Learning to signal: analysis of a micro-level reinforcement model

Raffaele Argiento [1]
Robin Pemantle [2,3]
Brian Skyrms [4]
Stanislav Volkov [5]

ABSTRACT:
We consider the following signaling game. Nature plays first from the set $\{1, 2\}$. Player 1 (the Sender) sees this and plays from the set $\{A, B\}$. Player 2 (the Receiver) sees only Player 1's play and plays from the set $\{1, 2\}$. Both players win if Player 2's play equals Nature's play and lose otherwise. Players are told whether they have won or lost, and the game is repeated. An urn scheme for learning coordination in this game is as follows. Each node of the desicion tree for Players 1 and 2 contains an urn with balls of two colors for the two possible decisions. Players make decisions by drawing from the appropriate urns. After a win, each ball that was drawn is reinforced by adding another of the same color to the urn. A number of equilibria are possible for this game other than the optimal ones. However, we show that the urn scheme achieves asymptotically optimal coordination.

Keywords: urn model, stochastic approximation, evolution, game, probability, stable, unstable, two-player game.

Subject classification: Primary: 60J20

[1]CNR-Istituto Matematica Applicata e Tecnologie Informatiche (IMATI), Via Bassini, 20133 Milano, Italy, raffaele@mi.imati.cnr.it
[2]Research supported in part by National Science Foundation grant # DMS 0103635
[3]Department of Mathematics, University of Pennsylvania, 209 S. 33rd Street, Philadelphia, PA 19104, USA, pemantle@math.upenn.edu
[4]School of Social Sciences, University of California at Irvine, Irvine, CA 92607
[5]Department of Mathematics, University of Bristol, Bristol BS8 1TW, England

1 Introduction

1.1 Motivation for the model

In recent decades, much attention has been given to repeated two-player, non-zero-sum games and the evolution of strategy. The evolutionary game theory paradigm, originating in the late 1970's in works such as [TJ78] has been thoroughly explored in a variety of contexts, with particular emphasis on explaining how cooperation might arise in games such as the Prisoner's dilemma for which there seems to be an inherent disincentive to cooperate.

Another recent line of inquiry has been the formation of reasonable strategies within a population under myopic, **bounded rationality** types of constraints. The emphasis here is on finding evolutionary pathways whose mechanisms are simple enough that they may be employed by unsophisticated individuals, without much conscious thought, across a wide variety of contexts. This kind of mechanism can, among other things, hope to explain the formation of social and moral norms (see [Sky04; Ale07]). These norms are heuristics that may be easily understood and employed in a wide variety of contexts, and their formation is possible if it results from individual-level mechanisms that are advantageous when averaged over the many contexts in which they are employed.

The emphasis in simplicity of micro-level mechanism has advantages beyond generalizability across contexts and levels of intelligence or conscious thought. It makes models more mathematically tractable. It also allows complexity to be increased in other dimesions, such as allowing the simultaneous evolution of strategy and network structure [SP00; Sky04]. Keeping the number of parameters to a minimum allows in principle for empirical testing and calibration of the micro-level parameters (see, e.g., the discussion of the discount rate parameter in [PS03]). On a theoretical level, finding the simplest, most parsimonious model to explain a phenomenon is generally thought to be a useful step in the investigation of the phenomenon in question.

The notion of an individual employing an urn model to govern plays in a repeated game is very old. The "two-armed bandit" problem, for instance [BJK56], features an individual trying to discover the state of nature and balance the considerations of optimal play under known information against play which will be most informative and thereby lead to gains in the future. One well known strategy for this takes the form of an urn model; see for example [Duf96]. This has been applied to commonly used protocols such as sequential sampling in medical trials [WD78]. Urn models are natural in the context of learning models for several reasons. It is not hard to posit micro-level psychological processes (such as extrapolating from available memories) that correspond well to the model. Urn models typically contain enough noise to avoid certain game-theoretically unstable equilibria while still possessing good convergence properties. Models that are neither quickly fixating (overly long memory) nor ergodic (overly short memory) correspond best to many qualitative learning phenomena.

Early instances of urn models arose as attempts to formulate reasonable strategies for one agent playing against Nature. More recent is the use of urns to model multi-player games in which simplicity at the micro-level is desired. A considerable number of formal interaction models have appeared in the last ten years in the fields of psychology, sociology and political science. One may find many of these, for insance, in the literature on **Agent-based modeling**. This refers to a much broader class of formal systems that includes urn models; for examples of agent-based urn models in sociology, see [BL03; FM02].

An important pre-cursor to the agent-based modeling paradigm is the evolutionary game theory paradigm, which improves the explanatory power of classical game theory by incorporation a Darwinian population dynamic along with the strategic interaction. A more recent twist, introduced in [SP00], is to allow the network of interactions to evolve as well. The explanatory power of such systems and the philosphical ramifications of this are discussed in [Sky04]. Types of classical games to which such an analysis has been applied include prisoner's dilemmas, cooperative games in the vein of Rousseau's Stag Hunt, and bargaining games. In the present work, we take up the application of urn models to signaling games. The model we analyze here is the first in a series of models that incoporate successively more features of signaling problems. The more complex models are described briefly in the final section.

1.2 A two-state, two-signal communication game

We consider the following game, in which the players are the Sender, the Receiver, and Nature. Nature plays first, choosing a state of nature which is either 1 or 2. The Sender sees this play and must choose a signal. In this simplest nontrivial model, there are only two legal signals, A and B. The last play belongs to the Receiver, who sees the Sender's play but not Nature's play. The Receiver chooses an action from the set $\{1, 2\}$. The mutual goal is to have the action match the state of nature. The game is completely cooperative, in the sense that either both players win or both players lose. The game tree has eight paths, which we may denote by 1A1, 1A2, 1B1, 1B2, 2A1, 2A2, 2B1 and 2B2. The first, third, sixth and eighth of these are wins and the other four are losses.

Of course, if the players are allowed to confer, they will simply decide on a code. One reasonable code for the messages sent from the Sender to the Receiver is "A means the state of nature is 1 and B means the state of nature is 2". There is another equally reasonable code, namely "B means the state of nature is 1 and A means the state of nature is 2". Agreeing on either of these beforehand will yield 100% efficiency. Even if the players are not allowed to confer, there are good protocols for minimizing the number of times one fails to get a win. One example, which seems likely to occur in real life, would be for the Sender to choose one of the two languages arbitrarily and never deviate, and for the Receiver eventually to conform to it. This requires differentiating the roles in advance but not breaking the symmetry. Such a protocol might be usable, for example, by nodes in a computer network, if the network is bipartite (there are two types of nodes and communication

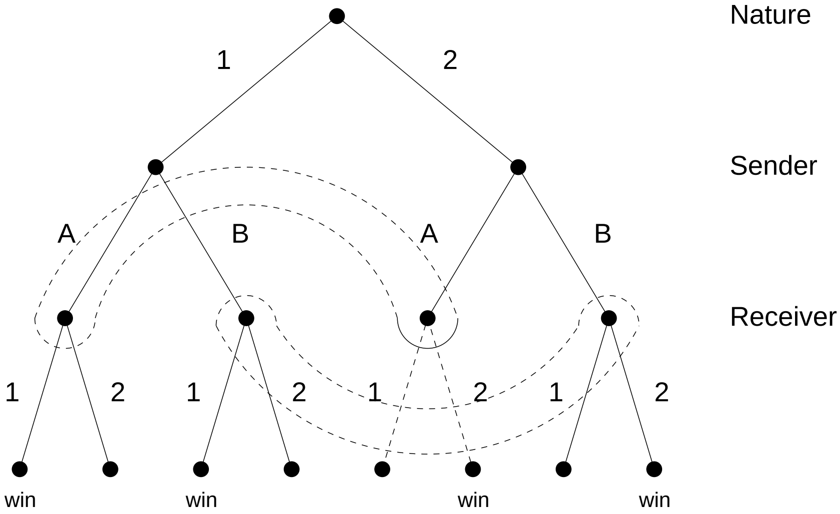

Figure 1: The game tree – dashed lines show information sets for the Receiver

only takes place between nodes of different types). A more general and symmetric protocol might begin with both players playing arbitarily; subsequent plays could be chosen by copying if the last play in the same situation was successful while choosing randomly otherwise. Here, the definition of "in the same situation" is context dependent but the protocol seems otherwise fairly general.

1.3 An urn scheme for playing the game

The Sender's information set is naturally indexed by Nature's plays: $\{1, 2\}$. Correspondingly, the Sender has two urns, call them Urn 1 and Urn 2. Each of these has two colors of ball, call them color A and color B. The contents are initialized to one ball of each color in each urn. Each time the Sender plays, if Nature has played j, then the Sender picks the signal by choosing a ball at random from urn j.

