A Natural Adaptive Process for Collective Decision-Making

Florian Brandl
Department of Economics
University of Bonn

Felix Brandt
Department of Computer Science
Technical University of Munich

Consider an urn filled with balls, each labeled with one of several possible collective decisions. Now, let a random voter draw two balls from the urn and pick her more preferred as the collective decision. Relabel the losing ball with the collective decision, put both balls back into the urn, and repeat. Once in a while, relabel a randomly drawn ball with a random collective decision. We prove that the empirical distribution of collective decisions produced by this process approximates a maximal lottery, a celebrated probabilistic voting rule proposed by Peter C. Fishburn (Rev. Econ. Stud., 51(4), 1984). In fact, the probability that the collective decision in round n is made according to a maximal lottery increases exponentially in n. The proposed procedure is more flexible than traditional voting rules and bears strong similarities to natural processes studied in biology, physics, and chemistry as well as algorithms proposed in machine learning.

1. Introduction

The question of how to collectively select one of many alternatives based on the preferences of multiple agents has occupied great minds from various disciplines. Its formal study goes back to the Age of Enlightenment, in particular during the French

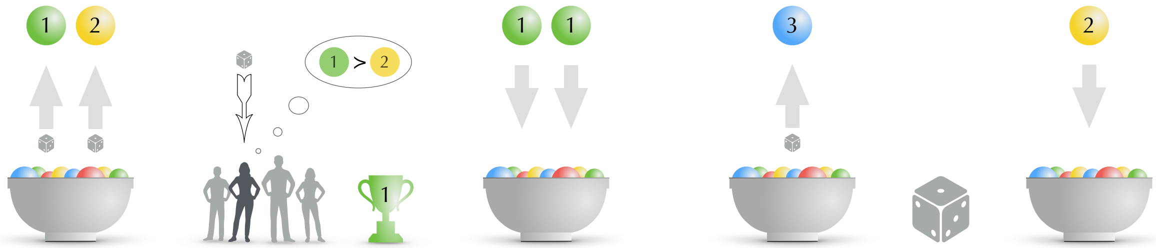
Revolution, and the important contributions by Jean-Charles de Borda and Marie Jean Antoine Nicolas de Caritat, better known as the Marquis de Condorcet. Borda and Condorcet agreed that plurality rule—then and now the most common collective choice procedure—has serious shortcomings. This observation remains a point of consensus among social choice theorists and is largely due to the fact that plurality rule merely asks each voter for her most-preferred alternative (see, e.g., Brams and Fishburn, 2002; Laslier, 2011).[1] When eliciting more fine-grained preferences such as complete rankings over all alternatives from the voters, much more attractive choice procedures are available. As a matter of fact, since Arrow's (1951) seminal work, the standard assumption in social choice theory is that preferences are given in the form of binary relations that satisfy completeness, transitivity, and often anti-symmetry. Despite a number of results which prove critical limitations of choice procedures for more than two alternatives (e.g., Arrow, 1951; Gibbard, 1973; Satterthwaite, 1975), there are many encouraging results (e.g. Young, 1974; Young and Levenglick, 1978; Brams and Fishburn, 1978; Laslier, 2000a). In particular, when allowing for randomization between alternatives, some of the traditional limitations can be avoided and there are appealing choice procedures that stand out (Gibbard, 1977; Brandl et al., 2016; Brandl and Brandt, 2020).

The standard framework in social choice theory rests on a number of rigid assumptions that confine its applicability: there is a fixed set of voters, a fixed set of alternatives, and a single point in time when preferences are to be aggregated; all voters are able to rank-order all alternatives; there is a central authority that collects all these rankings, computes the outcome, and convinces voters of the outcome's correctness, etc. On top of that, computing the outcome of many attractive choice procedures is a demanding task that requires a computer, which can render the process less transparent to voters.[2]

In this paper, we devise an ongoing process in which voters may arrive, leave, and change their preferences over time and collective decisions are made repeatedly at

[1]For example, plurality rule may select an alternative that an overwhelming majority of voters consider to be the worst of all alternatives.

[2]In some cases, computing the outcome was even shown to be NP-hard, i.e., the running time of all known algorithms for computing election winners increases exponentially in the number of alternatives (see, e.g., Bartholdi, III et al., 1989; Brandt et al., 2016).



(i) Draw two balls, random voter picks winner, relabel losing ball.

(ii) Randomly relabel one drawn ball with small probability r (mutation).

(iii) The distribution of winners converges to a maximal lottery w.r.t. the voters' preferences.

Figure 1: Illustration of one round of the urn process (i and ii) and the main result (iii).

intervals. Voters are never asked for their complete preference relations, but rather reveal minimal information about their preferences by choosing between two randomly drawn alternatives from time to time. No central voting authority is required. The process can be executed via a simple physical device: an urn filled with balls that allows for two primitive operations: *(i)* randomly sampling a ball and *(ii)* replacing a sampled ball of one kind with a ball of another kind. More precisely, the process works as follows (see Figure 1). There is an urn filled with balls that each carry the label of one alternative. The initial distribution of balls in the urn is arbitrary. In each round, a randomly selected voter will draw two balls from the urn at random. Say these two balls are labeled with alternatives 1 and 2, and the voter prefers 1 to 2. She will then change the label of the second ball to 1 and return both balls to the urn. Alternative 1 is declared the collective choice—or winner—of this round. After each round, with some small probability r which we call *mutation rate*, a randomly drawn ball is relabeled with a random alternative.

We show that if the number of balls in the urn is sufficiently large, then the limit of the empirical distribution of winners is almost surely close to a *maximal lottery*—a randomized extension of the Condorcet principle that was proposed by Fishburn (1984) and enjoys many desirable axiomatic properties. How far the limiting distribution will be from a maximal lottery depends on r. As r goes to 0, the limiting distribution converges to a maximal lottery. We can, however, not set r to 0 as then almost surely, all alternatives except one will permanently disappear from the urn and the limiting distribution will be degenerate. Our proof not only shows convergence of the limiting distribution but also that the probability that the urn distribution itself is close to a maximal lottery gets arbitrarily close to 1 and increases exponentially in the number of rounds. The winners of most rounds are thus selected according to approximate maximal lotteries.

1.1. Maximal Lotteries and Dynamic Voting

The basic idea of maximal lotteries is to avoid the Condorcet paradox—which lies at the heart of classic impossibility theorems—by extending the notion of a Condorcet winner to lotteries. A lottery p is a randomized Condorcet winner—or *maximal*—if for any other lottery q, a random voter is more likely to prefer the alternative sampled from p to that sampled from q than *vice versa*.[3] The minimax theorem guarantees that maximal lotteries exist. Maximal lotteries also have a natural interpretation in terms of electoral competition (see, e.g., Myerson, 1993; Laslier, 2000b; Carbonell-Nicolau and Ok, 2007). In fact, maximal lotteries are precisely the mixed Nash equilibrium (or maximin) strategies of the symmetric two-player zero-sum game given by the pairwise majority margins of the voters' preferences. When interpreting the two players as parties and the alternatives as possible positions of the parties, this can be seen as a game of electoral competition in which two parties aim at maximizing the number of voters who prefer their (mixed) position to that of the other party. For this reason, the social choice literature sometimes refers to the support of maximal lotteries as the *bipartisan set* (a term proposed by Roger Myerson).

