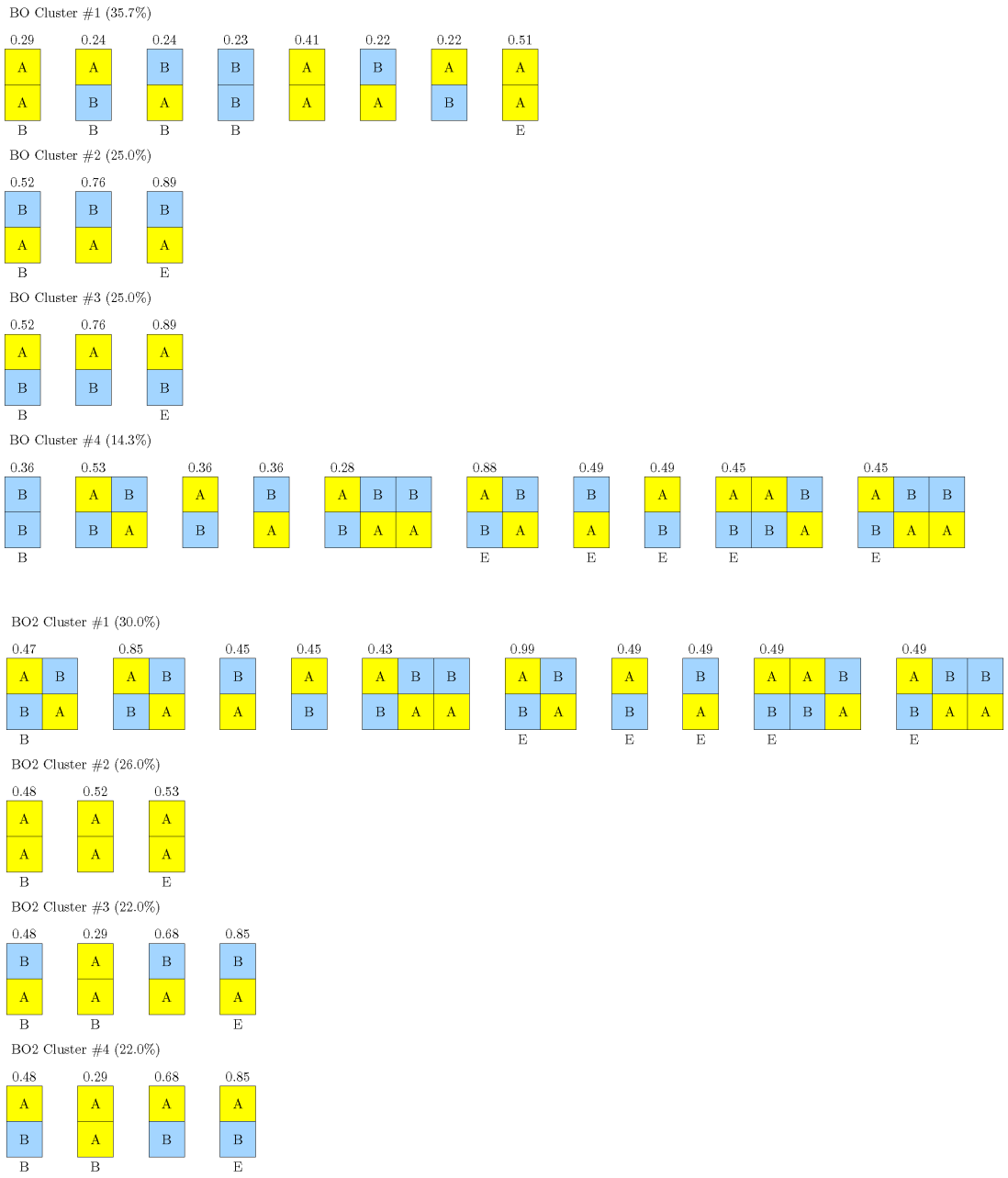


Notes: Figure 14 displays features of the clusters identified via the modified k-means approach in the SH, SH2 and SH2L games.

Figure 15: k-means in the SD, SD2, SD3 and SD3L Games

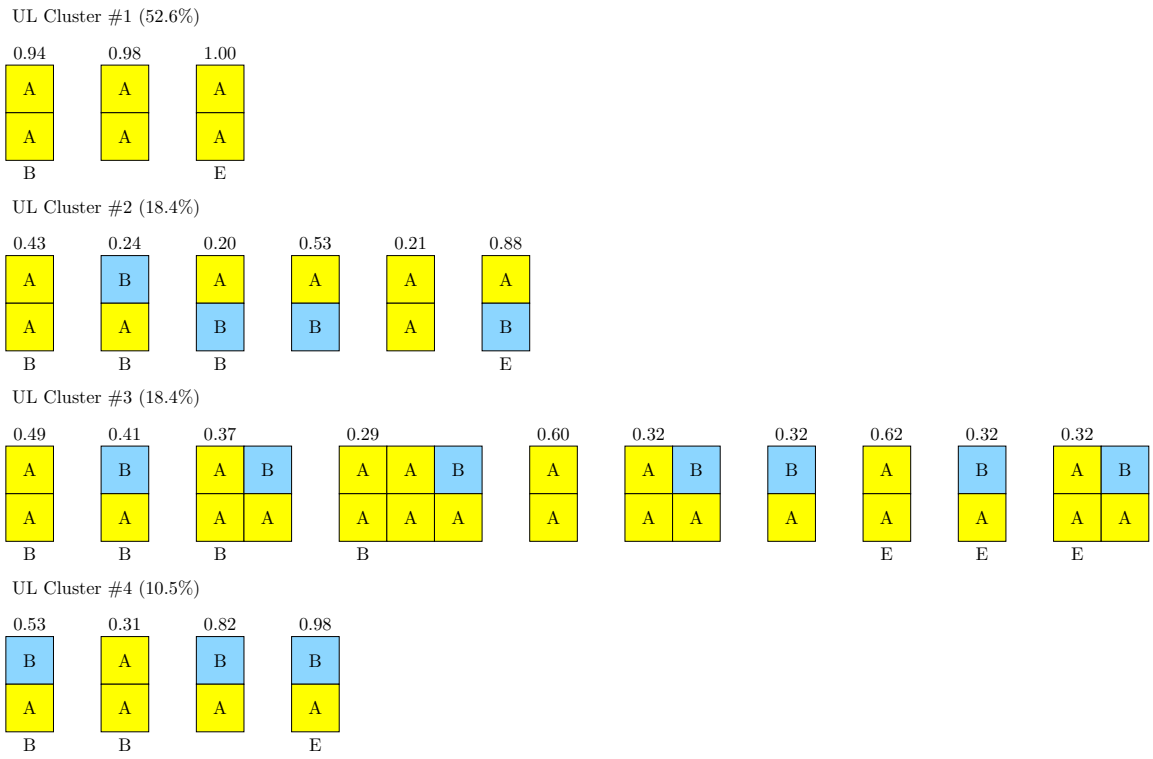
Notes: Figure 15 displays features of the clusters identified via the modified k-means approach in the SD, SD2, SD3 and SD3L games.

Figure 16: k-means in the BO and BO2 Games



Notes: Figure 16 displays features of the clusters identified via the modified k-means approach in the BO and BO2 games.