The Receiver plays at four different nodes but, like the Sender, has information partition of size two, indexed by the Sender's plays: $\{A, B\}$. Corresondingly, the Receiver has two urns, called urn A and urn B. Each initially contains one ball of color 1 and one ball of color 2. When the Receiver sees signal x, she chooses an action by picking at random from urn x. Once both players have played, it is revealed whether they won. If they lose, the contents of all urns remain the same, but if they win, the plays are reinforced: each player adds another copy of the ball they chose to the urn it came from. For example, if on the first play the state of nature is 2, the Sender signals A, and the Receiver plays 2 (which, under our model will happen 1/4 of the time that nature plays 2 initially), then a ball of color A is added to the Sender's urn 2 and a ball of color 2 is added to the Receiver's urn A.

We analyze the resulting random sequence of plays under the assumption that Nature chooses states according to independent fair coin flips. When viewed in the classical framework for signaling games, this game has multiple equilibria. In particular, there are two pareto-optimal Nash equilbria corresponding to play according to the two possible languages, and a family of "babbling" equilibria where the Sender ignores the state of nature, choosing signals according to independent coin flips and the Receiver ignores the signal, choosing plays according to her own sequence of independent coin flips. Our goals for this analysis are modest: we show that the urn model protocol converges to one of the two optimal languages in a sense to be made precise in the next section. Note that it is not *a priori* necessary that the urn model produce any Nash equilibrium at all, though we will see in the next section that the model must converge to an appropriately defined set of dynamic equilibria.

1.4 Formal construction of the model and statement of results

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a sufficiently rich source of randomness; for specificity, we take it to be a probability space on which there are defined random variables $\{U_{n,j} : j \in \{1,2,3\}, n \geq 1\}$, that are independent and uniform on the unit interval. Let $\mathcal{F}_n = \sigma(U_{k,j} : k < n, j \leq 3)$ be the σ-field of information up to time n. By induction on n, we may simultaneously define random variables corresponding to the contents of the urn at time n and the plays chosen at time n. The variable $V(n,i,x)$ is interpreted as the number of balls of color x in urn i at time n. The induction begins with the initialization $V(n,i,x) = 1$ for $i = 1,2$ and $x = A, B$ (this populates the two urns used by the Sender) and for $i = A, B$ and $x = 1, 2$ (populating the two urns used by the Receiver).

Given the eight values of $V(n,i,x)$, the plays N_n, S_n and R_n are constructed as follows. Let $N_n = 1$ if $U_{n,1} < 1/2$ and $N_n = 2$ otherwise. The interpretation of N_n is the play chosen by Nature at step n, which is always equally likely to be 1 or 2 independent of the past. Let $S_n = A$ if

$$U_{n,2} < \frac{V(n, N_n, A)}{V(n, N_n, A) + V(n, N_n, B)}$$

and $S_n = B$ otherwise. Thus, conditional on the past, the probability of the Sender choosing signal A is equal to the proportion of balls of color A in the urn N_n, that is, in the urn corresponding to the state that Nature has just played. Similarly, let $R_n = 1$ if

$$U_{n,3} < \frac{V(n, S_n, 1)}{V(n, S_n, 1) + V(n, S_n, 2)}$$

and $R_n = 2$ otherwise, so that the probabilities for the Receiver's plays are proportional to the contents of the urn with the label of the signal at time n.

To complete the induction, update the contents of the urns by defining

$$V(n+1, i, x) = V(n, i, x) + 1$$

if $N_n = i$ and $S_n = x$ and $R_n = i$ (updating the Sender's urns) or if $S_n = i$ and $R_n = x$ and $N_n = x$ (updating the Receiver's urns), and let $V(n+1,i,x) = V(n,i,x)$ otherwise.

Our main result may now be stated. Denote the number of wins up through the n^{th} play by $\mathsf{win}_n := \sum_{k=1}^{n} \delta(N_n, R_n)$ where δ is the usual delta function, namely 1 if the arguments are equal and zero otherwise.

Theorem 1.1. *With probability 1,* $\mathsf{win}_n/n \to 1$ *as* $n \to \infty$*. Furthermore, this occurs in one of two specific ways. With probability* $1/2$*, as* $n \to \infty$*,* $V(n,1,B)/V(n,1,A)$*,* $V(n,2,A)/V(n,2,B)$*,* $V(n,A,2)/V(n,A,1)$ *and* $V(n,B,1)/V(n,B,2)$ *all go to zero, while with probability* $1/2$*, the reciprocals of these all go to zero.*

Remark 1.2. If arbitrary initial conditions are permitted, that is, if $\{V(n,i,x)\}$ are allowed to be any real vector with strictly positive coordinates, then the same conclusions hold with some probability other than $1/2$, measureable in $\mathcal{F}_1$.

Before proceeding to the proof, we make one observation which, though it seems small, reduces the dimension of the problem and simplifies notation considerably. That is, we observe that for each n,

$$\begin{aligned}
V(n+1,1,A) = V(n,1,A)+1 &\quad\Leftrightarrow\quad V(n+1,A,1) = V(n,A,1)+1 \\
V(n+1,1,B) = V(n,1,B)+1 &\quad\Leftrightarrow\quad V(n+1,B,1) = V(n,B,1)+1 \\
V(n+1,2,A) = V(n,2,A)+1 &\quad\Leftrightarrow\quad V(n+1,A,2) = V(n,A,2)+1 \\
V(n+1,2,B) = V(n,2,B)+1 &\quad\Leftrightarrow\quad V(n+1,B,2) = V(n,B,2)+1 .
\end{aligned}$$

We may therefore keep track of the entire process by keeping track of the four quantities $\{V(n,i,x) : i = 1,2; x = A,B\}$ instead of all eight quantities. Denoting

$$\mathbf{V}_n := (V(n,1,A), V(n,1,B), V(n,2,A), V(n,2,B))$$

represents the process as a Markov chain $\{V_n\}$. Various formulae will appear more canonical if we refer to the coordinates of $\mathbf{V}_n$ in order as $1A, 1B, 2A, 2B$ instead of $1,2,3,4$, e.g., $(V_n)_{1A} = V(n,1,A)$ and so forth. If the initial conditions are altered as in Remark 1.2 so that $V(1,i,x) \neq V(1,x,i)$ for some (i,x), then instead of symmetry $V(n,i,x) = V(n,x,i)$, we have that $V(n,i,x) - V(n,x,i)$ is independent of n; the arguments are messier in this case, but the same conclusions hold.

Let $\bowtie$ denote the set $\{1A, 1B, 2A, 2B\}$ and let $T_n := \sum_{j \in \bowtie} V_j$ be the total number of balls in the Sender's urns. Let

$$\mathbf{X}_n := \left(\frac{V(n,1,A)}{T_n}, \frac{V(n,1,B)}{T_n}, \frac{V(n,2,A)}{T_n}, \frac{V(n,2,B)}{T_n} \right) \tag{1.1}$$

be the normalized proportion vector. The vector $\mathbf{X}_n$ is an element of the interior of the 3-simplex

$$\Delta := \{(x_{1A}, x_{1B}, x_{2A}, x_{2B}) \in \mathbb{R}^{\bowtie} : x_{1A}, x_{1B}, x_{2A}, x_{2B} \geq 0, \sum_{j \in \bowtie} x_j = 1\} .$$

Let us write $X_{n,1,A}$ instead of $(X_n)_{1A}$ and so forth. Let $\psi_n = \mathbf{V}_{n+1} - \mathbf{V}_n$ be the standard basis vector corresponding to the reinforcement due to the play at time n if there was a win, and the zero vector otherwise. Thus $|\psi_n| = \mathsf{win}_{n+1} - \mathsf{win}_n$, where we use the L^1-norm on $\mathbb{R}^{\bowtie}$ here and througout.

In this notation, Theorem 1.1 is a consequence of the reformulation

$$\mathbf{X}_n \to \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right) \text{ or } \left(0, \frac{1}{2}, \frac{1}{2}, 0\right) . \tag{1.2}$$

That these happen with equal probability when $\mathbf{V}_1 = (1,1,1,1)$ follows from symmetry.

2 Relation to stochastic approximation and an ODE

A common version of the **stochastic approximation** process is one that satisfies

$$\mathbf{X}_{n+1} - \mathbf{X}_n = \gamma_n (F(\mathbf{X}_n) + \xi_n)) \tag{2.1}$$

where $\{\gamma_n\}$ are constants such that $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$, and where ξ_n are bounded and $\mathbb{E}(\xi_n \,|\, \mathcal{F}_n) = 0$. Sometimes an extra, possibly random, remainder term R_n is added to $F(\mathbf{X}_n) + \xi_n$, with the condition that $\sum_n |R_n| < \infty$ almost surely. There is no precise definition for an urn model, but the normalized content vector in an urn model is typically a stochastic approximation processes with $\gamma_n = 1/n$. One sees this by computing $\mathbb{E}(\mathbf{X}_{n+1} - \mathbf{X}_n \,|\, \mathcal{F}_n)$ and seeing that when scaled by $1/n$ it converges to a vector function F.