Maximal lotteries are known to satisfy a number of desirable properties that are typically considered in social choice theory (see, e.g., Felsenthal and Machover, 1992; Laslier, 2000a; Rivest and Shen, 2010; Hoang, 2017; Brandl et al., 2022). For example, Condorcet winners (i.e., alternatives that defeat every other alternative in a pairwise majority comparison) will be selected with probability 1, and Condorcet losers (i.e., alternatives that are defeated in all pairwise majority comparisons) will never be se-

[3]This comparison of lotteries induces a binary relation on lotteries whose maximal elements are precisely the maximal lotteries.

lected. No group of voters benefits by abstaining from an election, removing losing alternatives does not affect maximal lotteries, and each alternative's probability is unaffected by cloning other alternatives. Maximal lotteries have been axiomatically characterized using Arrow's independence of irrelevant alternatives and Pareto efficiency (Brandl and Brandt, 2020) as well as population-consistency and composition-consistency (Brandl et al., 2016). The dynamic procedure described above implements maximal lotteries while providing

- *myopic strategyproofness* within each round,
- minimal *preference elicitation* and thus increased privacy protection,
- *verifiability* realized via a simple physical procedure, and
- all-round *flexibility*.

Myopic strategyproofness: Each round's decision is made by letting a randomly selected voter choose between two alternatives. Clearly, a voter who is only concerned with the outcome of the current round is best off by choosing the alternative that she truly prefers. If she also takes into account the outcomes of future rounds, however, she may be able to skew the distribution in the urn by choosing alternatives strategically.[4]

Preference elicitation: Eliciting pairwise preferences on an as-needed basis has several advantages. First, it spares the voters from the cognitive burden of having to rank-order all alternatives at once. If the number of voters is large, it may well be possible that the urn process yields satisfying results without ever querying some of the voters. Secondly, rather than submitting a complete ranking of all alternatives to a trusted authority, voters only reveal their preferences by making pairwise choices from time to time.[5]

[4]Maximal lotteries, like any *ex post* Pareto efficient randomized choice procedure other than random dictatorships, fail to be strategyproof (Gibbard, 1977). The simple notion of myopic strategyproofness could be strengthened by discounting future rounds.

[5]Privacy can be further increased by letting voters draw their balls privately, announce the winner, and put two balls of the same color back into the urn, without revealing the original color of the losing ball. Alternatively, the voters' preferences can be protected completely by letting the voter publicly draw both balls, make one copy of each ball, and let her privately put back two balls of her choice. The collective decision in each round can then be made by drawing a random ball from the urn.

Verifiability: Previously, the deployment of maximal lotteries required that a central authority collects the preferences of all voters, computes a maximal lottery by solving a linear program, and instantiates the lottery in some user-verifiable way. The urn process allows to achieve these goals via a simple physical device.

Flexibility: The urn process is oblivious to changes in the voters' preferences, the set of voters as well as the set of alternatives. Everything that has happened up to the current round is irrelevant. Since the process converges from any initial configuration, it will keep "walking in the right direction" (towards a maximal lottery of the current preference profile). If the preferences change slowly in the sense that only a small fraction of voters changes their preferences from one round to the next, collective choices will thus be made according to a maximal lottery for the current preferences in most rounds. This includes the case when the distribution of preferences converges.

We also note some disadvantages of the urn process. The convergence of the distribution of winners to an approximate maximal lottery is an asymptotic result. In particular, for a finite number of rounds, there is a non-zero probability that the chosen alternative is subpar for a significant fraction of rounds, for example, because it is Pareto dominated. To bound this probability below an acceptable threshold, it may be necessary to run the process for an excessively large number of rounds. Second, ensuring that the limit distribution is sufficiently close to a maximal lottery could require an urn with a large number of balls. We address the first concern by showing that the probability for the distribution of winners to be far from the limit distribution converges to 0 exponentially fast in the number of rounds. The rate of convergence is also evident in computational simulations we ran for various parameterizations of the process. When the preference profile admits a Condorcet winner, we can give tractable bounds on the number of balls in the urn required to achieve a good approximation in the limit. This partially mitigates the second concern since it has been observed that most real-world preference profiles admit Condorcet winners (see, e.g., Gehrlein and Lepelley, 2011).

The axiomatic characterizations of maximal lotteries not only imply that maximal lotteries satisfy desirable axioms, but also that any deviation from maximal lotteries leads to a violation of at least one of the axioms. Hence, a process that only guarantees an approximation of a maximal lottery will not enjoy the same axiomatic properties. However, rather than insisting on stringent axioms, one can relax them by only requiring them to hold in an approximate sense. For example, a natural notion of approximate Condorcet-consistency would require that a Condorcet winner receives probability close to 1 whenever one exists. Since the empirical distribution of winners according to our process is almost surely close to a maximal lottery and maximal lotteries are Condorcet-consistent, the process is approximately Condorcet-consistent in the above sense. More generally, approximate maximal lotteries satisfy approximate versions of many of the axioms enjoyed by maximal lotteries such as population-consistency, composition-consistency, agenda-consistency, and efficiency.[6]

Maximal lotteries have been repeatedly recommended for practical use (Felsenthal and Machover, 1992; Rivest and Shen, 2010; Brandl et al., 2016; Hoang, 2017). We believe that the benefits of the urn process described above extend the applicability of maximal lotteries. Rather than for traditional political elections, probabilistic rules like maximal lotteries seem more suitable for frequently repeated low-stakes elections where some degree of randomization may not only be tolerable but even desirable. Two example applications that have been suggested for maximal lotteries are to help a group of coworkers with the daily decision where to have lunch and to select music for a party or a radio station based on the preferences of the listeners (Brandl et al., 2016). The transparency and the flexibility of the urn process seem particularly effective in the music broadcasting example. Agents come and go, they only need to select from a pair of songs rather than rank-order all of them, and individual preferences, as well as the set of available songs, can be changed at any time. Our theorem shows that the sequence of simple pairwise choices results in a socially desirable distribution of songs: the more songs are being played, the less likely it becomes that another distribution of songs would have been preferred by an expected majority of listeners. It is plausible that, over time, the preferences of the listeners change depending on the songs that have been played so far. These changes will be reflected immediately in the selection of future songs.

[6]Details on how these statements can be formalized are given in an extended version of this paper (Brandl and Brandt, 2021).

