

Heinrich, Torsten et al.

Working Paper

Best-response dynamics, playing sequences, and convergence to equilibrium in random games

LEM Working Paper Series, No. 2021/02

Provided in Cooperation with:

Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Heinrich, Torsten et al. (2021) : Best-response dynamics, playing sequences, and convergence to equilibrium in random games, LEM Working Paper Series, No. 2021/02, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:

<https://hdl.handle.net/10419/243498>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Best-Response Dynamics, Playing Sequences, and Convergence to Equilibrium in Random Games

Torsten Heinrich ^{a,b,c}

Yoojin Jang ^{b,d}

Luca Mungo ^{b,e}

Marco Pangallo ^f

Alex Scott ^e

Bassel Tarbush ^g

Samuel Wiese ^{b,d}

^a Faculty for Economics and Business Administration, Chemnitz University of Technology, Germany.

^b Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK.

^c Oxford Martin Programme on Technological and Economic Change (OMPTEC), Oxford Martin School, University of Oxford, UK.

^d Department of Computer Science, University of Oxford, UK.

^e Mathematical Institute, University of Oxford, UK.

^f Institute of Economics and EmbeDS Department, Scuola Superiore Sant'Anna, Pisa, Italy.

^g Department of Economics, University of Oxford, UK.

2021/02

January 2021

ISSN(ONLINE) 2284-0400

BEST-RESPONSE DYNAMICS, PLAYING SEQUENCES, AND CONVERGENCE TO EQUILIBRIUM IN RANDOM GAMES

TORSTEN HEINRICH^{1,2,3}, YOOJIN JANG^{2,4}, LUCA MUNGO^{2,5}, MARCO PANGALLO⁶, ALEX
SCOTT⁵, BASSEL TARBUSH⁷, SAMUEL WIESE^{2,4}

ABSTRACT. We show that the playing sequence—the order in which players update their actions—is a crucial determinant of whether the best-response dynamic converges to a Nash equilibrium. Specifically, we analyze the probability that the best-response dynamic converges to a pure Nash equilibrium in random n -player m -action games under three distinct playing sequences: clockwork sequences (players take turns according to a fixed cyclic order), random sequences, and simultaneous updating by all players. We analytically characterize the convergence properties of the clockwork sequence best-response dynamic. Our key asymptotic result is that this dynamic almost never converges to a pure Nash equilibrium when n and m are large. By contrast, the random sequence best-response dynamic converges almost always to a pure Nash equilibrium when one exists and n and m are large. The clockwork best-response dynamic deserves particular attention: we show through simulation that, compared to random or simultaneous updating, its convergence properties are closest to those exhibited by three popular learning rules that have been calibrated to human game-playing in experiments (reinforcement learning, fictitious play, and replicator dynamics).

JEL CODES: C62, C72, C73, D83.

KEYWORDS: Best-response dynamics, equilibrium convergence, random games, learning models in games.

¹Faculty for Economics and Business Administration, Chemnitz University of Technology, Chemnitz, Germany

²Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, UK

³Oxford Martin Programme on Technological and Economic Change (OMPTEC), Oxford Martin School, University of Oxford, Oxford, UK

⁴Department of Computer Science, University of Oxford, Oxford, UK

⁵Mathematical Institute, University of Oxford, Oxford, UK

⁶Institute of Economics and EMbeDS Department, Sant’Anna School of Advanced Studies, Pisa, Italy

⁷Department of Economics, University of Oxford, Oxford, UK

Email addresses: `torsten.heinrich@wirtschaft.tu-chemnitz.de`, `yjluca98@gmail.com`,
`luca.mungo@maths.ox.ac.uk`, `marco.pangallo@santannapisa.it`, `scott@maths.ox.ac.uk`,
`bassel.tarbush@economics.ox.ac.uk`, `samuel.wiese@cs.ox.ac.uk`

Date: Monday 11th January, 2021.

We thank Doyné Farmer for useful comments at the early stages of this project. We acknowledge funding from Baillie Gifford (Luca Mungo), the James S Mc Donnell Foundation (Marco Pangallo) and the Foundation of German Business (Samuel Wiese).

CONTENTS

1. Introduction	3
2. Best-response dynamics in games	7
2.1. Games	7
2.2. Best-response digraphs	8
2.3. Best-response dynamics	9
2.4. Convergence	10
2.5. Best-response dynamics with random inputs	12
3. Theoretical results	14
3.1. m -action games with $n > 2$ players	16
3.2. m -action games with $n = 2$ players	18
4. Simulation results	21
4.1. Simulations of clockwork best-response dynamics	21
4.2. Simulations of best-response dynamics under clockwork, random, and simultaneous updating	23
4.3. Simulation of other learning rules	25
Appendix A. Proof of Theorem 2	32
Appendix B. Proofs of Theorem 3, Proposition 3, and Theorem 4	43
B.1. Proof of Theorem 3	43
B.2. Proof of Theorem 4	45
B.3. Proof of Proposition 3	47
Appendix C. Descriptions of the learning rules	48
C.1. Reinforcement learning	48
C.2. Fictitious play	50
C.3. Replicator dynamics	51
References	54

1. INTRODUCTION

The best-response dynamic is a ubiquitous iterative game-playing process in which, in each period, players myopically select actions that are a best-response to the actions last chosen by all other players. Most of the existing work on best-response dynamics and on learning rules, such as fictitious play, establishes sufficient conditions on a game’s payoff structure to guarantee convergence to a Nash equilibrium.¹ In this paper we also investigate the convergence properties of the best-response dynamic but, rather than restricting our attention to games with a particular structure, our focus is instead on the role of the *playing sequence*—the order in which players update their actions. Our key insight is that the playing sequence is a crucial determinant of whether the best-response dynamic converges to a pure Nash equilibrium.

We focus on three specific playing sequences: “clockwork” sequences, random sequences, and simultaneous updating. Under the clockwork playing sequence, players take turns to play one at a time according to a fixed cyclic order. Player 1 plays first, followed by player 2, and so on up to player n , and then the sequence returns to player 1, and so on. To our knowledge, the behavior of the best-response dynamic under this playing sequence has received relatively little attention in the literature.² Under the random playing sequence, players take turns to play one at a time and the next player to play is chosen uniformly at random from among all players. This playing sequence is the most well-studied in the literature.³ Finally, we also consider simultaneous updating by all players in each period.^{4,5}

To investigate the role of the playing sequence in determining the convergence properties of the best-response dynamic, we analyze the probability that the best-response dynamic converges to a pure Nash equilibrium in random n -player m -action games under clockwork, random, and simultaneous updating. In other words, we generate a game by drawing all payoffs at random (from atomless distributions to avoid payoff ties) and we determine the

¹For example, previous work has established that the best-response dynamic converges to a Nash equilibrium in weakly acyclic games (Fabrikant et al., 2013), potential games (Monderer and Shapley, 1996), aggregative games (Dindoš and Mezzetti, 2006), and quasi-acyclic games (Friedman and Mezzetti, 2001, Takahashi and Yamamori, 2002).

²Boucher (2017) analyzes the clockwork sequence best-response dynamic in potential games.

³The random sequence best-response dynamic has been analyzed in anonymous games (Babichenko, 2013), near-potential games (Candogan et al., 2013), potential games (Christodoulou et al., 2012, Coucheney et al., 2014, Durand and Gaujal, 2016, Swenson et al., 2018, Durand et al., 2019), and games on a lattice (Blume et al., 1993). “Sink” equilibria are studied in (Goemans et al., 2005, Mirrokni and Skopalik, 2009).

⁴This case is studied in Quint et al. (1997) for 2-player games and in Kash et al. (2011) for anonymous games.

⁵There are, of course, many other possible playing sequences. For example, Feldman and Tamir (2012) study the case in which the sequence of play depends on current payoffs.

probability that the best-response dynamic starting at a random initial action profile converges to a pure Nash equilibrium of the randomly drawn game. Our paper therefore builds on the growing literature on random games.⁶ Studying such games allows us to abstract from the specific structure of a given game, thereby allowing us to focus solely on the role of the playing sequence. Furthermore, random games are conceptually useful because they can be seen as null models for generic situations involving strategic interactions.⁷

The novel theoretical contributions of this paper are primarily about the convergence properties of the clockwork best-response dynamic in random games in which payoffs are drawn independently. Our main finding, which is presented in Section 3.1, is that the probability that the clockwork best-response dynamic converges to a pure Nash equilibrium is, up to a polynomial factor, of order $1/\sqrt{m^{n-1}}$. This has two implications: (i) when the number of players n and/or the number of actions m is large ($nm \rightarrow \infty$), the probability that the clockwork best-response dynamic converges to a pure Nash equilibrium goes to zero, and (ii) since the asymptotic convergence probability depends essentially only on the quantity m^{n-1} , we have that, when n and/or m are large, the probability of convergence to a pure Nash equilibrium in n -player m -action games is approximately the same as it is in 2-player m^{n-1} -action games. In fact, our simulations indicate that this asymptotic relationship between n -player m -action games and 2-player m^{n-1} -action games is also fairly accurate for small values of n and m . In Section 3.2 we focus exclusively on 2-player games. This allows us to provide more granular results on the convergence properties of the clockwork best-response dynamic. In particular, we provide results on game duration and we derive an exact expression for the probability that the best-response dynamic reaches a (best-response) cycle of given length at a particular period. As a special case, we obtain the exact probability that the clockwork best-response dynamic converges to a pure Nash equilibrium in 2-player m -action games (and we argue that, in the 2-player m -action case, this probability is the same for random playing sequences). Furthermore, we show that this probability is asymptotically $\sqrt{\pi/m}$ when m is large (and $\pi \approx 3.14$).

⁶The literature on randomly generated games starts with [Goldman \(1957\)](#), [Goldberg et al. \(1968\)](#), and [Dresher \(1970\)](#). Since then, a number of papers have analyzed the distribution of pure and mixed Nash equilibria in random games ([Powers, 1990](#), [Stanford, 1995, 1996, 1997, 1999](#), [McLennan, 2005](#), [McLennan and Berg, 2005](#), [Takahashi, 2008](#), [Kultti et al., 2011](#), [Daskalakis et al., 2011](#)). [Cohen \(1998\)](#) derives the probability that a pure Nash equilibrium is Pareto efficient. More recently, [Alon et al. \(2020\)](#) derive the probability that a random game is dominance-solvable.

⁷See [Pangallo et al. \(2019\)](#) for a general discussion on the usefulness of considering null models and statistical ensembles in game theory, and on how this approach is extensively used in other disciplines such as statistical mechanics and ecology.

We briefly comment on the approach that we adopted to derive the main result of Section 3.1. We represent the best-response structure of a game by a directed graph (or digraph) in which the vertices are the action profiles and the directed edges correspond to the players' best-responses. A pure Nash equilibrium corresponds to a sink of the digraph. The best-response dynamic can be represented by a path that starts at some initial profile in the digraph and travels along the directed edges in a direction that is determined by the playing sequence. Drawing payoffs independently at random (from atomless distributions) induces a uniform distribution over the best-response digraphs, so the probability of convergence to a pure Nash equilibrium can be reduced to working out the probability that the best-response path initiated at a random vertex reaches a sink of the randomly drawn digraph. The main theoretical challenge that we face when analyzing the best-response dynamic is that it exhibits some path-dependence: if a player encounters an environment that they had seen before, they must play the same action that they played when the environment was first encountered. We tackle this issue by relying on a coupling argument in which the best-response dynamic is coupled to a (memoryless) random walk through the digraph that is easier to analyze.

Our results for the convergence properties of the best-response dynamic under random and simultaneous updating rely mostly on simulations. Under a random sequence, conditional on the game having a pure Nash equilibrium, we show that the probability of convergence to a pure Nash equilibrium goes to one when n or m are large.⁸ This is in sharp contrast to our results for the clockwork sequence and highlights one of the key insights of this paper; namely, that the playing sequence is a crucial determinant of the probability of convergence to equilibrium. While the existing literature on best-response dynamics has focused primarily on identifying sufficient conditions on a game's payoff structure to guarantee convergence to equilibrium, our results indicate that the playing sequence must also be given careful consideration. To further corroborate this insight, we show that the playing sequence also determines the probability of convergence to equilibrium in random games with correlated payoffs.⁹ For example, in two player games with strongly positively correlated payoffs, we find that the best-response dynamic with simultaneous updating is unlikely to converge to equilibrium, whereas it is very likely to do so under a clockwork or a random sequence. Over all possible values for the payoff correlation parameter, the best-response dynamic tends to converge to a pure Nash equilibrium most frequently under a random playing sequence and least frequently under simultaneous updating.

⁸Amiet et al. (2019) prove this result analytically for the case $m = 2$ and $n \rightarrow \infty$.

⁹See Goldberg et al. (1968), Stanford (1999), Berg and Weigt (1999), Rinott and Scarsini (2000), Galla and Farmer (2013), Sanders et al. (2018) for work on random games with payoff correlations.

Among the three playing sequences, the clockwork playing sequence stands out as deserving particular attention. Through extensive simulations, we show that the frequency of convergence to equilibrium of the clockwork best-response dynamic most closely tracks the convergence frequency of three popular learning rules, namely the Bush-Mosteller reinforcement learning algorithm (Bush and Mosteller, 1953), fictitious play (Brown, 1951, Robinson, 1951), and replicator dynamics (Maynard Smith, 1982).¹⁰ The three learning algorithms are most naturally defined as involving simultaneous updating, yet when we vary n , m , or the payoff correlation parameter, the clockwork sequence best-response dynamic outperforms both the random sequence and the simultaneous updating best-response dynamics in most of our simulations. Additionally, when compared with the random sequence best-response dynamic, the paths traced by the clockwork sequence best-response dynamic in the space of all action profiles more closely resemble the paths traced by the three learning algorithms.

Our focus on reinforcement learning, fictitious play, and replicator dynamics is driven by the fact that these learning rules have been used to calibrate human game-play in experiments (Bush and Mosteller, 1953, Arthur, 1991, Erev and Roth, 1998, Sarin and Vahid, 2001, Van Huyck et al., 1995, Friedman, 1996, Cheung and Friedman, 1997).¹¹ Our results suggest that, to the extent that the learning algorithms are consistent with human game-play in randomly-generated games, the clockwork best-response dynamic could provide a first-order approximation for the evolution of play in such games.

The paper is structured as follows. In Section 2 we present our analytical framework. Section 3 contains our theoretical results on the probability that the clockwork sequence best-response dynamic converges to pure Nash equilibria in random games with independently drawn payoffs. The section also compares our findings to existing analytical results regarding the random playing sequence. Section 4 contains all our numerical simulation results. All proofs and detailed descriptions of the three learning rules (reinforcement learning, fictitious play, and replicator dynamics) are in the appendix.

¹⁰The convergence properties of these learning algorithms have been extensively studied (Fudenberg and Levine, 1998), but there is no general result about their probability of convergence to Nash equilibria in random games.

¹¹Of course, Bush-Mosteller reinforcement learning, fictitious play, and replicator dynamics are not representative of all learning algorithms that have been studied in game theory. For example, differently from these learning rules, regret testing (Foster and Young, 2006, Germano and Lugosi, 2007) converges to a Nash equilibrium in essentially every n -player, m -action game with high probability. Therefore, its convergence properties are better approximated by best-response dynamics under a random sequence rather than under a clockwork sequence.

2. BEST-RESPONSE DYNAMICS IN GAMES

In this section, we introduce the central concepts of our paper. For clarity, we summarize some of our keys terms in Table 1.

TABLE 1. Terminology

Game $g_{n,m}$	Game with n players and m actions per player
Environment \mathbf{a}_{-i}	Part of the action profile \mathbf{a} that is played by all players but i
Best-response $b_i(\mathbf{a}_{-i})$	Maps \mathbf{a}_{-i} to the actions giving highest payoff to i
Non-degenerate game	Game with no payoff ties, i.e. the best-response is unique for each i and \mathbf{a}_{-i}
Playing sequence s	The function $s : \mathbb{N} \rightarrow [n]$ determines whose turn it is to play
s -best-response dynamic on $g_{n,m}$ initiated at \mathbf{a}^0	Starting at profile \mathbf{a}^0 , in each period $t \in \mathbb{N}$, player $s(t)$ plays her myopic best-response to environment $\mathbf{a}_{-s(t)}^{t-1}$ in the game $g_{n,m}$
Path $\langle \bar{\mathbf{a}}, s \rangle$	Infinite sequence of action profiles $\bar{\mathbf{a}} = (\mathbf{a}^0, \mathbf{a}^1, \dots)$ satisfying $\mathbf{a}_{-s(t)}^t = \mathbf{a}_{-s(t)}^{t-1}$ for each $t \in \mathbb{N}$

2.1. Games. A game with $n \geq 2$ players and $m \geq 2$ actions is a tuple

$$g_{n,m} := ([n], [m], \{u_i\}_{i \in [n]}),$$

where $[n] := \{1, \dots, n\}$ is the set of players and each player $i \in [n]$ has a set of actions $[m] := \{1, \dots, m\}$ and a payoff function $u_i : [m]^n \rightarrow \mathbb{R}$.

An *action profile* is a vector of actions $\mathbf{a} = (a_1, \dots, a_n)$ belonging to the set $[m]^n$ that lists the action taken by each player. An *environment* for player i is a vector \mathbf{a}_{-i} belonging to the set $[m]^{n-1}$ that lists the action taken by each player but i . A *best-response correspondence* b_i for player i is a mapping from the set of environments for player i to the set of all non-empty subsets of i 's actions and is defined by

$$b_i(\mathbf{a}_{-i}) := \arg \max_{a_i \in [m]} u_i(a_i, \mathbf{a}_{-i}).$$

A game is *non-degenerate* if for each player i and environment \mathbf{a}_{-i} , the best-response action is unique. Games in which there are no ties in payoffs are non-degenerate games.¹² In the rest of this paper, we focus only on non-degenerate games, so each instance of “game” will be taken to mean “non-degenerate game”. Since best-responses are unique in non-degenerate games, we write $a_i = b_i(\mathbf{a}_{-i})$ whenever $a_i \in b_i(\mathbf{a}_{-i})$.

¹²There are no ties in payoffs if for all players i , all environments \mathbf{a}_{-i} , and all $a_i \neq a'_i$, $u_i(a_i, \mathbf{a}_{-i}) \neq u_i(a'_i, \mathbf{a}_{-i})$.

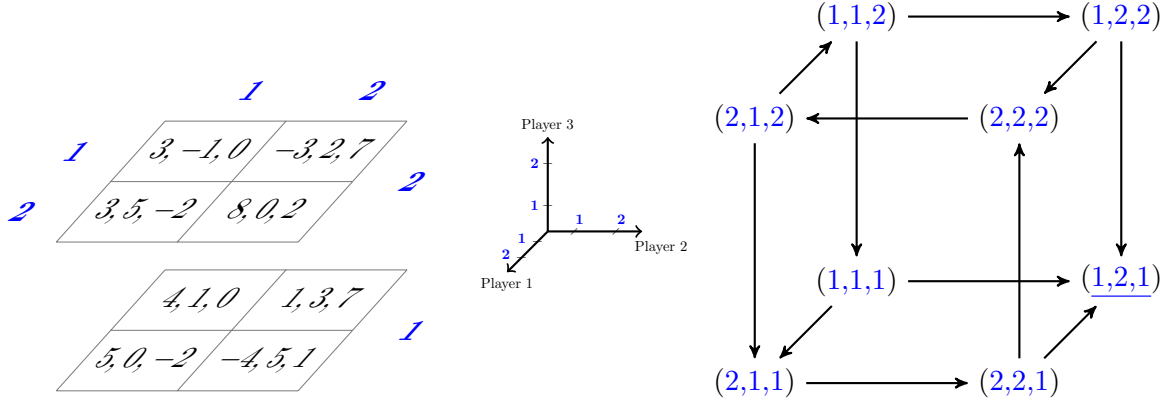


FIGURE 1. Illustration of a 3-player 2-action non-degenerate game (left) and its associated best-response digraph (right). The axes shown in the center give us our coordinate system.

An action profile $\mathbf{a} \in [m]^n$ is a *pure Nash equilibrium* (PNE) if for all $i \in [n]$ and all $a_i \in [m]$,

$$u_i(\mathbf{a}) \geq u_i(a_i, \mathbf{a}_{-i}).$$

Equivalently, $\mathbf{a} \in [m]^n$ is a PNE if each player $i \in [n]$ is playing their best-response action i.e. $a_i = b_i(\mathbf{a}_{-i})$. Denote the set of PNE of the game $g_{n,m}$ by $\text{PNE}(g_{n,m})$ and let $\#\text{PNE}(g_{n,m})$ denote the cardinality of this set.