To analyze the particluar chain $\{\mathbf{V}_n\}$, or equivalently the time-inhomogeneous chain $\{\mathbf{X}_n\}$, begin by writing down the transition probabilities.

$$\begin{array}{rclcl}
\mathbb{P}(\psi_n = (1,0,0,0)) & = & \mathbb{P}(1A1) & = & \frac{1}{2} \frac{X_{n,1,A}}{X_{n,1,A}+X_{n,1,B}} \frac{X_{n,1,A}}{X_{n,1,A}+X_{n,2,A}} ; \\
\mathbb{P}(\psi_n = (0,1,0,0)) & = & \mathbb{P}(1B1) & = & \frac{1}{2} \frac{X_{n,1,B}}{X_{n,1,B}+X_{n,1,A}} \frac{X_{n,1,B}}{X_{n,1,B}+X_{n,2,B}} ; \\
\mathbb{P}(\psi_n = (0,0,1,0)) & = & \mathbb{P}(2A2) & = & \frac{1}{2} \frac{X_{n,2,A}}{X_{n,2,A}+X_{n,2,B}} \frac{X_{n,2,A}}{X_{n,2,A}+X_{n,1,A}} ; \\
\mathbb{P}(\psi_n = (0,0,0,1)) & = & \mathbb{P}(2B2) & = & \frac{1}{2} \frac{X_{n,2,B}}{X_{n,2,B}+X_{n,2,A}} \frac{X_{n,2,B}}{X_{n,2,B}+X_{n,1,B}} ; \\
\mathbb{P}(\psi_n = (0,0,0,0)) & = & \mathbb{P}(1*2, 2*1) & = & 1 - \mathbb{P}(|\psi_n| = 1) ,
\end{array} \tag{2.2}$$

where $*$ denotes a symbol that can be either A or B. Since ψ_n denotes $\mathbf{V}_{n+1} - \mathbf{V}_n$, we have

$$\mathbf{X}_{n+1} - \mathbf{X}_n = \frac{\mathbf{V}_{n+1}}{1+T_n} - \frac{\mathbf{V}_n}{1+T_n} + \frac{\mathbf{V}_n}{1+T_n} - \frac{\mathbf{V}_n}{T_n} = \frac{1}{1+T_n}(\psi_n - \mathbf{X}_n) \tag{2.3}$$

if $|\psi_n| = 1$, and $\mathbf{X}_{n+1} - \mathbf{X}_n = \mathbf{0}$ otherwise.

Taking expectations gives

$$\mathbb{E}(\mathbf{X}_{n+1} - \mathbf{X}_n \,|\, \mathcal{F}_n) = \frac{1}{1+T_n} F(\mathbf{X}_n) \tag{2.4}$$

where $F(\mathbf{x}) := \mathbb{E}(|\psi_n|(\psi_n - \mathbf{X}_n) \,|\, \mathbf{X}_n = \mathbf{x})$ is a function from Δ to the tangent space $T\Delta := \{\mathbf{x} \in \mathbb{R}^{\bowtie} : \sum_{j \in \bowtie} x_j = 0\}$ given by the formula (written as a column vector so as to fit better)

$$\frac{1}{2}\begin{bmatrix} \frac{(1-x_{1A})x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} - \frac{x_{1A}x_{1B}^2}{(x_{1A}+x_{1B})(x_{1B}+x_{2B})} - \frac{x_{1A}x_{2A}^2}{(x_{2A}+x_{2B})(x_{2A}+x_{1A})} - \frac{x_{1A}x_{2B}^2}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\ -\frac{x_{1B}x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} + \frac{(1-x_{1B})x_{1B}^2}{(x_{1B}+x_{1A})(x_{1B}+x_{2B})} - \frac{x_{1B}x_{2A}^2}{(x_{2A}+x_{2B})(x_{2A}+x_{1A})} - \frac{x_{1B}x_{2B}^2}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\ -\frac{x_{2A}x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} - \frac{x_{2A}x_{1B}^2}{(x_{1B}+x_{1A})(x_{1B}+x_{2B})} + \frac{(1-x_{2A})x_{2A}^2}{(x_{2A}+x_{2B})(x_{2A}+x_{1A})} - \frac{x_{2A}x_{2B}^2}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\ -\frac{x_{2B}x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} - \frac{x_{2B}x_{1B}^2}{(x_{1B}+x_{1A})(x_{1B}+x_{2B})} - \frac{x_{2B}x_{2A}^2}{(x_{2A}+x_{2B})(x_{2A}+x_{1A})} + \frac{(1-x_{2B})x_{2B}^2}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \end{bmatrix}.$$

Letting $\xi_n = (1+T_n)(\mathbf{X}_{n+1} - \mathbf{X}_n - F(\mathbf{X}_n))$ be the noise term, we see that (2.4) is variant of (2.1) with non-determinstic γ_n.

For processes obeying (2.1) or (2.4), the heuristic is that the trajectories of the process should approximate trajectories of the corresponding differential equation $\mathbf{X}' = F(\mathbf{X})$. Let $Z(F)$ denote the set of zeros of the vector field F. The heuristic says that if there are no cycles in the vector field F, then the process should converge to the set $Z(F)$. A sufficient condition for nonexistence of cycles is that there be a **Lyapunov function**, namely a function L such that $\nabla L \cdot F \geq 0$ with equality only where F vanishes. When $Z(F)$ is larger enough to contain a curve, there is a question unsettled by the heuristic, as to whether the process can continue to move around in $Z(F)$. There is however, a nonconvergence heuristic saying that the process should not converge to an unstable equilibrium.

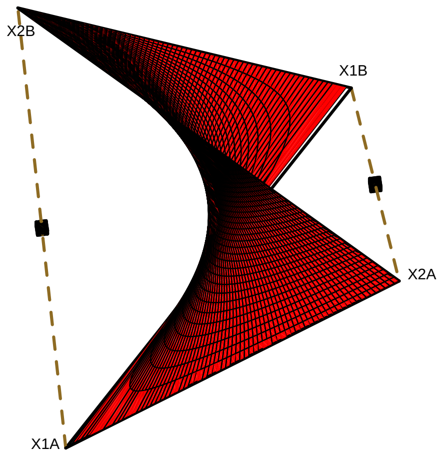

Figure 2: the surface $Q = 0$ in barycentric coordinates, and the two other zeros of F

Proposition 2.1 (Zero set of F). *Let Q be the polynomial $x_{1A}x_{2B} - x_{1B}x_{2A}$. The zero set $Z(F)$ of F on Δ consists of the zero set $Z(Q) := \{Q = 0\}$ together with the two points $(\frac{1}{2}, 0, 0, \frac{1}{2})$ and $(0, \frac{1}{2}, \frac{1}{2}, 0)$.*

PROOF: It is routine to check that F vanishes on the surface and the two points. It suffices, therefore, to check that these are the only solutions to $F = \mathbf{0}$ on Δ. Let Z' be the subset of the simplex where

$x_{1A}x_{1B}x_{2A}$ vanishes. In other words, Z' is a union of three of the four faces of Δ. We claim that $Z(F)$ is contained in the set $Z(Q) \cup Z'$.

First, clearing denominators, we let $P_1, \ldots, P_4$ denote the four polynomials obtained by multiplying the components of F by $(x_{1A} + x_{1B})(x_{1A} + x_{2A})(x_{1B} + x_{2B})(x_{2A} + x_{2B})$. Let $P_5 := 1 - x_{1A} - x_{1B} - x_{2A} - x_{2B}$. We will check that $g := Qx_{1A}x_{1B}x_{2A}$ is contained in the **ideal** generated by $P_1, \ldots P_5$. This is defined as the set of $\sum_{i=1}^5 q_i P_i$ as q_i range over polynomials, and we denote it by $\mathfrak{I}$. Assuming this for the moment, let us see how the claim is proved. On $Z(F)$, we know that P_5 vanishes because $Z(F) \in \Delta$ and $P_1, \ldots, P_4$ vanishes because F vanishes. Hence every polynomial in $\mathfrak{I}$ vanishes, and in particular g vanishes. The set where g vanishes contains $Z(Q) \cup Z'$, which establishes the claim.

Checking that $g \in \mathfrak{I}$ is easy with the aid of a computer algebra system. For example, in Maple 11 with the `Groebner` package loaded, the command

`B := Basis (`$[P_1, P_2, P_3, P_4, P_5]$` , tdeg(`$x_{1A}, x_{1B}, x_{2A}, x_{2B}$`));`

produces a **Gröbner basis** for $\mathfrak{I}$ (with respect to the term order `tdeg(`$x_{1A}, x_{1B}, x_{2A}, x_{2B}$`)`), this being a canonical representation of $\mathfrak{I}$ for which an algorithm exists to test membership. Specifically, given a polynomial g and a Gröbner basis B, the command to produce a **remainder** r for which $g - r \in B$ and r is small (with respect to the same term order) is

`NormalForm(`g` , B , tdeg(`$x_{1A}, x_{1B}, x_{2A}, x_{2B}$`));`

When we try this, we find that $r = 0$, implying that $g \in \mathfrak{I}$ and verifying the claim.