1.2. Applications Beyond Collective Decision-Making

Interestingly, dynamic processes similar to the process we describe here have recently been studied in population biology, quantum physics, chemical kinetics, and plasma physics to model phenomena such as the coexistence of species, the condensation of bosons, the reactions of molecules, and the scattering of plasmons. In each of these cases, simple interactions between randomly sampled entities result in distributions that correspond to equilibrium strategies of symmetric zero-sum games. Since the definition of maximal lotteries and our dynamic process merely rely on this comparison matrix, describing with which probability one entity will be replaced with another in a pairwise encounter, our results are also of relevance to these areas. We discuss these connections, as well as those to equilibrium learning and evolutionary game theory, in more detail in Section 5.

An alternative interpretation of our result can be used to describe the formation of opinions. In this model, there is a population of agents, each of which entertains one of many possible opinions. Agents come together in random pairwise interactions, in which they try to convince each other of their opinion. The probabilities with which one opinion beats another are given as a square matrix and, with some small probability, an agent randomly changes her opinion. In other words, the agents correspond to the balls in the urn, the opinions correspond to the alternatives, and there are neither voters nor preference profiles as transition probabilities are given explicitly. Our theorem then shows that, if the population is large enough, the distribution of opinions within the population is close to a maximal lottery of the probability matrix most of the time. Other models of opinion formation based on different processes were for example considered by DeGroot (1974), Holley and Ligget (1975), and Goel and Lee (2014).

The process we describe approximately computes a mixed Nash equilibrium of a symmetric zero-sum game. This problem is known to be equivalent to linear programming. In fact, deciding whether an action is played with positive probability in an equilibrium of a symmetric zero-sum game is P-complete (Brandt and Fischer, 2008, Theorem 5), which, loosely speaking, means that the problem is at least as hard as any problem that can be solved in polynomial time. The urn process can thus be seen as a probabilistic algorithm that approximates polynomial-time computable functions. In contrast to traditional computing devices such as Turing machines, the urn process is based on unordered elementary entities that randomly interact according to very simple replacement rules.[7]

The remainder of the paper is structured as follows. After defining our model in Section 2, we state the main result (Theorem 1) and a rough proof sketch in Section 3. The full proof is given in the Appendix. In Section 4, we analyze the instructive special case of preference profiles that admit a Condorcet winner, which allows for a more elementary proof. In Section 5, we extensively discuss the connections between our work and results from equilibrium learning, evolutionary game theory, and the natural sciences. We also state a continuous version of our main result (Theorem 2) that may be of independent interest.

2. The Model

Let $[d] = \{1, \dots, d\}$ be a set of alternatives and Δ the $d-1$-dimensional unit simplex in $\mathbb{R}^d$, that is, $\Delta = \{x \in \mathbb{R}^d_{\geq 0} \colon \sum_{i=1}^d x_i = 1\}$. We refer to elements of Δ as lotteries. By $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ we denote the sets of positive and non-negative integers, respectively. Throughout the paper, for a vector $x \in \mathbb{R}^k$ for some k, $|x| = \sum_{l=1}^k |x_l|$ denotes its L^1-norm. For $\delta > 0$ and $S \subset \mathbb{R}^d$, let $B_\delta(S) = \{x \in \Delta \colon |x - y| < \delta \text{ for some } y \in S\}$ be the δ-ball around S. For a finite set S, we write $|S|$ for the number of elements of S.

A *preference relation* $\succ$ is an asymmetric binary relation over $[d]$.[8] By $\mathcal{R}$ we denote

[7]Related decentralized models of computation with applications to sensor networks and molecular computing are studied under the name "population protocols" in computer science (e.g., Angluin et al., 2006; Aspnes and Ruppert, 2009). While the urn process has the same *modus operandi* as population protocols, the input-output behavior is different. The input of population protocols is given by the initial distribution of balls in the urn and the output has been reached if all balls belong to a certain subset of types. By contrast, the input for our urn process is encoded in the matrix describing the replacement rules and the (approximate) output is given by the distribution of balls in the urn after sufficiently many rounds.

[8]Preferences need not be transitive or complete. The definition of maximal lotteries and the urn process we describe only depend on the fractions of voters who prefer one alternative to another.

the set of all preference relations. Let V be a finite set of voters. A *preference profile* $R \in \mathcal{R}^V$ specifies a preference relation for each voter. With each preference profile R, we can associate a comparison matrix $M_R \in [0,1]^{d \times d}$ that states for each ordered pair of alternatives the fraction of voters who prefer the first to the second. That is, $M_R(i,j) = |\{v \in V : i \succ_v j\}|/|V|$. This matrix induces a skew-symmetric matrix $\tilde{M}_R = M_R - M_R^\intercal$, which we call the skew-comparison matrix.[9]

2.1. Maximal Lotteries

A lottery $p \in \Delta$ is a *maximal lottery* for a profile R if $\tilde{M}_R p \leq 0$. The minimax theorem implies that every profile admits at least one maximal lottery. By $ML(R)$ we denote the set of all lotteries that are maximal for R. Most profiles admit a unique maximal lottery. For example, when the number of voters is odd and voters have strict preferences, there is always a unique maximal lottery (Laffond et al., 1997).

Example 1 (Condorcet winner)**.** Consider, for example, 900 voters, three alternatives, and a preference profile R given by the following table. Each column header contains the number of voters with the corresponding preference ranking.

300	300	300
1	1	2
2	3	3
3	2	1

Then,

$$M_R = \begin{pmatrix} 0 & ^2\!/_3 & ^2\!/_3 \\ ^1\!/_3 & 0 & ^2\!/_3 \\ ^1\!/_3 & ^1\!/_3 & 0 \end{pmatrix} \quad \text{and} \quad \tilde{M}_R = \begin{pmatrix} 0 & ^1\!/_3 & ^1\!/_3 \\ -^1\!/_3 & 0 & ^1\!/_3 \\ -^1\!/_3 & -^1\!/_3 & 0 \end{pmatrix}.$$

The set of maximal lotteries $ML(R) = \{(1,0,0)^\intercal\}$ only contains the degenerate lottery with probability 1 on the first alternative. This alternative is a *Condorcet winner*,

In particular, indifferences can easily be accommodated by randomly selecting which of the two balls will be relabelled.

[9] A matrix M is skew-symmetric if $M = -M^\intercal$.

i.e., an alternative that is preferred to every other alternative by some majority of voters.

2.2. Markov Chains

Let S be a finite set and $\{X(n) \colon n \in \mathbb{N}_0\}$ be a discrete-time, time-homogeneous *Markov chain* with state space S. The *transition probability matrix* $P \in [0,1]^{S \times S}$ of $\{X(n) \colon n \in \mathbb{N}_0\}$ is given by

$$P(p, p') = \mathbb{P}\left(X(n+1) = p' \mid X(n) = p\right)$$

for all $p, p' \in S$. We will frequently write $X(n, p_0)$ for $X(n)$ conditioned on $X(0) = p_0 \in S$ and call p_0 the *initial state*.