2.2. Best-response digraphs. The best-response structure of a non-degenerate game $g_{n,m}$ can be represented by a *best-response digraph* $\mathcal{D}(g_{n,m})$ whose vertex set is the set of action profiles $[m]^n$ and whose edges are constructed as follows: for each $i \in [n]$ and each pair of distinct vertices $\mathbf{a} = (a_i, \mathbf{a}_{-i})$ and $\mathbf{a}' = (a'_i, \mathbf{a}_{-i})$, place a directed edge from \mathbf{a} to \mathbf{a}' if and only if a'_i is player i 's best-response to environment \mathbf{a}_{-i} , i.e. $a'_i = b_i(\mathbf{a}_{-i})$. There are edges only between action profiles that differ in exactly one coordinate. A profile \mathbf{a} is a PNE of $g_{n,m}$ if and only if it is a sink of the best-response digraph $\mathcal{D}(g_{n,m})$.

Example (Best-response digraphs). Panel (A) of Figure 1 illustrates a 3-player 2-action game (on the left) and its associated best-response digraph (on the right). Player 1 selects rows (along the depth), player 2 selects columns (along the width), and player 3 selects levels (along height). In the left-hand panel, the payoffs of players 1, 2 and 3 are listed in that order. The vertices of the best-response digraph are the action profiles. The unique PNE at the profile (1, 2, 1) is underlined and is a sink of the digraph. ■

2.3. Best-response dynamics. We now consider games played over time, with each player in turn myopically best-responding to their current environment. A *playing sequence* function $s : \mathbb{N} \rightarrow [n]$ determines whose turn it is to play at each time period $t \in \mathbb{N}$, where \mathbb{N} denotes the set of positive integers $\{1, 2, \dots\}$. A *path* $\langle \vec{\mathbf{a}}, s \rangle$ is an infinite sequence of action profiles $\vec{\mathbf{a}} = (\mathbf{a}^0, \mathbf{a}^1, \dots)$ and an associated playing sequence function $s : \mathbb{N} \rightarrow [n]$ satisfying the constraint that only one player changes her action at a time, $\mathbf{a}_{-s(t)}^t = \mathbf{a}_{-s(t)}^{t-1}$ for each $t \in \mathbb{N}$. So only the action of player $s(t)$ is allowed to differ between profiles \mathbf{a}^{t-1} and \mathbf{a}^t along a path.

Note that this set up rules out simultaneous updating from our theoretical analysis because we allow only one player to play in any given period. A more general framework would allow subsets of players to update their actions simultaneously in each period – in other words, a playing sequence would be a sequence of non-empty subsets of $[n]$ – but we avoid this generality here. As mentioned in our introduction, we focus exclusively on clockwork and random playing sequences for our theoretical results.

The best-response dynamic with playing sequence $s : \mathbb{N} \rightarrow [n]$ on a game $g_{n,m}$ initiated at the action profile \mathbf{a}^0 generates a path $\langle \vec{\mathbf{a}}, s \rangle$ according to Algorithm 1. Namely, set the initial action profile to \mathbf{a}^0 and, in each period $t \in \mathbb{N}$, player $s(t)$ myopically plays the best-response to her current environment $\mathbf{a}_{-s(t)}^{t-1}$.

Algorithm 1 s -sequence best-response dynamic on $g_{n,m}$ initiated at \mathbf{a}^0

- (1) For $t \in \mathbb{N}$:
 - (a) Set $i = s(t)$
 - (b) Set $\mathbf{a}_{-i}^t = \mathbf{a}_{-i}^{t-1}$
 - (c) Set $a_i^t = b_i(\mathbf{a}_{-i}^{t-1})$ where $b_i(\mathbf{a}_{-i}^{t-1}) := \arg \max_{x_i \in [m]} u_i(x_i, \mathbf{a}_{-i}^{t-1})$
-

Algorithm 1 generates a path by traveling along the edges of the best-response digraph $\mathcal{D}(g_{n,m})$ in direction $s(t)$ at step t starting from the initial profile \mathbf{a}^0 . More precisely, the infinite sequence of actions $\vec{\mathbf{a}}$ is determined as follows: if player $s(t)$ is already best responding then \mathbf{a}^{t-1} does not point to any vertex $(a'_{s(t)}, \mathbf{a}_{-s(t)}^{t-1}) \neq \mathbf{a}^{t-1}$ and the next profile in the sequence is \mathbf{a}^{t-1} itself, i.e. $\mathbf{a}^t = \mathbf{a}^{t-1}$; otherwise, if player $s(t)$ is not already playing her best response then travel to the vertex that corresponds to her playing her best-response action, i.e. set $\mathbf{a}^t = (a'_{s(t)}, \mathbf{a}_{-s(t)}^{t-1})$ where $(a'_{s(t)}, \mathbf{a}_{-s(t)}^{t-1}) \neq \mathbf{a}^{t-1}$ is the unique vertex that \mathbf{a}^{t-1} points to.

2.4. Convergence. For any path $\langle \vec{\mathbf{a}}, s \rangle$ and any set of action profiles $\mathcal{A} \subseteq [m]^n$ define $\tau_{\langle \vec{\mathbf{a}}, s \rangle}(\mathcal{A})$ as the first period $t \geq 1$ in which some element of the sequence $\vec{\mathbf{a}}$ is in the set \mathcal{A} :

$$\tau_{\langle \vec{\mathbf{a}}, s \rangle}(\mathcal{A}) := \inf\{t \in \mathbb{N} : \mathbf{a}^t \in \mathcal{A}\},$$

where \inf is the infimum operator and we use the convention that $\inf \emptyset = \infty$ (i.e. we take $\tau_{\langle \vec{\mathbf{a}}, s \rangle}(\mathcal{A})$ to be infinite if no element of the sequence $\vec{\mathbf{a}}$ is in \mathcal{A}). The path $\langle \vec{\mathbf{a}}, s \rangle$ *reaches* the set \mathcal{A} (in period t) if $t = \tau_{\langle \vec{\mathbf{a}}, s \rangle}(\mathcal{A})$ and t is finite.¹³

Definition 1. The s -sequence best-response dynamic on game $g_{n,m}$ initiated at \mathbf{a}^0 *converges* to a PNE if the path $\langle \vec{\mathbf{a}}, s \rangle$ generated according to Algorithm 1 reaches $\text{PNE}(g_{n,m})$.

Clearly, if the path reaches a PNE in some period, it stays there forever.

2.4.1. Convergence for the clockwork playing sequence. There are infinitely many possible playing sequences. We will be particularly interested in the *clockwork* playing sequence which is defined by $s(t) = s_c(t) := 1 + (t - 1) \bmod n$. In other words, player 1 plays in period 1, followed by player 2, then 3, and so on until player n , and then the sequence returns to player 1, and so on.

Definition 1 applies to all playing sequences but, when the sequence is clockwork, we can characterize convergence (and non-convergence) more simply in terms of path properties. We refer to one complete rotation of the clockwork sequence as a *round* of play; for example, if a round starts at player i then each player plays once and the round is complete when it is once again i 's turn to play. For any $k \in \mathbb{N}$ define

$$T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k) := \inf \left\{ t \in \mathbb{N} : \mathbf{a}^t = \mathbf{a}^{t+nk} \text{ and } \mathbf{a}^t \neq \mathbf{a}^{t+nk'} \text{ for all } k' \in \mathbb{N} \text{ such that } k' < k \right\},$$

to be the first period in which an action profile is repeated k rounds later (and at no earlier round). If $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$ is finite then the path $\langle \vec{\mathbf{a}}, s_c \rangle$ has the property that from period $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$ onwards, the sequence of nk possibly non-distinct action profiles $\mathbf{a}^t, \dots, \mathbf{a}^{t+nk-1}$ repeats itself forever. We therefore say that the path $\langle \vec{\mathbf{a}}, s_c \rangle$ reaches a *nk -cycle* in period $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$ or, equivalently, that the clockwork sequence best-response dynamic converges to a *nk -cycle* (in period $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$).

Notice that if the action profile is \mathbf{a}^t in some period t and no one deviates from this profile in a single round (i.e. $\mathbf{a}^t = \mathbf{a}^{t+n}$), then \mathbf{a}^t must be a PNE. Therefore, if the path $\langle \vec{\mathbf{a}}, s_c \rangle$ reaches a nk -cycle in period $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$ and $k = 1$ then we say that the clockwork sequence best-response dynamic reaches a PNE in that period. However, if the path $\langle \vec{\mathbf{a}}, s_c \rangle$

¹³We also say that the path reaches the set \mathcal{A} *by* period t if it reaches \mathcal{A} in period τ with $\tau \leq t$ and the path reaches \mathcal{A} *before* (after) period t if it reaches \mathcal{A} in period $\tau < t$ ($\tau > t$).

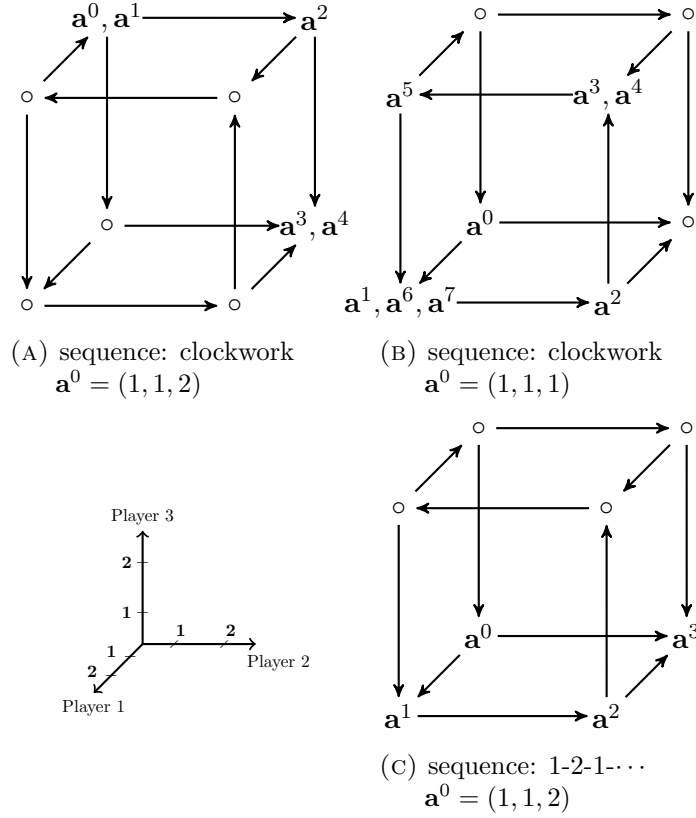


FIGURE 2. The digraphs in panels (A)-(C) above are all identical and correspond to the best-response digraph of the game shown in Figure 1. In the panels we show the first few elements of paths generated according to the best-response dynamic for different initial profiles and playing sequences.

reaches a nk -cycle in period $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$ and $k > 1$ then we say that the clockwork sequence best-response dynamic converges to a *best-response cycle* (of length nk) in that period.

Because the number of action profiles is finite, $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)$ must be finite for some k , so the clockwork best-response dynamic must always converge either to a PNE or to a best-response cycle. In dynamical systems language, nk -cycles (including PNE) are *attractors* of the clockwork best-response dynamic. Clearly,

$$T_{\langle \vec{\mathbf{a}}, s_c \rangle} := \inf \{k \in \mathbb{N} : T_{\langle \vec{\mathbf{a}}, s_c \rangle}(k)\},$$

is the period in which the path $\langle \vec{\mathbf{a}}, s_c \rangle$ reaches a PNE or a best-response cycle.

Example (Best-response dynamics and convergence). The digraphs in panels (A)-(C) of Figure 2 are all identical and correspond to the best-response digraph of the game shown in Figure 1. Each vertex of a best-response digraph is a point in $[m]^n$ but, unlike the right-hand panel of Figure 1, we no longer show the explicit coordinate of each vertex in our illustrations to avoid clutter. In panels (A)-(C) of Figure 2 we show the first few elements of paths generated according to the best-response dynamic for different initial profiles and playing sequences.

In panel (A) the initial profile is set to $\mathbf{a}^0 = (1, 1, 2)$ and the playing sequence is clock-work. The first few elements of the infinite sequence $\vec{\mathbf{a}}$ are shown in the figure. The path stays at $(1, 1, 2)$ in period 1 because player 1 does not change her action. The path then moves to $\mathbf{a}^2 = (1, 2, 2)$ in period 2 because player 2 plays action 2. In period 3, player 3 plays action 1 which takes the path to $\mathbf{a}^3 = (1, 2, 1)$. Once at this profile, which is the unique PNE, the path remains there forever. Furthermore, $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(1) = 3$.

In panel (B) the initial profile is set to $\mathbf{a}^0 = (1, 1, 1)$ and the playing sequence is clock-work. This time, the path moves to the bottom left corner on the front face of the cube in period 1. The path then cycles forever among the four profiles on the front face of the cube. Since period $t = 1$ is the first period in which $\mathbf{a}^t \neq \mathbf{a}^{t+3}$ but $\mathbf{a}^t = \mathbf{a}^{t+3k}$ for some integer k (namely $k = 2$) we have that $T_{\langle \vec{\mathbf{a}}, s_c \rangle}(2) = 1$. In other words, the path reaches a best-response cycle in period 1. In fact, the path reaches a 6-cycle in period 1: once reached, the action profile sequence $\mathbf{a}^1, \dots, \mathbf{a}^6$ is repeated forever. Note that not all action profiles in the sequence $\mathbf{a}^1, \dots, \mathbf{a}^6$ are distinct.

In panel (C) the initial profile is once again set to $\mathbf{a}^0 = (1, 1, 1)$ but the playing sequence is 1-2-1- \dots . This time, the path reaches the PNE in period 3. The playing sequence from period 4 onwards is irrelevant: once at the PNE, the path will remain forever regardless of the playing sequence.

The examples in panels (B) and (C) illustrate how changes to the playing sequence and to the initial profile can affect convergence to a PNE. ■

2.5. Best-response dynamics with random inputs. How likely is it for best-response dynamics to converge to pure Nash equilibria? As discussed in the introduction, several papers have analyzed the convergence properties of best-response dynamics in games with a specific payoff structure (e.g. in potential games, aggregative games, etc) and our examples above illustrate how difficult it is to obtain general results for all games. In particular, the dynamics depend on the details of the game, the playing sequence, and the initial condition. Rather than imposing restrictions on the game itself, we answer our question by working

out the probability that the best-response dynamic converges to a PNE when the inputs to the dynamic are drawn at random.

We generate random games by drawing all payoffs at random: for each $\mathbf{a} \in [m]^n$ and $i \in [n]$, the payoff $U_i(\mathbf{a})$ is a random number that is drawn from an atomless distribution \mathbb{P} . The draws are independent across all $i \in [n]$ and $\mathbf{a} \in [m]^n$. The distribution \mathbb{P} ensures that any ties in payoffs have zero measure, so any resulting game is non-degenerate almost surely. A random n -player m -action game drawn according to this process is denoted by $G_{n,m} := ([n], [m], \{U_i\}_{i \in [n]})$.

In addition to the clockwork playing sequence, we will also consider the *random* playing sequence s_r which is determined as follows: for each $t \in \mathbb{N}$, draw $s_r(t)$ uniformly at random from $[n]$. So, in each period, the player playing in that period is drawn uniformly at random from among all players. In what follows, we will take a playing sequence s to be an element of $\{s_c, s_r\}$ because we will compare our results concerning the clockwork sequence against existing analytical results concerning the random playing sequence (Amiet et al., 2019).

Finally, we will draw the initial profile \mathbf{A}^0 uniformly at random from among all profiles. Since the game itself is drawn at random, the choice of initial condition is actually irrelevant, i.e. our results would not change if we had arbitrarily fixed the initial profile to some specific value. The advantage of drawing the initial profile at random is that it allows us to drop the dependence on the initial profile in our description of the best-response dynamic.

The best-response dynamic on a random game, playing sequence, and initial condition is described by Algorithm 2. We randomly draw the game and initial condition and then essentially run Algorithm 1. Doing so induces a distribution over paths and PNE sets. Our definitions of convergence given in Section 2.4 all apply here. For example, we say that the s -sequence best-response dynamic on game $G_{n,m}$ (and initial condition \mathbf{A}^0) converges to a PNE if the path $\langle \vec{\mathbf{A}}, s \rangle$ generated according to Algorithm 2 reaches $\text{PNE}(G_{n,m})$.

Algorithm 2 s -sequence best-response dynamic on $G_{n,m}$

- (1) For all $i \in [n]$ and $\mathbf{a} \in [m]^n$ draw $U_i(\mathbf{a})$ at random according to \mathbb{P}
 - (2) Draw \mathbf{A}^0 uniformly at random from $[m]^n$
 - (3) For $t \in \mathbb{N}$:
 - (a) Set $i = s(t)$
 - (b) Set $\mathbf{A}_{-i}^t = \mathbf{A}_{-i}^{t-1}$
 - (c) Set $A_i^t = B_i(\mathbf{A}_{-i}^{t-1})$ where $B_i(\mathbf{A}_{-i}^{t-1}) := \arg \max_{x_i \in [m]} U_i(x_i, \mathbf{A}_{-i}^{t-1})$
-

Step (1) of Algorithm 2 effectively creates a best-response digraph $\mathcal{D}(G_{n,m})$ on the vertices $[m]^n$ according to the following stochastic process: for each $i \in [n]$ and environment

\mathbf{a}_{-i} select an action a'_i uniformly at random from $[m]$ and then for each $a_i \neq a'_i$ create a directed edge from (a_i, \mathbf{a}_{-i}) to (a'_i, \mathbf{a}_{-i}) . This follows from the manner in which the payoffs are drawn: there is a zero probability of ties because \mathbb{P} is atomless and for each $i \in [n]$ the probability that action $a_i \in [m]$ is a best-response to environment \mathbf{a}_{-i} is given by

$$\Pr \left[U_i(a_i, \mathbf{a}_{-i}) \geq \max_{x_i \in [m]} U_i(x_i, \mathbf{a}_{-i}) \right] = \frac{1}{m}.$$

Step (2) of Algorithm 2 then selects an initial profile and step (3) essentially traces a path by traveling along the edges of the best-response digraph in direction $s(t)$ at step t starting from the initial profile.

3. THEORETICAL RESULTS

In this section we show that there is a striking difference between the probability of convergence to a PNE of the clockwork sequence vs. the random sequence best-response dynamic. Roughly, while the clockwork sequence best-response dynamic never converges to a PNE in large games, the random sequence best-response dynamic always converges in large games (conditional on there being a PNE). Our next example develops some intuition for this result.

Example (Best-response dynamics for clockwork vs. random playing sequence). In the best-response digraph of Figure 2, we saw that the clockwork playing sequence may not converge to the PNE depending on the initial profile. By contrast, the random playing sequence best-response dynamic converges to the PNE at $(1, 2, 1)$ with probability 1 given sufficient time. The path cannot be stuck in the best-response cycle on the front face of the cube for example because there is a positive probability that the path will escape to the PNE.

Figure 3 provides further examples of best-response digraphs. Panels (A) and (B) illustrate possible best-response digraphs for 3-player 2-action games. In the digraph of panel (A), there are two PNE which are represented by black dots at the profiles $(1, 1, 1)$ and $(2, 2, 1)$. In other words, these profiles are the PNE of any game whose best-response digraph is the one illustrated in panel (A). While the random sequence best-response dynamic converges to one of the PNE with probability 1, the clockwork sequence best-response dynamic will converge if and only if the initial profile is one of the four profiles at the bottom of the cube.

Since there is no PNE in the digraph of panel (B), the best-response dynamic does not converge to a PNE regardless of the playing sequence.

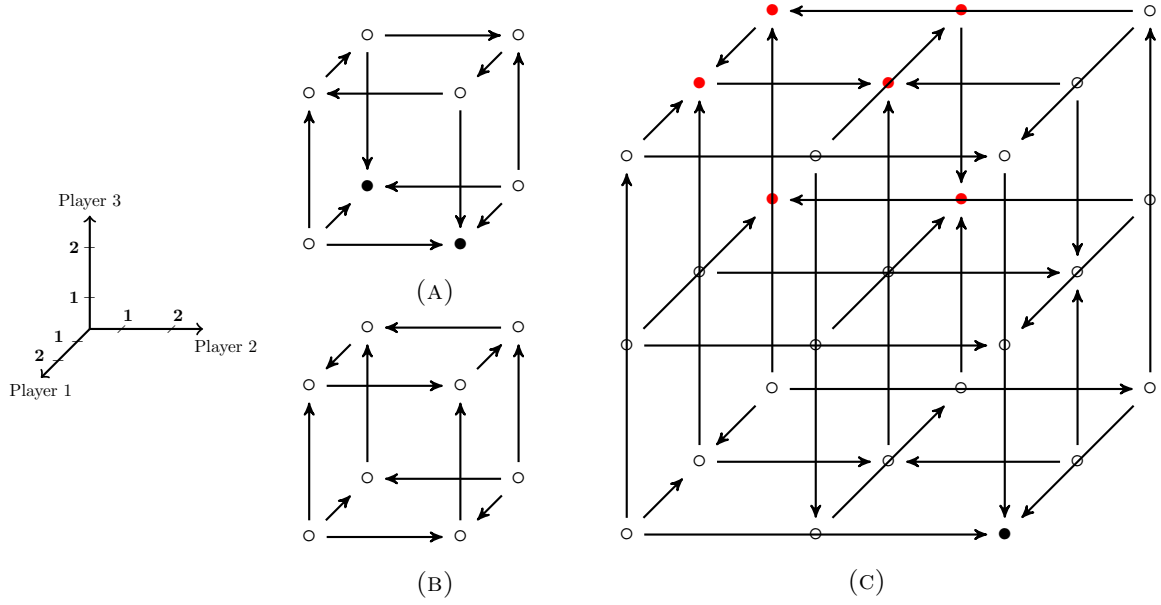


FIGURE 3. Panels (A) and (B) provide examples of possible best-response digraphs for $n = 3$ and $m = 2$. Panel (C) illustrates a possible best-response digraph for $n = m = 3$.