Finally, having seen that $Z(F) \subseteq Z(Q) \cup Z'$, identical arguments show that $Z(F) \subseteq Z(Q) \cup Z''$ where Z'' is the zero set in Δ of the product of any three of the four variables $x_{1A}, x_{1B}, x_{2A}, x_{2B}$. Taking the intersection over the four possible sets Z'' shows that $Z(F) \subseteq Z(Q) \cap Z_*$ where Z_* is the intersection of the zero sets in Δ of the four monomials $x_{1A}x_{1B}x_{2A}, x_{1A}x_{1B}x_{2B}, x_{1A}x_{2A}x_{2B}$ and $x_{1B}x_{2A}x_{2B}$. In other words, Z_* is the 1-skeleton of Δ (the 1-skeleton being the union of all one-dimensional edges). The set $Z(Q)$ already contains four of the six edges in the 1-skeleton. Checking the edge $(\alpha, 0, 0, 1 - \alpha)$ produces exactly one solution to $F = 0$ in the interior of the edge, namely $(\frac{1}{2}, 0, 0, \frac{1}{2})$. Checking the edge $(0, \alpha, 1 - \alpha, 0)$ produces the point $(0, \frac{1}{2}, \frac{1}{2}, 0)$. This finishes the proof of the proposition. $\square$

We now check that $Z(Q)$ is a geometrically unstable set for the vector field F.

Proposition 2.2 (instability of $Z(Q)$).

$$\operatorname{sgn}(\nabla Q \cdot F) = \operatorname{sgn}(Q)$$

at all points of Δ, except $(\frac{1}{2}, 0, 0, \frac{1}{2})$ and $(0, \frac{1}{2}, \frac{1}{2}, 0)$.

PROOF: The previous proposition shows that F vanishes when Q vanishes, so the conclusion is true when $Q = 0$. By symmetry it suffices to prove that $Q > 0$ implies $\nabla Q \cdot F > 0$ on Δ.

Let $\mathbf{x} = (x_{1A}, x_{1B}, x_{2A}, x_{2B})$ be any point of Δ with $Q(\mathbf{x}) > 0$ and with at most one vanishing coordinate. Then the following relations hold:

$$\begin{aligned}
\frac{x_{1A}}{x_{1A}+x_{1B}} &> \frac{x_{2A}}{x_{2A}+x_{2B}}\,; \\
\frac{x_{1A}}{x_{1A}+x_{2A}} &> \frac{x_{1B}}{x_{2B}+x_{1B}}\,; \\
\frac{x_{2B}}{x_{2B}+x_{2A}} &> \frac{x_{1B}}{x_{1B}+x_{1A}}\,; \\
\frac{x_{2B}}{x_{2B}+x_{1B}} &> \frac{x_{2A}}{x_{2A}+x_{1A}}\,;
\end{aligned} \tag{2.5}$$

We may write

$$2\nabla Q \cdot F = x_{1A}x_{2B}H - x_{1B}x_{2A}\tilde{H} \tag{2.6}$$

where

$$\begin{aligned}
H(\mathbf{x}) &= \frac{x_{1A}}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} + \frac{x_{2B}}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} - 4\psi(\mathbf{x}) \\
\tilde{H}(\mathbf{x}) &= \frac{x_{2A}}{(x_{2A}+x_{2B})(x_{2A}+x_{1A})} + \frac{x_{1B}}{(x_{1B}+x_{1A})(x_{1B}+x_{2B})} - 4\psi(\mathbf{x})
\end{aligned}$$

where $\psi(\mathbf{x}) := \mathbb{P}(|\psi_n| = 1 \mid \mathbf{X}_n = \mathbf{x})$.

By the inequalities (2.5),

$$\begin{aligned}
4\psi(\mathbf{x}) &= \frac{2x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} + \frac{2x_{1B}^2}{(x_{1B}+x_{1A})(x_{1B}+x_{2B})} \\
&\qquad + \frac{2x_{2A}^2}{(x_{2A}+x_{2B})(x_{2A}+x_{1A})} + \frac{2x_{2B}^2}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\
&< \frac{2x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} + \frac{x_{1A}(x_{1B}+x_{2A})}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} \\
&\qquad + \frac{x_{2B}(x_{2A}+x_{1B})}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} + \frac{2x_{2B}^2}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\
&= \frac{x_{1A}}{x_{1A}+x_{1B}} + \frac{x_{1A}}{x_{1A}+x_{2A}} + \frac{x_{2B}}{x_{2B}+x_{2A}} + \frac{x_{2B}}{x_{2B}+x_{1B}}\,.
\end{aligned} \tag{2.7}$$

Denote the common denominator

$$D := (x_{1A}+x_{1B})(x_{1A}+x_{2A})(x_{2B}+x_{2A})(x_{2B}+x_{1B})\,. \tag{2.8}$$

It follows (using $x_{1A}+x_{1B}+x_{2A}+x_{2B}=1$ in the second line) that

$$\begin{aligned}
H(\mathbf{x}) &> \frac{x_{1A}(1-(x_{1A}+x_{1B})-(x_{1A}+x_{2A}))}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})}+\frac{x_{2B}(1-(x_{2B}+x_{2A})-(x_{2B}+x_{1B}))}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\
&= \frac{x_{1A}(x_{2B}-x_{1A})}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})}+\frac{x_{2B}(x_{1A}-x_{2B})}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})} \\
&= (x_{1A}+x_{2B})\left[\frac{x_{2B}}{(x_{2B}+x_{2A})(x_{2B}+x_{1B})}-\frac{x_{1A}}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})}\right] \\
&= \frac{(x_{1A}-x_{2B})^2 Q}{D} \\
&> 0\,.
\end{aligned}$$

Analogous computations show that $\tilde{H}<0$. Since at most one of the coordinates vanishes, it follows that the left-hand side of (3.5) is strictly positive.

Finally, if more than one of the coordinates of $\mathbf{x}$ vanishes but $Q\neq 0$, then $\mathbf{x}$ is interior to one of the two line segments $(\alpha,0,0,1-\alpha)$ or $(0,\alpha,1-\alpha,0)$. Plugging in these parametrizations shows the only interior zeros of $\nabla Q\cdot F$ to be at the midpoints. $\square$

3 Probabilistic analysis

Lemma 3.1. *With probability 1,*

$$\frac{1}{2}\leq \liminf \frac{T_n}{n}\leq \limsup \frac{T_n}{n}\leq 1\,.$$

Proof: The upper is trivial because $T_n\leq n-1+T_1$. The lower bound follows from the conditional Borel-Cantelli lemma [Dur04, Theorem I.6] once we show that $\psi(\mathbf{x})$ is always at least $1/2$. To prove the lower bound, multiply the expression (2.7) for ψ by D to clear the denominators, and double. The result is easily seen to be $D+Q^2$. Thus

$$\psi-\frac{1}{2}=\frac{Q^2}{2D}\,,$$

which is clearly a nonnegative quantity. $\square$

With this preliminary result out of the way, the remainder of the proof of Theorem 1.1 may be broken into three pieces, namely Propositions 3.2, 3.3 and 3.4 below. We have seen that $L:=Q^2$ is a Lyapunov function for the stochastic process $\{\mathbf{X}_n\}$; this is implied by Proposition 2.2 and the fact that $\nabla(Q^2)$ is parallel to ∇Q. The minimum value of zero occurs exactly on the surface $Z(Q)$ and the maximum value of $1/16$ occurs at the two other points of $Z(F)$. Let

$$Z_0(Q):=Z(Q)\cap\partial\Delta=Z(D)\,.$$

Proposition 3.2 (Lyapunov function implies convergence). *The stochastic process* $\{L(\mathbf{X}_n)\}$ *converges almost surely to 0 or* $1/16$.

Proposition 3.3 (instability implies nonconvergence). *The probability that* $\lim_{n\to\infty} \mathbf{X}_n$ *exists and is in* $Z(Q) \setminus Z_0(Q)$ *is zero.*

Proposition 3.4 (no convergence to boundary). *The limit* $\lim_{n\to\infty} \mathbf{X}_n$ *exists with probability 1. Furthermore,* $\mathbb{P}\left(\lim_{n\to\infty} \mathbf{X}_n \in Z_0(Q)\right) = 0$.