The period of a state $p \in S$ is the greatest common divisor of the return times with positive probability $\{n \in \mathbb{N} \colon (P^n)(p,p) > 0\}$. A Markov chain is *aperiodic* if every state has period 1. Note that any Markov chain with $P(p,p) > 0$ for all $p \in S$ is aperiodic. A Markov chain is *irreducible* if every state is reached from any other state with positive probability. That is, for any two states $p, p' \in S$, there is a positive integer n so that $(P^n)(p, p') > 0$. If $\{X(n) \colon n \in \mathbb{N}_0\}$ is irreducible and aperiodic, it has a unique *stationary distribution* $\pi \in \Delta S$ so that $\pi^\intercal = \pi^\intercal P$.

2.3. The Urn Process

Consider an urn with $N \in \mathbb{N}$ balls, each labeled with some alternative. Viewing balls with the same label as indistinguishable, we can identify each state of the urn with an element of the discrete unit simplex $\Delta^{(N)} = \{p \in \Delta \colon Np \in \mathbb{N}_0^d\}$. Fix a *mutation rate* $r \in [0,1]$.

We are interested in a Markov chain with state space $\Delta^{(N)}$ that can be informally described as follows. First, we flip a coin that has probability $1 - r$ of landing heads. If the coin shows heads, we choose one voter $v \in V$ uniformly at random and ask the voter to draw two balls from the urn. Say these two balls are labeled with alternatives 1 and 2. If $1 \succ_v 2$, the label of the second ball is changed to 1. Likewise, if $2 \succ_v 1$, the first ball is relabeled with label 2. If both balls carry the same label, the labels remain unchanged. If the coin shows tails, we draw a single ball from the urn, relabel it with an alternative chosen uniformly at random, and put it back into the urn.

This description of the process assumes that two alternatives are sampled from the urn distribution *without* replacement. For the formal description, we will assume that drawing is *with* replacement. This corresponds to sampling one alternative by drawing one ball, putting the ball back into the urn, and sampling a second alternative by again drawing one ball (which may be the same as the first).[10] Doing so avoids a lot of clumsy notation. If the number of balls in the urn is large, there is no significant difference between drawing with and without replacement. In the proof, we point out why the same arguments also carry through with minor adaptions for drawing without replacement.

We define a transition probability matrix $P^{(N,r)}$ that specifies for every pair of states the probability that the distribution of the urn transitions from the first to the second. Denote by e_i the ith unit vector in $\mathbb{N}_0^d$. For $p \in \Delta^{(N)}$ and $i, j \in [d]$ with $p' = p + \frac{e_i}{N} - \frac{e_j}{N} \in \Delta^{(N)}$, let

$$P^{(N,r)}(p,p') = \begin{cases} (1-r)2p_i p_j M_R(i,j) + \dfrac{r}{d} p_j & \text{if } i \neq j \\ (1-r)\displaystyle\sum_{k=1}^{d} p_k^2 + \dfrac{r}{d} & \text{if } i = j \end{cases}$$

be the probability of transitioning from p to p'. For the remaining pairs of states $p, p' \in \Delta^{(N)}$, let $P^{(N,r)}(p,p') = 0$. Then, $P^{(N,r)}$ has non-negative values and its rows sum to 1 so that it is a valid transition probability matrix. For an initial state $p_0 \in \Delta^{(N)}$, we consider a Markov chain $\{X^{(N,r)}(n,p_0) \colon n \in \mathbb{N}_0\}$ with transition probability matrix $P^{(N,r)}$. The distribution of $X^{(N,r)}(n,p_0)$ over $\Delta^{(N)}$ is given by the row of $\left(P^{(N,r)}\right)^n$ with index p_0. If $r > 0$, this Markov chain is irreducible and aperiodic (since it remains in the same state with positive probability). It corresponds to the urn process described above when the initial state of the urn is p_0.

Continuing Example 1, consider an urn with $N = 5$ balls and recall that $d = 3$.

[10]In practice, it would the be infeasible to relabel the first drawn ball since it is "lost" in the urn after putting it back. However, one could modify the process by relabeling the second ball if the voter prefers the first sampled alternative and change nothing otherwise. This eliminates the factor of 2 in the transition probabilities below but does not change the results.

Then, the transition probability matrix $P^{(N,r)}$ is an $\binom{3+5-1}{5} = 21$-dimensional square matrix. Let the mutation rate be $r = 0.1$ and the initial state $p_0 = \frac{1}{5}(1,2,2)^\intercal$. The probability that one of the balls of the second type is replaced with one of the first type is

$$P^{(5,0.1)}(p_0, \frac{1}{5}(2,1,2)^\intercal) = 0.9 \cdot \frac{4}{25} \cdot \frac{2}{3} + 0.1 \cdot \frac{1}{3} \cdot \frac{2}{5} \sim 0.109.$$

3. The Result

We prove the following:

> *For any small enough mutation rate $r > 0$, there is a maximal lottery p^* so that for any initial state p_0, $X^{(N,r)}(k, p_0)$ is close to p^* for all but a small fraction of rounds k provided that the number of balls N is large enough.*

More precisely, for any $\delta, \tau > 0$, there is an upper bound on the mutation rate $r_0 > 0$ so that for every $0 < r \leq r_0$, there is a maximal lottery p^* and a lower bound on the number of balls $N_0 \in \mathbb{N}$ such that for every $N \geq N_0$ and every $p_0 \in \Delta^{(N)}$, the fraction of rounds k in which $X^{(N,r)}(k, p_0)$ is no more than δ away from p^* is almost surely at least $1 - \tau$.[11]

Theorem 1. *Let $\delta, \tau > 0$. Then, there is $r_0 > 0$ such that for all $0 < r \leq r_0$, there are $p^* \in ML(R)$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and $p_0 \in \Delta^{(N)}$, almost surely*

$$\lim_{n\to\infty} \frac{1}{n} \left|\left\{k \leq n \colon \left|X^{(N,r)}(k, p_0) - p^*\right| \leq \delta\right\}\right| \geq 1 - \tau.$$

Moreover, there is $C > 0$ such that for all $n \in \mathbb{N}_0$,

$$\mathbb{P}\left(\left|X^{(N,r)}(n, p_0) - p^*\right| \leq \delta\right) \geq 1 - \tau - e^{-\lfloor Cn \rfloor}.$$

[11] Theorem 1 implies that the stationary distribution of $X^{(N,r)}$ assigns probability at least $1 - \tau$ to states that are in a δ-neighborhood of p^*. Conversely, this property of the stationary distribution implies Theorem 1 by the ergodic theorem for Markov chains. The proof does however not derive the above property of the stationary distribution as an intermediate step. It is only a by-product of the final result. For more than two alternatives, our result is stronger than proving that the expectation of the stationary distribution, or, equivalently, the temporal average of the urn distribution, is close to a maximal lottery.