Finally, panel (C) illustrates a possible best-response digraph for 3-player 3-action games. The unique PNE at $(3, 3, 1)$ is represented by a black dot. Note that there is a directed edge from $(3, 1, 1)$ to $(3, 3, 1)$ as well as a directed edge from $(3, 2, 1)$ to $(3, 3, 1)$ but these edges overlap in our illustration so appear as a single long edge at the bottom of the front face of the cube. This digraph shows us a situation in which the random sequence best-response dynamic may not converge to a PNE even if there is one: indeed, if the path reaches one of the action profiles illustrated as a red dot then the path can never escape to the PNE regardless of the playing sequence. As implied by the results below, cases like this one become vanishingly rare in large games. ■

We start by noting that in random n -player m -action games, the probability that there is a pure Nash equilibrium is asymptotically $1 - \exp\{-1\} \approx 0.63$ as either n or m (or both) get large.

Proposition 1 (Rinott and Scarsini, 2000).

$$\lim_{nm \rightarrow \infty} \Pr [\# \text{PNE}(G_{n,m}) \geq 1] = 1 - \exp\{-1\}.$$

In fact, [Rinott and Scarsini \(2000\)](#) prove a much stronger result: they characterize the asymptotic distribution of the number of pure Nash equilibria in random games, showing that $\#PNE(G_{n,m})$ is asymptotically Poisson(1) as $nm \rightarrow \infty$. The probability that a PNE exists in a random game has been studied by [Goldberg et al. \(1968\)](#) in the 2-player case and by [Dresher \(1970\)](#) in the n -player case as $m \rightarrow \infty$. [Powers \(1990\)](#) and [Stanford \(1995\)](#) noted that the distribution of the number of PNE approaches a Poisson(1) as $m \rightarrow \infty$. Building on [Arratia et al. \(1989\)](#), [Rinott and Scarsini \(2000\)](#) show that the Poisson(1) limit holds as $nm \rightarrow \infty$ (i.e. when m or n get large).

Next, we present the theoretical results for best-response dynamics in random games.¹⁴ In Section 3.1 we focus on games with $n > 2$ players. In this case, we find that best-response dynamics behave very differently under clockwork vs. random playing sequences. Most of our results on the probability of convergence to equilibrium are asymptotic. In Section 3.2 we focus on games with $n = 2$ players. In this case, the probability of convergence to equilibrium is the same under both clockwork and random playing sequences. Furthermore, we are able to provide asymptotic as well as *exact* results for game duration and for the probability of convergence to equilibrium.

3.1. m -action games with $n > 2$ players. The following result shows that, in large 2-action games, the random sequence best-response dynamic converges with high probability to a PNE if there is one.

Proposition 2 ([Amiet et al., 2019](#)).

$$\lim_{n \rightarrow \infty} \Pr[s_r\text{-best-response dynamic on } G_{n,2} \text{ converges to a PNE} \mid \#PNE(G_{n,2}) \geq 1] = 1.$$

Combined with Proposition 1, it follows that over the class of all (non-degenerate) 2-action games, the random sequence best-response dynamic converges to a PNE in $(1 - \exp\{-1\}) \times 100\% \approx 63\%$ of those games when the number of players is large.

A generalization of Proposition 2 to m -action games is non-trivial and we are not aware of existing analytical results for $m > 2$.¹⁵ However, we conjecture that the random sequence

¹⁴Note that all the convergence results hold modulo any relabelling of the players. For example, while we described the clockwork playing sequence as ordering the players according to 1-2- \dots - n , our results would equally hold if the sequence had ordered the players according to $n-\dots-2-1$ or any such permutation.

¹⁵When $m = 2$, the random digraph $\mathcal{D}(G_{n,2})$ is a random n -cube in which, independently, for each pair of profiles \mathbf{a} and \mathbf{a}' that differ in exactly one coordinate, there is a directed edge from \mathbf{a} to \mathbf{a}' with probability $1/2$; otherwise, there is a directed edge from \mathbf{a}' to \mathbf{a} with complementary probability $1/2$. For such an n -cube, [Amiet et al. \(2019\)](#) show that when n is large, every pure Nash equilibrium belongs to the set of vertices that are reachable by some directed path from the initial action profile \mathbf{a}^0 . This is sufficient to show that, when the number of players is large, the random sequence best-response dynamic converges with

best-response dynamic converges to a PNE with high probability if there is one as $mn \rightarrow \infty$. Consistent with this conjecture, in the simulations of Section 4 we show that the random sequence best-response dynamic does converge to a PNE with probability close to $1 - \exp\{-1\}$ when m or n get large, provided that $n > 2$.

Our main result for the clockwork sequence best-response dynamic in games with $n > 2$ players is given below.

Theorem 1.

$$\lim_{nm \rightarrow \infty} \Pr[s_{\text{c}}\text{-best-response dynamic on } G_{n,m} \text{ converges to a PNE}] = 0.$$

So, with high probability, the clockwork sequence best-response dynamic does not converge to a PNE as the number of players or actions gets large. This is in sharp contrast with the asymptotic behavior of the random sequence best-response dynamic.

Theorem 1 is an immediate consequence of the result below, which gives us bounds on the probability of convergence to equilibrium:

Theorem 2.

$$\frac{1}{4\sqrt{n}} \frac{1}{\sqrt{m^{n-1}}} \leq \Pr \left[\begin{array}{c} s_{\text{c}}\text{-best-response dynamic} \\ \text{on } G_{n,m} \text{ converges to a PNE} \end{array} \right] \leq \frac{7n^{3/2}\sqrt{\log m}}{\sqrt{m^{n-1}}}.$$

Theorem 2 also gives us the following corollary:

Corollary 1. *The probability that the clockwork sequence best-response dynamic converges to a PNE is, up to a polynomial factor, of order $1/\sqrt{m^{n-1}}$.*

This result gives us a clear “scaling” law: since the asymptotic convergence probability depends essentially only on the quantity m^{n-1} , we have that, when n and/or m are large, the probability of convergence to a pure Nash equilibrium in n -player m -action games is approximately the same as it is in 2-player m^{n-1} -action games.¹⁶ This scaling is reflected in our simulations even for relatively small values of m and n .

We briefly comment on the approach that we take in the appendix to derive Theorem 2. As is clear from the discussion following Algorithm 2, drawing payoffs independently

probability 1 to a pure Nash equilibrium if there is one: indeed, the random playing sequence ensures that some path to equilibrium is played given sufficient time. The edges in $\mathcal{D}(G_{n,2})$ are oriented in one way or the other independently of each other but this is no longer true when $m > 2$. In $\mathcal{D}(G_{n,m})$ with $m > 2$, if there is a directed edge from (a_i, \mathbf{a}_{-i}) to (a'_i, \mathbf{a}_{-i}) for some $a_i \neq a'_i$ then the graph must also have the directed edges (a''_i, \mathbf{a}_{-i}) to (a'_i, \mathbf{a}_{-i}) for all $a''_i \neq a'_i$. This dependence and the more complex graph structure renders a generalization of Proposition 2 to $m > 2$ non-trivial.

¹⁶Indeed, consider G_{n_1, m_1} and G_{n_2, m_2} where $n_2 = 2$ and $m_2 = m_1^{n_1-1}$. Then, $\sqrt{m_2^{n_2-1}} = \sqrt{m_1^{n_1-1}}$.

at random (from atomless distributions) induces a uniform distribution over best-response digraphs. So, the probability of convergence to a pure Nash equilibrium can be reduced to working out the probability that the path generated by Algorithm 2 initiated at a random vertex reaches a sink of the randomly drawn digraph. The main theoretical challenge that we face when analyzing such paths is that they exhibit some path-dependence: if a player encounters an environment that they had seen before, they must play the same action that they played when the environment was first encountered. We tackle this issue by relying on a coupling argument in which the clockwork best-response dynamic is coupled to a (memoryless) random walk through the digraph that is easier to analyze.

We are not aware of any analytical results on the probability of convergence to equilibrium in random games for the best-response dynamic under simultaneous updating. Obtaining results for simultaneous updating is non-trivial because the pattern of path-dependence is more complex than it is for the clockwork best-response sequence. See footnote 28 for a more detailed comment.

3.2. m -action games with $n = 2$ players. When there are $n = 2$ players, we are able to provide detailed results on both game duration and on the probability of convergence to equilibrium.

The following theorem gives us an *exact* expression for the probability that the clockwork sequence best-response dynamic converges to a $2k$ -cycle in period t .¹⁷

Theorem 3. *For any $k \in \{1, \dots, m\}$ and $t \in \{1, \dots, 2(m - k + 1)\}$,*

$$(1) \quad \Pr \left[\begin{array}{l} s_{\text{c}}\text{-best-response dynamic on } G_{2,m} \\ \text{converges to a } 2k\text{-cycle in period } t \end{array} \right] = \frac{1}{m} \prod_{i=1}^{t+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right).$$

For any $k \in \{1, \dots, m\}$, if $t > 2(m - k + 1)$ then the probability that the clockwork sequence best-response dynamic reaches a $2k$ -cycle is zero.¹⁸ Setting $k = 1$ in (1) yields the exact probability that the clockwork sequence best-response dynamic on $G_{2,m}$ converges to a PNE in period t .

As a straightforward corollary of Theorem 3, we can sum (1) over all $t \in \{1, \dots, 2(m - k + 1)\}$ to obtain the exact probability that the clockwork sequence best-response dynamic converges to a $2k$ -cycle:

¹⁷See also Pangallo et al. (2019) for an exact formula giving the probability of existence of cycles of any length in 2-player games.

¹⁸Since the number of action profiles is finite, a path cannot reach a $2k$ -cycle only after $2(m - k + 1)$ periods.

Corollary 2.

$$(2) \quad \Pr \left[\begin{array}{l} s_c\text{-best-response dynamic on} \\ G_{2,m} \text{ converges to a } 2k\text{-cycle} \end{array} \right] = \frac{1}{m} \sum_{t=1}^{2(m-k+1)} \prod_{i=1}^{t+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right).$$

Setting $k = 1$ in (2) yields the exact probability that the clockwork sequence best-response dynamic on $G_{2,m}$ converges to a PNE.

In order to get a better sense of the behavior of (2), we now study the expression when m is large. To do so, let $\lfloor x \rfloor$ be the floor operator of x and let $\Phi(\cdot)$ denote the standard normal cumulative distribution function:

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left\{ -\frac{z^2}{2} \right\} dz.$$

The asymptotics of (2) are given below.

Proposition 3. *If $k = o(m^{2/3})$ then, as $m \rightarrow \infty$, (2) is asymptotically¹⁹*

$$2\sqrt{\frac{\pi}{m}} \left(1 - \Phi \left(\frac{2k-1}{\sqrt{2m}} \right) \right).$$

And if $k = o(\sqrt{m})$ then, as $m \rightarrow \infty$, (2) is asymptotically $\sqrt{\pi/m}$.

The asymptotics given in Proposition 3 help us to better understand the behavior of the clockwork sequence best-response dynamic in large 2-player games. (i) The probability of convergence to a PNE, which corresponds to setting $k = 1$, goes to zero when $m \rightarrow \infty$. (ii) Short cycles all have about the same probability. Indeed, for $k = o(\sqrt{m})$ the probability is asymptotically $\sqrt{\pi/m}$. Finally, (iii) it is very unlikely that the best-response dynamic converges to a very long cycle: if $k/\sqrt{m} \rightarrow \infty$ then the probability that the dynamic converges to a cycle of length at least $2k$ tends to 0.²⁰

Our results are illustrated in Figure 4, which shows the probability of convergence to cycles of given length as calculated from the exact formula in Theorem 3. The panels on the left and in the center plot (2) for $m = 5$ and $m = 10$ and show that the probability of convergence is lower for long cycles than for short cycles. The panel on the right plots (2) for $m = 1000$. Since $31 \approx \sqrt{1000}$, we placed a vertical line at $2k = 62$. The probability

¹⁹ $f(n) = o(g(n))$ denotes $f(n)/g(n) \rightarrow 0$ as $n \rightarrow \infty$.

²⁰When $k = o(\sqrt{m})$, the argument of $\Phi(\cdot)$ goes to zero. Since $\Phi(0) = 1/2$ we have that the convergence probability goes to $\sqrt{\pi/m}$ which is independent of k . If, instead, $k/\sqrt{m} \rightarrow \infty$ then the argument of $\Phi(\cdot)$ grows large and since $\Phi(\infty) = 1$, the convergence probability goes zero. Our proof of Proposition 3 allows us to derive the asymptotics only for the range $k = o(m^{2/3})$, but this is sufficient to obtain some insight into the behavior of (2).

of convergence is more or less uniform up to that point, which is consistent with our observations above.

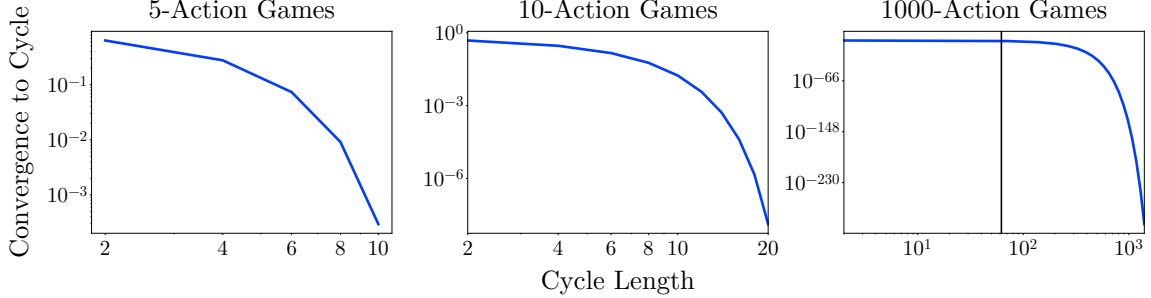


FIGURE 4. Convergence of the clockwork sequence best-response dynamic to $2k$ -cycles in 2-player games, using the exact formula from Theorem 3.

We now compare the behavior of the clockwork sequence best-response dynamic in 2-player games with the behavior of the random sequence best-response dynamic in 2-player games. (i) The probability of convergence to a PNE is the same for clockwork and for random playing sequences in 2-player games. The reason is that, under the random playing sequence, players' actions do not change whenever the sequence asks the same player to play several times in a row. The profiles that are therefore visited along the path are the same under both playing sequences, which induces the same probability of convergence to equilibrium. However, (ii) the game duration will be different since the random playing sequence introduces delays. In fact, the result below allows us to more precisely pin down game duration.

Theorem 4. *The probability that the s_c -best-response dynamic on $G_{2,m}$ does not converge to a PNE or to a best-response cycle before period $x\sqrt{2m}$ is $\exp\{-x^2/2\}$ as $m \rightarrow \infty$.*

This result shows that the clockwork sequence best-response dynamic in 2-player games is likely to converge to a PNE or to a best-response cycle within $\sqrt{2m}$ periods when m is large. The game duration for the random playing sequence should be greater than for the clockwork playing sequence by a factor of 2. The reason is that, under the clockwork playing sequence, the players alternate at the tick of each period whereas, under the random playing sequence, the number of periods that it takes for the playing sequence to turn to the other player is $\text{Geometric}(\frac{1}{2})$. Thus the random playing sequence can be considered as a slowing down of the clockwork playing sequence in which the expected time to play the next step is 2.

4. SIMULATION RESULTS

In this section, we run simulations of the clockwork sequence best-response dynamic. This allows us to compare its behavior against the best-response dynamic under random and simultaneous updating and against other learning rules (reinforcement learning, fictitious play, and replicator dynamics).

For our theoretical results, we limited ourselves to analyzing best-response dynamics in random games where the payoffs are drawn independently across players and action profiles. For our simulations, we allow the payoffs to be correlated across players (Goldberg et al., 1968, Stanford, 1999, Berg and Weigt, 1999, Rinott and Scarsini, 2000, Galla and Farmer, 2013, Sanders et al., 2018). To do so, at initialization, for each action profile \mathbf{a} we draw the vector $U(\mathbf{a}) = (U_1(\mathbf{a}), \dots, U_n(\mathbf{a}))$ at random from a multivariate normal distribution with mean zero, unit variance, and covariance matrix with 1s on the diagonal and $\frac{\Gamma}{n-1}$ on the other entries. So $\Gamma \in [-1, n-1]$ parametrizes the degree to which payoffs are correlated. Once drawn, the payoffs are kept fixed for the rest of the simulation. If $\Gamma = 0$ then the payoffs are chosen independently, so this recovers the case for which we derived our theoretical results. At one extreme, if $\Gamma = n-1$ then at each action profile all players receive the same payoff and, at the other extreme, if $\Gamma = -1$ then we have a zero-sum game. More generally, when $\Gamma > 0$ the game is “cooperative” and when $\Gamma < 0$ the game is “competitive” (Rinott and Scarsini, 2000).

In all simulations, for each combination of n , m , and Γ , we draw 500 games and simulate for 5000 time steps starting from randomly chosen initial conditions. In Section 4.1 we simulate the clockwork sequence best-response dynamic in random n -player m -action games with independent payoffs (i.e. $\Gamma = 0$). This allows us to verify our theoretical results. In Section 4.2 we compare the behavior of the clockwork sequence best-response dynamic against the best-response dynamic under random and simultaneous updating for different values of n , m , and Γ . In Section 4.3 we compare the behavior of the best-response dynamic (with clockwork, random, and simultaneous updating) against other learning rules (reinforcement learning, fictitious play, and replicator dynamics) for different values of n , m , and Γ .

4.1. Simulations of clockwork best-response dynamics. We simulate the clockwork sequence best-response dynamic in n -player m -action games with $\Gamma = 0$. We find good agreement with our main theoretical results and we show that Corollary 1, which states that the asymptotic probability of convergence to a PNE is, up to a polynomial factor, of order $1/\sqrt{m^{n-1}}$, is also reflected in our simulations for relatively small values of m and n .

The blue markers in Figure 5 show the frequency of convergence to a PNE in our simulations for different values of n and m . Clearly, the frequency of convergence to a PNE decreases as the number of players and/or actions increases. The solid black line in the top panel is the analytical probability of convergence to a PNE in 2-player games which is calculated using equation (2). Up to sampling noise, our analytical result perfectly matches the numerical simulations.

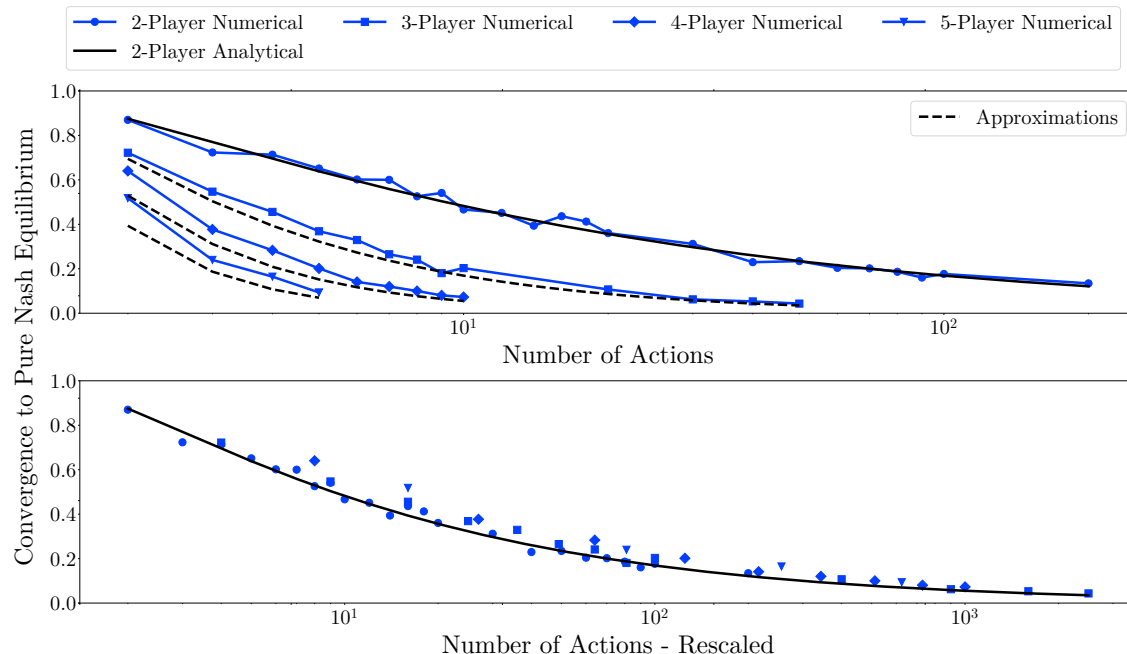


FIGURE 5. Frequency of convergence to a PNE for the clockwork best-response dynamic with $\Gamma = 0$. In the bottom panel, the horizontal axis is rescaled according to Corollary 1.