Figure 17: k-means in the UL Game

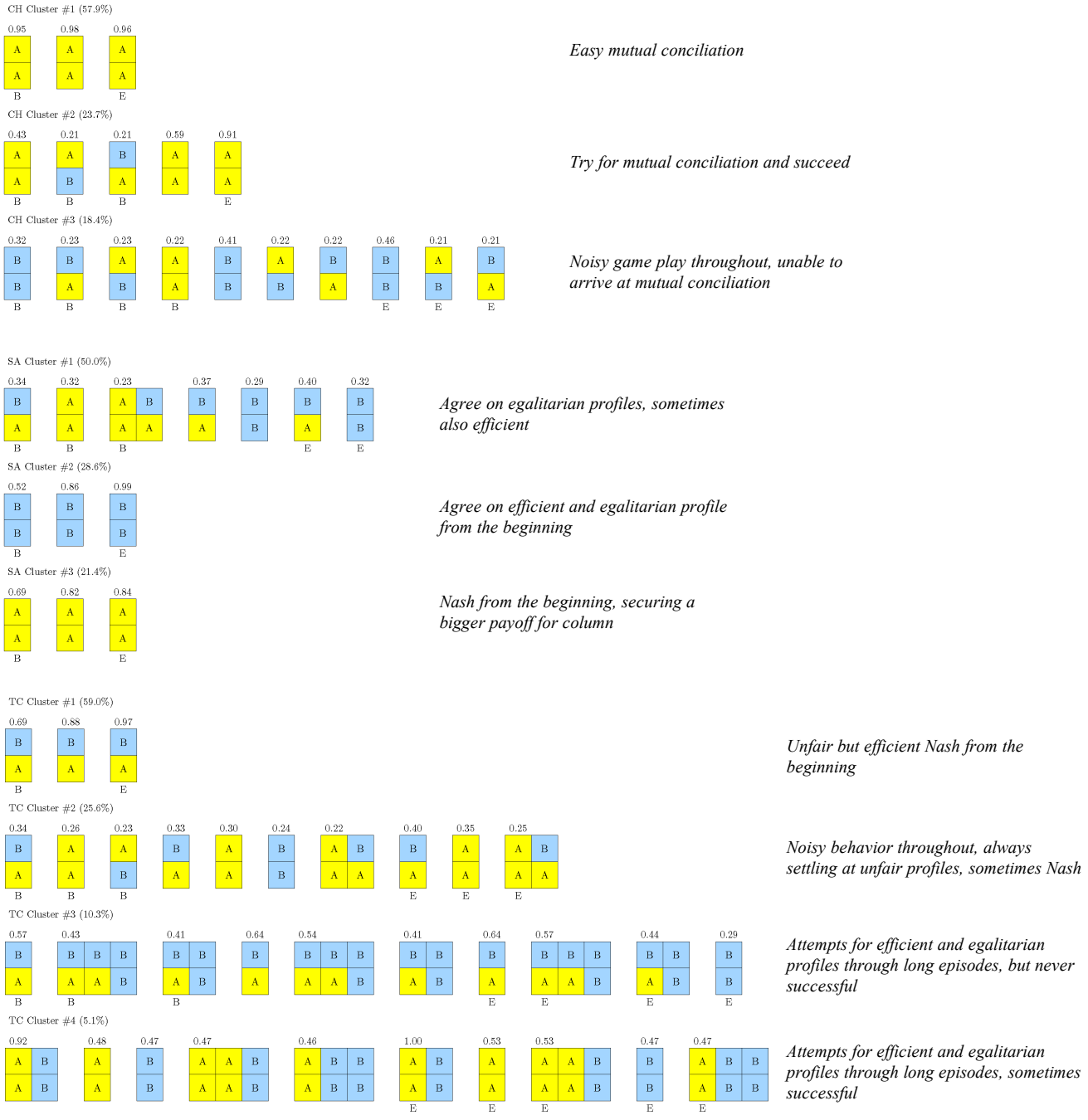


Notes: Figure 17 displays features of the clusters identified via the modified k-means approach in the UL game.

Figure 18: k-means in the UM Game

Notes: Figure 18 displays features of the clusters identified via the modified k-means approach in the UM game.

Figure 19: k-means in the CH, SA and TC Games



Notes: Figure 19 displays features of the clusters and their storylines as identified via the modified k-means approach in the CH, SA and TC games.

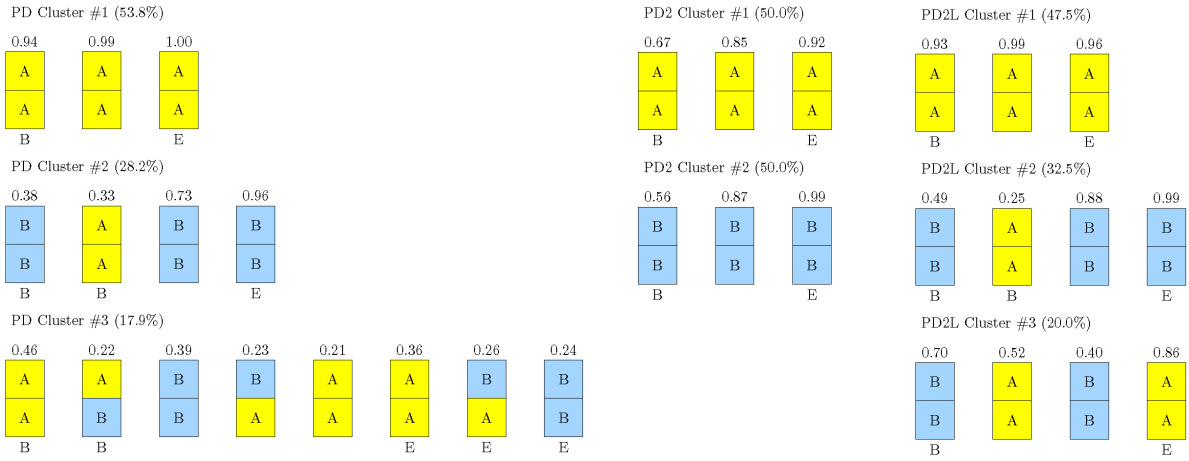
Figure 20: k-means in the CI and CO Games

Notes: Figure 20 displays features of the clusters and their storylines as identified via the modified k-means approach in the CI and CO games.

C k-Means Clustering in Symmetric Games Without Player-Swap Augmentation

Recall that in the symmetric payoff matrices (i.e. PD, BO, SH, CH, CI, PD2, BO2), swapping the row player and the column player leaves the strategic environment unchanged. The clusters are constructed though from ordered action pairs (e.g. (A,B) versus (B,A)); hence, a single behavioral class is split into two clusters depending on the arbitrary and random role assignment of players. As a result the prevalence of one single behavioral class would be understated. To ensure our findings are invariant to player assignment, in the analysis in the main text, we augmented the dataset for symmetric games by adding a player-swapped duplicate for each observed sequence, where the players' actions are exchanged in every period. We report next results based on the original, non-augmented dataset to illustrate how arbitrary role assignment can confound the emergent cluster structure.