These three results together imply Theorem 1.1. The first is an easy result; it is shown via martingale methods that $\{\mathbf{X}_n\}$ cannot continue to cross regions where F does not vanish. The second result, fashioned after the non-convergence results of [Pem90, Theorem 1] and generalizations such as [Ben99, Theorem 9.1], follows the argument, by now standard, given in condensed form in [Pem07, Theorem 2.9]. The third result is the trickiest, relying on special properties of the process $\{\mathbf{X}_n\}$. This is necessary because the nonconvergence method of [Pem90] fails near the boundary of an urn scheme due to diminished variance of the increments; a more general rubric for proving nonconvergence to unstable points in such cases and proving convergence of the process (and not just the Lyapunov function) would be desirable.

Proof of Proposition 3.2

Denote $Y_n := L(\mathbf{X}_n)$. Decompose $\{Y_n\}$ into a martingale and a predictable process $Y_n = M_n + A_n$ where $A_{n+1} - A_n = \mathbb{E}(Y_{n+1} - Y_n \,|\, \mathcal{F}_n)$. The increments in Y_n are $O(1/T_n) = O(1/n)$ almost surely by Lemma 3.1, so the martingale $\{M_n\}$ is in L^2 and hence almost surely convergent. To evaluate A_n, use the Taylor expansion

$$L(\mathbf{x} + \mathbf{y}) = L(\mathbf{x}) + \mathbf{y} \cdot \nabla L(\mathbf{x}) + R_{\mathbf{x}}(\mathbf{y})$$

with $R_{\mathbf{x}}(\mathbf{y}) = O(|\mathbf{y}|^2)$ uniformly in $\mathbf{x}$. Then

$$\begin{aligned} A_{n+1} - A_n &= \mathbb{E}\left[L(F(\mathbf{X}_{n+1})) - L(F(\mathbf{X}_n)) \,|\, \mathcal{F}_n\right] \\ &= \mathbb{E}\left[\nabla L(\mathbf{X}_n) \cdot (\mathbf{X}_{n+1} - \mathbf{X}_n) + R_{\mathbf{X}_n}(\mathbf{X}_{n+1} - \mathbf{X}_n) \,|\, \mathcal{F}_n\right] \\ &= \frac{1}{1+T_n} (\nabla L \cdot F)(\mathbf{X}_n) + \mathbb{E}[R_{\mathbf{X}_n}(\mathbf{X}_{n+1} - \mathbf{X}_n) \,|\, \mathcal{F}_n] \,. \end{aligned}$$

Since the $R_{\mathbf{X}_n}(\mathbf{X}_{n+1} - \mathbf{X}_n) = O(T_n^{-2}) = O(n^{-2})$ is summable, this gives

$$A_n = \eta(n) + \sum_{k=1}^{n} \frac{1}{1+T_k} (\nabla L \cdot F)(\mathbf{X}_k)$$

for some almost surely convergent η.

We may now use the usual argument by contradiction: if $\mathbf{X}_n$ is found infinitely often away from the critical values of the Lyapunov function, then the drift would cause the Lyapunov function to blow up. To set this up, observe first that boundedness of $\{Y_n\}$ and $\{M_n\}$ imply that $\{A_n\}$ is bounded. For any $\epsilon \in (0, 1/32)$, let Δ_ϵ denote $L^{-1}[\epsilon, 1/16 - \epsilon]$. On Δ_ϵ, the function $\nabla L \cdot F$, which is always nonnegative, is bounded below by some constant c_ϵ. Let δ be the distance from Δ_ϵ to the complement of $\Delta_{\epsilon/2}$. Suppose $\mathbf{X}_n, \mathbf{X}_{n+1}, \ldots, \mathbf{X}_{n+k-1} \in \Delta_\epsilon$ and $\mathbf{X}_{n+k} \notin \Delta_{\epsilon/2}$. Then, since $|\psi_n|$ and $|\mathbf{X}_n|$ are at most 1, from (2.3) we see that

$$\begin{aligned}
\delta &\leq \sum_{j=n}^{n+k-1} |\mathbf{X}_{j+1} - \mathbf{X}_j| \\
&\leq \sum_{j=n}^{n+k-1} \frac{2}{1+T_j} \\
&\leq \frac{4}{\epsilon} \left[A_{n+k} - A_n - (\eta(n+k) - \eta(n)) \right].
\end{aligned}$$

Thus, if $\mathbf{X}_n \in \Delta_\epsilon$ infinitely often, it follows that $\{A_n\}$ increases without bound. By contradiction, for each ϵ, $\{\mathbf{X}_n\}$ eventually remains outside of Δ_ϵ, which proves the proposition. □

Proof of Proposition 3.3

The idea of this proof appeard first in [Pem88, page 103], but the hypotheses there, as well as those of [Pem90, theorem 1] and [Ben99, Theorem 9.1] require deterministic step sizes $\{\gamma_n\}$ and analyses of isolated unstable fixed points or entire unstable orbits. We therefore take some care here to document what is minimally required of the process $\{\mathbf{X}_n\}$ and its Lyapunov function Q.

For any process $\{Y_n\}$ we let ΔY_n denote $Y_{n+1} - Y_n$. Let $\mathcal{N} \subseteq \mathbb{R}^d$ be any closed set, let $\{\mathbf{X}_n : n \geq 0\}$ be a process adapted to a filtration $\{\mathcal{F}_n\}$ and let $\sigma := \inf\{k : \mathbf{X}_k \notin \mathcal{N}\}$ be the time the process exits $\mathcal{N}$. Let $\mathbb{P}_n$ and $\mathbb{E}_n$ denote conditional probability and expectation with respect to $\mathcal{F}_n$. We will impose several hypotheses, (3.1), (3.2) and (3.3), on $\{\mathbf{X}_n\}$ and then check that the process $\{\mathbf{X}_n\}$ defined in (1.1) satisfies these conditions. We require

$$\mathbb{E}_n |\Delta \mathbf{X}_n|^2 \leq c_1 n^{-2} \tag{3.1}$$

for some constant $c_1 > 0$, which also implies $\mathbb{E}_n |\Delta \mathbf{X}_n| \leq \sqrt{c_1} n^{-1}$. Let Q be a twice differentiable real function on a neighborhood $\mathcal{N}'$ of $\mathcal{N}$. We require that

$$\operatorname{sgn}(Q(\mathbf{X}_n)) \left[\nabla Q(\mathbf{X}_n) \cdot \mathbb{E}_n \Delta \mathbf{X}_n \right] \geq -c_2 n^{-2} \tag{3.2}$$

whenever $\mathbf{X}_n \in \mathcal{N}'$. Let c_3 be an upper bound for the determinant of the matrix of second partial derivatives of Q on $\mathcal{N}'$. We require a lower bound on the incremental variance of the process $\{Q(\mathbf{X}_n)\}$:

$$\mathbb{E}_n (\Delta Q(X_n))^2 \geq c_4 n^{-2} \tag{3.3}$$

when $n < \sigma$. The relation between these assumptions and the process $\{\mathbf{X}_n\}$ defined in (1.1) is as follows.

Lemma 3.5. *Suppose there is a function $F : \mathcal{N} \to T\Delta$ and there are nonnegative quantities $\gamma_n \in \mathcal{F}_n$ and $c' > 0$ such that*

$$|\mathbb{E}_n \Delta \mathbf{X}_n - \gamma_n F(\mathbf{X}_n)| \leq c' n^{-2} ; \tag{3.4}$$

$$\operatorname{sgn}(\nabla Q \cdot F) = \operatorname{sgn}(Q) . \tag{3.5}$$

Then (3.2) *is satisfied. When $\mathcal{N}$ is disjoint from $\partial\Delta$, it follows that the particular process $\{\mathbf{X}_n\}$ defined in* (1.1) *satisfies* (3.1) *and* (3.3) *as well as* (3.2)*.*

PROOF: Let $R := \mathbb{E}_n \Delta \mathbf{X}_n - \gamma_n F(\mathbf{X}_n)$. Then

$$\begin{aligned} \nabla Q(X_n) \cdot \mathbb{E}_n \Delta \mathbf{X}_n &= \nabla Q(\mathbf{X}_n) \cdot [\gamma_n F(\mathbf{X}_n) + R] \\ &\geq 0 - |\nabla Q(\mathbf{X}_n)|\, c' n^{-2} \end{aligned}$$

and (3.2) follows by picking $c_2 \geq c' \sup_{x \in \mathcal{N}} |\nabla Q(x)|$. The process $\{\mathbf{X}_n\}$ of (1.1) satisfies (3.1) because $|\Delta \mathbf{X}_n|$ is bounded from above by n^{-1}. Finally, to see (3.3), note that $|\nabla Q| \geq \epsilon > 0$ on any closed set disjoint from $\partial\Delta$, and also that on such a set $\mathbb{P}(\psi_n = e_j)$ is bounded from below for any elementary basis vector e_j; the lower bound on the second moment of $\Delta Q(\mathbf{X}_n)$ follows. □

Proposition 3.3 now follows from a more general result:

Proposition 3.6. *Let $\{X_n\}$, Q, $\mathcal{N} \subseteq \mathcal{N}'$ and the exit time σ from $\mathcal{N}$ be defined as in the proof of Proposition 3.3 and satisfy* (3.1), (3.2), *and* (3.3), *with constants c_1, c_2 and c_4 appearing there and the bound c_3 on the Hessian determinant of Q on $\mathcal{N}'$. Assume further that there is an N_0 such that for $n \geq N_0$, $X_n \in \mathcal{N} \Rightarrow X_{n+1} \in \mathcal{N}'$. Then*

$$\mathbb{P}\left[\sigma = \infty \text{ and } Q(\mathbf{X}_n) \to 0\right] = 0 .$$

Remark. Proposition 3.3 follows by applying this to a countable cover of $Z(Q) \setminus Z_0(Q)$ by compact sets.