To prove this, we approximate our discrete and stochastic urn process by a continuous and deterministic process. The latter can be viewed as a version of the urn process with a continuum of balls. Using analytical tools, it can be shown that this process converges to an approximate maximal lottery for every initial state, where the approximation can be made arbitrarily precise if r is made small (see Theorem 2 in Section 5). The approximation only works for a finite number of rounds (respectively, bounded time interval) and only with probability close to 1 (rather than almost surely). However, on long enough time intervals, the deterministic process is close to an approximate maximal lottery most of the time (since it converges to such a lottery). Moreover, on any such interval, the deterministic process is a good approximation to the stochastic process with probability close to 1 (provided that the number of balls is large enough). By a variant of the strong law of large numbers, it then follows that the stochastic process is close to an approximate maximal lottery for most rounds almost surely. This is the first statement of Theorem 1. We give a more detailed outline and a complete proof in the Appendix. The second statement follows from the first using the standard result that the distribution of an irreducible and aperiodic Markov chain converges exponentially fast to its stationary distribution in the total variation norm.

Theorem 1 is a statement about the distribution in the urn. Recall that the collective decision in each round is the winner of the pairwise comparison between the two drawn balls. It is not hard to show that the empirical distribution of winners is also close to a maximal lottery.[12]

Another straightforward corollary of Theorem 1 is that the temporal average of the urn distribution is almost surely close to a maximal lottery (provided that r is small

[12]Suppose the distribution of balls in the urn is $p \in \Delta^{(N)}$. Then the probability that $i \in [d]$ is the collective decision is

$$w_i = p_i\Big(p_i + 2\sum_{j\neq i} M_R(i,j)p_j\Big) = p_i\Big(p_i + \sum_{j\neq i}(\tilde{M}_R(i,j)+1)p_j\Big) = p_i\Big(1+\tilde{M}_R p\Big)$$

where we used that $2M_R(i,j) = \tilde{M}_R(i,j)+1$, $\sum_{j\in[d]} p_j = 1$, and $M_R(i,i) = 0$. If p^* is a maximal lottery and $|p-p^*| \leq \delta$, then $(\tilde{M}_R p)_i \leq \delta$ for all $i \in [d]$. Hence, $w_i \in [p_i - \delta, p_i + \delta]$ for all i, so that $|w - p^*| \leq (d+1)\delta$. For every $\delta' > 0$, choosing $\delta = \tau = \frac{\delta'}{2(d+1)}$ in Theorem 1 thus shows that the empirical distribution of collective decisions is almost surely no more than δ' away from p^*.

and N is large). Let

$$Z^{(N,r)}(n, p_0) = \frac{1}{n} \cdot \sum_{k=0}^{n-1} X^{(N,r)}(k, p_0)$$

be the temporal average of $X^{(N,r)}(k, p_0)$ over the first n rounds. Then, we have the following.

Corollary 1. *Let $\delta > 0$. Then, there is $r_0 > 0$ such that for all $0 < r \leq r_0$, there are $p^* \in ML(R)$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and $p_0 \in \Delta^{(N)}$,*

$$\mathbb{P}\left(\left|\lim_{n\to\infty} Z^{(N,r)}(n, p_0) - p^*\right| \leq \delta\right) = 1$$

Proof. For some τ to be determined later, let r_0 and, depending on $0 < r \leq r_0$, p_* and N_0 be as obtained from Theorem 1. By the triangle inequality, we have

$$\left|Z^{(N,r)}(n, p_0) - p^*\right| \leq \frac{1}{n} \cdot \sum_{k=0}^{n-1} \left|X^{(N,r)}(k, p_0) - p^*\right| .$$

By Theorem 1, in the limit when n goes to infinity, all but a $1 - \tau$ fraction of the summands on the right-hand side are smaller than δ. The remaining summands are bounded by 2. Hence, choosing $\tau = \frac{\delta}{2}$ gives that almost surely, $|\lim_{n\to\infty} Z^{(N,r)}(n, p_0) - p^*| \leq 2\delta$. The ergodic theorem for Markov chains ensures that the limit exists. $\square$

Before illustrating these results via examples, we discuss variations of the urn process.

Remark 1 (Decoupling collective decisions)**.** We assume that in each round, a collective decision is made by selecting the winner of the pairwise comparison. It may however be more practical to decouple collective decisions from the preference elicitation process and draw winners less frequently. For example, collective decisions could be made by drawing a random ball after any fixed number of rounds or at random times. Corollary 1 shows that the resulting distribution would approximate a maximal lottery.

Remark 2 (Non-uniform mutation rates)**.** The results still hold if we let the probability of a random mutation from alternative i to alternative j depend on the pair (i, j). It suffices that every alternative in the support of a maximal lottery can escape permanent depletion via some path of mutations. More explicitly, it suffices if for any two alternatives i and j, there is a path of alternatives from i to j so that the mutation rate is positive from any alternative on the path to the next. The proof can be adapted at the expense of more book-keeping.

Remark 3 (Mutation rate vs. urn size)**.** Corollary 1 shows that the temporal average of the urn distribution converges to a maximal lottery if we let N go to infinity and then take r to 0 (see also Theorem 2). This is in contrast to other works on evolutionary dynamics that take limits in the reverse order.[13] While the frameworks are similar, these results are conceptually different. In our model, if r is too small compared to $\frac{1}{N}$, it will in general not be the case that the distribution in the urn is close to a maximal lottery for most rounds. For any long enough time interval, the distribution in the urn will for all r degenerate within the interval with high probability, that is, it will only contain balls of one type. If r is very small, it will stay in a degenerate state for a long time (compared to the chosen interval) with high probability. When the process leaves the degenerate state, the same will repeat itself (possibly with a different degenerate state), so that the process spends most rounds in degenerate states. As a consequence, decreasing r over time does not work unless N is increased as well. When increasing N at an appropriate rate, the urn distribution will converge exactly to the set of maximal lotteries by Theorem 1.

Remark 4 (Majority voting)**.** Rather than letting only a single voter decide on the pairwise comparison between the two randomly drawn balls, it is possible to ask all voters which alternative they prefer and replace the alternative which is less preferred by a *majority* of voters. This variant is equivalent to the original process for a single voter with possibly intransitive preferences (given by the majority relation of

[13]For example, Fudenberg and Imhof (2008) study imitation dynamics with mutations in symmetric two-player games (not necessarily zero-sum). They consider the case when the mutation rate goes to 0 for a fixed population size N. For small but positive mutation rates, the dynamics spend most of the time in degenerate states where all but a small fraction of individuals play the same strategy. Letting the mutation rate go to 0 thus induces a distribution over actions. Their main result determines the limit of this distribution as the population size goes to infinity.

the entire population of voters) and converges to a so-called C1 maximal lottery of the preference profile (see Brandl et al., 2022, for more information on C1 maximal lotteries).