Figure 21: k-means in the PD, PD2 and PD2L Games

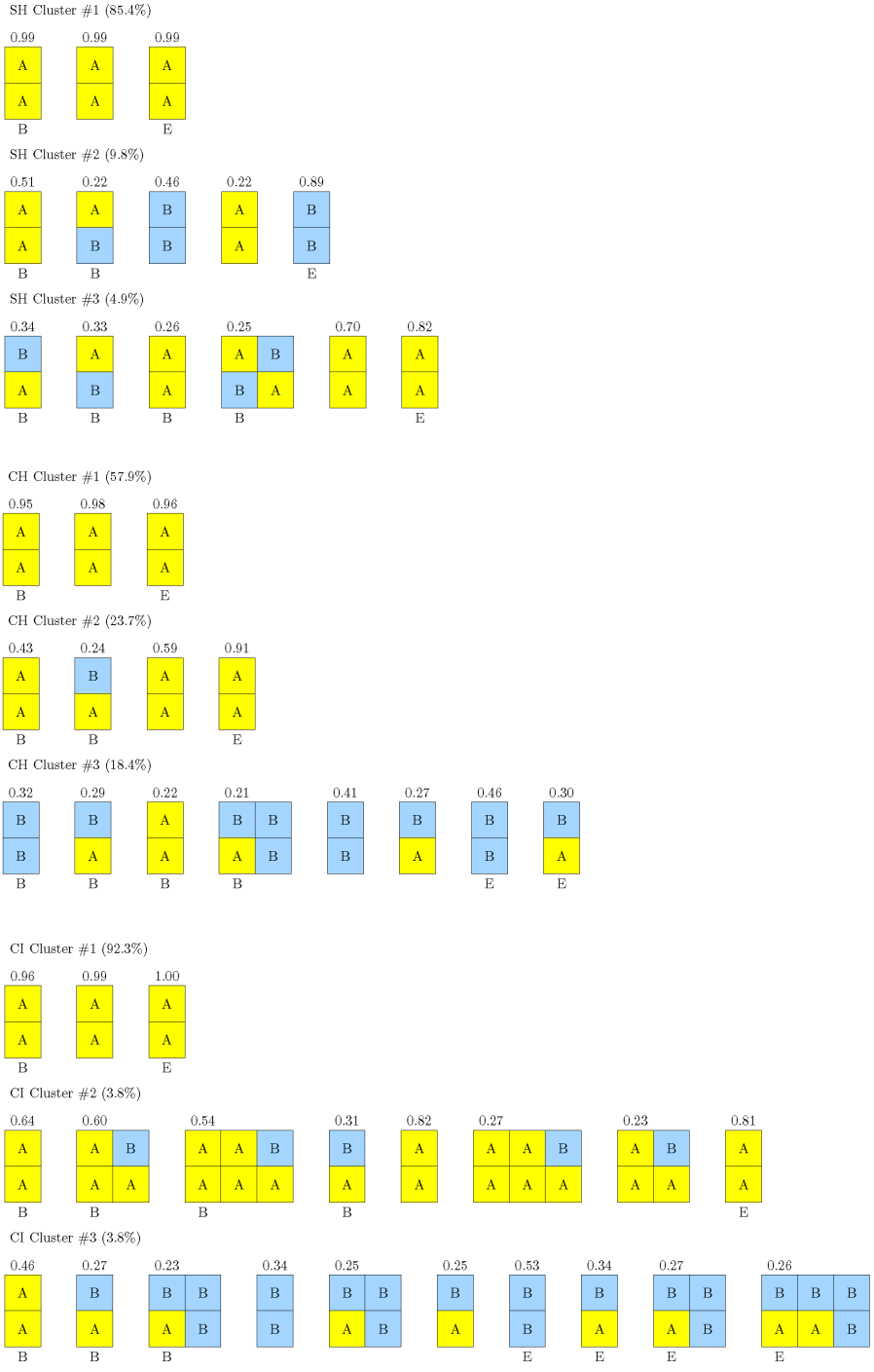


Notes: Figure 21 displays features of the clusters identified via the modified k-means approach, without the player-swap augmentation, in the PD, PD2 and PD2L games.

Figure 22: k-means in the BO and BO2 Games

Notes: Figure 22 displays features of the clusters identified via the modified k-means approach, without the player-swap augmentation, in the BO and BO2 games.

Figure 23: k-means in the SH, CH and CI Games



Notes: Figure 23 displays features of the clusters as identified via the modified k-means approach, without the player-swap augmentation, in the SH, CH and CI games.

D Comparative Evidence Across Related Games

This section uses the action-convergence criterion to compare games that differ in the incentives to choose a dominant action, the salience of static equilibria, the alignment of efficiency and equality, and the complexity of implementing an efficient and egalitarian outcome.

Dominant-action incentives We first compare PD with PD2 and PD2L. Relative to PD, PD2 lowers the sucker's payoff and therefore strengthens the incentive to choose the inefficient dominant action. Among long-run outcomes, the share that are efficient falls from 69% in PD (22/32)to 46% in PD2 (17/37). With the longer expected horizon in PD2L, the corresponding share rises to 67% (24/36). Stronger incentives to defect therefore reduce long-run efficiency, while a longer interaction offsets part of this reduction.

Equilibrium salience We next compare BO with BO2, in which the two static Nash equilibria are made more salient. The share of long-run outcomes that alternate between these equilibria rises from 20% in BO (5/25)to 47% in BO2 (14/30). Thus, alternation is more common when the static equilibria are more sharply distinguished from unilateral deviations.[26]

Weak dominance and equality UL has two static Nash equilibria. At (A, B) the column player chooses a weakly-dominated action and both players receive 2. At (B, A), neither player chooses a dominated action, but the payoff vector is $(3, 1)$. The unequal equilibrium occurs in only 4 of the 30 long-run outcomes. The remaining 26 yield equal payoffs: 5 occur at (A, B), and 21 at the payoff-equivalent profile (A, A). Long-run play in UL is therefore concentrated on the egalitarian payoff vector, although it need not implement the corresponding equilibrium action profile.

Result 10 (Incentives, salience, and equality)

(i) When a strictly dominant outcome is inefficient, stronger incentives to play the dominant action lower the incidence of efficient outcomes. However, when interactions last longer on average, efficiency partially recovers.

[26] A NE a^* is more salient in one game than in another if the loss from deviating from a^* (as defined in the main text) is larger in the former game.

(ii) Alternations between static Nash equilibria increase when they become more salient.

(iii) An egalitarian payoff vector that might not implement the payoff-equivalent static Nash equilibrium in a weakly-dominated action, is chosen over an unfair static Nash Equilibrium.

We next consider environments in which efficiency and equality point to different outcomes. In SH2 and SH2L, (A, A) is efficient but unequal, whereas (B, B) is egalitarian but Pareto inefficient. A third possibility is the three-period alternation of action profiles (A, A), (A, A) and (A, B), which yields expected payoffs of $(8/3, 8/3)$. This episode Pareto dominates (B, B), but it remains inefficient relative to (A, A). The observed outcomes favor efficiency. In SH2, 29 of the 33 long-run outcomes are efficient but unequal, while 4 are egalitarian but inefficient. The corresponding counts in SH2L are 31 and 8. Moreover, every long-run outcome in these two games is a singleton episode. Extending the expected horizon therefore does not induce players to use a multi-period compromise when efficiency and equality remain in conflict.

The Samaritan's Dilemma variants provide a useful contrast because, in these games, alternating play can produce outcomes that are both efficient and egalitarian. In SD, only two long-run outcomes implement the relevant two-period alternation. In SD2, where that alternation is more salient, 10 of the 20 long-run outcomes implement it. In SD3 and SD3L, attaining both objectives requires the more complex three-period episode $((B, A)$, (B, A) and $(B, B))$, which yields average payoffs of $(10/3, 10/3)$. This episode occurs in 4 of the 22 long-run outcomes in SD3, and in 3 of the 26 long-run outcomes in SD3L. Thus, some pairs coordinate successfully on the longer episode, but its prevalence is lower than that of the corresponding two-period episode in SD2.[27]

Result 11 (Efficiency, equality, and complexity)

(i) When efficiency and equality point to different outcomes, long-run play predominantly selects the efficient outcome.

(ii) Efficient and egalitarian long-run outcomes are less frequent when their implementation requires a longer episode.

[27]Relative to SD2, in SD3 and SD3L, the egalitarian profiles (A, A) and (A, B) are less attractive.

(iii) Increasing the expected horizon does not materially alter (i) and (ii).

E Independent-Play Benchmark

We provide next statistical analysis using dynamic programming techniques to assess the validity of the frequent episodes identified via the action-convergence criterion. Recent works in the frequent-pattern-mining framework connect episode discovery to traditional, statistical-hypothesis testing. Indeed, a frequent episode may not be meaningful if it is statistically expected; conversely, an infrequent episode may be noteworthy if it is statistically unexpected (Low-Kam et al. (2013)). Given a 0.9-frequent episode $\hat{p}$ from Table 2, we seek to determine the probability that k out of n sequences 0.9-converge to $\hat{p}$ under the null hypothesis of independent play across periods. This probability, or p-value, helps assess whether the observed occurrences of $\hat{p}$ are likely due to chance or indicate intentional coordination. To compute the p-value, we restrict our analysis to sequences of the last 20 periods, where each action profile follows its empirical frequency independently across periods.