Figure 5 also allows us to verify Corollary 1; namely, that the frequency of convergence to a PNE in a n -player m -action game is roughly the same as in a 2-player m^{n-1} action game. While in Section 3 this result was only proved asymptotically ($mn \rightarrow \infty$), we investigate the extent to which the result holds for small values of m and n . Using Corollary 1, we approximate the frequency of convergence to a PNE in a n -player m -action game by replacing the number of actions m in equation (2) by m^{n-1} , which is the equivalent number of actions in the corresponding 2-player game. We plot these approximate frequencies as dashed lines in the top panel of Figure 5. As can be seen, there is a good match between the approximation and our simulation results, particularly when the number of actions m is

relatively large. The bottom panel of Figure 5 gives us another way to illustrate Corollary 1. Here, we rescale the number of actions for n -player games to match the number of actions of the equivalent 2-player game. After rescaling, a point corresponding to m actions in an n -player game is moved on the horizontal axis to a number of actions given by m^{n-1} . For example, the point giving the convergence frequency for 4-player 10-action games is translated to the right to the horizontal coordinate corresponding to $10^3 = 1000$ actions in a 2-player game. The re-scaled markers all lie relatively close to the black line – which corresponds to the analytical probability of convergence to a PNE in 2-player m -action games.

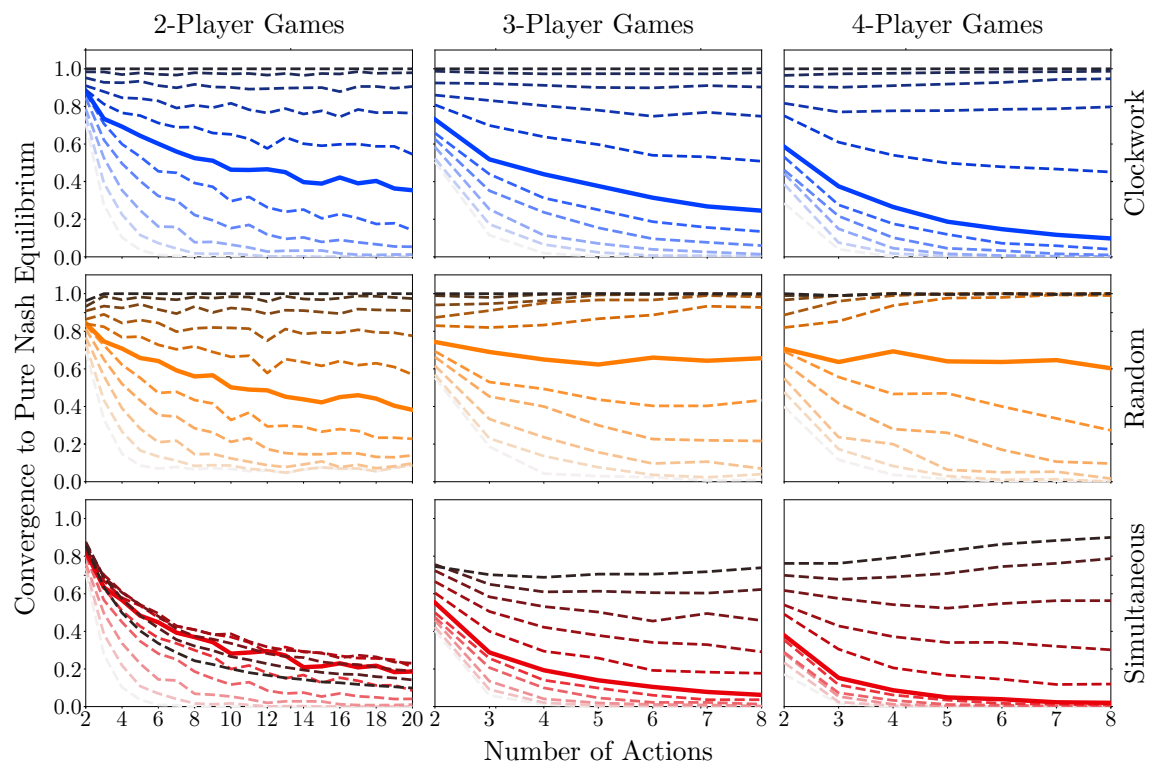


FIGURE 6. Frequency of convergence to pure Nash equilibria under clockwork, random, and simultaneous best-response dynamics. The solid line corresponds to $\Gamma = 0$, the darker dashed lines to $\Gamma = \frac{n-1}{5}, 2\frac{n-1}{5}, \dots, (n-1)$, and the lighter dashed lines to $\Gamma = -0.2, -0.4, \dots, -1$.

4.2. Simulations of best-response dynamics under clockwork, random, and simultaneous updating. We simulate best-response dynamics under clockwork, random,

and simultaneous updating. We find that there are significant differences in the probability of convergence to a PNE when $\Gamma = 0$. When comparing clockwork against random sequences, the differences are consistent with the theoretical findings of Section 3. When $\Gamma \neq 0$, we find that the differences in the probability of convergence to a PNE become more muted but, overall, best-response dynamics converge to a PNE most frequently under a random sequence and least frequently under simultaneous updating, with the clockwork case lying somewhere in between.

Figure 6 shows the frequency of convergence to a PNE under clockwork, random, and simultaneous best-response dynamics. (i) The solid line corresponds to $\Gamma = 0$, (ii) the darker lines correspond to positive correlations, $\Gamma = \frac{n-1}{5}, 2\frac{n-1}{5}, \dots, (n-1)$, and (iii) the lighter lines correspond to negative correlations, $\Gamma = -0.2, -0.4, \dots, -1$. We discuss each case in turn:

- (i) Uncorrelated payoffs: $\Gamma = 0$. The frequency of convergence to a PNE is decreasing in n and m for the clockwork playing sequence and for simultaneous updating. The random playing sequence is different. When there are only $n = 2$ players, the random playing sequence has the same convergence probability as the clockwork playing sequence – as argued in Section 3.2. Amiet et al. (2019) proved that the random sequence best-response dynamic always converges to a PNE if there is one when $m = 2$ and $n \rightarrow \infty$. As argued in Section 3.1, this gives us an unconditional probability of convergence of $1 - 1/e \approx 63\%$. Our simulations show that the result of Amiet et al. (2019) also appears to hold for games with more than two actions provided $n > 2$. In fact, the random sequence best-response dynamic almost always converges to a PNE in games that have a PNE (even for relatively small values of n and m).
- (ii) Positively correlated payoffs: $\Gamma > 0$. For any value of m and n convergence tends to be more likely than with $\Gamma = 0$ under all playing sequences. The reason is that under positively correlated payoffs (especially if the correlation is very strong) there is a proliferation of pure Nash equilibria (Goldberg et al., 1968, Stanford, 1999, Berg and Weigt, 1999, Rinott and Scarsini, 2000). The best-response dynamic is therefore very likely to converge to one of these equilibria. The only exception is the simultaneous best-response dynamic in 2-player games with highly correlated payoffs (e.g. $\Gamma = 0.9, 1$). For some intuition, consider a 2-player 2-action coordination game (in which payoffs are strongly positively correlated). If players start at one of the two action profiles that are not pure Nash equilibria, they keep bouncing back and forth between these two action profiles forever.

- (iii) Negatively correlated payoffs: $\Gamma < 0$. For any value of m and n convergence to a PNE tends to be less likely than with $\Gamma = 0$ under all playing sequences. When $\Gamma \approx -1$ we essentially generate zero-sum games (Goldberg et al., 1968, Rinott and Scarsini, 2000). Such games do not have pure Nash equilibria, so no version of the best-response dynamics converges to a PNE.

4.3. Simulation of other learning rules. In this section we compare the behavior of best-response dynamics (under clockwork, random, and simultaneous updating) against three classic learning rules: Bush-Mosteller reinforcement learning (Bush and Mosteller, 1953), fictitious play (Brown, 1951, Robinson, 1951), and replicator dynamics (Maynard Smith, 1982). Our interest in these rules stems from the fact that they are well-known, they embody different behavioral assumptions about learning in games, and they have been calibrated to human game-play in experiments. The upshot of our simulation results is that, compared with random and simultaneous updating, the convergence properties of the clockwork best-response dynamic most closely match the convergence properties of the three learning rules.

4.3.1. Description of the learning rules. Here, we provide high-level descriptions of our three learning rules (reinforcement learning, fictitious play, and replicator dynamics) and of the convergence criteria that we use in our simulations. More detailed descriptions of the rules and of the convergence criteria are given in Appendix C.

Bush-Mosteller reinforcement learning is based on the idea that players are more likely to play actions that yielded a better payoff in the past. It is a standard learning algorithm that is used to model game playing under limited information and/or without sophisticated reasoning, such as in animal learning. Variants of reinforcement learning models have been calibrated to human game-play in experiments in Arthur (1991), Erev and Roth (1998), and Sarin and Vahid (2001). Under Bush-Mosteller learning, in each period, each player chooses their action by sampling according to a mixed strategy vector whose evolution is governed by reinforcement learning. We assess convergence of these vectors, i.e. whether the difference from one period to the next falls below a threshold and becomes indistinguishable from sampling noise. Tracking the mixed strategy vectors rather than the actions played makes it possible for us to determine whether the dynamic converges to mixed Nash equilibria.

Fictitious play requires more sophistication, as it assumes that the players construct a mental model of their opponent. Each player assumes that the empirical distribution of her opponent’s past actions is her mixed strategy, and plays the best response to this belief.

Classical experiments with human players in which fictitious play is used as a learning model are those by [Cheung and Friedman \(1997\)](#) who consider coordination, dominance-solvable, and cyclic 2-player 2-action games. To assess convergence, we follow [Fudenberg and Levine \(1998\)](#) in tracking the convergence of the belief vectors rather than the convergence of actions played. As with Bush-Mosteller learning, this choice makes it possible to include convergence to mixed equilibria in our analysis.

The replicator equation is commonly used in ecology and population biology, but it has also been viewed as a learning algorithm in which each population trait corresponds to an action ([Börger and Sarin, 1997](#)).²¹ In our implementation, in each period, each player chooses their action by sampling according a mixed strategy vector whose evolution is governed by the replicator equation. [Van Huyck et al. \(1995\)](#) study two tacit bargaining games and show that players' behavior is in line with what they would do if they were playing replicator dynamics. ([Friedman \(1996\)](#) comes to a similar conclusion in a larger sample of games.) As above, we track the convergence of the mixed strategy vectors. Note, however, that the multi-population replicator dynamic never converges to mixed equilibria in random games. In other words, if the dynamic converges to an equilibrium, each mixed strategy vector will assign all the mass to a single action.

The convergence properties of our three learning algorithms have been studied theoretically. It is well-known, for instance, that fictitious play converges to Nash equilibrium in certain classes of games such as potential, zero-sum, and supermodular games ([Fudenberg and Levine, 1998](#)). It is also well-known that evolutionarily stable strategies are locally stable fixed points of replicator dynamics ([Hofbauer and Sigmund, 1998](#)). However, there is no general result about the probability of convergence of these learning rules to pure Nash equilibria in random games.

4.3.2. Results. We compare the probability of convergence of best-response dynamics (under clockwork, random, and simultaneous updating) against each of the three learning rules. In [Figure 7](#) we do this for uncorrelated payoffs so $\Gamma = 0$, for $n = 2$ and 3 players, and for a varying number of actions m . In [Figure 8](#) we allow Γ to vary but fix the number of actions to $m = 5$. Finally, in [Figure 9](#) we allow n , m , and Γ to all vary.

In our comparisons, we distinguish between cases in which Bush-Mosteller learning and fictitious play converge to pure vs. mixed Nash equilibria. In [Figures 7 and 8](#), we use a dashed line to indicate the frequency of convergence to pure equilibria only, while a solid line indicates convergence to any type of equilibrium. Thus, for any number of actions,

²¹We consider a multi-population version of the replicator dynamic because our randomly drawn payoff matrices are, in general, not symmetric.

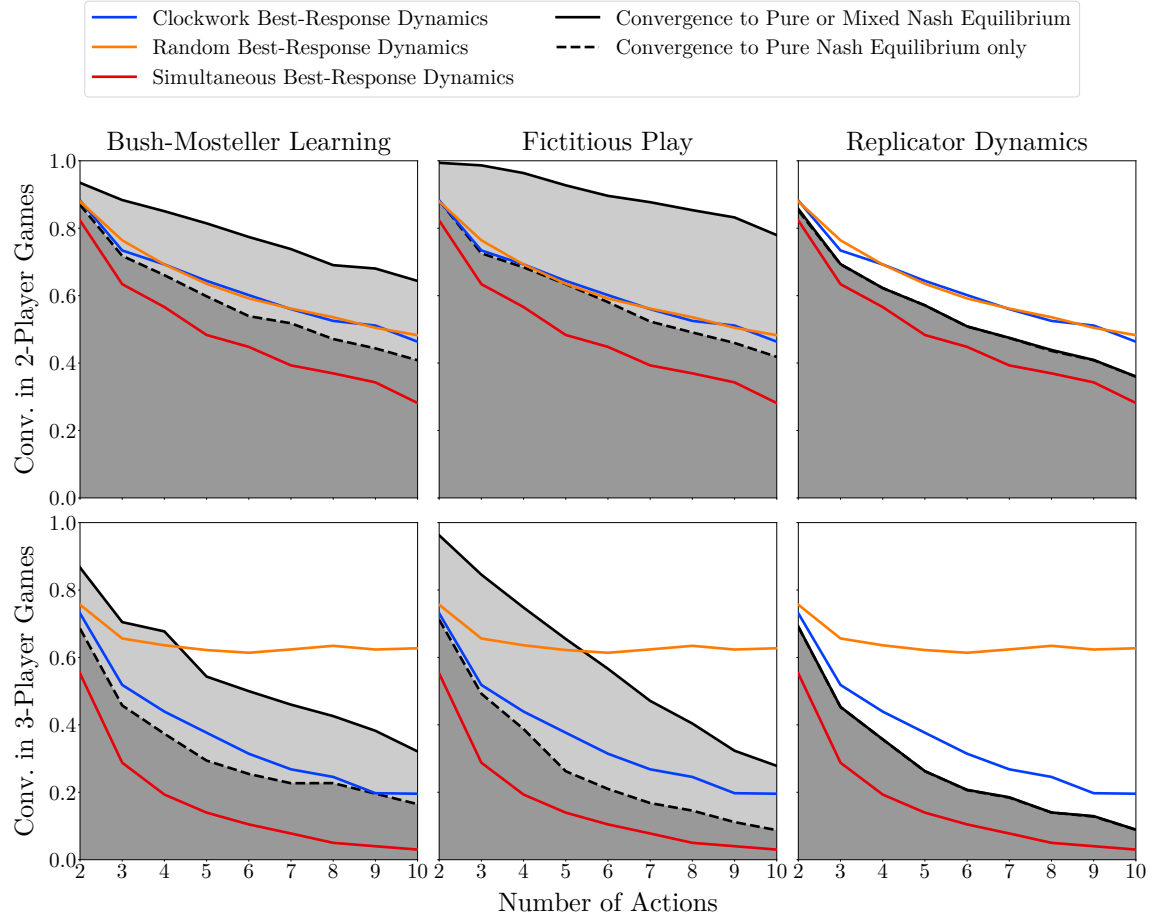


FIGURE 7. The frequency of convergence to PNE under best-response dynamics compared to the frequency of convergence to (mixed and pure) Nash equilibria under the other learning rules for $\Gamma = 0$.

the vertical distance between the dashed line and the solid line indicates the frequency of convergence to mixed equilibria only.

As can be seen in Figure 7, the frequency of convergence to a PNE under the clockwork best-response dynamic most closely tracks the frequency of convergence to PNE under the three learning rules. In games with 3 or more players, the random sequence best-response dynamic almost always converges to a PNE if there is one, which is why the orange line rapidly flattens at $1 - 1/e \approx 0.63$. By contrast, the frequency of convergence to pure (or mixed) Nash equilibria under the three learning rules decreases as the number of actions

grows. The best-response dynamic under simultaneous updating converges to a PNE too infrequently relative to the three learning rules.²²

When also considering convergence to mixed Nash equilibria, we see that the clockwork best-response dynamic converges too infrequently compared to the three learning rules, especially in the case of fictitious play with two players. However, (i) as the number of actions increases, it tracks the trend in convergence of the learning rules to mixed or pure Nash equilibria better than random best-response dynamics when there are three or more players, and (ii) its frequency of convergence to equilibrium is closer to that of the three learning rules as compared to the simultaneous best-response dynamic.

In Figure 8 we fix the number of actions to $m = 5$ and vary Γ .²³ The clockwork and random sequence best-response dynamics tend to track the other learning rules relatively well (though the clockwork sequence appears to outperform the random sequence). Note, however, that the best-response dynamic under simultaneous updating converges too infrequently. This is consistent with our previous observation that even when there are many Nash equilibria (under strongly positively correlated payoffs), this version of the dynamic will not converge as often as the other versions.

Figure 9 shows us scatter plots of the frequency of convergence to a PNE for the best-response dynamic (under each playing sequence) against the frequency of convergence to pure or mixed equilibria for each of the three learning rules. Each dot corresponds to the convergence frequency for a particular value of n , m , and Γ .²⁴ The identity line is plotted for reference. The frequency of convergence to a PNE for the clockwork sequence best-response dynamic does not perfectly match the frequency of convergence to pure or mixed equilibria for the three learning rules, but it does appear to outperform the other versions of the best-response dynamic that we have considered in this paper. We emphasize that, in Figure 9, Bush-Mosteller learning and fictitious play are allowed to converge to either pure or mixed equilibria. If we had considered convergence to pure equilibria *only* for each of our learning rules, then the clockwork sequence best-response dynamic would match the outcomes of the three learning rules even more closely.

²²There is an offset between the frequency of convergence to a PNE for the replicator dynamic and the frequency of convergence to a PNE for the clockwork best-response dynamic. As we explain in the appendix, this is mainly due to numerical limitations: the replicator dynamic has infinite memory, so a trajectory might hit the machine precision limit without having reached a PNE.

²³Note that for the 2-player case with $\Gamma = -1$, the frequency of convergence to any equilibrium (pure or mixed) for fictitious play is close to one. This is consistent with existing theoretical results regarding the convergence of fictitious play in 2-player zero-sum games (Fudenberg and Levine, 1998).

²⁴The 418 combinations for the values of n , m , and Γ that we consider are: $n = 2$ with $m = 2, \dots, 15$; $n = 3$ with $m = 2, \dots, 10$; $n = 4$ with $m = 2, \dots, 8$; $n = 5$ with $m = 2, \dots, 5$; $n = 6$ with $m = 2, 3$; $n = 7$ with $m = 2, 3$; each for $\Gamma = -1, -0.8, \dots, 0, \frac{n-1}{5}, 2\frac{n-1}{5}, \dots, (n-1)$.

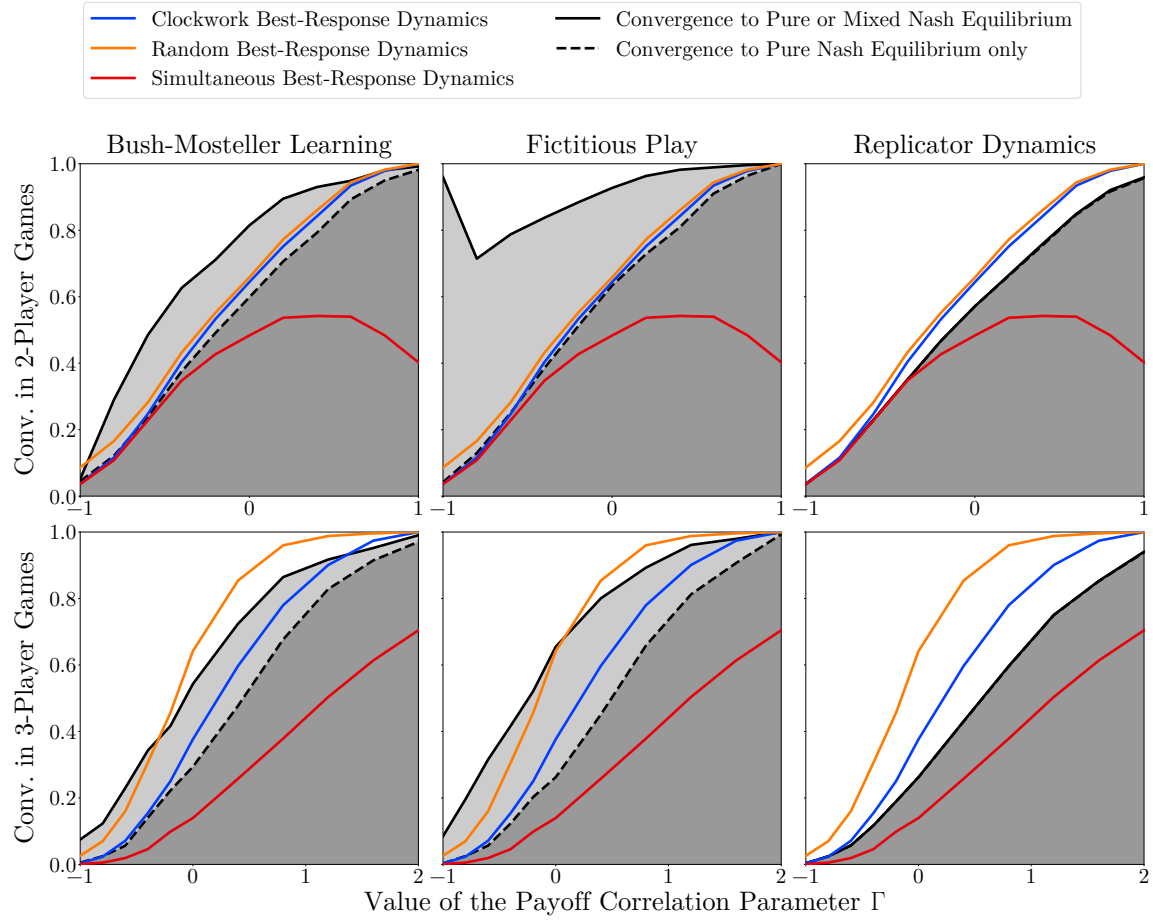


FIGURE 8. The frequency of convergence to PNE under best-response dynamics compared to the frequency of convergence to (mixed and pure) Nash equilibria under the three learning rules, with varying Γ and $m = 5$.