Remark 5 (Static or growing urn)**.** When the initial distribution of balls in the urn is uniform and remains fixed (i.e., no balls are replaced over time), then the empirical distribution of winners converges to the lottery returned by the *proportional Borda rule* (see, e.g., Barberà, 1979; Heckelman, 2003).[14] This rule violates Condorcet-consistency and Pareto efficiency. It can put probability $\frac{1}{d}$ on Pareto-dominated alternatives and almost as little as $\frac{1}{d}$ on Condorcet winners for large numbers of voters (Brandt et al., 2022). When adding a new ball labeled with the winning alternative rather than replacing the losing one (i.e., the number of balls increases over time), neither the relative distribution in the urn nor the temporal average converges (see Section 5).

Figure 2 (left) shows a simulation of the urn process for the preference profile and corresponding skew-comparison matrix given in Example 1. The urn process corresponds to a random walk within the shown triangle starting from the center (an almost uniform distribution). The first alternative in this profile is a Condorcet winner. From round 177 on, at least 90% of the balls (45 of the 50) are labeled with the Condorcet winner except for three rounds. At this point, only 160 of the 900 voters were asked for their preferences. The path is tilted to the left because a majority of voters prefer alternative 2 to alternative 3. Note that the process only depends on the fractions of voters who prefer one alternative to another and is thus independent of the number of voters. Hence, if there are nine million—rather than nine hundred—voters whose preferences are distributed as in Example 1, the process could turn out exactly as shown in Figure 2. In particular, the overwhelming majority of voters would never be queried for their preferences.

We now give another example, for which the unique maximal lottery is not degenerate.

[14]The proportional Borda rule assigns to each alternative a probability that is proportional to its Borda score. For example, for one voter with lexicographic preferences over a, b, and c, the Borda scores are 2, 1, and 0, respectively. The proportional Borda rule thus returns the lottery $(2/3, 1/3, 0)$.

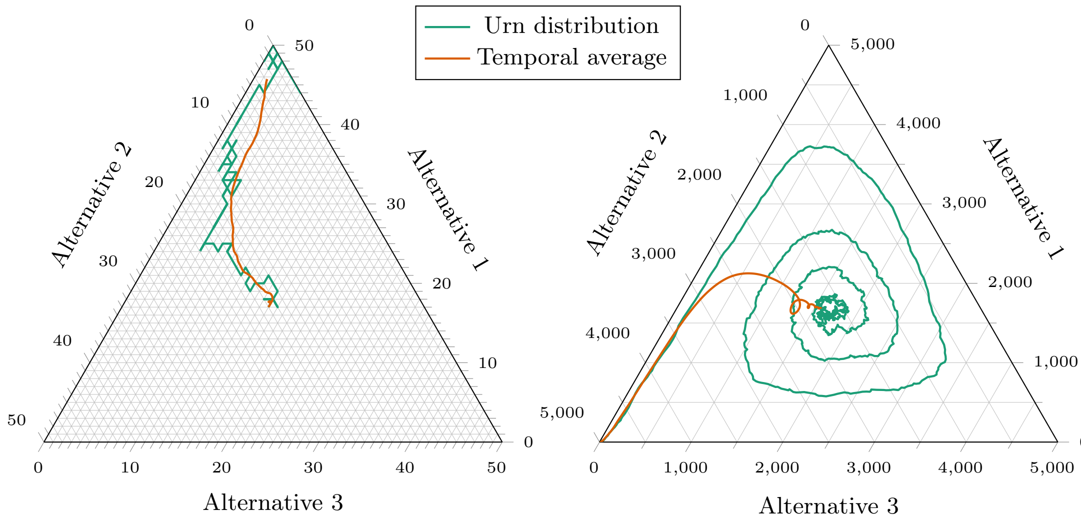


Figure 2: Simulations of the urn process.
The left diagram shows the urn process for the profile given in Example 1 using an urn with $N = 50$ balls for 1,000 rounds and mutation rate $r = 0.02$, starting from an almost uniform distribution. Each intersection of the grid lines corresponds to a configuration of the urn. The right diagram shows the urn process for the profile given in Example 2 using an urn with $N = 5,000$ balls for 500,000 rounds and mutation rate $r = 0.04$, starting from the degenerate distribution in which all balls are labeled with Alternative 2. The green lines depict the actual distribution of balls while the red lines depict the temporal average of urn distributions until the given round.

Example 2 (Condorcet cycle)**.** Consider 900 voters, three alternatives, and the following preference profile R, leading to a so-called Condorcet cycle or Condorcet paradox.

300	300	300
1	2	3
2	3	1
3	1	2

Then,

$$M_R = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 2/3 & 1/3 & 0 \end{pmatrix} \quad \text{and} \quad \tilde{M}_R = \begin{pmatrix} 0 & 1/3 & -1/3 \\ -1/3 & 0 & 1/3 \\ 1/3 & -1/3 & 0 \end{pmatrix}.$$

The set of maximal lotteries $ML(R) = \{(1/3, 1/3, 1/3)\}$ consists of the uniform lottery over the three alternatives. A simulation of the urn process for this profile is given in Figure 2 (right). This time, the initial distribution is degenerate with all balls being of type 2. It can be seen how the distribution of balls in the urn closes in on the maximal lottery and remains in its neighborhood for most of the time while the temporal average converges to the maximal lottery.

4. The Case of a Condorcet Winner

We give an elementary proof of Theorem 1 for profiles that admit a Condorcet winner. For those profiles, the unique maximal lottery assigns probability 1 to the Condorcet winner. To analyze the stationary distribution $\pi \in \Delta(\Delta^{(N)})$ of the Markov chain induced by the urn process, it suffices to examine the fraction of balls labeled with the Condorcet winner. This allows us to relate the Markov chain to a process that is one-dimensional in the sense that each state can only transition to two different states, and is thus easy to analyze. It also enables us to give a concrete lower bound on the number of balls N required for given $\delta, \tau > 0$ for the conclusion of Theorem 1 to hold.

Let $M = M_R$ be the majority matrix of a profile R with Condorcet winner $i \in [d]$. Hence, $M_{ij} > \frac{1}{2}$ for all $j \in [d] \setminus \{i\}$. Let $\alpha = \min\{M_{ij} \colon j \in [d] \setminus \{i\}\} - \frac{1}{2}$. We slice up $\Delta^{(N)}$ into the level sets of the map $p \mapsto p_i$. For $k \in \{0, \dots, N\}$, let $S_k = \{p \in \Delta^{(N)} \colon p_i = \frac{k}{N}\}$ be the states corresponding to distributions with k of the N balls of type i. Then $\sigma_k := \sum_{p \in S_k} \pi(p)$ is the limit probability that the urn is in a state in S_k as the number of rounds goes to infinity. We want to show that if r is sufficiently small and N sufficiently large, π has most of the probability on states in S_k with k close to N.