PROOF: The structure of the proof mimics the nonconvergence proofs of [Pem88; Pem90; Ben99]. We show that the incremental quadratic variation of the process $\{Q(\mathbf{X}_n)\}$ is of order at least n^{-2}; this is (3.7). Then we show that conditional on any past at time n, the probability is bounded away from zero that the process $\{Q(\mathbf{X}_n)\}$ wanders away from zero by at least a constant multiple of $n^{-1/2}$ (this is Lemma 3.7) and that the subsequent probability of never returning much nearer to zero is also bounded from below (this is Lemma 3.8).

To begin in earnest, we fix $\epsilon > 0$ and $N \geq N_0$ also satisfying

$$N \geq \frac{16(c_2 + c_1 c_3)^2}{c_4^2} . \tag{3.6}$$

Let $\tau := \inf\{k \geq N : |Q(\mathbf{X}_k)| \geq \epsilon k^{-1/2}\}$. Suppose that $N \leq n \leq \sigma \wedge \tau$. From the Taylor estimate

$$|Q(x+y) - Q(x) - \nabla Q(x) \cdot y| \leq C|y|^2 ,$$

where C is an upper bound on the Hessian determinant for Q on the ball of radius $|y|$ about x, we see that

$$\begin{aligned}
& \mathbb{E}_n \Delta(Q(\mathbf{X}_n)^2) \\
= \;& \mathbb{E}_n 2Q(\mathbf{X}_n)\Delta Q(\mathbf{X}_n) + \mathbb{E}_n(\Delta Q(\mathbf{X}_n))^2 \\
\geq \;& 2Q(\mathbf{X}_n)\nabla Q(\mathbf{X}_n) \cdot \mathbb{E}_n \Delta \mathbf{X}_n - 2c_3 Q(\mathbf{X}_n)\mathbb{E}_n|\Delta \mathbf{X}_n|^2 + \mathbb{E}_n|\Delta Q(\mathbf{X}_n)|^2 \\
\geq \;& \left[-2Q(\mathbf{X}_n)(c_2 + c_3 c_1) + c_4\right] n^{-2} .
\end{aligned}$$

By (3.6), we have $n^{-1/2} \leq c_4/((4(c_2 + c_1 c_3))$. Hence

$$\mathbb{E}_n \Delta(Q(\mathbf{X}_n)^2) \geq \frac{c_4}{2} n^{-2} . \tag{3.7}$$

Lemma 3.7. *If ϵ is taken to equal $c_4/2$ in the definition of τ, then $\mathbb{P}_n(\tau \wedge \sigma < \infty) \geq 1/2$.*

Proof: For any $m \geq n$ it is clear that $|Q(\mathbf{X}_{m\wedge\tau\wedge\sigma})| \leq \epsilon n^{-1/2}$. Thus,

$$\begin{aligned}
\epsilon n^{-1} \;&\geq\; \mathbb{E}_n Q(\mathbf{X}_{m\wedge\tau\wedge\sigma})^2 \\
&\geq\; \mathbb{E}_n \left[Q(\mathbf{X}_{m\wedge\tau\wedge\sigma})^2 - Q(\mathbf{X}_n)^2\right] \\
&=\; \sum_{k=n}^{m-1} \mathbb{E}_n \Delta(Q(\mathbf{X}_k)^2)\mathbf{1}_{k<\tau\wedge\sigma} \\
&\geq\; \sum_{k=n}^{m=1} c_4 n^{-2}\mathbb{P}_n(\sigma \wedge \tau > k) \\
&\geq\; \frac{c_4}{2}(n^{-1} - m^{-1})\mathbb{P}_n(\sigma \wedge \tau = \infty) .
\end{aligned}$$

Letting $m \to \infty$ we conclude that $\epsilon \leq c_4/2$ implies $\mathbb{P}(\tau \wedge \sigma = \infty) \leq 1/2$. □

Lemma 3.8. *There is an N_0 and a $c_5 > 0$ such that for all $n \geq N_0$,*

$$\mathbb{P}_n\left(\sigma < \infty \textit{ or for all } m \geq n,\ |Q(\mathbf{X}_m)| \geq \frac{c_4}{5} n^{-1/2}\right) \geq c_5$$

whenever $|Q(\mathbf{X}_n)| \geq (c_4/2)n^{-1/2}$.

Let us now see that Lemmas 3.7 and 3.8 prove Proposition 3.6. Let $\mathcal{N}$ be any closed ball in the interior of Δ and let $\mathcal{N}'$ be any convex neighborhood of $\mathcal{N}$ whose closure is still in the interior of Δ. For $n \geq N_0$, we have

$$\mathbb{P}_n\left[\sigma = \infty \text{ and } Q(\mathbf{X}_n) \to 0\right] \leq \frac{1}{2} + \frac{1}{2}(1 - c_5) < 1 .$$

But $\mathbb{P}_n(A) \to \mathbf{1}_A$ almost surely for any event $A \in \sigma(\bigcup_n \mathcal{F}_n)$. Thus

$$\mathbb{P}_n\left[\sigma = \infty \text{ and } Q(\mathbf{X}_n) \to 0\right] \to 1$$

almost surely on the event $\{\sigma = \infty \text{ and } Q(\mathbf{X}_n) \to 0\}$, and it follows that the probability of this event is zero. It remains to prove Lemma 3.8.

Let $\phi(x) := \phi_\lambda(x) := x + \lambda x^2$ and let $\tilde{Q}(x) := \phi(Q(x))$. First, we establish that there is a $\lambda > 0$ such that $\tilde{Q}(\mathbf{X}_n)$ is a submartingale when $Q \geq 0$ and $n \geq N_0$.

$$\begin{aligned}
\mathbb{E}_n \Delta \tilde{Q}(\mathbf{X}_n) &= \mathbb{E}_n \Delta Q(\mathbf{X}_n) + \lambda \mathbb{E}_n \Delta (Q(\mathbf{X}_n)^2) \\
&\geq \nabla Q(\mathbf{X}_n) \cdot \mathbb{E}_n \Delta \mathbf{X}_n - c_3 \mathbb{E}_n |\Delta \mathbf{X}_n|^2 + \lambda \frac{c_4}{2} n^{-2}
\end{aligned}$$

Choosing $\lambda \geq (2/c_4)(c_2 + c_1 c_3)$ then yields a submartingale when $Q(\mathbf{X}_n) \geq 0$.

Next, let $M_n + A_n$ denote the Doob decomposition, of $\{\tilde{Q}(\mathbf{X}_n)\}$; in other words, $\{M_n\}$ is a martingale and A_n is predictable and increasing. An upper bound on $|\phi'_\lambda(Q(x))|$ is $c_7 := 1 + 2\lambda \sup |Q| = 1 + 2\lambda$. From the definition of Q, we see that $|\nabla Q| \leq 1$. It follows from these two facts that

$$\frac{|\tilde{Q}(x+y) - \tilde{Q}(x)|}{|y|} \leq 1 + 2\lambda \,.$$

It is now easy to estimate that

$$\begin{aligned}
\mathbb{E}_n (\Delta M_n)^2 &\leq \mathbb{E}_n (\Delta \tilde{Q}(\mathbf{X}_n))^2 \\
&\leq \left(\sup \frac{|\tilde{Q}(x+y) - \tilde{Q}(x)|}{|y|} \right) \mathbb{E}_n |\Delta \mathbf{X}_n|^2 \\
&\leq c_1 c_7 n^{-2} \sup \frac{d\tilde{Q}}{dQ} \,.
\end{aligned}$$

We conclude that there is a constant $c_6 > 0$ such that $\mathbb{E}_n (\Delta M_n)^2 \leq c_6 n^{-2}$ and consequently

$$\mathbb{E}_n (M_{n+m} - M_n)^2 \leq c_6 n^{-1} \tag{3.8}$$

for all $m \geq 0$ on the event $\{Q(\mathbf{X}_n) \geq 0\}$.