We now compare how the clockwork best-response dynamic performs against the three learning rules not only in terms of convergence probability but also in terms of the evolution of play itself. Figure 10 shows a best-response digraph as well as the paths traced by Bush-Mosteller learning, fictitious play, and the replicator dynamic starting from various initial conditions.²⁵ The paths show the evolution of the mixed strategy vectors for the learning rules, and these appear to follow the directions of the edges in the best-response digraph.

²⁵The underlying game has binary payoffs of zero or one. The trajectories in the action profile space would be distorted if the payoff values had been different, though Pangallo et al. (2019) show that the patterns exhibited by the learning algorithms in two player games are quite robust to general payoff values.

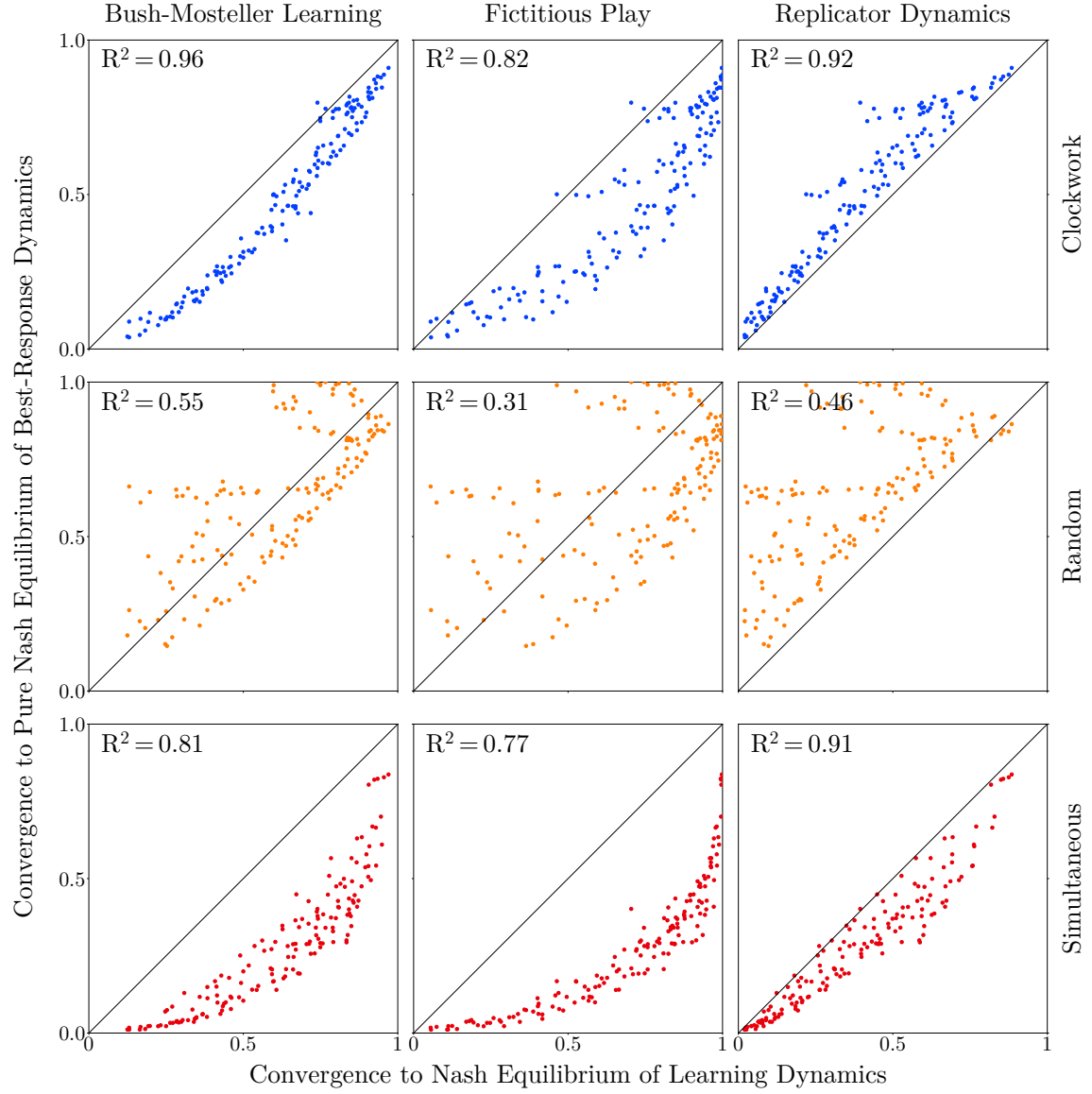


FIGURE 9. The frequency of convergence to PNE under best-response dynamics against the frequency of convergence to (mixed or pure) Nash equilibria under the three learning rules, for varying values of n , m , and Γ and $m = 5$.

These edges also govern the evolution of play in the clockwork best-response dynamic, but they do not govern the evolution of play under a random sequence. In fact, the random sequence best-response dynamic would eventually converge to the pure Nash equilibrium

Bush-Mosteller Learning

Fictitious Play

Replicator Dynamics

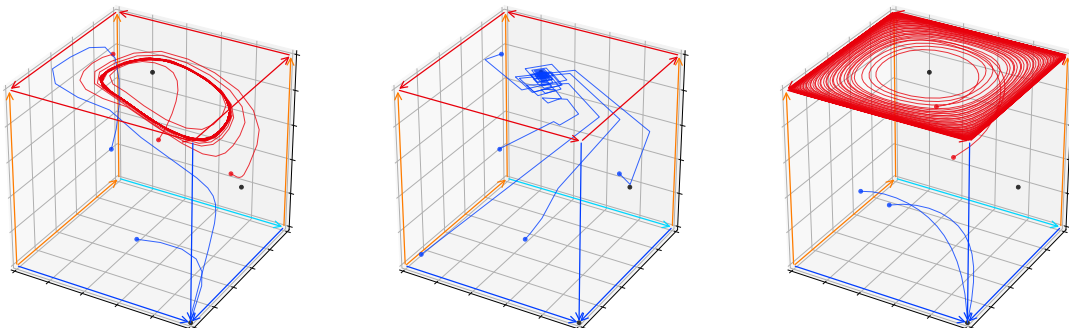


FIGURE 10. Trajectories of Bush-Mosteller learning, fictitious play, and replicator dynamics. Blue trajectories converge to Nash equilibria (pure or mixed), red trajectories do not. Blue arrows correspond to pure Nash equilibria, light blue arrows lead there; red arrows correspond to cycles, orange arrows lead there.

in this digraph, given sufficient time. So, the paths traced by the clockwork sequence best-response dynamic more closely resemble the paths traced by the three learning algorithms than those traced by the random sequence best-response dynamic, and this is true in spite of the fact the three learning algorithms are most naturally defined as involving simultaneous updating.²⁶ Our conclusion regarding how “close” the paths of the clockwork vs. random sequence best-response dynamics are to those exhibited by the learning algorithms is based on our observations in a number of games.²⁷ And, our results corroborate [Pangallo et al. \(2019\)](#) who find that the prevalence of $2k$ -cycles is a good predictor of the frequency of convergence to Nash equilibrium of the learning algorithms in 2-player random games. More generally our findings suggest that, to the extent that the learning algorithms are consistent with human game-play in randomly-generated games, the clockwork best-response dynamic could provide a first-order approximation for the evolution of play in such games.

²⁶The paths traced by the learning algorithms are likely to have features resembling elements of the paths traced by the best-response dynamic under both clockwork and simultaneous updating. The degree to which the learning algorithms have “memory” is likely to modulate the extent to which the paths resemble those generated by the best-response dynamic under clockwork vs. simultaneous updating.

²⁷We do not carry out a comprehensive quantitative analysis of “path closeness” though we expect our finding regarding clockwork vs. random sequence best-response dynamics to be robust, particularly for large games.

APPENDIX A. PROOF OF THEOREM 2

We start by stating two lemmas that will be used to prove Theorem 2. Lemma 1 bounds the probability that the clockwork sequence best-response dynamic converges to a pure Nash equilibrium or to a best-response cycle only after period t . Lemma 2 bounds the probability that the clockwork sequence best-response dynamic converges to a pure Nash equilibrium by period t .

Lemma 1. *Let $\langle \vec{\mathbf{A}}, s_c \rangle$ be generated according to Algorithm 2. For any $t \in \mathbb{N}$,*

$$\Pr [T_{\langle \vec{\mathbf{A}}, s_c \rangle} > t] \leq \exp \left\{ -\frac{(\lfloor \frac{t}{n} - 1 \rfloor)^2}{2m^{n-1}} \right\}.$$

Recall that $T_{\langle \vec{\mathbf{A}}, s_c \rangle}$ is the period in which $\langle \vec{\mathbf{A}}, s_c \rangle$ reaches $\text{PNE}(G_{n,m})$ or a best-response cycle.

Lemma 2. *Let $\langle \vec{\mathbf{A}}, s_c \rangle$ be generated according to Algorithm 2. For any $t \in \mathbb{N}$,*

$$\left\lceil \frac{t}{n} \right\rceil \frac{1}{m^{n-1}} \left(1 - \frac{n}{m^{n-1}} \frac{(\lceil \frac{t}{n} \rceil)^2}{2} \right) \leq \Pr [\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \text{PNE}(G_{n,m}) \text{ by } t] \leq \frac{t}{m^{n-1}}.$$

We now show how Theorem 2 follows from Lemmas 1 and 2. In what remains of this section we provide proofs for the lemmas themselves.

Proof of Theorem 2. Let $\langle \vec{\mathbf{A}}, s_c \rangle$ be generated according to Algorithm 2. The probability that the s_c -best-response dynamic on $G_{n,m}$ converges to a PNE is equal to the probability that $\langle \vec{\mathbf{A}}, s_c \rangle$ reaches $\text{PNE}(G_{n,m})$. Let us start with the upper bound. For any $t \in \mathbb{N}$,

$$\begin{aligned} & \Pr [\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \text{PNE}(G_{n,m})] \\ &= \Pr [\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \text{PNE}(G_{n,m}) \text{ by } t] + \Pr [\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \text{PNE}(G_{n,m}) \text{ after } t] \\ &\leq \Pr [\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \text{PNE}(G_{n,m}) \text{ by } t] + \Pr [T_{\langle \vec{\mathbf{A}}, s_c \rangle} > t] \\ (3) \quad &\leq \frac{t}{m^{n-1}} + \exp \left\{ -\frac{(\lfloor \frac{t}{n} - 1 \rfloor)^2}{2m^{n-1}} \right\}. \end{aligned}$$

Equation (3) follows from Lemmas 1 and 2. Now, set

$$t = n \left(\left\lceil \sqrt{2m^{n-1} \log(m^{n-1})} \right\rceil + 1 \right).$$

Since $x \leq \lceil x \rceil \leq x + 1$ and $\sqrt{2m^{n-1} \log(m^{n-1})} > 1$ for $m \geq 2$ and $n \geq 2$, we obtain

$$n \left(\sqrt{2m^{n-1} \log(m^{n-1})} + 1 \right) \leq t \leq n \left(\sqrt{2m^{n-1} \log(m^{n-1})} + 2 \right) < 3n \sqrt{2m^{n-1} \log(m^{n-1})}.$$

It follows that

$$(4) \quad \frac{t}{m^{n-1}} < \frac{3n\sqrt{2m^{n-1}\log(m^{n-1})}}{m^{n-1}} < \frac{5n^{3/2}\sqrt{\log m}}{\sqrt{m^{n-1}}},$$

and that

$$(5) \quad \exp\left\{-\frac{(\lfloor \frac{t}{n} \rfloor - 1)^2}{2m^{n-1}}\right\} \leq \frac{1}{m^{n-1}} < \frac{2n^{3/2}\sqrt{\log m}}{\sqrt{m^{n-1}}}.$$

Adding the upper bounds in (4) and (5) yields the desired result.

Let us now turn to the lower bound. For any $t \in \mathbb{N}$,

$$(6) \quad \begin{aligned} & \Pr\left[\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches PNE}(G_{n,m})\right] \\ & \geq \Pr\left[\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches PNE}(G_{n,m}) \text{ by } t\right] \\ & \geq \left\lceil \frac{t}{n} \right\rceil \frac{1}{m^{n-1}} \left(1 - \frac{n}{m^{n-1}} \frac{(\lceil \frac{t}{n} \rceil)^2}{2}\right). \end{aligned}$$

Equation (6) follows from Lemma 2. Now, set

$$t = n \left\lfloor \frac{\sqrt{m^{n-1}}}{\sqrt{n}} \right\rfloor.$$

Since $m \geq 2$ and $n \geq 2$, we obtain $\frac{1}{2}\sqrt{n}\sqrt{m^{n-1}} \leq t \leq \sqrt{n}\sqrt{m^{n-1}}$. It follows that

$$(7) \quad 1 - \frac{n}{m^{n-1}} \frac{(\lceil \frac{t}{n} \rceil)^2}{2} \geq \frac{1}{2},$$

and that

$$(8) \quad \left\lceil \frac{t}{n} \right\rceil \frac{1}{m^{n-1}} \geq \frac{1}{2\sqrt{n}} \frac{1}{\sqrt{m^{n-1}}}.$$

Multiplying the lower bounds in (7) and (8) together yields the desired result. \square

We now turn to the proofs of Lemmas 1 and 2. The main challenge posed by paths generated according to Algorithm 2 is that they have “memory”: whenever player $s_c(t)$ encounters an environment that she has encountered before (i.e. $\mathbf{A}_{-s_c(t)}^{t-1} = \mathbf{A}_{-s_c(u)}^{u-1}$ and $s_c(t) = s_c(u)$) then, in period t , the player must play the same action that she played when she previously encountered the environment (i.e. $A_{s_c(t)}^t = A_{s_c(u)}^u$). This path-dependence complicates the analysis of the clockwork best-response dynamic. We therefore study a simpler (random walk) process that is “memoryless” to which we couple a dynamic that induces the same distribution over paths as Algorithm 2. The coupled dynamic follows the

random walk process until an environment is encountered by some player for the second time and becomes deterministic thereafter.

The coupled system is described by Algorithms 3 and 4 below.

Algorithm 3 Clockwork random walk

- (1) Draw an initial profile \mathbf{X}^0 uniformly at random from $[m]^n$
 - (2) For $t \in \mathbb{N}$:
 - (a) Set $i = s_c(t)$
 - (b) Set $\mathbf{X}_{-i}^t = \mathbf{X}_{-i}^{t-1}$
 - (c) Independently draw X_i^t uniformly at random from $[m]$
-

Algorithm 4 Coupled dynamic

- (1) Set $R_i(\mathbf{a}_{-i}) = 0$ for all $i \in [n]$ and $\mathbf{a}_{-i} \in [m]^{n-1}$
 - (2) Set the initial action profile to $\mathbf{Y}^0 = \mathbf{X}^0$
 - (3) For $t \in \mathbb{N}$:
 - (a) Set $i = s_c(t)$
 - (b) Set $\mathbf{Y}_{-i}^t = \mathbf{Y}_{-i}^{t-1}$
 - (c) If $R_i(\mathbf{Y}_{-i}^{t-1}) = 0$: set $Y_i^t = X_i^t$ and $R_i(\mathbf{Y}_{-i}^{t-1}) = Y_i^t$
 If $R_i(\mathbf{Y}_{-i}^{t-1}) \neq 0$: set $Y_i^t = R_i(\mathbf{Y}_{-i}^{t-1})$
-

$\langle \vec{\mathbf{X}}, s_c \rangle$ and $\langle \vec{\mathbf{Y}}, s_c \rangle$ denote paths generated according to Algorithms 3 and 4 respectively.

Algorithm 3 is a “clockwork random walk” on the set of action profiles $[m]^n$. The walk starts at some randomly drawn initial profile \mathbf{X}^0 and, in each period t , moves in direction $s_c(t)$ to a profile chosen uniformly at random from among the m profiles in that direction. A path generated according to this process does not have memory.

Algorithm 4 describes the coupled dynamic. The process starts at the same initial profile as the clockwork random walk. For each player i and environment \mathbf{a}_{-i} , we set the initial “response” value $R_i(\mathbf{a}_{-i})$ to zero. The crucial step to how the process evolves is (3c): if the response value to the current environment \mathbf{Y}_{-i}^{t-1} is zero, then the environment was never encountered before and, in that case, player i ’s response value is set to X_i^t , the action drawn by the clockwork random walk in period t . If, on the other hand, the response value to the current environment \mathbf{Y}_{-i}^{t-1} is non-zero (i.e. the environment was encountered before), then this value is the action that i takes in period t . In other words, $\langle \vec{\mathbf{Y}}, s_c \rangle$ has the same memory property that is characteristic of paths generated according to Algorithm 2.²⁸

²⁸The pattern of path-dependence is more complex for simultaneous updating than for the clockwork playing sequence. In the case of simultaneous updating, the choices of the players who have encountered an

Recall that Algorithm 2 essentially draws a best-response digraph, selects an initial profile, and then traces a path by traveling along the edges of the digraph starting at the initial profile and moving in direction $s_c(t)$ at step t . Under Algorithm 2, the entire best-response digraph is drawn up front. In contrast, Algorithm 4 starts with an empty digraph and then generates its edges in an “online” manner. Nevertheless, both algorithms induce the same distribution over paths, as summarized in the following remark.

Remark 1. Let $\langle \vec{\mathbf{A}}, s_c \rangle$ and $\langle \vec{\mathbf{Y}}, s_c \rangle$ be generated according to Algorithms 2 and 4 respectively. Then $\langle \vec{\mathbf{A}}, s_c \rangle$ and $\langle \vec{\mathbf{Y}}, s_c \rangle$ have the same distribution.

For any path $\langle \vec{\mathbf{a}}, s_c \rangle$ and for each $t \in \mathbb{N}$ define

$$f_{\langle \vec{\mathbf{a}}, s_c \rangle}(t) := \min \left\{ u \leq t : \mathbf{a}_{-s_c(u)}^{u-1} = \mathbf{a}_{-s_c(t)}^{t-1} \text{ and } s_c(u) = s_c(t) \right\}.$$

So $f_{\langle \vec{\mathbf{a}}, s_c \rangle}(t)$ is the first period along the path $\langle \vec{\mathbf{a}}, s_c \rangle$ that player $s_c(t)$ encounters the environment $\mathbf{a}_{-s_c(t)}^{t-1}$. Notice that if $s_c(t)$ encounters $\mathbf{a}_{-s_c(t)}^{t-1}$ for the first time in period t then $f_{\langle \vec{\mathbf{a}}, s_c \rangle}(t) = t$, and if $s_c(t)$ encountered $\mathbf{a}_{-s_c(t)}^{t-1}$ for the first time in some period $u < t$ then $f_{\langle \vec{\mathbf{a}}, s_c \rangle}(t) < t$. We also define

$$F_{\langle \vec{\mathbf{a}}, s_c \rangle} := \inf \{ t \in \mathbb{N} : f_{\langle \vec{\mathbf{a}}, s_c \rangle}(t) < t \}.$$

So $F_{\langle \vec{\mathbf{a}}, s_c \rangle}$ is the first period in which some player encounters an environment that they encountered previously along the path. The value $F_{\langle \vec{\mathbf{a}}, s_c \rangle}$ is bounded above by $1 + nm^{n-1}$ for any path.