For 4 alternatives, one can illustrate the ensuing argument as follows. The set of states $\Delta^{(N)}$ corresponds to rooms in a tetrahedral-shaped pyramid. The rooms on the kth floor correspond to S_k, so that the tip of the pyramid is the state where all balls are of type i. The urn process is a random walk through the pyramid, moving from one room to an adjacent one (which could be on the same floor, the floor below, or the floor above). With the exception of few floors close to the tip, the probability of going up is always larger than the probability of going down. It is then intuitively clear that if the pyramid is large enough, one should expect to find the random walk close to the tip of the pyramid most of the time.[15]

Recall that $P^{(N,r)}(p,q)$ is the probability of transitioning from state p to state q. Since π is a stationary distribution, we have $\pi^\intercal P^{(N,r)} = \pi^\intercal$. Consider any partition of $\Delta^{(N)}$ into two sets. For the stationary distribution, the probability of transitioning from the first set to the second is equal to the probability of transitioning from the second set to the first since the probabilities of both sets are conserved. Applying this to the sets $\bigcup_{l=0}^{k-1} S_l$ and $\bigcup_{l=k}^{N} S_l$ for $k \in [N]$ and noticing that the only transitions between the two sets with positive probability are from S^{k-1} to S^k and vice versa, we get

$$\sum_{p\in S_{k-1}} \pi(p) \sum_{q\in S_k} P^{(N,r)}(p,q) = \sum_{p\in S_k} \pi(p) \sum_{q\in S_{k-1}} P^{(N,r)}(p,q). \tag{1}$$

That is, the probability of being in a state in S_{k-1} and transitioning to a state in S_k equals the probability of being in a state in S_k and transitioning to a state in S_{k-1}.

Now observe that for $p \in S_k$, $k \in \{0,\dots,N-1\}$, we have

$$\sum_{q\in S_{k+1}} P^{(N,r)}(p,q) \geq 2(1-r)\frac{k(N-k)}{N^2}\left(\frac{1}{2}+\alpha\right) + \frac{r}{d}\frac{N-k}{N} =: u_k$$

where the left hand side is the probability of replacing a ball of type other than i by one of type i in state $p \in S_k$ (moving up one floor in the pyramid). Similarly, we find that for $p \in S_k$, $k \in [N]$, we have

$$\sum_{q\in S_{k-1}} P^{(N,r)}(p,q) \leq 2(1-r)\frac{k(N-k)}{N^2}\left(\frac{1}{2}-\alpha\right) + r\frac{d-1}{d}\frac{k}{N} =: d_k$$

for the probability of replacing a ball of type i by one of type other than i in state $p \in S_k$ (moving down one floor in the pyramid). Plugging this into (1), we get

$$\sigma_{k-1}u_{k-1} \leq \sigma_k d_k. \tag{2}$$

[15] In the analysis of the general case, the number of balls of type i is replaced by the entropy of the urn distribution relative to a maximal lottery. The fact that the number of balls not of type i more likely than not decreases corresponds to the fact that the expected entropy relative to a maximal lottery decreases.

All terms in (2) are strictly positive if $r > 0$.

Let N be so that $\frac{r}{Nd} \geq 2\frac{1-r}{N^2}$ (we choose $r > 0$ later). Then,

$$\begin{aligned} u_k &\geq 2(1-r)\frac{k(N-k)}{N^2}\left(\frac{1}{2}+\alpha\right) + 2(1-r)\frac{N-k}{N^2} \\ &\geq 2(1-r)\frac{(k+1)(N-k-1)}{N^2}\left(\frac{1}{2}+\alpha\right) \end{aligned}$$

where the last inequality uses $1 \geq \frac{1}{2} + \alpha$. Similarly, we find that for $r \leq \frac{1}{d}$ and $k \leq N\left(1 - \frac{r}{\alpha}\right)$,

$$d_k \leq 2(1-r)\frac{k(N-k)}{N^2}\frac{1-\alpha}{2}.$$

Hence, with this bound on k, we have

$$\frac{d_k}{u_{k-1}} \leq \frac{1-\alpha}{2\left(\frac{1}{2}+\alpha\right)} = \frac{1-\alpha}{1+2\alpha} =: \beta.$$

Thus, by (2), $\frac{\sigma_{k-1}}{\sigma_k} \leq \beta < 1$. We have shown that the cumulative probability σ_k of the states S_k decreases at least as fast as the terms of the geometric series with parameter β from some k (close to N) downwards.

The maximal lottery for R is the degenerate lottery with probability 1 on i. For given $\delta, \tau > 0$, we are aiming for a lower bound on N so that the probability on states with at least a $1-\delta$ fraction of balls of type i in the stationary distribution π is at least $1-\tau$. That is,

$$\sum_{k=\lceil N(1-\delta)\rceil}^{N} \sigma_k \geq 1-\tau.$$

First observe that

$$\sum_{k \geq k_0} \beta^k = \beta^{k_0}\frac{1}{1-\beta} \leq \tau \tag{3}$$

for $k_0 \geq \frac{\log(\tau(1-\beta))}{\log \beta}$. For our bound, N needs to be large enough so that there are at least k_0 integers in the interval $\{\lceil (1-\delta) N \rceil, \ldots, \lfloor (1-\frac{r}{\alpha}) N \rfloor\}$. The probability on states in S_k with $k < (1-\delta) N$ will then be below τ by (3) and the choice of k_0 (since the bound on d_k assumes that $k \leq N(1-\frac{r}{\alpha})$). Choosing $r \leq \frac{\alpha\delta}{2}$ and

$$N \geq \frac{k_0}{\delta - \frac{r}{\alpha}} \geq \frac{1}{\delta} \left\lceil \frac{\log(\tau(1-\beta))}{\log \beta} \right\rceil$$

achieves this.

In Example 1, there are three alternatives and 900 voters. Alternative 1 is a Condorcet winner as it is preferred to every other alternative by 600 of the voters ($\alpha = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$, $\beta = \frac{5}{8}$). Suppose we want that at least 90% of the balls in the urn are of type 1 in at least 90% of rounds ($\delta = 0.2$, $\tau = 0.1$). Choosing $r = \frac{\alpha\delta}{2} = \frac{1}{60}$, we need $N \geq 70$ balls in the urn. These calculations suggest that, when a Condorcet winner exists, a reasonable choice of the parameters is $N \geq -\frac{1}{\delta}\log(\tau)$ and $\frac{1}{N} \leq r \leq \delta$.

5. Discussion

Since the urn process described in this paper only depends on the comparison matrix M_R and the mutation rate r, it is connected to various problems unrelated to collective decision-making. In particular, the literature on equilibrium learning and evolutionary game theory has extensively studied dynamics based on payoff matrices and their convergence behavior.