Consider, for example, the length-1 episode (B, B) in the PD game. In the last 20 periods of all PD plays, the empirical distributions of the four action profiles (A, A), (A, B), (B, A), and (B, B) are 0.6, 0.04, 0.04, and 0.32, respectively. The probability of a single sequence 0.9-converging to the length-1 episode $\hat{p} = (B, B)$ is given by

$$\alpha(\hat{p}, 2) = \sum_{x=18}^{20} \binom{20}{x} P(\hat{p})^x (1 - P(\hat{p}))^{20-x}, \tag{2}$$

where the difference in the summation bounds is based on the allowed number of errors. Substituting $P(\hat{p}) = 0.32$, we obtain $\alpha((B, B), 2) = 1.66 \times 10^{-10}$. The probability of observing at least $k = 10$ convergent sequences in $N = 39$ (see Table 2) is then

$$p\text{-value } (\hat{p}, N, k) = \sum_{i=k}^{N} \binom{N}{i} (\alpha(\hat{p}, 2))^i (1 - \alpha(\hat{p}, 2))^{N-i}, \tag{3}$$

which yields an extremely small p-value (i.e. 1.02×10^{-89}).

While computing p-values for length-1 episodes is straightforward, the complexity increases significantly for episodes of greater length. Consider the length-2 episode $\hat{p} = ((A, B), (B, A))$ in the BO game, where the empirical relative frequencies are: $P(A, A) = 0.21, P(A, B) = 0.28, P(B, A) = 0.46, P(B, B) = 0.05$. Errors can occur at any position in the sequence; furthermore, after an error, alignment with the episode may restart from any action profile (recall Subsection 2.1). This dynamic

aspect makes direct enumeration infeasible beyond length-1 episodes.

To accurately compute $\alpha(\hat{p}, 2)$ for episodes longer than 1, we employ Dynamic Programming (DP). Once $\alpha(\hat{p}, 2)$ is determined, the corresponding p-value can be computed using equation (3).[28] The DP approach efficiently accounts for all possible ways a sequence can align with an episode while considering allowed errors and potential realignments to maximize matching opportunities. By systematically handling dependencies and permitting flexible restarts after an error, DP ensures precise and scalable probability estimation for episodes of varying lengths and complexities.[29] Applying the DP method, we find that $\alpha((A, B), (B, A), 2) = 6.08 \times 10^{-7}$ in the BO game. Thus, the p-value of observing at least 5 convergent sequences $((A, B), (B, A))$ in 42 plays is 7.07×10^{-26}.

Based on all of our computations, we conclude, perhaps unsurprisingly, that the frequent episodes mined in Table 2 are highly unlikely to have resulted from independent action choices across periods ($p < 0.001$). Instead, the identified episodes exhibit strong intertemporal dependence that is inconsistent with independent period-by-period play.

E.1 Dynamic Programming

We consider a finite set of action profiles,

$$\mathcal{A} \;=\; \{(A, A),\, (A, B),\, (B, A),\, (B, B)\},$$

each occurring i.i.d. across periods with probabilities

$$p_{(A,A)}, \quad p_{(A,B)}, \quad p_{(B,A)}, \quad p_{(B,B)},$$

respectively.

Let $\hat{p} = (\hat{p}^1 \hat{p}^2 \dots \hat{p}^n)$ be a finite *episode* of length $n \leq 6$. We denote the *equivalence*

[28] Monte Carlo simulation can approximate these probabilities, but its effectiveness diminishes due to the exceedingly low probability of convergence. A prohibitively large number of simulations would then be required to achieve reliable accuracy, making DP computationally more efficient. However, if we only need to check whether this probability exceeds a given threshold, Monte Carlo simulation remains a viable option.

[29] A detailed description of the DP approach is provided next.

class of $\hat{p}$ as

$$[\hat{p}] := \big\{ q \in \mathcal{A}^{|\hat{p}|} : \text{there exists some } 0 < \ell < n \text{ such that } \hat{p}^k = q^{\bmod_{|\hat{p}|}(k+\ell)} \text{ for all } 1 \le k \le |\hat{p}| \big\}.$$

An equivalence class thus contains all permutations of the same elements that preserve the order between them. Given $T = 20$ periods, we say a sequence *s 0.9-converges* to $\hat{p}$ if it has at most 2 errors.

Goal
Compute the exact probability

$$\alpha(\hat{p}, 2) \;=\; \Pr\big(\text{a random i.i.d. } T = 20 \text{ sequence has} \le 2 \text{ errors w.r.t. } [\hat{p}]\big).$$

If $\alpha(\cdot, \cdot)$ is extremely small, Monte Carlo methods become inaccurate. Instead, we use the following Dynamic Programming (DP) approach, which is exact and handles overlaps between consecutive windows of length n.

E.1.1 Pseudocode

Key Idea
At each period $t = 0, \ldots, T$ keep track of

- the last $n - 1$ profiles (the *partial overlap*),
- the current error count e,

so that when we *add* a new profile at time $t + 1$, we form a new window of length n and can check if it belongs to $[\hat{p}]$.

Algorithmic Pseudocode
Let $W = T - n + 1$ be the number of potential errors (windows). We maintain a DP table DP[t] for $t = 0, \ldots, T$ where DP[t](*partialseq*, e) is the probability mass of having read t profiles so far, ending with the $(n - 1)$-profile *partialseq*, and having accrued e errors up to that point. The pseudocode is given next.

Algorithm 1 DP for Exact Probability of 0.9-Convergence to $\hat{p}$ in a $T = 20$ Sequence

Input:

- Episode $\hat{p} = (\hat{p}^1 \dots \hat{p}^n)$ in $\mathcal{A}^n$ with length $n \leq 6$.
- Probabilities $p_{(A,A)}, p_{(A,B)}, p_{(B,A)}, p_{(B,B)}$ for each period (i.i.d.).
- Total length $T = 20$.

Compute: the set $[\hat{p}]$ of all permutations of $\hat{p}$ stored as Python-tuples for quick membership checks.
Initialize DP[0](($\,$), 0) = 1.0 (probability mass = 1 when no profiles have been read), and all other states = 0.
for $t = 0$ **to** $T - 1$ **do**
  Create a fresh map DP$[t+1]$ set to zero everywhere.
  for all states (*partialseq*, e) **with** DP$[t]$(*partialseq*, e) > 0 **do**
    let α = DP$[t]$(*partialseq*, e)
    if $e \leq 2$ **then**
      for all nextprofile $\in \{(A,A),(A,B),(B,A),(B,B)\}$ **do**
        probdraw = $p_{\text{nextprofile}}$
        extended = *partialseq* + [nextprofile]
        if $|$*extended*$| < n$ **then**
          DP$[t+1]$(*extended*, e) += $\alpha \times$ *probdraw*
        else
          if *extended* $\in [\hat{p}]$ **then**
            $e' := e$
          else
            $e' := e + 1$
          end if
          newpartial := *extended*$[1 : n]$ (drop the first of the n profiles)
          DP$[t+1]$(*newpartial*, e') += $\alpha \times$ *probdraw*
        end if
      end for
    end if
  end for
end for
Output:

$$p = \sum_{(partialseq,\, e)} \Big[\text{DP}[T](partialseq,\, e) \Big] \quad \text{where } e \leq 2.$$

E.1.2 Illustrative Example

Below is a step-by-step illustration of how this approach works. We choose the simplest case with at most one error and compute the probability that a sequence of only 3 periods converges to the episode $\hat{p} = \big((A,B),(B,A)\big)$ with at most 1 error.
The equivalence class of $\hat{p}$ is

$$[\hat{p}] = \Big\{\big((A,B),(B,A)\big),\ \big((B,A),(A,B)\big)\Big\}.$$

An error occurs at position t if the next two consecutive draws $\big(s^t, s^{t+1}\big)$ do *not* match any permutation in $[\hat{p}]$. Suppose further that each period $s^t \in \{(A,A),(A,B),(B,A),(B,B)\}$ occurs i.i.d. with empirical probabilities

$$p_{(A,A)} = 0.21, \quad p_{(A,B)} = 0.28, \quad p_{(B,A)} = 0.46, \quad p_{(B,B)} = 0.05.$$

Dynamic Programming States

Define the dynamic programming (DP) table:

$$\mathrm{DP}[t]\big(\mathit{partialseq}, e\big),$$

where

- t: number of action profiles drawn so far ($0 \le t \le 3$),
- *partialseq*: the last $n-1 = 1$ profile observed (or empty if fewer than 1 has been read)
- e: the total error count ($0 \le e \le 1$).