By construction, the sequences $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ must agree at least up to (but not including) the period at which some player encounters an environment for the second time. In that period, under Algorithm 4, the player must play the action determined by their response function evaluated at that environment but, under Algorithm 3, the next action may be any of the available actions for that player. Remark 2 summarizes the key relationship between the clockwork random walk and the coupled dynamic.

Remark 2. $F_{\langle \vec{\mathbf{X}}, s_c \rangle} = F_{\langle \vec{\mathbf{Y}}, s_c \rangle}$.

Example (Illustration of Algorithms 3 and 4). Figure 11 illustrates the relationship between $\langle \vec{\mathbf{X}}, s_c \rangle$ and $\langle \vec{\mathbf{Y}}, s_c \rangle$ by plotting the first few elements of $\vec{\mathbf{X}}$ and of $\vec{\mathbf{Y}}$. Panel (A)

environment twice is deterministic but the choices of the players who have never encountered an environment twice remain random. For example, suppose $\mathbf{A}^0 = (1, 1, 1)$ and $\mathbf{A}^1 = (2, 1, 1)$. Then, player 1 must repeat action 2 in period 2, but the actions of players 2 and 3 in period 2 are not deterministic. Keeping track of the evolution of the path under simultaneous updating is therefore complex.

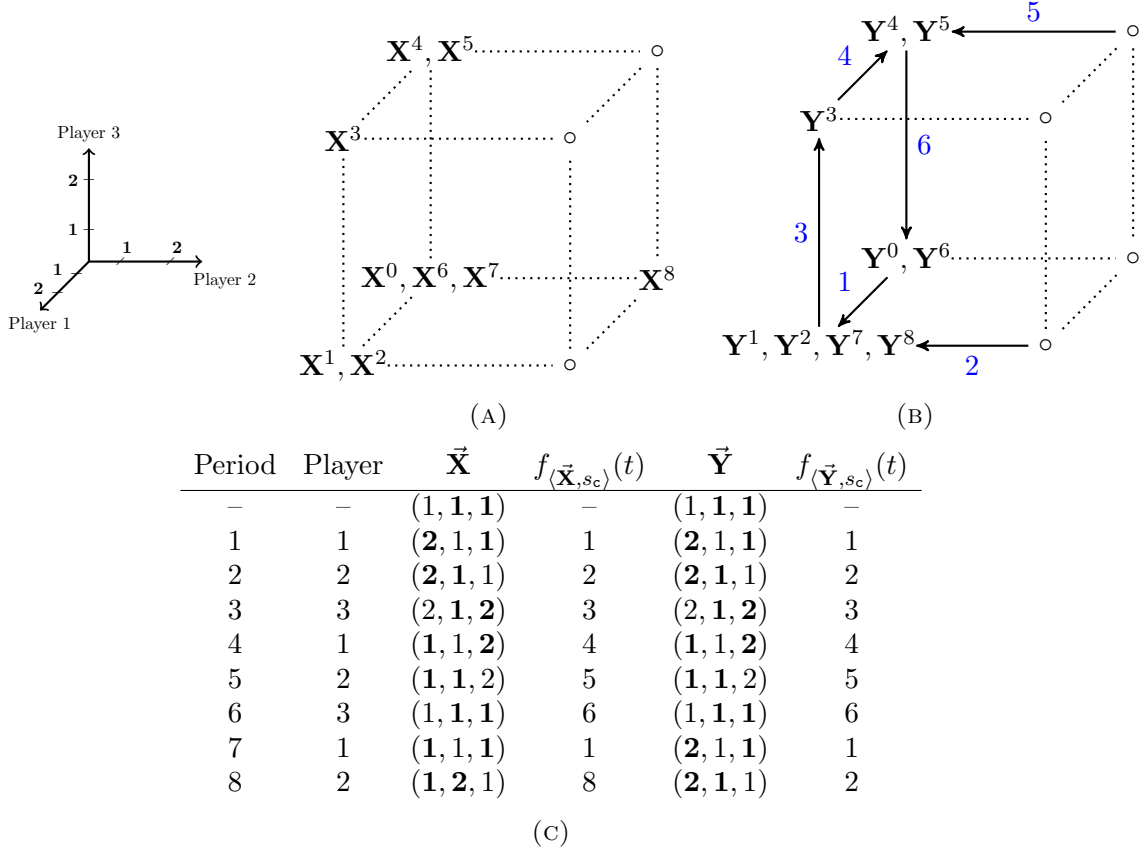


FIGURE 11. Illustration of Algorithms 3 and 4. Panel (A) shows the first elements of a possible path generated according to Algorithm 3 and panel (B) shows the corresponding path generated according to Algorithm 4. The table in panel (C) provides details, with environments highlighted in bold.

shows the first few elements of a possible path generated according to the clockwork random walk starting at the profile $\mathbf{X}^0 = (1, 1, 1)$. In panel (B), we represent the coupled dynamic, starting with an empty digraph and numbering the directed edges according to the period in which they are first placed.

In the clockwork random walk, player 1 chooses action 2 in period 1. This generates the directed edge from \mathbf{Y}^0 to \mathbf{Y}^1 in period 1 in panel (B). The paths are identical up to and including period 6. In period 7, however, player 1 encounters the same environment that she had encountered in period 1 (namely, players 2 and 3 each choosing action 1). The first time that player 1 encountered this environment, she responded by playing action 2, so she must play action 2 again in period 7. In other words, the path must follow the edge

Period	Player	Panel (A)		Panel (B)	
		\vec{a}	$f_{\langle \vec{a}, s_c \rangle}(t)$	\vec{a}	$f_{\langle \vec{a}, s_c \rangle}(t)$
—	—	(1, 1, 2)	—	(1, 1, 1)	—
1	1	(1, 1, 2)	1	(2, 1, 1)	1
2	2	(1, 2, 2)	2	(2, 2, 1)	2
3	3	(1, 2, 1)	3	(2, 2, 2)	3
4	1	(1, 2, 1)	4	(2, 2, 2)	4
5	2	(1, 2, 1)	5	(2, 1, 2)	5
6	3	(1, 2, 1)	3	(2, 1, 1)	6
7	1	(1, 2, 1)	4	(2, 1, 1)	1
8	2	(1, 2, 1)	5	(2, 2, 1)	2

TABLE 2. First few elements of the paths in panels (A) and (B) of Figure 2.

that was placed in period 1 and therefore $\mathbf{Y}^7 = (2, 1, 1)$. This is not true of the clockwork random walk. Since the process is memoryless, it can remain at $\mathbf{X}^7 = (1, 1, 1)$ in period 7 and travel to $\mathbf{X}^8 = (1, 2, 1)$ in period 8.

The path in panel (B) will keep cycling among the action profiles on the left-hand side of the cube forever whereas the path in panel (A) is allowed to freely wander. Note here that $F_{\langle \vec{\mathbf{X}}, s_c \rangle} = F_{\langle \vec{\mathbf{Y}}, s_c \rangle} = 7$. ■

Remark 3. $T_{\langle \vec{\mathbf{A}}, s_c \rangle} < F_{\langle \vec{\mathbf{A}}, s_c \rangle}$.

Remark 3 notes that any path $\langle \vec{\mathbf{A}}, s_c \rangle$ generated according to Algorithm 2 must reach $\text{PNE}(G_{n,m})$ or a best-response cycle before any player encounters an environment for the second time.

Example (Illustration of Remark 3). Table 2 shows the values of the function $f_{\langle \vec{a}, s_c \rangle}(t)$ for the first few elements of the paths generated according to the clockwork sequence best-response dynamic in panels (A) and (B) of Figure 2. Recall that the path in panel (A) reaches the pure Nash equilibrium in period 3 and that the path in panel (B) reaches a best-response cycle in period 1. Furthermore, from Table 2 we can see that the value of $F_{\langle \vec{a}, s_c \rangle}$ is 6 for panel (A) and 7 for panel (B). We therefore conclude that, for panel (A), $T_{\langle \vec{a}, s_c \rangle} = 3 < 6 = F_{\langle \vec{a}, s_c \rangle}$, and for panel (B), $T_{\langle \vec{a}, s_c \rangle} = 1 < 7 = F_{\langle \vec{a}, s_c \rangle}$. ■

The lemma below, which concerns paths $\langle \vec{\mathbf{X}}, s_c \rangle$ that are generated by the clockwork random walk, is useful for proving Lemmas 1 and 2. Under the clockwork sequence, player $i \in [n]$ plays in period $h_i(k) := i + (k - 1)n$ for $k \in \mathbb{N}$. For any $i \in [n]$ and any period $t \in \mathbb{N}$, define

$$k_i^*(t) := 1 + \left\lfloor \frac{t - i}{n} \right\rfloor.$$

So $k_i^*(t)$ is the largest $k \in \mathbb{N}$ such that $h_i(k) \leq t$. The environments that player $i \in [n]$ encounters on her turns between (and including) periods 1 and t are given in the sequence $(\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t))-1})$. Lemma 3 establishes bounds on the probability that these environments are all distinct.

Lemma 3. *For any $i \in [n]$ and $t \in \mathbb{N}$,*

$$1 - \frac{1}{m^{n-1}} \frac{(\lceil \frac{t}{n} \rceil)^2}{2} \leq \Pr \left[\mathbf{X}_{-i}^{h_i(k)-1} \text{ for } k \in \{1, \dots, k_i^*(t)\} \text{ are all distinct} \right] \leq \exp \left\{ -\frac{(\lfloor \frac{t}{n} \rfloor - 1)^2}{2m^{n-1}} \right\}.$$

Proof of Lemma 3. For any $i \in [n]$, the environments $\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t))-1}$ are independent because they are disjoint subsets of the draws of the clockwork random walk. Each environment is distributed uniformly on $[m]^{n-1}$. Therefore,

$$(9) \quad \Pr \left[\mathbf{X}_{-i}^{h_i(k)-1} \text{ for } k \in \{1, \dots, k_i^*(t)\} \text{ are all distinct} \right] = \prod_{k=1}^{k_i^*(t)-1} \left(1 - \frac{k}{m^{n-1}} \right).$$

If $k_i^*(t) > 1 + m^{n-1}$ then equation (9) is zero, and the lemma holds trivially ($k_i^*(t) > 1 + m^{n-1}$ implies $\lfloor \frac{t-i}{n} \rfloor > m^{n-1}$ which, in turn, implies $\lceil \frac{t}{n} \rceil > m^{n-1}$, so the lower bound in the statement of the lemma is negative and the upper bound is positive). We will therefore consider the case in which $k_i^*(t) \leq 1 + m^{n-1}$.

We obtain the following upper bound:

$$\prod_{k=1}^{k_i^*(t)-1} \left(1 - \frac{k}{m^{n-1}} \right) \leq \prod_{k=1}^{k_i^*(t)-1} \exp \left\{ -\frac{k}{m^{n-1}} \right\} \leq \exp \left\{ -\frac{(k_i^*(t)-1)^2}{2m^{n-1}} \right\} \leq \exp \left\{ -\frac{(\lfloor \frac{t}{n} \rfloor - 1)^2}{2m^{n-1}} \right\}.$$

The first step follows from $\exp\{x\} \geq 1 + x$ for all x . The final inequality follows from $k_i^*(t) - 1 = \lfloor \frac{t-i}{n} \rfloor \geq \lfloor \frac{t-n}{n} \rfloor = \lfloor \frac{t}{n} \rfloor - 1$.

We now turn to the lower bound:

$$\prod_{k=1}^{k_i^*(t)-1} \left(1 - \frac{k}{m^{n-1}} \right) \geq 1 - \sum_{k=1}^{k_i^*(t)-1} \frac{k}{m^{n-1}} = 1 - \frac{1}{m^{n-1}} \frac{k_i^*(t)^2}{2} \geq 1 - \frac{1}{m^{n-1}} \frac{(\lceil \frac{t}{n} \rceil)^2}{2}.$$

The first step is an application of the Weierstrass product inequality. The final inequality follows from the fact that $k_i^*(t) = 1 + \lfloor \frac{t-i}{n} \rfloor \leq 1 + \lfloor \frac{t-1}{n} \rfloor = \lceil \frac{t}{n} \rceil$. \square

Proof of Lemma 1. $T_{\langle \vec{\mathbf{A}}, s_c \rangle} > t$ is the event that $\langle \vec{\mathbf{A}}, s_c \rangle$ reaches $\text{PNE}(G_{n,m})$ or a best-response cycle only after period t . Remark 3 implies that $F_{\langle \vec{\mathbf{A}}, s_c \rangle} > t$. So

$$\Pr \left[T_{\langle \vec{\mathbf{A}}, s_c \rangle} > t \right] \leq \Pr \left[F_{\langle \vec{\mathbf{A}}, s_c \rangle} > t \right].$$

By Remarks 1 and 2,

$$\Pr \left[F_{\langle \bar{\mathbf{A}}, s_c \rangle} > t \right] = \Pr \left[F_{\langle \bar{\mathbf{Y}}, s_c \rangle} > t \right] = \Pr \left[F_{\langle \bar{\mathbf{X}}, s_c \rangle} > t \right].$$

Now, let us focus on the path $\langle \bar{\mathbf{X}}, s_c \rangle$ and on player 1. The environments that player 1 faces between periods 1 and t are given in the sequence $(\mathbf{X}_{-1}^{h_1(1)-1}, \mathbf{X}_{-1}^{h_1(2)-1}, \dots, \mathbf{X}_{-1}^{h_1(k_1^*(t))-1})$. The event $F_{\langle \bar{\mathbf{X}}, s_c \rangle} > t$ implies that the environments in this sequence are all distinct. Hence

$$\Pr \left[F_{\langle \bar{\mathbf{X}}, s_c \rangle} > t \right] \leq \Pr \left[\mathbf{X}_{-1}^{h_1(k)-1} \text{ for } k \in \{1, \dots, k_1^*(t)\} \text{ are all distinct} \right] \leq \exp \left\{ -\frac{(\lfloor \frac{t}{n} - 1 \rfloor)^2}{2m^{n-1}} \right\},$$

where the final step follows from Lemma 3. \square

To prove Lemma 2, we introduce Algorithm 5 which describes a dynamic that is also coupled with the clockwork random walk. $\langle \bar{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ denotes a path generated according to Algorithm 5.

Algorithm 5 Coupled dynamic with sink \mathbf{x}

- (1) Set $R_i(\mathbf{a}_{-i}) = 0$ for all $i \in [n]$ and $\mathbf{a}_{-i} \in [m]^{n-1}$
 - (2) Set $R_i(\mathbf{x}_{-i}) = x_i$ for all $i \in [n]$
 - (3) Set the initial action profile to $\mathbf{Z}^0 = \mathbf{X}^0$
 - (4) For $t \in \mathbb{N}$:
 - (a) Set $i = s_c(t)$
 - (b) Set $\mathbf{Z}_{-i}^t = \mathbf{Z}_{-i}^{t-1}$
 - (c) If $R_i(\mathbf{Z}_{-i}^{t-1}) = 0$: set $Z_i^t = X_i^t$ and $R_i(\mathbf{Z}_{-i}^{t-1}) = Z_i^t$
If $R_i(\mathbf{Z}_{-i}^{t-1}) \neq 0$: set $Z_i^t = R_i(\mathbf{Z}_{-i}^{t-1})$
-

Algorithm 5 is identical to Algorithm 4 except that for some particular profile \mathbf{x} the algorithm is initialized with $R_i(\mathbf{x}_{-i}) = x_i$ for all $i \in [n]$. Algorithm 5 therefore initializes the digraph with the directed edges (x'_i, \mathbf{x}_{-i}) to (x_i, \mathbf{x}_{-i}) for all i and $x'_i \neq x_i$, so that the profile \mathbf{x} is a sink. In the remaining steps, the algorithm selects a random initial profile and starts tracing a path by traveling along edges that (other than those edges already pointing to \mathbf{x} in the initialization) are generated in an online manner. The paths traced by the clockwork random walk and this coupled dynamic with a sink at \mathbf{x} must agree at least up to (but not including) the period at which *either* an environment is encountered by a player for the second time *or* the environment is \mathbf{x}_{-i} for some player i .

Example (Illustration of Algorithms 3 and 5). Figure 12 illustrates the relationship between $\langle \bar{\mathbf{X}}, s_c \rangle$ and $\langle \bar{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ by plotting the first few elements of $\bar{\mathbf{X}}$ and of $\bar{\mathbf{Z}}$. Panel (A) shows the first few elements of a possible path generated according to the clockwork random

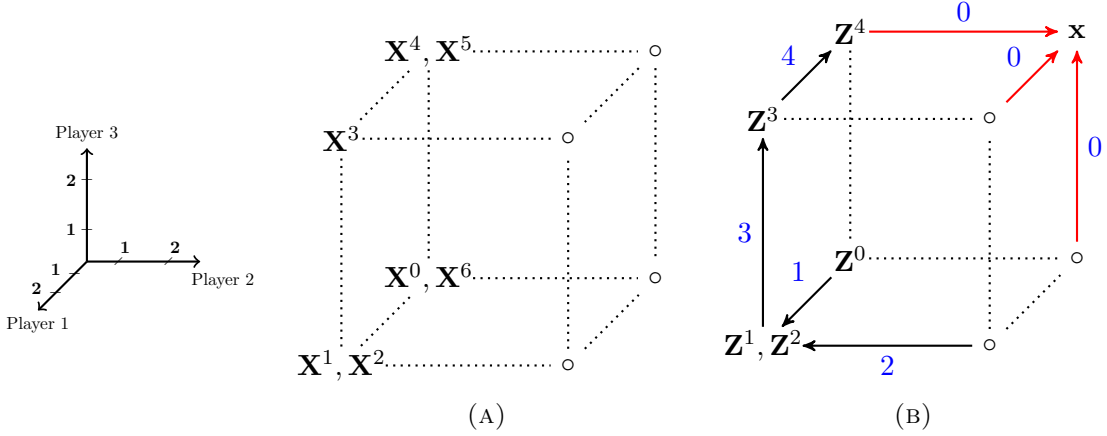


FIGURE 12. Illustration of Algorithms 3 and 5. Panel (A) shows the first elements of a possible path generated according to Algorithm 3 and panel (B) shows the corresponding path generated according to Algorithm 5.

walk starting at the profile $\mathbf{X}^0 = (1, 1, 1)$. In panel (B), we represent the corresponding path generated according Algorithm 5. This time, rather than starting with an empty digraph, the profile \mathbf{x} is made a sink (with the red edges placed in period 0). The remaining directed edges are numbered according to the period in which they are first placed.

The clockwork random walk takes the path $\vec{\mathbf{Z}}$ to $\mathbf{Z}^4 = (1, 1, 2)$ in period 4. While the random walk can continue wandering through the action profiles according to the clockwork sequence, the path $\vec{\mathbf{Z}}$ must end up at $\mathbf{Z}^5 = \mathbf{x}$ in period 5. ■

Remark 4. Let $\langle \vec{\mathbf{A}}, s_c \rangle$ and $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ be generated according to Algorithms 2 and 5 respectively. Then the distribution of $\langle \vec{\mathbf{A}}, s_c \rangle$ conditional on $\mathbf{x} \in \text{PNE}(G_{n,m})$ is the same as the distribution of $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$.

Proof of Lemma 2. For any $t \in \mathbb{N}$,

$$\begin{aligned}
 & \Pr \left[\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \text{PNE}(G_{n,m}) \text{ by } t \right] \\
 &= \sum_{\mathbf{x} \in [m]^n} \Pr \left[\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \{\mathbf{x}\} \text{ by } t \text{ and } \mathbf{x} \in \text{PNE}(G_{n,m}) \right] \\
 &= \sum_{\mathbf{x} \in [m]^n} \Pr \left[\langle \vec{\mathbf{A}}, s_c \rangle \text{ reaches } \{\mathbf{x}\} \text{ by } t \mid \mathbf{x} \in \text{PNE}(G_{n,m}) \right] \Pr [\mathbf{x} \in \text{PNE}(G_{n,m})] \\
 (10) \quad &= \sum_{\mathbf{x} \in [m]^n} \underbrace{\Pr \left[\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle \text{ reaches } \{\mathbf{x}\} \text{ by } t \right]}_{(10.1)} \underbrace{\Pr [\mathbf{x} \in \text{PNE}(G_{n,m})]}_{(10.2)}
 \end{aligned}$$

The first step follows from the definition of reaching a pure Nash equilibrium. The final step follows from Remark 4; namely, the probability that $\langle \vec{\mathbf{A}}, s_c \rangle$ reaches $\{\mathbf{x}\}$ by period t conditional on $\mathbf{x} \in \text{PNE}(G_{n,m})$ is equal to the probability that $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ reaches $\{\mathbf{x}\}$ by period t . We now analyze the expressions (10.1) and (10.2).