For any $a, n, V > 0$ and any martingale $\{M_k\}$ satisfying $M_n \geq a$ and $\sup_m \mathbb{E}_n (M_{n+m} - M_n)^2 \leq V$, there holds an inequality

$$\mathbb{P}\left(\inf_m M_{n+m} \leq \frac{a}{2} \right) \leq \frac{4V}{4V + a^2} \,.$$

To see this, let $\tau = \inf\{k \geq n : M_k \leq a/2\}$ and let $p := \mathbb{P}_n(\tau < \infty)$. Then

$$V \geq p \left(\frac{a}{2}\right)^2 + (1-p) \mathbb{E}_n \left(M_\infty - M_n \,|\, \tau = \infty \right)^2 \geq p \left(\frac{a}{2}\right)^2 + (1-p) \left(\frac{p(a/2)}{1-p} \right)^2$$

which is equivalent to $p \leq 4V/(4V + a^2)$. It follows, with $a = \frac{c_4}{2} n^{-1/2}$ and $V = c_6 n^{-1}$, that

$$\mathbb{P}_n \left(\inf_{k \geq n} M_k \leq \frac{c_4}{4} n^{-1/2} \right) \leq c_5 := \frac{4 c_6}{4 c_6 + (1/4) c_4^2} \,.$$

But $M_k \leq \tilde{Q}(\mathbf{X}_k)$ for $k \geq n$, so $Q(\mathbf{X}_k) \leq (c_4/5)n^{-1/2}$ implies $\tilde{Q}(\mathbf{X}_k) \leq (c_4/4)n^{-1/2}$ for $n \geq N_0$, which implies $M_k \leq (c_4/4)n^{-1/2}$. Thus the conclusion of the lemma is established in the positive case, $Q(\mathbf{X}_n) \geq (c_4/2)n^{-1/2}$. An entirely analogous computation shows that $Q(\mathbf{X}_n) - \lambda Q(\mathbf{X}_n)^2$ is a supermartingale when $Q(\mathbf{X}_n) \leq 0$, and the conclusion follows as well in the negative case, that is, the case $Q(\mathbf{X}_n) \geq (c_4/2)n^{-1/2}$. The lemma is established, and along with it, Proposition 3.3. $\square$

Proof of Proposition 3.4

The following lemma compares an urn process to a Pólya urn, deducing from the known properties of Pólya's urn that the compared urn satisfies an inequality. The proof is easy and consumes space only in order to spell out certain couplings.

Lemma 3.9. *Suppose an urn has balls of two colors, white and black. Suppose that the number of balls increases by precisely 1 at each time step, and denote the number of white balls at time n by W_n and the number of black balls by B_n. Let $X_n := W_n/(W_n + B_n)$ denote the fraction of white balls at time n and let $\mathcal{F}_n$ denote the σ-field of information up to time n. Suppose further that there is some $0 < p < 1$ such that the fraction of white balls is always attracted toward p:*

$$(\mathbb{P}(X_{n+1} > X_n \,|\, \mathcal{F}_n) - X_n) \cdot (p - X_n) \geq 0\,.$$

Then the limiting fraction $\lim_{n\to\infty} X_n$ almost surely exists and is strictly between zero and one.

Proof: Let $\tau_N := \inf\{k \geq N : X_k \leq p\}$ be the first time after N that the fraction of white balls drops below p. The process $\{X_{k\wedge\tau_N} : k \geq N\}$ is a bounded supermartingale, hence converges almost surely. Let $\{W'_k, B'_k) : k \geq N\}$ be a Pólya urn process coupled to $\{(W_k, B_k)\}$ as follows. Let $(W'_N, B'_N) = (W_N, B_N)$. We will verify inductively that $X_k \leq X'_k := W'_k/(W'_k + B'_k)$ for all $k \leq \tau_N$. If $k < \tau_N$ and $W_{k+1} - W_k = 1$ then let $W'_{k+1} = W'_k + 1$. If $k < \tau_N$ and $W_{k+1} = W_k$ then let Y_{k+1} be a Bernoulli random variable independent of everything else with $\mathbb{P}(Y_{k+1} = 0 \,|\, \mathcal{F}_k) = (1 - X'_k)/(1 - X_k)$, which is nonnegative. Let $W'_{k+1} = W_k + Y_{k+1}$. The construction guarantees that $X'_{k+1} \geq X_{k+1}$, completing the induction, and it is easy to see that $\mathbb{P}(W'_{k+1} > W_k) = X'_k$, so that $\{X'_k : N \leq k \leq \tau_N\}$ is a Pólya urn process.

Complete the definition by letting $\{X'_k\}$ evolve independently as a Pólya urn process once $k \geq \tau_N$. It is well known that X'_k converges almost surely and that the conditional law of $X'_\infty := \lim_{k\to\infty} X_k$ given $\mathcal{F}_N$ is a beta distribution, $\beta(W_N, B_N)$. For later use, we remark that beta distributions satisfy the estimate

$$\mathbb{P}\left(|\beta(xn, (1-x)n) - x| > \delta\right) \leq c_1 e^{-c_2 n\delta} \tag{3.9}$$

uniformly for x in a compact sub-interval of $(0, 1)$ Since the beta distribution has no atom at 1, we see that $\lim_{k\to\infty} X_k$ is strictly less than 1 on the event $\{\tau_N = \infty\}$. An entirely analogous argument with τ_N replaced by $\sigma_N := \inf\{k \geq N : X_k \geq p\}$ shows that $\lim_{k\to\infty} X_k$ is strictly greater than 0 on the event $\{\sigma_N = \infty\}$. Taking the union over N shows that $\lim_{k\to\infty} X_k$ exists on the event $\{(X_k - p)(X_{k+1} - p) < 0$ finitely often$\}$ and is strictly between zero and one. The proof of the lemma will therefore be finished once we show that $X_k \to p$ on the event that $X_k - p$ changes sign infinitely often.

Let $G(N,\epsilon)$ denote the event that $X_{N-1} < p < X_N$ and there exists $k \in [N, \tau_N]$ such that $X_k > p + \epsilon$. Let $H(N,\epsilon)$ denote the event that $X_{N-1} > p > X_N$ and there exists $k \in [N, \sigma_N]$ such that $X_k < p + \epsilon$. It suffices to show that for every $\epsilon > 0$, the sums $\sum_{N=1}^\infty \mathbb{P}(G(N,\epsilon))$ and $\sum_{N=1}^\infty \mathbb{P}(H(N,\epsilon))$ are finite; for then by Borel-Cantelli, these occur finitely often, implying $p - \epsilon \le \liminf X_k \le \limsup X_k \le p + \epsilon$ on the event that $X_k - p$ changes sign infinitely often; since ϵ is arbitrary, this suffices. Recall the Pólya urn coupled to $\{X_k : N \le k \le \tau_N\}$. On the event $G(N,\epsilon)$, either $X'_\infty \ge \epsilon/2$ or $X'_\infty - X_\rho \le -\epsilon/2$ where $\rho \ge k$ is the least $m \ge N$ such that $S'_m \ge \epsilon$. The conditional distribution of $X'_\infty - X_\rho$ given $\mathcal{F}_\rho$ is $\beta(W'_\rho, B'_\rho)$. Hence

$$\mathbb{P}(G(N,\epsilon)) \le \mathbb{E}\mathbf{1}_{X_{N-1}<p<X_N}\mathbb{P}(\beta(W_N,B_N) \ge \frac{\epsilon}{2} + \mathbb{E}\mathbf{1}_{\rho<\infty}\mathbb{P}(\beta(W'_\rho,B'_\rho) \le -\frac{\epsilon}{2}). \tag{3.10}$$

Combining this with the estimate (3.9) establishes summability of $\mathbb{P}(G(N,\epsilon))$. An entirely analogous argument establishes summability of $\mathbb{P}(H(N,\epsilon))$, finishing the proof of the lemma. $\square$

Proof of Proposition 3.4: Color the urn process $\{\mathbf{V}_n\}$, by coloring balls of types 1A and 1B white and coloring balls of type 2A and 2B black. Let $\tau_k := \inf\{k : T_n = k\}$ denote the times of increase of $\{T_n\}$. We let $W_k := V(\tau_k, 1, A) + V(\tau_k, 2, B)$ denote the number of white balls at time τ_k and $B_k := V(\tau_k, 2, A) + V(\tau_k, 1, B)$ denote the number of black balls. We claim that the urn process $\{(W_k, B_k)\}$ satisfies the hypotheses of Lemma 3.9 with $p = 1/2$. To verify this, let $(x_{1A}, x_{1B}, x_{2A}, x_{2B})$ denote $\mathbf{X}_{\tau_n}$ and write $\mathbb{P}(X_{n+1} > X_n \,|\, \mathcal{F}_n) - X_n$ as `Num/Den` where

$$\begin{aligned}
\texttt{Num} &= \frac{x_{1A}^2}{(x_{1A}+x_{1B})(x_{1A}+x_{2A})} + \frac{x_{1B}^2}{(x_{1A}+x_{1B})(x_{2B}+x_{1B})};\\
\texttt{Den} &= \texttt{Num} + \frac{x_{2A}^2}{(x_{2B}+x_{2A})(x_{1A}+x_{2A})} + \frac{x_{2B}^2}{(x_{2B}+x_{1A})(x_{2B}+x_{1B})}.
\end{aligned}$$

Simplifying and using $x_{1A} + x_{1B} + x_{2A} + x_{2B} = 1$ shows that

$$\mathbb{P}(X_{n+1} > X_n \,|\, \mathcal{F}_n) - X_n = -\frac{(x_{1,A} + x_{1,B} - x_{2,A} - x_{2,B})\, Q^2}{(x_{1,A} + x_{1,B} + x_{2,A} + x_{2,B})\,(Q^2 + D)}$$

where, as before, D denotes the common denominator (2.8). This is clearly nonpositive when $x_{1A} + x_{1B} \ge x_{2A} + x_{2B}$. This is the same condition as $x_{1A} + x_{1B} \ge 1/2$, so the claim is proved.