The DP is initialized as

$$\mathrm{DP}[0]\big((), 0\big) = 1, \quad \text{all other states} = 0.$$

We update the DP for $t = 0 \to 1$, then $t = 1 \to 2$, and finally $t = 2 \to 3$.

DP Updates

Step 1: From $t = 0$ to $t = 1$

We draw the first action profile. Each profile becomes the new *partialseq* without causing an error since no length-n block is formed yet. Thus, we have

$$\begin{aligned}
\mathrm{DP}[1]\big((A, A), 0\big) &= 0.21, \\
\mathrm{DP}[1]\big((A, B), 0\big) &= 0.28, \\
\mathrm{DP}[1]\big((B, A), 0\big) &= 0.46, \\
\mathrm{DP}[1]\big((B, B), 0\big) &= 0.05.
\end{aligned}$$

All other DP[1] states remain zero.

Step 2: From $t = 1$ to $t = 2$

At $t = 1$, we have four possible 'last profiles' with no errors ($e = 0$)

Last Profile	Probability
(A, A)	0.21
(A, B)	0.28
(B, A)	0.46
(B, B)	0.05

Each of the 16 possible pairs (first, second) is formed by multiplying the two probabilities. The only 'matched' blocks are

$$(A, B) \to (B, A): \quad 0.28 \times 0.46 = 0.1288, \quad (B, A) \to (A, B): \quad 0.46 \times 0.28 = 0.1288.$$

Thus the no-error states at $t = 2$ are

$$\begin{aligned}
\mathrm{DP}[2]\big((B, A), 0\big) &= 0.1288, \\
\mathrm{DP}[2]\big((A, B), 0\big) &= 0.1288.
\end{aligned}$$

All other 14 cases with 1 error, summing to $1 - 0.2576 = 0.7424$, are distributed by the ending profile

$$\begin{aligned}
\mathrm{DP}[2]\big((A,A),1\big) &= 0.21,\\
\mathrm{DP}[2]\big((A,B),1\big) &= 0.28 - 0.1288 = 0.1512,\\
\mathrm{DP}[2]\big((B,A),1\big) &= 0.46 - 0.1288 = 0.3312,\\
\mathrm{DP}[2]\big((B,B),1\big) &= 0.05.
\end{aligned}$$

Therefore, at $t = 2$, we have these six surviving states

State	**Prob.**	e
$\mathrm{DP}[2]\big((B,A),0\big)$	0.1288	0
$\mathrm{DP}[2]\big((A,B),0\big)$	0.1288	0
$\mathrm{DP}[2]\big((A,A),1\big)$	0.21	1
$\mathrm{DP}[2]\big((A,B),1\big)$	0.1512	1
$\mathrm{DP}[2]\big((B,A),1\big)$	0.3312	1
$\mathrm{DP}[2]\big((B,B),1\big)$	0.05	1

Step 3: From $t = 2$ to $t = 3$

For each parent state at $t = 2$, we draw a new profile $r \in \{(A,A),(A,B),(B,A),(B,B)\}$.

- From a state with $e = 0$: matching the length-2 block stays at $e = 0$, errors move to $e = 1$.
- From a state with $e = 1$: only matching is allowed (to avoid $e = 2$); it stays at $e = 1$.

From $\mathrm{DP}[2]((B,A),0) = 0.1288$

Drawn r	$P(r)$	Block $\big((B,A),r\big)$	Match?	Contribution
(A,B)	0.28	$((B,A),(A,B))$	Yes	$0.1288 \times 0.28 = 0.036064 \to \mathrm{DP}[3]((A,B),0)$
(A,A)	0.21	$((B,A),(A,A))$	No	$0.1288 \times 0.21 = 0.027048 \to \mathrm{DP}[3]((A,A),1)$
(B,A)	0.46	$((B,A),(B,A))$	No	$0.1288 \times 0.46 = 0.059248 \to \mathrm{DP}[3]((B,A),1)$
(B,B)	0.05	$((B,A),(B,B))$	No	$0.1288 \times 0.05 = 0.00644 \to \mathrm{DP}[3]((B,B),1)$

From $\mathrm{DP}[2]((A,B),0) = 0.1288$

Drawn r	$P(r)$	Block $\big((A,B),r\big)$	Match?	Contribution
(B,A)	0.46	$((A,B),(B,A))$	Yes	$0.1288 \times 0.46 = 0.059248 \to \mathrm{DP}[3]((B,A),0)$
(A,A)	0.21	$((A,B),(A,A))$	No	$0.1288 \times 0.21 = 0.027048 \to \mathrm{DP}[3]((A,A),1)$
(A,B)	0.28	$((A,B),(A,B))$	No	$0.1288 \times 0.28 = 0.036064 \to \mathrm{DP}[3]((A,B),1)$
(B,B)	0.05	$((A,B),(B,B))$	No	$0.1288 \times 0.05 = 0.00644 \to \mathrm{DP}[3]((B,B),1)$

From $\mathrm{DP}[2]((A,B),1) = 0.1512$

Only $(A,B) \to (B,A)$ matches

$$0.1512 \times 0.46 = 0.069552 \quad \to \mathrm{DP}[3]\big((B,A),1\big).$$

From $\mathrm{DP}[2]((B,A),1) = 0.3312$

Only $(B,A) \to (A,B)$ matches

$$0.3312 \times 0.28 = 0.092736 \quad \to \mathrm{DP}[3]\big((A,B),1\big).$$

State	From Contributions	Value
$\mathrm{DP}[3]((A,B),0)$	0.036064	0.036064
$\mathrm{DP}[3]((B,A),0)$	0.059248	0.059248
$\mathrm{DP}[3]((A,A),1)$	0.027048+0.027048	0.054096
$\mathrm{DP}[3]((A,B),1)$	0.036064+0.092736	0.128800
$\mathrm{DP}[3]((B,A),1)$	0.059248+0.069552	0.128800
$\mathrm{DP}[3]((B,B),1)$	0.00644+0.00644	0.012880

Summing across all valid transitions, we get

$$\begin{aligned}
\mathrm{DP}[3]\big((A,B),0\big) &= 0.036064, & \mathrm{DP}[3]\big((B,A),0\big) &= 0.059248,\\
\mathrm{DP}[3]\big((A,A),1\big) &= 0.054096, & \mathrm{DP}[3]\big((A,B),1\big) &= 0.1288,\\
\mathrm{DP}[3]\big((B,A),1\big) &= 0.1288, & \mathrm{DP}[3]\big((B,B),1\big) &= 0.01288.
\end{aligned}$$

Final Probability

Summing all probabilities for $e \leq 1$

$$\alpha(\hat{p},1) = 0.036064 + 0.059248 + 0.054096 + 0.128800 + 0.128800 + 0.012880 \approx 0.42.$$

F Sensitivity

We provide next our sensitivity analysis on both the number of errors (1, 2 and 3) and intervals of game-play periods (15, 20 and 25 periods) based on the action-convergence criterion. The analysis confirms that the principal early-late patterns are robust to these alternative specifications.