For (10.2), since payoffs are drawn identically and independently according to the atomless distribution \mathbb{P} , we have that

$$(11) \quad \Pr[\mathbf{x} \in \text{PNE}(G_{n,m})] = \prod_{i=1}^n \Pr \left[U_i(\mathbf{x}) \geq \max_{x'_i \in [m]} U_i(x'_i, \mathbf{x}_{-i}) \right] = \frac{1}{m^n}.$$

We now find upper and lower bounds on (10.1) by relating $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ to the clockwork random walk path $\langle \vec{\mathbf{X}}, s_c \rangle$. We start with the upper bound. Notice that $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ cannot reach $\{\mathbf{x}\}$ by period t unless $\mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)}$ for some $\tau \leq t$. Therefore

$$(12) \quad \begin{aligned} \Pr \left[\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle \text{ reaches } \{\mathbf{x}\} \text{ by } t \right] &\leq \Pr \left[\bigcup_{\tau=1}^t \{ \mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)} \} \right] \\ &\leq \sum_{\tau=1}^t \Pr \left[\mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)} \right] = \frac{t}{m^{n-1}}. \end{aligned}$$

The final step follows from the fact that $\mathbf{X}_{-s_c(\tau)}^{\tau-1}$ consists of $n-1$ independent random variables, each uniformly distributed on $[m]$.

We now turn to the lower bound. If $F_{\langle \vec{\mathbf{X}}, s_c \rangle} > t$ and $\mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)}$ for some $\tau \leq t$ then $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ must reach $\{\mathbf{x}\}$ by period t . In other words, if no environments are repeated for any player and the environment is \mathbf{x}_{-i} for some player i by period t , then $\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle$ must reach $\{\mathbf{x}\}$ by period t . Therefore,

$$(13) \quad \begin{aligned} &\Pr \left[\langle \vec{\mathbf{Z}}, \mathbf{x}, s_c \rangle \text{ reaches } \{\mathbf{x}\} \text{ by } t \right] \\ &\geq \Pr \left[\bigcup_{\tau=1}^t \{ \mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)} \} \text{ and } F_{\langle \vec{\mathbf{X}}, s_c \rangle} > t \right] \\ &= \Pr \left[\bigcup_{\tau=1}^t \{ \mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)} \} \mid F_{\langle \vec{\mathbf{X}}, s_c \rangle} > t \right] \Pr \left[F_{\langle \vec{\mathbf{X}}, s_c \rangle} > t \right]. \end{aligned}$$

To bound the first term in (13), notice that $\mathbf{X}_{-1}^{h_1(k)-1} = \mathbf{x}_{-1}$ for some $k \in \{1, \dots, k_1^*(t)\}$ implies that $\mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)}$ for some $\tau \leq t$. Therefore

$$\begin{aligned}
\Pr \left[\bigcup_{\tau=1}^t \{ \mathbf{X}_{-s_c(\tau)}^{\tau-1} = \mathbf{x}_{-s_c(\tau)} \} \mid F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t \right] &\geq \Pr \left[\bigcup_{k=1}^{k_1^*(t)} \{ \mathbf{X}_{-1}^{h_1(k)-1} = \mathbf{x}_{-1} \} \mid F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t \right] \\
&= \sum_{k=1}^{k_1^*(t)} \Pr \left[\mathbf{X}_{-1}^{h_1(k)-1} = \mathbf{x}_{-1} \mid F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t \right] \\
&= \sum_{k=1}^{k_1^*(t)} \frac{1}{m^{n-1}} \\
(14) \qquad &= \left\lceil \frac{t}{n} \right\rceil \frac{1}{m^{n-1}}.
\end{aligned}$$

The first summation follows from the fact that since all the environments for player 1 are distinct, the events in the union are mutually exclusive. The next step follows from the fact that our process is invariant under symmetry. So for any $k \in \{1, \dots, k_1^*(t)\}$ and for all \mathbf{x}_{-1} and \mathbf{y}_{-1} , $\Pr[\mathbf{X}_{-1}^{h_1(k)-1} = \mathbf{x}_{-1} \mid F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t] = \Pr[\mathbf{X}_{-1}^{h_1(k)-1} = \mathbf{y}_{-1} \mid F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t]$ which implies that $\Pr[\mathbf{X}_{-1}^{h_1(k)-1} = \mathbf{x}_{-1} \mid F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t] = \frac{1}{m^{n-1}}$. The last step follows from $k_1^*(t) = 1 + \lfloor \frac{t-1}{n} \rfloor = \lceil \frac{t}{n} \rceil$.

To bound the second term in (13), notice that if for each $i \in [n]$ the environments $\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t))-1}$ are all distinct then $F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t$. Therefore

$$\begin{aligned}
\Pr[F_{\langle \vec{\mathbf{x}}, s_c \rangle} > t] &\geq \Pr \left[\bigcap_{i \in [n]} \{ \mathbf{X}_{-i}^{h_i(k)-1} \text{ for } k \in \{1, \dots, k_i^*(t)\} \text{ are all distinct} \} \right] \\
&= 1 - \Pr \left[\bigcup_{i \in [n]} \{ \mathbf{X}_{-i}^{h_i(k)-1} \text{ for } k \in \{1, \dots, k_i^*(t)\} \text{ are not all distinct} \} \right] \\
&\geq 1 - \sum_{i \in [n]} \Pr \left[\mathbf{X}_{-i}^{h_i(k)-1} \text{ for } k \in \{1, \dots, k_i^*(t)\} \text{ are not all distinct} \right] \\
(15) \qquad &\geq 1 - \frac{n}{m^{n-1}} \frac{(\lceil \frac{t}{n} \rceil)^2}{2}.
\end{aligned}$$

The final step follows from Lemma 3.

Gathering the results (10), (11), (12), (14), and (15) together yields the desired conclusion. \square

APPENDIX B. PROOFS OF THEOREM 3, PROPOSITION 3, AND THEOREM 4

In this section, we focus exclusively on the clockwork sequence best-response dynamic in 2-player games. We first explicitly work out the exact probability that a path generated by Algorithm 4 reaches a $2k$ -cycle in period t . We then turn to the asymptotic behavior of our formulas of interest.

Recall the definitions of $h_i(k)$ and $k_i^*(t)$ preceding Lemma 3. On a path $\langle \vec{\mathbf{X}}, s_c \rangle$ generated by Algorithm 3, the environments that player $i \in \{1, 2\}$ encounters on her turns between (and including) periods 1 and t are given in the sequence $(\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t))-1})$.

B.1. Proof of Theorem 3. In order for a path generated by Algorithm 4 to reach neither a pure Nash equilibrium nor a best-response cycle by period t it must be the case that, by period $t + 1$ (inclusive), no player encounters an environment that they have seen before, and the action taken by player $s_c(t + 1)$ in period $t + 1$ must not repeat any of the environments encountered by period t by player $s_c(t)$. To put it differently, for each $i \in \{1, 2\}$ it must be the case that the environments $(\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t+1))-1})$ are all distinct, and the action $X_{s_c(t+1)}^{t+1}$ taken by player $s_c(t + 1)$ in period $t + 1$ is distinct from each of the environments encountered by period t by player $s_c(t)$. It follows that the probability that the clockwork sequence best-response dynamic converges to neither a pure Nash equilibrium nor a best-response cycle by period t for $t \in [2m]$ is

$$(16) \quad \prod_{i=1}^t \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right).$$

In order for the path to reach a pure Nash equilibrium in period t , for each $i \in \{1, 2\}$ it must be the case that the environments $(\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t+1))-1})$ are all distinct, and the action $X_{s_c(t+1)}^{t+1}$ taken by player $s_c(t + 1)$ in period $t + 1$ is equal to the environment $\mathbf{X}_{-s_c(t)}^{t-1}$ encountered by player $s_c(t)$ in period t . Therefore, the probability that the clockwork sequence best-response dynamic converges to a pure Nash equilibrium in period $t \in [2m]$ is

$$(17) \quad \frac{1}{m} \prod_{i=1}^t \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right).$$

More generally, in order for the path to reach a $2k$ -cycle in period t , for each $i \in \{1, 2\}$ it must be the case that the environments $(\mathbf{X}_{-i}^{h_i(1)-1}, \mathbf{X}_{-i}^{h_i(2)-1}, \dots, \mathbf{X}_{-i}^{h_i(k_i^*(t+2k-1))-1})$ are all distinct, and the action $X_{s_c(t+2k-1)}^{t+2k-1}$ taken by player $s_c(t + 2k - 1)$ in period $t + 2k - 1$ is equal to the environment $\mathbf{X}_{-s_c(t)}^{t-1}$ encountered by player $s_c(t)$ in period t . Therefore, the

probability that the clockwork sequence best-response dynamic converges to a $2k$ -cycle for $k \in [m]$ in period $t \in [2(m - k + 1)]$ is

$$(18) \quad \frac{1}{m} \prod_{i=1}^{t+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right).$$

This is precisely formula (1) given in the statement of Theorem 3. Notice that setting $k = 1$ in (18) recovers formula (17).

The probability that the clockwork sequence best-response dynamic reaches a $2k$ -cycle for $k \in [m]$ is obtained by summing (18) over all $t \in [2(m - k + 1)]$:

$$(19) \quad \frac{1}{m} \sum_{t=1}^{2(m-k+1)} \prod_{i=1}^{t+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right).$$

This is expression (2) in Corollary 2.

Example (Illustration of formulas (16) to (18)). To illustrate our results we schematically map out all the possible paths of the clockwork sequence best-response dynamic in 2-player m -action games in the tree shown in Figure 13. The initial profile, in period 0, is arbitrarily set to $(1, 1)$.

In order to reach neither a pure Nash equilibrium nor a best-response cycle by period 3, for example, we must travel down the tree along the sequence $(1, 1), (1, 1), (1, 2), (2, 2), (2, 3)$ or the sequence $(1, 1), (2, 1), (2, 2), (1, 2), (1, 3)$. Let us look at the first sequence more closely. Up to (and including) period 4, the environments encountered by each player respectively are all distinct. Player 2 must select action 3 in period 4 to ensure that the sequence does not end up revisiting a previously encountered environment. The probability of traveling along $(1, 1), (1, 1), (1, 2), (2, 2), (2, 3)$ or $(1, 1), (2, 1), (2, 2), (1, 2), (1, 3)$ is given by formula (16) with $t = 3$.

Let us consider the probability of reaching a pure Nash equilibrium in period 4. In Figure 13, we must travel along the sequence $(1, 1), (1, 1), (1, 2), (2, 2), (2, 3), (2, 3)$ or the sequence $(1, 1), (2, 1), (2, 2), (1, 2), (1, 3), (1, 3)$. Let us look at the first sequence more closely. Up to (and including) period 5, the environments encountered by each player respectively are all distinct. Player 1 must select action 2 in period 5 (this was the environment encountered by player 2 in period 4) to ensure that the pure Nash equilibrium $(2, 3)$ was reached in period 4. The probability of traveling along $(1, 1), (1, 1), (1, 2), (2, 2), (2, 3), (2, 3)$ or $(1, 1), (2, 1), (2, 2), (1, 2), (1, 3), (1, 3)$ is given by formula (17) with $t = 4$.

Let us consider the probability of reaching a 4-cycle in period 2. In Figure 13 we must travel down the tree along the sequence $(1, 1), (1, 1), (1, 2), (2, 2), (2, 3), (1, 3)$ or the sequence $(1, 1), (2, 1), (2, 2), (1, 2), (1, 3), (2, 3)$. Let us look at the first sequence more closely. The action profile is $(1, 2)$ in period 2, and this is the first action profile of our 4-cycle. A further $2k - 1$ periods must pass before looping back to $(1, 2)$. Up to (and including) period 5, the environments encountered by each player respectively are all distinct. In order to “close” the cycle, player 1 must select action 1 in period 5 (this was the environment encountered by player 2 in period 2). The probability of traveling along $(1, 1), (1, 1), (1, 2), (2, 2), (2, 3), (1, 3)$ or $(1, 1), (2, 1), (2, 2), (1, 2), (1, 3), (2, 3)$ is given by formula (18) with $t = k = 2$. ■

B.2. Proof of Theorem 4. To prove Theorem 4 we now work out the asymptotic behavior of formula (16). Note that (16) can be written as

$$\prod_{i=1}^t \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right) = \begin{cases} \frac{m!^2}{(m - \frac{t+1}{2})!^2 m^{t+1}} & \text{if } t \text{ is odd} \\ \left(\frac{m - \frac{t}{2}}{m} \right) \frac{m!^2}{(m - \frac{t}{2})!^2 m^t} & \text{if } t \text{ is even} \end{cases}.$$

Using Stirling’s formula which states that

$$n! \sim \sqrt{2\pi n} \cdot n^n \exp\{-n\},$$

as $n \rightarrow \infty$, we obtain²⁹

$$(20) \quad \frac{m!^2}{(m - \frac{t+1}{2})!^2 m^{t+1}} \sim \left(\frac{m - \frac{t+1}{2}}{m} \right)^{t-2m} \exp\{-(t+1)\}.$$

and

$$(21) \quad \left(\frac{m - \frac{t}{2}}{m} \right) \frac{m!^2}{(m - \frac{t}{2})!^2 m^t} \sim \left(\frac{m - \frac{t}{2}}{m} \right)^{t-2m} \exp\{-t\}.$$

whenever $m - t \rightarrow \infty$. Taking a logarithm of the last expression,³⁰

$$\begin{aligned} & -t + (t - 2m) \ln \left(1 - \frac{1}{m} \frac{t}{2} \right) \\ &= -t + (t - 2m) \left(-\frac{1}{2} \frac{t}{m} - \frac{1}{8} \frac{t^2}{m^2} + O\left(\frac{t^3}{m^3}\right) \right) \\ &= -\frac{1}{4} \frac{t^2}{m} + O\left(\frac{t^3}{m^2}\right). \end{aligned}$$

²⁹ $f(n) \sim g(n)$ denotes $f(n)/g(n) \rightarrow 1$ as $n \rightarrow \infty$.

³⁰ $f(n) = O(g(n))$ if there is $M > 0$ and N such that $|f(n)| \leq Mg(n)$ for all $n \geq N$.

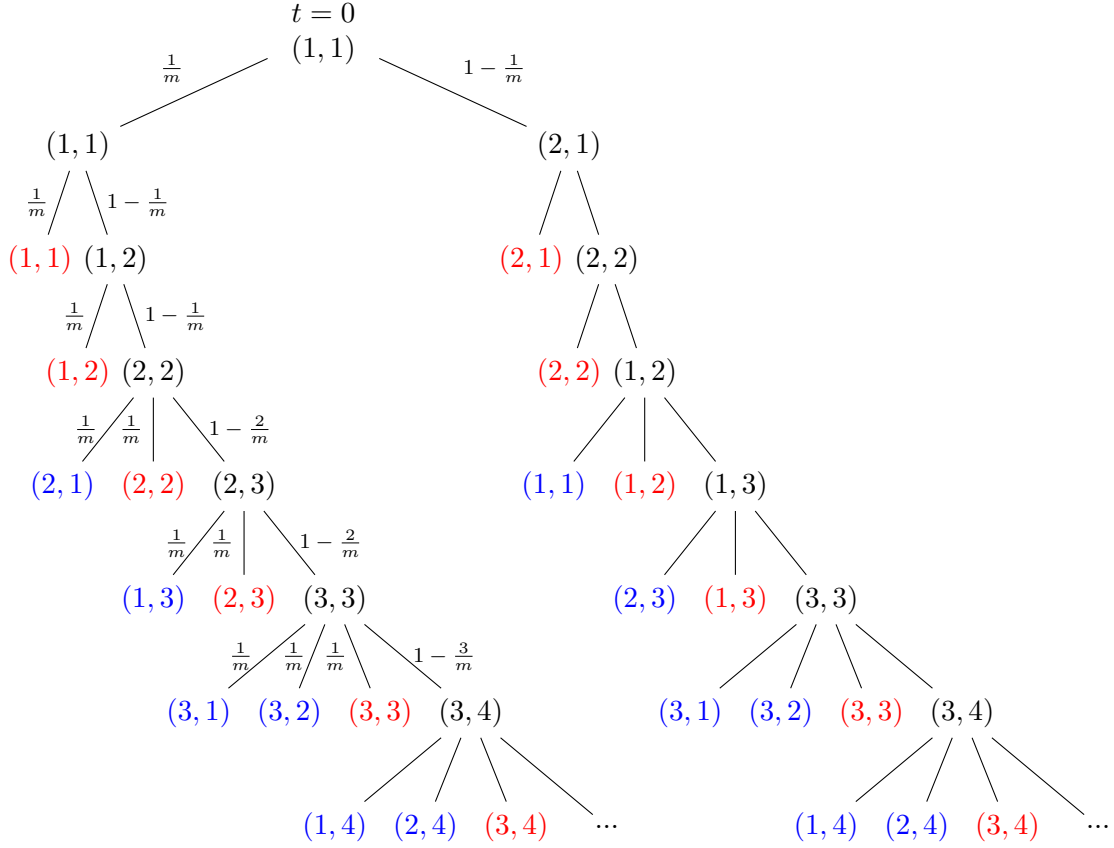


FIGURE 13. Illustration of possible paths for 2-player m -action games. We arbitrarily set the initial action profile to be $(1, 1)$ in period 0. In period 1, player 1 either plays action 1 (left branch) or some other action (right branch) which we arbitrarily call action 2. Player 2 then responds in period 2, and so on. All red leaves are Nash equilibria and all blue leaves are profiles that belong to best-response cycles.

Provided that $t = o(m^{2/3})$, the second term goes to zero and therefore equation (21) behaves asymptotically like $\exp\{-t^2/(4m)\}$. An identical argument shows that, under the same conditions, (20) is also asymptotically $\exp\{-t^2/(4m)\}$. Hence,

$$(22) \quad \prod_{i=1}^t \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor\right) \sim \exp\left\{-\frac{t^2}{4m}\right\}.$$

This completes the proof of Theorem 4. Note that approximation (22) holds uniformly in the range $[1, o(m^{2/3})]$.

B.3. Proof of Proposition 3. To prove Proposition 3, we now turn to the asymptotic behavior of (19). Let $T = T(m)$ satisfy $T = o(m^{2/3})$ and $k = o(T)$. We assume that $T \geq \frac{m^{2/3}}{\ln(m)}$ so that T is not too small, and we split the summation in (19) into two ranges: $t \leq T$ and $t > T$. Since (22) holds uniformly in our first range, we have

$$\frac{1}{m} \sum_{t=1}^T \prod_{i=1}^{t+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor\right) \sim \frac{1}{m} \sum_{t=1}^T \exp \left\{ -\frac{(t+2(k-1))^2}{4m} \right\}.$$

We now approximate the summation on the right-hand side with an integral. Firstly, note that

$$\begin{aligned} \frac{1}{m} \int_1^{T+1} \exp \left\{ -\frac{(t+2(k-1))^2}{4m} \right\} dt &= \sqrt{\frac{2}{m}} \int_{\frac{2k-1}{\sqrt{2m}}}^{\frac{T+1+2(k-1)}{\sqrt{2m}}} \exp \left\{ -\frac{x^2}{2} \right\} dx \\ &\sim \sqrt{\frac{2}{m}} \int_{\frac{2k-1}{\sqrt{2m}}}^{\infty} \exp \left\{ -\frac{x^2}{2} \right\} dx \\ &= 2\sqrt{\frac{\pi}{m}} \left(1 - \Phi \left(\frac{2k-1}{\sqrt{2m}} \right) \right), \end{aligned} \tag{23}$$

where the first step uses the transformation $x = (t+2(k-1))/\sqrt{2m}$. Furthermore,

$$\frac{1}{m} \int_0^1 \exp \left\{ -\frac{(t+2(k-1))^2}{4m} \right\} dt \leq \frac{1}{m},$$

which goes to zero faster than (23). Since

$$\int_1^{T+1} f(t) dt \leq \sum_{t=1}^T f(t) \leq \int_0^T f(t) dt \leq \int_1^{T+1} f(t) dt + \int_0^1 f(t) dt,$$

for any positive and decreasing function $f(\cdot)$, it follows that

$$\frac{1}{m} \sum_{t=1}^T \exp \left\{ -\frac{(t+2(k-1))^2}{4m} \right\} \sim 2\sqrt{\frac{\pi}{m}} \left(1 - \Phi \left(\frac{2k-1}{\sqrt{2m}} \right) \right).$$

It remains for us to show that the summation (19) over the second range is negligible. Since $\exp\{x\} \geq 1 + x$ and $\lfloor x \rfloor > x - 1$ for all x , we obtain the following upper bound:

$$\begin{aligned} \frac{1}{m} \sum_{t=T+1}^{2(m-k+1)} \prod_{i=1}^{t+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right) &\leq \frac{1}{m} \sum_{t=T+1}^{2(m-k+1)} \prod_{i=1}^{T+1+2(k-1)} \left(1 - \frac{1}{m} \left\lfloor \frac{i}{2} \right\rfloor \right) \\ &\leq \frac{1}{m} \sum_{t=T+1}^{2(m-k+1)} \exp \left\{ -\frac{1}{m} \sum_{i=1}^{T+1+2(k-1)} \left(\frac{i}{2} - 1 \right) \right\} \\ &\leq \frac{2m - 2k - T + 1}{m} \exp \left\{ -\frac{1}{4m} (T + 2(k-1) - 2)^2 \right\}. \end{aligned}$$

This expression is small compared to the other half of the sum.