Lemma 3.9 now allows us to conclude that $(V(n,1,A) + V(n,1,B))/T_n$ converges to a nonzero value. The process $\{\mathbf{V}_n : n \ge 0\}$ is invariant under transposing the first and fourth coordinates, as also under transposing the second and third coordinates. We conclude that the four quantities

$$\begin{array}{cc}
\dfrac{V(n,1,A)+V(n,1,B)}{T_n}, & \dfrac{V(n,1,A)+V(n,2,A)}{T_n},\\[2mm]
\dfrac{V(n,2,B)+V(n,2,A)}{T_n}, & \dfrac{V(n,2,B)+V(n,1,B)}{T_n}
\end{array} \tag{3.11}$$

all converge almost surely to nonzero values.

Combining this with Proposition 3.2, we see that there is almost surely a pair of numbers $a, b \in (0,1)$ such that the limit set of $\mathbf{V}_n$ is contained in the set

$$\Xi_{a,b} := \{x := (x_{1A}, x_{1B}, x_{2A}, x_{2B}) \in \Delta : L(x) \in \{0, \frac{1}{16}\} \text{ and } x_{1A} + x_{1B} = a \text{ and } x_{1A} + x_{2A} = b\}\,.$$

When $a = b = 1/2$, the set $\Xi_{a,b}$ consists of the three points $(1/2, 0, 0, 1/2), (0, 1/2, 1/2, 0)$ and $(1/4, 1/4, 1/4, 1/4)$. In any other case, the set $\{x_{1A} + x_{1B} = a, x_{1A} + x_{2A} = b\}$ in the simplex Δ is a line segment parallel to $(1, -1, -1, 1)$, and can never intersect $\{Q = 0\}$ in more than one point, hence the set $\Xi_{a,b}$ consists of at most one point. Almost sure convergence of $\mathbf{V}_n$ follows.

On $Z_0(Q)$, one of the four quantities $x_{1A} + x_{1B}, x_{1A} + x_{2A}, x_{2B} + x_{2A}, x_{2B} + x_{1B}$ always vanishes; according to (3.11), the law of $\lim_{n\to\infty} \mathbf{V}_n$ must therefore give zero probability to the set $Z_0(Q)$, establishing the final conclusion of the proposition, and finishing the proof of Theorem 1.1. $\square$

4 Discussion

We have analyzed what we consider to be the simplest nontrivial model of a coordination game. There are a number of natural extensions to the model, all of which raise interesting questions and none of which has been rigorously analyzed. A list of extensions for which we have both simulations and heuristics (via an ODE) but no rigorous analyses is as follows: states not equally probable; number of states, signals or acts greater than two (the problems differ depending on which of the numbers is greatest); more than two agents interacting in a signaling network. We consider these in turn.

Suppose we have 2 states, 2 acts and 3 signals. Do we still get efficient signaling? Does one signal fall out of use so that we end up with essentially a 2 signal system, or does one signal comes to stand for one state and the other two persist as synonyms for the other state? Heuristics and simulations suggest that synonyms form, with no signal falling out of use. Suppose we have 3 states, 3 acts and 2 signals. There is now an informational bottleneck and efficient signaling is right only 2/3 of the time. Again efficiency could be achieved in different ways. It appears that one signal is shared bewteen two states, rather than one state being left without a signal. Moving beyond two agents, suppose that there are two senders and one receiver. There are 4 states, but each sender only observes the correct member of a partition. Sender 1 observes the partition $\{\{1,2\},\{3,4\}\}$ and Sender 2 observes the partition $\{\{1,3\},\{2,4\}\}$. Each sender has 2 signals and the receiver has 4 acts, each paying off in exactly one state, and in that case everyone is reinforced. On the other hand we can have one sender and two receivers. The sender observes one of 4 states, and sends one of 2 signals to each receiver. The receivers each choose one of two acts, and the pair of acts chosen must be right for the state for everyone to be reinforced. We can have a chain, where the sender observes one of 2 states, sends one of 2 signals to an intermediary and the intermediary sends one of 2 signals to the receiver. The receiver must do one of two acts, and if it is right for the state all get reinforced. Simulations suggest that in each of the models described in this paragraph, individuals always learn to signal.

However, even simpler variations may introduce new complexity. With 3 states, 3 signals and 3 acts, there is a new class of equilibria of partial information transfer, which combines bottlenecks and synonyms. For example, the sender always sends signal 1 in states 1 and 2 and mixes between signals 2 and 3 in state 3. The receiver always does acts 3 when getting signals 2 and 3, and mixes between acts 1 and 2 when getting signal 1. Simulations suggest that reinforcement sometimes converges to such equilibria and sometimes to signaling systems. The slow convergence of such systems to equilibrium behavior casts some doubt on whether these mixed equilibria are in fact possible (cf. [PV99, Theorem 1.2] and the remark following; see also [PS01]).

Problem. *Determine whether mixed equilibria are possible in the case of 3 states, 3 signals and 3 acts.*

Finally, if we lift the assumption that states are equiprobable, simulations suggest that even in the 2 state, 2 signal, 2 act case it is possible for reinforcement to converge to a state where the receiver ignores the signal and always chooses the act that is right for the most probable state. In these cases, recovery of almost sure convergence to efficient signaling may require some perturbation of the learning dynamics.

References

[Ale07] J. Alexander. *The Structural Evolution of Morality.* Cambridge University Press, Cambridge, 2007.

[Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Seminaires de Probabilités XXXIII*, volume 1709 of *Lecture notes in mathematics*, pages 1–68. Springer-Verlag, Berlin, 1999.

[BJK56] R. Bradt, S. Johnson, and S. Karlin. On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.*, 27:1060–1074, 1956.

[BL03] P. Bonacich and T. Liggett. Asymptotics of a matrix valued Markov chain arising in sociology. *Stochastic Proc. Appl.*, 104:155–171, 2003.

[Duf96] M. Duflo. *Algorithmes Stochastiques.* Springer, Berlin, 1996.

[Dur04] R. Durrett. *Probability: Theory and Examples.* Duxbury Press, Belmont, CA, third edition, 2004.

[FM02] A. Flache and M. Macy. Stochastic collusion and the power law of learning. *J. Conflict Res.*, 46:629–653, 2002.

[Pem88] R. Pemantle. *Random processes with reinforcement*. Doctoral Dissertation. M.I.T., 1988.

[Pem90] R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Ann. Probab.*, 18:698–712, 1990.

[Pem07] Robin Pemantle. A survey of random processes with reinforcement. *Probab. Surv.*, 4(0):1–79, 2007.

[PS01] R. Pemantle and B. Skyrms. Reinforcement schemes may take a long time to exhibit limiting behavior. *Preprint*, 2001.

[PS03] R. Pemantle and B. Skyrms. Time to absorption in discounted reinforcement models. *Stochastic Proc. Appl.*, 109:1–12, 2003.

[PV99] R. Pemantle and S. Volkov. Vertex-reinforced random walk on Z has finite range. *Ann. Probab.*, 27:1368–1388, 1999.

[Sky04] B. Skyrms. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge, 2004.

[SP00] B. Skyrms and R. Pemantle. A dynamic model of social network formation. *Proc. Nat. Acad. Sci. U.S.A.*, 97:9340–9346, 2000.

[TJ78] P. Taylor and L. Jonker. Evolutionary stable strategies and game dynamics. *Math. Biosci.*, 40:145–146, 1978.

[WD78] L. Wei and S. Durham. The randomized play-the-winner rule in medical trials. *J. Amer. Stat. Assoc.*, 73:840–843, 1978.