Table 4: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	20 → 32	1 → 25	34 → 38	17 → 28	7 → 14	5 → 20	2 → 25	21 → 24	→ 6	6 → 22	12 → 37	19 → 37	1 → 34	19 → 31	27 → 32	25 → 39	14 → 20	5 → 21	11 → 28
(A,A)	20 → 23 (-1)	→ 4	34 → 35 (-1) *	17 → 25 (-1)		3 → 6 (-1) *	→ 1	21 → 24 *	→ 2	6 → 19 *	10 → 18 (-1)	17 → 24 (-2)	1 → 1 (-1)	18 → 22 (-1)	27 → 28 (-3) *	25 → 31 (-2) *			
(A,B)		→ 4 *		→ 1 *	→ 1							→ 1	→ 8 *	→ 5 *			→ 1	→ 2	→ 4
(B,A)		1 → 12 *		→ 1 *	1 → (-1)	→ 3	2 → 22 *		→ 3				→ 8 *	1 → 4 *				→ 1	→ 1
(B,B)	→ 9 *		→ 3 *	→ 2	4 → 11 (-1) *	2 → 11			→ 1		2 → 19 *	2 → 12 (-1) *	→ 1		→ 4 *	→ 8 *	7 → 9 (-1) *	3 → 11 (-1) *	10 → 17 (-3) *
(A,B,A,A)													→ 1						
(A,B,A,B)			→ 2												*	*			
(A,B,B,A)		→ 5 *											→ 14 *	*					
(B,B,A,B)					2 → 2 (-1)								→ 1				7 → 10 (-1)	→ 3	1 → 2 (-1)
(A,A,A,B)										→ 3									
(B,B,A,B)																		2 → 4 (-1)	→ 4
Divergent	19 → 7	41 → 17	7 → 3	21 → 10	31 → 24	37 → 22	37 → 14	5 → 2	40 → 34	34 → 18	28 → 3	21 → 3	49 → 16	19 → 7	11 → 6	19 → 5	27 → 21	31 → 15	32 → 15
Entropy	3.1 → 1.9	5.4 → 3.7	1.1 → 0.9	3.4 → 2.1	5.0 → 4.2	5.2 → 4.0	5.2 → 2.7	1.2 → 0.5	5.3 → 5.2	4.9 → 3.2	4.4 → 1.4	3.5 → 1.5	5.6 → 3.6	3.3 → 2.2	1.9 → 1.5	2.8 → 1.4	4.4 → 3.9	5.0 → 3.7	4.7 → 3.4

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 1 error within an interval of 15 periods.

Table 5: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	19 → 31	1 → 24	34 → 38	17 → 27	7 → 14	5 → 16	1 → 22	21 → 24	→ 5	6 → 18	10 → 36	19 → 36	1 → 30	19 → 29	25 → 32	24 → 39	13 → 19	5 → 21	9 → 24
(A,A)(A,A)	19 → 22 (-1)	→ 4	34 → 35 (-1)*	17 → 25 (-1)		3 → 6 (-1)*	→ 1	21 → 24*	→ 1	6 → 16*	8 → 17 (-1)	17 → 24 (-2)	1 → 1 (-1)	18 → 21 (-1)	25 → 28 (-3)*	24 → 31 (-2)*			
(A,A)(A,B)		→ 4*			→ 1								→ 8*	→ 4*			→ 1	→ 2	→ 3
(B,A)(A,A)		1 → 12*		→ 1*	1 → (-1)		1 → 19*		→ 3				→ 5*	1 → 4*				→ 1	→ 1
(B,B)(A,A)	→ 9*		→ 3*	→ 1	4 → 11 (-1)*	2 → 10			→ 1		2 → 19*	2 → 12 (-1)*	→ 1		→ 4*	→ 8*	6 → 8 (-1)*	3 → 11 (-1)*	9 → 17 (-2)*
(A,B)(A,A)													→ 1						
(A,B)(A,B)			→ 3*				→ 2								*	*			
(A,B)(B,A)		→ 4*											→ 14*	*					
(B,B)(A,B)					2 → 2 (-1)												7 → 10 (-1)	→ 3	→ 1
(A,A,A)(A,A,B)										→ 2									
(B,B,B)(A,A,B)																		2 → 4 (-1)	→ 2
Divergent	20 → 8	41 → 18	7 → 3	21 → 11	31 → 24	37 → 26	38 → 17	5 → 2	40 → 35	34 → 22	30 → 4	21 → 4	49 → 20	19 → 9	13 → 6	20 → 5	28 → 22	31 → 15	34 → 19
Entropy	3.2 → 2.0	5.4 → 3.8	1.1 → 0.9	3.4 → 2.2	5.0 → 4.2	5.2 → 4.2	5.3 → 3.2	1.2 → 0.5	5.3 → 5.2	4.9 → 3.7	4.7 → 1.6	3.5 → 1.5	5.6 → 3.9	3.3 → 2.4	2.2 → 1.5	3.0 → 1.4	4.5 → 4.0	5.0 → 3.7	4.8 → 3.7

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 1 error within an interval of 20 periods.

Table 6: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	19 → 30	1 → 20	34 → 35	17 → 26	7 → 12	4 → 15	1 → 19	21 → 24	→ 5	6 → 18	10 → 34	19 → 36	1 → 27	19 → 29	25 → 31	24 → 38	13 → 19	3 → 20	9 → 23
A A	19 → 21 (-1)	→ 4	34 → 35 (-1)*	17 → 24 (-1)		2 → 5 (-1)*	→ 1	21 → 24*	→ 1	6 → 16*	8 → 17 (-1)	17 → 24 (-2)	1 → 1 (-1)	18 → 21 (-1)	25 → 28 (-3)*	24 → 31 (-2)*			
A B		→ 3*		→ 1*	→ 1								→ 8*	→ 4*			→ 1	→ 1	→ 3
B A		1 → 11*		→ 1*	1 → (-1)		1 → 16 (-1)*		→ 3				→ 4*	1 → 4*				→ 1	→ 1
B B	→ 9*		→ 1*	→ 1	4 → 9 (-1)*	2 → 10			→ 1		2 → 17*	2 → 12 (-1)*	→ 1		→ 3*	→ 7*	6 → 8 (-1)*	2 → 11*	9 → 16 (-3)*
A B A A													→ 1						
A B A B			→ 1*				→ 2								*	*			
A B B A		→ 2*		→ 1*									→ 12*	*					
B B A B					2 → 2 (-1)												7 → 10 (-1)	→ 3	→ 1
A A A A B										→ 2									
B B B A B																		1 → 4	→ 2
Divergent	20 → 9	41 → 22	7 → 6	21 → 12	31 → 26	38 → 27	38 → 20	5 → 2	40 → 35	34 → 22	30 → 6	21 → 4	49 → 23	19 → 9	13 → 7	20 → 6	28 → 22	33 → 16	34 → 20
Entropy	3.2 → 2.2	5.4 → 4.1	1.1 → 1.0	3.4 → 2.4	5.0 → 4.4	5.3 → 4.3	5.3 → 3.6	1.2 → 0.5	5.3 → 5.2	4.9 → 3.7	4.7 → 1.8	3.5 → 1.5	5.6 → 4.1	3.3 → 2.4	2.2 → 1.6	3.0 → 1.5	4.5 → 4.0	5.1 → 3.8	4.8 → 3.8

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 1 error within an interval of 25 periods.

Table 7: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	24 → 33	4 → 25	34 → 39	18 → 30	8 → 17	6 → 23	7 → 26	22 → 25	→ 8	11 → 22	14 → 38	20 → 38	3 → 36	20 → 32	27 → 34	28 → 39	21 → 22	9 → 24	14 → 32
(A,A)(A,A)	22 → 23 (-2)	→ 4	34 → 35 (-1)*	18 → 27 (-1)		3 → 6 (-1)*	→ 1	22 → 25*	→ 2	10 → 19 (-1)*	11 → 18 (-1)	17 → 24 (-2)	1 → 1 (-1)	19 → 23 (-2)	27 → 30 (-2)*	28 → 31 (-3)*			
(A,A)(A,B)		→ 4*			→ 1							→ 1	1 → 8*	→ 5*			→ 2	2 → 2 (-2)	→ 4
(B,A)(B,A)		4 → 12*		→ 1*	1 → (-1)	→ 4	7 → 23*		→ 4				→ 8*	1 → 4*				→ 1	→ 1
(B,A)(B,B)	2 → 10*		→ 4*	→ 2	5 → 13 (-2)*	2 → 13			→ 1		3 → 20*	3 → 13 (-2)*	→ 1		→ 4*	→ 8*	13 → 10 (-5)*	4 → 13 (-1)*	13 → 20 (-3)*
(A,A)(A,B)						1 → (-1)			→ 1										
(A,B)(A,A)													→ 1						
(A,B)(A,B)			→ 2												*	*			
(A,B)(B,A)		→ 5*											1 → 15*	*					
(B,B)(A,B)					2 → 3 (-1)								→ 1				8 → 10 (-1)	→ 4	1 → 3 (-1)
(A,A,A)(A,A,B)										1 → 3									
(B,B,B)(A,A,B)													→ 1					3 → 4 (-1)	→ 4
Divergent	15 → 6	38 → 17	7 → 2	20 → 8	30 → 21	36 → 19	32 → 13	4 → 1	40 → 32	29 → 18	26 → 2	20 → 2	47 → 14	18 → 6	11 → 4	16 → 5	20 → 19	27 → 12	29 → 11
Entropy	2.7 → 1.8	5.2 → 3.7	1.1 → 0.8	3.3 → 1.8	4.9 → 3.9	5.2 → 3.7	4.8 → 2.6	0.9 → 0.2	5.3 → 5.1	4.5 → 3.2	4.3 → 1.3	3.5 → 1.4	5.6 → 3.5	3.1 → 2.0	1.9 → 1.2	2.4 → 1.4	3.6 → 3.7	4.8 → 3.3	4.3 → 2.9

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 2 errors within an interval of 15 periods.