APPENDIX C. DESCRIPTIONS OF THE LEARNING RULES

We compare best-response dynamics to three more complicated learning dynamics: *Bush-Mosteller learning* as an example of *reinforcement learning*, *fictitious play* as an example of *belief learning*, and *replicator dynamics* as the most important equation in evolutionary biology.

There are unavoidable arbitrary choices in the specification of the learning dynamics, the values of the parameters and the criteria that determine convergence to mixed or pure Nash equilibria, however the overall picture is robust to the specific implementation for all sensible parametrizations. The dynamics here are all in discrete-time or had to be converted to discrete-time.

We now describe the learning dynamics in detail, as well as the convergence criteria and our choice of parameters. We use similar convergence criteria to those used by Pangallo et al. (2019) in the two-player case.

C.1. Reinforcement learning. We consider the Bush-Mosteller learning algorithm as an example of reinforcement learning (Bush and Mosteller, 1953), using the specifications in Macy and Flache (2002) and Galla and Farmer (2013).

Each player has an *aspiration* level that corresponds to a weighted average of the payoffs that the player has received while playing the game. Each player then associates a level of satisfaction with each action, which is positive if the payoff the player gets when choosing this action is higher than the player's aspiration level, and negative otherwise. The probability of playing an action is increased if the satisfaction was positive and decreased if it was negative.

Formal description. In each period, each player $i \in [n]$ chooses an action $x \in [m]$ with probability $p_i^t(x)$. The evolution of the mixed strategy of each player i , $\mathbf{p}_i^t = (p_i^t(1), \dots, p_i^t(m))$, is governed by reinforcement learning, as we describe below. The learning rule generates a mapping from $\mathbf{p}^t = (\mathbf{p}_1^t, \dots, \mathbf{p}_n^t)$ to \mathbf{p}^{t+1} .

Let \aleph_i^t be the *aspiration level* of player i in period t . It evolves according to

$$\aleph_i^{t+1} = (1 - \alpha)\aleph_i^t + \alpha u_i(x, \mathbf{a}_{-i}^t).$$

when (x, \mathbf{a}_{-i}^t) is the profile played in period t .

The updated aspiration level is therefore a weighted average of the payoff received at time t and the player's past aspiration level. Payoffs received in the past are discounted by a factor of $(1 - \alpha)$, where α stands for the rate of *memory loss*. Player i 's *satisfaction* with action $x \in [m]$ in period t is defined by

$$\sigma_i^t(x) = \frac{u_i(x, \mathbf{a}_{-i}^t) - \aleph_i^t}{\max_{\mathbf{y} \in [m]^n} |u_i(\mathbf{y}) - \aleph_i^t|}.$$

Note that $\sigma_i^t(x)$ lies within -1 and 1 . If player i chooses action x in period t , player i associates positive satisfaction with this action if the payoff they received in period t is higher than the player's aspiration level.

If player i played action x in period t , then the probability that i plays x again in period $t + 1$ is updated as

$$(24) \quad p_i^{t+1}(x) = \begin{cases} p_i^t(x) + \beta \sigma_i^t(x)(1 - p_i^t(x)) & \sigma_i^t(x) \geq 0 \\ p_i^{t+1}(x) + \beta \sigma_i^t(x)p_i^t(x) & \sigma_i^t(x) < 0 \end{cases},$$

and the probability of choosing a different action $y \neq x$ in period $t + 1$ is updated as

$$(25) \quad p_i^{t+1}(y) = \begin{cases} p_i^t(y) - \beta \sigma_i^t(x)p_i^t(y) & \sigma_i^t(x) \geq 0 \\ p_i^t(y) - \beta \sigma_i^t(x) \frac{p_i^t(x)p_i^t(y)}{1 - p_i^t(x)} & \sigma_i^t(x) < 0 \end{cases}.$$

In the equations above, β represents the learning rate. Positive satisfaction for action x leads to an increase of the probability to choose action x , negative satisfaction has the opposite effect. Note that the actions of all players but i , \mathbf{a}_{-i}^t , only enter the learning process of i through the payoff that i receives from playing action x against \mathbf{a}_{-i}^t , $u_i(x, \mathbf{a}_{-i}^t)$. Player i need not know the actions of the other players. Rather i needs only to keep track of her own actions and of her own payoffs in order to update her aspiration, satisfaction, and mixed strategy vector. This implementation of the Bush-Mosteller dynamic is therefore a classic example of reinforcement learning in which limited information is required.

Convergence criteria. To assess convergence, we check whether \mathbf{p}^t converges to a fixed point of the mapping $\mathbf{p}^t \mapsto \mathbf{p}^{t+1}$. This choice makes it possible to also assess convergence to mixed Nash equilibria, which would be missed if we only looked at the actions played. Of course, because players play actions by randomly sampling from their mixed strategy vectors, the evolution of \mathbf{p}^t is stochastic, and so we need to allow for noise in our assessment of convergence. Additionally, \mathbf{p}^t never reaches a fixed point of $\mathbf{p}^t \mapsto \mathbf{p}^{t+1}$ within simulation time. The reason for that is that equations (24) and (25) have no memory loss term, so the probability for playing an unsuccessful action keeps decreasing over time without ever reaching a steady state. To address these issues, we use the same heuristic as in Pangallo et al. (2019):

- (1) Only consider the last 20% time steps, to avoid transient effects.
- (2) Only keep the actions that have been played with a probability larger than 0.05, averaged over the time interval.
- (3) If the average standard deviation, calculated over the time interval for each selected action and averaged over the selected actions is larger than 0.01, the simulation run will be regarded as non-convergent, otherwise as convergent. We identify a convergent simulation run as having reached a pure Nash equilibrium if each belief vector \mathbf{p}_i^t has a component that is larger than 0.98.

Parameter values. We perform the simulations with $\alpha = 0.2$ and $\beta = 0.5$, but could not observe much sensitivity to the parameter values. We simulate for 5000 time steps with randomly chosen initial conditions.

C.2. Fictitious play. Fictitious play is an example of belief learning and was first proposed as an algorithm to calculate Nash equilibria. It was later interpreted as a learning algorithm (Brown, 1951, Robinson, 1951). Each player takes the empirical distribution of actions taken by the opponents as an estimate of their mixed strategies, calculates the expected payoff of each action based on this estimate, and chooses the (pure) action with the highest expected payoff. Variants include *weighted fictitious play* (Fudenberg and Levine, 1998), in which the players discount opponents’ past actions and give higher weight to more recent actions, and *stochastic fictitious play*, where the players choose the best performing action with a certain probability, and the other actions with a smaller probability.

Formal description. In period $t \geq 0$, each player's belief $p_j^t(x)$ that player j will play action x in period $t + 1$ is given by the fraction of times that player j chose action x in the past:

$$p_j^t(x) = \frac{1}{t+1} \sum_{\tau=0}^t \mathbf{1}[a_j^\tau = x],$$

where $\mathbf{1}[a_j^\tau = x] = 1$ if j played action x in period τ and $\mathbf{1}[a_j^\tau = x] = 0$ otherwise. In each period, each player i then deterministically selects the action with the highest expected payoff given their belief about their opponents, \mathbf{p}_{-i}^t :

$$a_i^{t+1} = \arg \max_{x \in [m]} \sum_{\mathbf{x}_{-i} \in [m]^{n-1}} u_i(x, \mathbf{x}_{-i}) \prod_{j \in [n] \setminus \{i\}} p_j^t(x_j).$$

Convergence criteria. To study convergence to mixed equilibria, we follow [Fudenberg and Levine \(1998\)](#) in considering convergence of beliefs $\mathbf{p}^t = (\mathbf{p}_1^t, \dots, \mathbf{p}_n^t)$ rather than convergence of the actions played. Our convergence criteria for \mathbf{p}^t are the same as those described above for reinforcement learning. A minor difference is that we identify a convergent simulation run as having reached a pure Nash equilibrium if each belief vector \mathbf{p}_i^t has a component that is larger than 0.99.

Parameter values. Fictitious play has no parameters.

C.3. Replicator dynamics. Replicator dynamics are the most basic evolutionary model ([Maynard Smith, 1982](#)). They play an important role in describing evolutionary game dynamics and population dynamics. Following the interpretation in [Börger and Sarin \(1997\)](#), we view replicator dynamics as a learning algorithm for individual players.³¹ Because our randomly generated payoff matrices are not necessarily symmetric, we consider the multi-population version of the replicator dynamic ([Taylor and Nowak, 2006](#), [Gokhale and Traulsen, 2010](#)).

Formal description. In each period t , each player i chooses an action x with probability $p_i^t(x)$, and the probability vector $\mathbf{p}_i^t = (p_i^t(1), \dots, p_i^t(m))$ evolves according to the replicator equation, as described below.

When all other players sample their actions according to \mathbf{p}_{-i}^t , the expected payoff of player i when choosing action x in period t is

$$\tilde{u}_i(x, \mathbf{p}_{-i}^t) = \sum_{\mathbf{x}_{-i} \in [m]^{n-1}} u_i(x, \mathbf{x}_{-i}) \prod_{j \in [n] \setminus \{i\}} p_j^t(x_j).$$

³¹Replicator dynamics are also obtained as the continuous time limit of discrete time reinforcement-learning algorithms ([Börger and Sarin, 1997](#), [Sato and Crutchfield, 2003](#), [Tuyts et al., 2006](#), [Pangallo et al., 2017](#)).

The average expected payoff for player i is then

$$\bar{u}_i(\mathbf{p}^t) = \sum_{x \in [m]} \tilde{u}_i(x, \mathbf{p}_{-i}^t) p_i^t(x).$$

For our simulation, the usual continuous replicator equation

$$\dot{p}_i^t(x) = p_i^t(x) (\tilde{u}_i(x, \mathbf{p}_{-i}^t) - \bar{u}_i(\mathbf{p}^t)),$$

must be discretized. We use the discretization proposed in [Maynard Smith \(1982\)](#), where δ takes small values:

$$p_i^{t+1}(x) = p_i^t(x) \frac{1 + \delta \tilde{u}_i(x, \mathbf{p}_{-i}^t)}{1 + \delta \bar{u}_i(\mathbf{p}^t)}.$$

Convergence criteria. Similarly to the other learning rules, we consider the convergence of $\mathbf{p}^t = (\mathbf{p}_1^t, \dots, \mathbf{p}_n^t)$. There are several technical problems associated with simulating replicator dynamics, including the fact that all stable fixed points are on the boundary of the strategy space and therefore cannot be reached in finite simulation time, and that the period of cycles increases over time, due to the infinite memory of the process.

Additionally, we must stop the simulation run as soon as one component of one of the players' mixed strategy vector reaches the machine precision limit and is taken to be zero by the simulator. Indeed, by the properties of replicator dynamics, if $p_i^t(x)$ reaches zero, it remains at zero forever. However, it often happens in simulations of replicator dynamics that an action whose probability had been decreasing for a long time suddenly becomes advantageous due to changes in what other players are playing, leading to a reversal of the dynamics. This reversal will not be reflected in our simulations if $p_i^t(x)$ is stuck at zero due to the machine precision limit being reached, leading to an unfaithful numerical representation of the dynamics.

To address all these issues, and to specifically account for the behavior of replicator dynamics, we choose the following simulation criteria:

- (1) Only consider the last 20% time steps.
- (2) For each player, find the action with the highest probability and verify whether this probability has been increasing over the full time interval.
- (3) Check that the probabilities of all other actions have been decreasing.
- (4) If conditions 2-3 are satisfied for all players, identify the solution run as convergent.

Note that the issue of machine precision unavoidably creates biases when the replicator dynamics take long to reach an attractor, be it a fixed point or a cycle. In particular, it could lead us to consider as non-convergent a simulation run that would eventually converge, because the replicator dynamics hit the machine precision limit while still in a

transient phase. Empirically, it turns out that transient dynamics are longer as the number of players or actions increases, thus these biases are likely to be more serious in “large” games than in games with just a few actions and players.

Parameter values. We choose $\delta = 0.1$.

REFERENCES

- Alon, N., K. Rudov, and L. Yariv (2020). Dominance solvability in random games. <https://lyariv.mycpanel.princeton.edu/papers/DominanceSolvability.pdf>.
- Amiet, B., A. Collevocchio, and M. Scarsini (2019). Pure Nash equilibria and best-response dynamics in random games. arXiv:1905.10758.
- Arratia, R., L. Goldstein, L. Gordon, et al. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *The Annals of Probability* 17(1), 9–25.
- Arthur, W. B. (1991). Designing economic agents that act like human agents: A behavioral approach to bounded rationality. *The American economic review* 81(2), 353–359.
- Babichenko, Y. (2013). Best-reply dynamics in large binary-choice anonymous games. *Games and Economic Behavior* 81, 130–144.
- Berg, J. and M. Weigt (1999). Entropy and typical properties of nash equilibria in two-player games. *EPL (Europhysics Letters)* 48(2), 129–135.
- Blume, L. E. et al. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior* 5(3), 387–424.
- Börger, T. and R. Sarin (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory* 77(1), 1–14.
- Boucher, V. (2017). Selecting equilibria using best-response dynamics. *Economics Bulletin* 37(4), 2728–2734.
- Brown, G. W. (1951). *Iterative solutions of games by fictitious play*. Activity Analysis of Production and Allocation. New York: Wiley.
- Bush, R. R. and F. Mosteller (1953). A stochastic model with applications to learning. *The Annals of Mathematical Statistics* 24(4), 559–585.
- Candogan, O., A. Ozdaglar, and P. A. Parrilo (2013). Dynamics in near-potential games. *Games and Economic Behavior* 82, 66–90.
- Cheung, Y.-W. and D. Friedman (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior* 19(1), 46–76.
- Christodoulou, G., V. S. Mirrokni, and A. Sidiropoulos (2012). Convergence and approximation in potential games. *Theoretical Computer Science* 438, 13–27.
- Cohen, J. E. (1998). Cooperation and self-interest: Pareto-inefficiency of Nash equilibria in finite random games. *Proceedings of the National Academy of Sciences* 95(17), 9724–9731.
- Coucheney, P., S. Durand, B. Gaujal, and C. Touati (2014). General revision protocols in best response algorithms for potential games. In *2014 7th International Conference on NETwork Games, COntrol and OPTimization (NetGCoop)*, pp. 239–246. IEEE.
- Daskalakis, C., A. G. Dimakis, E. Mossel, et al. (2011). Connectivity and equilibrium in random games. *The Annals of Applied Probability* 21(3), 987–1016.
- Dindoš, M. and C. Mezzetti (2006). Better-reply dynamics and global convergence to Nash equilibrium in aggregative games. *Games and Economic Behavior* 54(2), 261–292.
- Dresher, M. (1970). Probability of a pure equilibrium point in n -person games. *Journal of Combinatorial Theory* 8(1), 134–145.

- Durand, S., F. Garin, and B. Gaujal (2019). Distributed best response dynamics with high playing rates in potential games. *Performance Evaluation* 129, 40–59.
- Durand, S. and B. Gaujal (2016). Complexity and optimality of the best response algorithm in random potential games. In *International Symposium on Algorithmic Game Theory*, pp. 40–51. Springer.
- Erev, I. and A. E. Roth (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, 848–881.
- Fabrikant, A., A. D. Jaggard, and M. Schapira (2013). On the structure of weakly acyclic games. *Theory of Computing Systems* 53(1), 107–122.
- Feldman, M. and T. Tamir (2012). Convergence of best-response dynamics in games with conflicting congestion effects. In *International Workshop on Internet and Network Economics*, pp. 496–503. Springer.
- Foster, D. P. and H. P. Young (2006). Regret testing: Learning to play nash equilibrium without knowing you have an opponent. *Theoretical Economics* 1(3), 341–367.
- Friedman, D. (1996). Equilibrium in evolutionary games: Some experimental results. *Economic Journal*, 1–25.
- Friedman, J. W. and C. Mezzetti (2001). Learning in games by random sampling. *Journal of Economic Theory* 98(1), 55–84.
- Fudenberg, D. and D. K. Levine (1998). *The theory of learning in games*. MIT Press.
- Galla, T. and J. D. Farmer (2013). Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences* 110(4), 1232–1236.
- Germano, F. and G. Lugosi (2007). Global nash convergence of foster and young’s regret testing. *Games and Economic Behavior* 60(1), 135–154.
- Goemans, M., V. Mirrokni, and A. Vetta (2005). Sink equilibria and convergence. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*, pp. 142–151. IEEE.
- Gokhale, C. S. and A. Traulsen (2010). Evolutionary games in the multiverse. *Proceedings of the National Academy of Sciences* 107(12), 5500–5504.
- Goldberg, K., A. Goldman, and M. Newman (1968). The probability of an equilibrium point. *Journal of Research of the National Bureau of Standards* 72(2), 93–101.
- Goldman, A. (1957). The probability of a saddlepoint. *The American Mathematical Monthly* 64(10), 729–730.
- Hofbauer, J. and K. Sigmund (1998). *Evolutionary games and population dynamics*. Cambridge university press.
- Kash, I. A., E. J. an, and J. Y. Halpern (2011). Multiagent learning in large anonymous games. *Journal of Artificial Intelligence Research* 40, 571–598.
- Kultti, K., H. Salonen, and H. Vartiainen (2011). Distribution of pure Nash equilibria in n-person games with random best responses. Technical Report 71, Aboa Centre for Economics. Discussion Papers.
- Macy, M. W. and A. Flache (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7229–7236.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.
- McLennan, A. (2005). The expected number of Nash equilibria of a normal form game. *Econometrica* 73(1), 141–174.

- McLennan, A. and J. Berg (2005). Asymptotic expected number of Nash equilibria of two-player normal form games. *Games and Economic Behavior* 51(2), 264–295.
- Mirroknii, V. S. and A. Skopalik (2009). On the complexity of Nash dynamics and sink equilibria. In *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 1–10.
- Monderer, D. and L. S. Shapley (1996). Potential games. *Games and economic behavior* 14(1), 124–143.
- Pangallo, M., T. Heinrich, and J. D. Farmer (2019). Best reply structure and equilibrium convergence in generic games. *Science Advances* 5(2), eaat1328.
- Pangallo, M., J. Sanders, T. Galla, and D. Farmer (2017). A taxonomy of learning dynamics in 2×2 games. arxiv.org/abs/1701.09043.
- Powers, I. Y. (1990). Limiting distributions of the number of pure strategy Nash equilibria in n -person games. *International Journal of Game Theory* 19(3), 277–286.
- Quint, T., M. Shubik, and D. Yan (1997). Dumb bugs vs. bright noncooperative players: A comparison. In W. Albers, W. Güth, P. Hammerstein, B. Moldvanu, and E. van Damme (Eds.), *Understanding Strategic Interaction*, pp. 185–197. Springer.
- Rinott, Y. and M. Scarsini (2000). On the number of pure strategy Nash equilibria in random games. *Games and Economic Behavior* 33(2), 274–293.
- Robinson, J. (1951). An iterative method of solving a game. *The Annals of Mathematics* 54(2), 296.
- Sanders, J. B., J. D. Farmer, and T. Galla (2018). The prevalence of chaotic dynamics in games with many players. *Scientific reports* 8(1), 4902.
- Sarin, R. and F. Vahid (2001). Predicting how people play games: a simple dynamic model of choice. *Games and Economic Behavior* 34(1), 104–122.
- Sato, Y. and J. P. Crutchfield (2003). Coupled replicator equations for the dynamics of learning in multiagent systems. *Physical Review E* 67(1), 1–5.
- Stanford, W. (1995). A note on the probability of k pure Nash equilibria in matrix games. *Games and Economic Behavior* 9(2), 238–246.
- Stanford, W. (1996). The limit distribution of pure strategy Nash equilibria in symmetric bimatrix games. *Mathematics of Operations Research* 21(3), 726–733.
- Stanford, W. (1997). On the distribution of pure strategy equilibria in finite games with vector payoffs. *Mathematical Social Sciences* 33(2), 115–127.
- Stanford, W. (1999). On the number of pure strategy Nash equilibria in finite common payoffs games. *Economics Letters* 62(1), 29–34.
- Swenson, B., R. Murray, and S. Kar (2018). On best-response dynamics in potential games. *SIAM Journal on Control and Optimization* 56(4), 2734–2767.
- Takahashi, S. (2008). The number of pure Nash equilibria in a random game with nondecreasing best responses. *Games and Economic Behavior* 63(1), 328–340.
- Takahashi, S. and T. Yamamori (2002). The pure Nash equilibrium property and the quasi-acyclic condition. *Economics Bulletin* 3(22), 1–6.
- Taylor, C. and M. A. Nowak (2006). Evolutionary game dynamics with non-uniform interaction rates. *Theoretical Population Biology* 69(3), 243–252.
- Tuyts, K., P. J. T. Hoen, and B. Vanschoenwinkel (2006). An evolutionary dynamical analysis of multi-agent learning in iterated games. *Autonomous Agents and Multi-Agent Systems* 12(1), 115–153.

Van Huyck, J., R. Battalio, S. Mathur, P. Van Huyck, and A. Ortmann (1995). On the origin of convention: Evidence from symmetric bargaining games. *International Journal of Game Theory* 24(2), 187–212.