Table 8: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	22 → 31	3 → 24	34 → 35	18 → 26	8 → 15	5 → 17	4 → 22	21 → 25	→ 6	9 → 20	12 → 37	20 → 36	3 → 29	20 → 29	26 → 33	25 → 38	17 → 19	6 → 21	13 → 24
(A,A)/(A,A)	20 → 21 (-1)	→ 4	34 → 35 (-1)*	18 → 24 (-1)		2 → 6 (-1)*	→ 1	21 → 25*	→ 1	8 → 17 (-1)*	9 → 17 (-1)	17 → 24 (-2)	1 → 1 (-1)	19 → 21 (-2)	26 → 29 (-2)*	25 → 31 (-2)*			
(A,A)/(A,B)		→ 3*		*	→ 1								1 → 8*	→ 4*			→ 1	1 → 1 (-1)	→ 3
(A,B)/(A,A)		3 → 12*		→ 1*	1 → (-1)		4 → 19 (-1)*		→ 3				→ 5*	1 → 4*				→ 1	→ 1
(A,B)/(A,B)	2 → 10*		*	→ 1	5 → 12 (-2)*	2 → 11			→ 1		3 → 20*	3 → 12 (-2)*	→ 1		→ 4*	→ 7*	9 → 8 (-2)*	3 → 11*	12 → 16 (-4)*
(A,A)/(A,B)						1 → (-1)			→ 1										
(A,B)/(A,A)													→ 1						
(A,B)/(A,B)			*				→ 2								*	*			
(A,B)/(B,A)		→ 5*		*									1 → 13*	*					
(B,B)/(A,B)					2 → 2 (-1)												8 → 10 (-1)	→ 4	1 → 1 (-1)
(A,A,A)/(A,A,B)										1 → 3									
(B,B,B)/(A,A,B)																		2 → 4	→ 3
Divergent	17 → 8	39 → 18	7 → 6	20 → 12	30 → 23	37 → 25	35 → 17	5 → 1	40 → 34	31 → 20	28 → 3	20 → 4	47 → 21	18 → 9	12 → 5	19 → 6	24 → 22	30 → 15	30 → 19
Entropy	3.0 → 2.1	5.3 → 3.8	1.1 → 1.0	3.3 → 2.4	4.9 → 4.1	5.3 → 4.1	5.1 → 3.2	1.2 → 0.2	5.3 → 5.2	4.7 → 3.5	4.5 → 1.4	3.5 → 1.5	5.6 → 4.0	3.1 → 2.4	2.0 → 1.3	2.8 → 1.5	4.1 → 4.0	5.0 → 3.7	4.4 → 3.7

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 2 errors within an interval of 25 periods.

Table 9: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	26 → 35	6 → 27	36 → 39	20 → 33	10 → 19	10 → 25	13 → 28	22 → 25	2 → 10	16 → 24	19 → 38	27 → 38	12 → 38	21 → 34	28 → 34	30 → 40	24 → 25	13 → 26	18 → 35
A A / A A	23 → 24 (-2)	→ 5	36 → 35 (-2) *	20 → 30 (-1)		4 → 7 (-1) *	→ 1	22 → 25 *	1 → 3 (-1)	15 → 21 (-2) *	12 → 18 (-1)	19 → 24 (-2)	4 → 2 (-4)	20 → 23 (-3)	28 → 30 (-2) *	29 → 31 (-4) *			1 → (-1)
A A / A B		1 → 4 *			→ 1		1 → (-1)					→ 1	1 → 8 *	→ 6 *			1 → 2	2 → 3 (-2)	→ 4
B A / B A		4 → 12 *		→ 1 *	1 → (-1)	1 → 4 (-1)	12 → 25 (-1) *		1 → 5 (-1)				2 → 9 (-1) *	1 → 4 *				→ 1	1 → 1 (-1)
B B / B B	2 → 11 *		→ 4 *	→ 2	7 → 15 (-4) *	3 → 14			→ 1		7 → 20 *	8 → 13 (-4) *	→ 1		→ 4 *	1 → 9 *	14 → 11 (-5) *	6 → 14 (-1) *	15 → 22 (-2) *
A A / A B						1 → (-1)			→ 1										
A B / A A													→ 1	→ 1					
A B / A B			*				→ 2								*	*	1 → (-1)		
A B / B A		→ 5 *		*									5 → 15 *	*					
A B / B B		→ 1																	
B B / A B					2 → 3 (-1)								→ 1				8 → 12	2 → 4	1 → 4 (-1)
A A A / A A B										1 → 3									
A A B / A A A	1 → (-1)	1 → (-1)				1 → (-1)													
B B B / A A B													→ 1					3 → 4 (-1)	→ 4
Divergent	13 → 4	36 → 15	5 → 2	18 → 5	28 → 19	32 → 17	26 → 11	4 → 1	38 → 30	24 → 16	21 → 2	13 → 2	38 → 12	17 → 4	10 → 4	14 → 4	17 → 16	23 → 10	25 → 8
Entropy	2.6 → 1.5	5.2 → 3.6	0.8 → 0.8	3.0 → 1.3	4.7 → 3.6	5.1 → 3.5	4.2 → 2.3	0.9 → 0.2	5.3 → 4.9	3.9 → 2.9	3.8 → 1.3	2.7 → 1.4	5.2 → 3.4	3.0 → 1.9	1.7 → 1.2	2.3 → 1.3	3.5 → 3.3	4.5 → 3.1	4.1 → 2.6

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 3 errors within an interval of 15 periods.

Table 10: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	24 → 33	5 → 25	35 → 38	20 → 30	8 → 18	9 → 22	7 → 24	21 → 25	→ 8	13 → 22	17 → 38	26 → 36	9 → 33	21 → 32	26 → 34	27 → 40	21 → 22	13 → 24	15 → 28
(A,A)/(A,A)	22 → 22 (-2)	→ 4	35 → 35 (-1) *	20 → 27 (-1)		4 → 6 (-1) *	→ 1	21 → 25 *	→ 3	12 → 19 (-2) *	11 → 18	19 → 24 (-2)	2 → 2 (-2)	20 → 23 (-3)	26 → 30 (-2) *	26 → 31 (-3) *			
(A,A)/(A,B)		1 → 4 *		*	→ 1								1 → 8 *	→ 5 *			→ 2	2 → 2 (-2)	→ 3
(B,A)/(A,A)		4 → 12 *		→ 1 *	1 → (-1)	1 → 3 (-1)	7 → 21 *		→ 3				1 → 6 *	1 → 4 *				→ 1	1 → 1 (-1)
(B,A)/(B,A)	2 → 11 *		→ 3 *	→ 2	5 → 14 (-2) *	3 → 13			→ 1		6 → 20 *	7 → 12 (-4) *	→ 1		→ 4 *	1 → 9 *	13 → 10 (-5) *	6 → 13 (-1) *	13 → 17 (-3) *
(A,A)/(A,B)						1 → (-1)			→ 1										
(A,B)/(A,A)													→ 1						
(A,B)/(A,B)			*				→ 2								*	*			
(A,B)/(B,A)		→ 5 *		*									5 → 15 *	*					
(B,B)/(A,B)					2 → 3 (-1)												8 → 10 (-1)	2 → 4	1 → 4 (-1)
(A,A,A)/(A,A,B)										1 → 3									
(B,B,B)/(A,A,B)																		3 → 4 (-1)	→ 3
Divergent	15 → 6	37 → 17	6 → 3	18 → 8	30 → 20	33 → 20	32 → 15	5 → 1	40 → 32	27 → 18	23 → 2	14 → 4	41 → 17	17 → 6	12 → 4	17 → 4	20 → 19	23 → 12	28 → 15
Entropy	2.7 → 1.8	5.2 → 3.7	1.0 → 0.9	3.0 → 1.8	4.9 → 3.7	5.1 → 3.8	4.8 → 2.9	1.2 → 0.2	5.3 → 5.1	4.2 → 3.2	4.0 → 1.3	2.8 → 1.5	5.4 → 3.6	3.0 → 2.0	2.0 → 1.2	2.7 → 1.3	3.6 → 3.7	4.5 → 3.3	4.3 → 3.4

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 3 errors within an interval of 20 periods.

Table 11: Action Convergence

	PD	BO	SH	CH	SD	SA	TC	CI	UM	CO	PD2	PD2L	BO2	UL	SH2	SH2L	SD2	SD3	SD3L
Data Points	39	42	41	38	38	42	39	26	40	40	40	40	50	38	38	44	41	36	43
Convergent	23 → 31	4 → 25	35 → 35	20 → 27	8 → 16	9 → 20	7 → 23	21 → 25	→ 6	10 → 20	16 → 38	24 → 36	9 → 31	21 → 31	26 → 33	27 → 38	20 → 19	9 → 21	15 → 25
AA / AA	21 → 21 (-1)	→ 4	35 → 35 (-1) *	20 → 25 (-2)		4 → 6 (-1) *	→ 1	21 → 25 *	→ 1	9 → 17 (-1) *	10 → 18	19 → 24 (-2)	2 → 2 (-2)	20 → 23 (-3)	26 → 29 (-2) *	26 → 31 (-3) *			
AA / AB		1 → 4 *		*	→ 1								1 → 8 *	→ 4 *			→ 1	2 → 1 (-2)	→ 3
BB / AA		3 → 12 *		→ 1 *	1 → (-1)	1 → 1 (-1)	7 → 20 (-1) *		→ 3				1 → 5 *	1 → 4 *				→ 1	1 → 1 (-1)
BB / BB	2 → 10 *		*	→ 1	5 → 13 (-2) *	3 → 13			→ 1		6 → 20 *	5 → 12 (-3) *	→ 1		→ 4 *	1 → 7 *	12 → 8 (-5) *	4 → 11 *	13 → 16 (-4) *
AA / AA / AB						1 → (-1)			→ 1										
AB / AA / AA													→ 1						
AB / AB / AB			*				→ 2								*	*			
AB / AB / BA		→ 5 *		*									5 → 14 *	*					
BB / AB / AB					2 → 2 (-1)												8 → 10 (-1)	1 → 4	1 → 2 (-1)
AAA / AAB										1 → 3									
BBB / AAB																		2 → 4	→ 3
Divergent	16 → 8	38 → 17	6 → 6	18 → 11	30 → 22	33 → 22	32 → 16	5 → 1	40 → 34	30 → 20	24 → 2	16 → 4	41 → 19	17 → 7	12 → 5	17 → 6	21 → 22	27 → 15	28 → 18
Entropy	2.9 → 2.1	5.3 → 3.7	1.0 → 1.0	3.0 → 2.2	4.9 → 3.9	5.1 → 3.9	4.8 → 3.0	1.2 → 0.2	5.3 → 5.2	4.6 → 3.5	4.1 → 1.3	3.0 → 1.5	5.4 → 3.8	3.0 → 2.1	2.0 → 1.3	2.7 → 1.5	3.7 → 4.0	4.8 → 3.7	4.3 → 3.7

* Stage Nash Pareto Efficient Equal Payoff

Notes: In this Table, we allow for 3 errors within an interval of 25 periods.

G MTurk Experiments

G.1 Experimental Instructions [PD]

General Information

- You will be randomly matched with a participant and interact in real time.
- You will be working with a fictitious currency called Francs. Francs will be converted to US Dollars at an exchange rate of 100 Francs = $1.20.
- The task will proceed as fast as the slower participant.
- **If you do not make your decision within 1 minute in each round of the interaction, you will be marked as a dropout.**
- The interaction will end for the two participants matched together if one of them is declared a dropout.
- A dropout participant does not get paid. The other participant affected will be paid the earnings accumulated up to the point where the dropout participant was declared and the participation fee.
- It is therefore important to make your decisions in a timely manner.
- With the completion of the experimental instructions, there will be a quiz, where you will have 5 minutes and 5 chances to answer all questions correctly.
- You will still be able to read the instructions once you start the quiz.
- If you fail to pass the quiz, you will not be able to continue with the experiment or get paid.

Number of Rounds

- The number of rounds in the interaction is determined randomly.
- At the end of each round, a number will be chosen randomly from $\{1, 2, 3, ..., 99, 100\}$ where each number is equally likely.
- If the number is 1, then the interaction will end.
- In other words, there is a 1% chance that the interaction will end after each round.

Your Decision in Each Round

- In each round, you will be asked to choose one of two options: W and Y.
- Your payoff is based on your choice and the choice of the other participant that you are paired with.
- A payoff table will be given to you at the start of the interaction. These payoffs will remain the same throughout the interaction.
- The following is an example of a payoff table. In this example, if you choose W and the other chooses Y in a round, your payoff will be 4 Francs in this round, and the other's payoff will be 2 Francs.

Your choice	Other's choice	Your payoff	Other's payoff
Y	Y	6	5
Y	W	7	3
W	Y	4	2
W	W	1	8

- After both you and the other participant have made a choice, you will see your payoff as well as the choice and payoff of the other participant.

Payoffs

- Your payoff for the interaction will be the sum of the payoffs in each round.
- The following example provides the interface you will see during the interaction. The payoffs are left blank.
- You need to make your choice Y or W.
- The history of the interaction is shown at the bottom. The history of payoffs is left blank.

Round 5

Time left: 27

Your choice	Other's choice	Your payoff	Other's payoff
Y	Y		
Y	W		
W	Y		
W	W		

Please make your choice

Y
W

History

Round	Your choice	Other's choice	Your payoff	Other's payoff	Your cumulative payoff	Other's cumulative payoff
4	Y	W				
3	W	Y				
2	Y	W				
1	Y	Y				

Quiz

Provide your answers to the questions before clicking "Check my answers." The answers that are incorrect will be marked in red. Check and correct the wrong answers before checking your answers again.

You have at most 5 minutes and 5 attempts in total to complete the quiz. Otherwise, you will not be able to participate in the rest of the experiment. Each time you click "Check my answers" will count as an attempt.

Answer the questions based on the following payoff table using the radio buttons.

Your choice	Other's choice	Your payoff	Other's payoff
Y	Y	6	5
Y	W	7	3
W	Y	4	2
W	W	1	8

- What is your payoff if you choose Y and the other participant chooses Y? 6
- What is the other participant's payoff if you choose Y and the other participant chooses W? 3
- What is your payoff if you choose W and the other participant chooses Y? 4
- What is the other participant's payoff if you choose W and the other participant also chooses W? 8
- When do you receive the highest possible payoff? I choose Y, the other chooses W

- When does the other participant receive the highest possible payoff? I choose W, the other also chooses W
- When do the two of you receive the highest joint payoff? I choose Y, the other also chooses Y
- Suppose the interaction is currently in round 53; what is the chance that the interaction will end after this round? 1%
- Suppose the interaction is currently in round 71; what is the chance that the interaction will end after this round? 1%

Instructions for the Match

You have just been matched with another participant. You will interact with the same participant throughout the interaction.

The payoff table is fixed from round to round in the interaction and is shown below.

Your choice	Other's choice	Your payoff	Other's payoff
Y	Y	3	3
Y	W	1	4
W	Y	4	1
W	W	2	2

- The experiment proceeds as fast as the slower participant.
- The experiment may take longer than expected if you or the other participant take a long time to make a decision.
- If you do not make your decision within 1 minute in a given round, you will be marked as a dropout and not get paid.