5.1. Equilibrium learning

When interpreting $\tilde{M}_R$ as a symmetric two-player zero-sum game and maximal lotteries as equilibrium strategies, our result can be phrased as a result about a learning procedure for equilibrium play. Such procedures have been extensively studied in game theory and, in particular for zero-sum games, a number of simple and attractive procedures have been proposed. The earliest of these is *fictitious play* (Brown, 1951; Robinson, 1951) and its variant *stochastic fictitious play* (Fudenberg and Kreps, 1993).[16] More recently, the *multiplicative weights update algorithm* (e.g., Freund and

[16] Hofbauer and Sandholm (2002) show that under stochastic fictitious play, players' strategies and beliefs converge to a Nash equilibrium in several classes of games, including two-player zero-sum

Schapire, 1999; Arora et al., 2012) and *regret matching* (Hart and Mas-Colell, 2000, 2013) have been celebrated in game theory, optimization, and machine learning. When translating the multiplicative weights update algorithm to our setting, one obtains a dynamic urn process, in which voters need to compare a drawn ball to all possible alternatives and adjust the distribution in the urn accordingly. It does not suffice to replace a single ball and the total number of balls does not remain constant. Also, the multiplicative weights update algorithm only guarantees convergence of the temporal average. The actual distribution does not converge, even for self-play in symmetric zero-sum games (Bailey and Piliouras, 2018).

A notable subarea of machine learning is concerned with *multi-armed bandits*, a simple model of learning optimal sequential decisions when only very limited information is available (see, e.g., Bubeck and Cesa-Bianchi, 2012; Slivkins, 2019). The theory of adversarial bandits is closely connected to learning in repeated multi-player games and it turns out that the prototypical algorithm for adversarial bandits, Exp3 (which stands for “exponential-weight algorithm for exploration and exploitation”), bears some similarities to the urn process we describe in this paper. Exp3 can be formulated as an algorithm that learns an equilibrium strategy of a symmetric zero-sum game in self-play by iteratively updating a probability distribution merely based on the payoff associated with two actions randomly drawn from the current distribution. Auer et al. (2002) prove strong bounds on the expected average regret and average regret achieved by Exp3 after a finite number of rounds, which imply that the temporal average of the distributions converges to a strategy close to an equilibrium. How close it gets to an equilibrium depends on a parameter that is roughly related to our mutation rate. Exp3 updates a probability distribution rather than the contents of a discrete urn and we are not aware of convergence results beyond the temporal average.

The literature on equilibrium learning often focusses on minimizing *regret* rather than relative entropy with respect to an equilibrium distribution (see, e.g., Foster and Vohra, 1999; Auer et al., 2002). In our context, the regret of the urn distribution at round n is $\max_{i\in[d]}(\tilde{M}_R X^{(N,r)}(n,p_0))_i$. It follows from Theorem 1 that for sufficiently

games. While best-response dynamics are conceptually different from our urn process, their technical approach bares similarities with ours in that they use a deterministic process obtained as a solution to a differential equation to approximate a stochastic process.

large n, the regret is close to zero with high probability. Our simulations show that the regret of the urn distribution converges faster than its relative entropy. This is interesting insofar as in order to approximately satisfy the desirable axiomatic properties of maximal lotteries, low regret is sufficient. It can be shown that a lottery has small regret if and only if it is a maximal lottery of a nearby preference profile. In other words, even if the urn distribution is still far from a maximal lottery, the distribution can perform almost as well as a maximal lottery. We have identified preference profiles where this effect is quite noticeable.

5.2. Evolutionary Game Theory

The *replicator equation* in evolutionary game theory (see, e.g., Taylor and Jonker, 1978; Schuster and Sigmund, 1983; Hofbauer and Sigmund, 1998) describes how the distribution of different species changes continuously over time based on the individuals' fitnesses. In its basic form, it states that the change in the relative frequency of a species equals the relative fitness of the species (that is, its fitness relative to the entire population) minus the change in the size of the entire population. When the fitness depends linearly on the relative frequencies of the species and the population size is constant, the replicator equation defines the continuous deterministic process $y\colon \mathbb{R}_{\geq 0} \to \Delta$ with fitness function $f^{(r)}\colon \Delta \to \mathbb{R}^d$ below when setting r to 0. When $r > 0$, this process corresponds to a continuous and deterministic version of the urn process described in this paper (see Theorem 2).

$$\begin{aligned} &\frac{d}{dt} y(t) = f^{(r)}(y(t)) \quad \text{and} \quad y(0) = p_0, \quad \text{where} \\ &f_i^{(r)}(p) = 2(1-r)p_i(\tilde{M}p)_i + r\left(\frac{1}{d} - p_i\right). \end{aligned} \tag{4}$$

Solutions of this equation for $r = 0$ are connected to *evolutionary stable* distributions as introduced by Maynard Smith and Price (1973). A distribution of species is evolutionary stable if its relative fitness exceeds that of every other distribution in a fixed neighborhood of it. Hence, evolutionary stable distributions are attractors of the dynamics defined by Equation (4) (with $r = 0$) in the sense that they are limit points of solutions when the initial distribution p_0 is in the respective neighborhood.

	Model	Interaction	Mutations	Pop. Size	Convergence
Knebel et al. (2015)	continuous	pairs, det.	no[a]	fixed	temporal average
Laslier et al. (2017)	discrete	triples, det.	no	increasing	distribution
Grilli et al. (2017)	continuous	triples, stoch.	no	fixed	distribution
Theorem 1	discrete	pairs, stoch.	yes	fixed	fraction of rounds
Corollary 1	discrete	pairs, stoch.	yes	fixed	temporal average
Theorem 2	continuous	pairs, det.	yes	fixed	distribution

Table 1: Comparison of related models and results.
a: While Knebel et al. (2015) consider a discrete process with mutations, the continuous process they study has no mutations.

Mixed equilibria of zero-sum games such as Rock-Paper-Scissors usually fail to be evolutionary stable. As a consequence, results that prove convergence of dynamics to equilibrium strategies, either modify the underlying process or settle for weaker notions of convergence such as convergence of the temporal average.

In the following, we briefly discuss three results that are closest to ours.[17]

Knebel et al. (2015) study a dynamic process that involves quantum particles and is equivalent to a deterministic version of our urn process. Leveraging a classic result from evolutionary game theory (Hofbauer and Sigmund, 1998, Theorem 5.2.3), they show that the *temporal average* of this process converges to an equilibrium strategy (i.e., a maximal lottery) of the zero-sum game induced by the transition probabilities between quantum states. Even though their model allows for mutations, Knebel et al. neglect mutations when analyzing the continuous process, which may cause the process to cycle around the equilibrium strategy without converging to it.

Laslier and Laslier (2017) consider a discrete urn process that is similar to ours, but in which the number of balls in the urn increases over time. Given a binary comparison matrix that specifies which alternative wins against which alternative, they consider a process, in which three balls are drawn from the urn. Whenever one of three balls beats both other balls, a new ball of the same type is added to the urn. Otherwise, one of the three types is chosen at random and a ball of that type is added. Their main result shows that the distribution in the urn converges towards the (unique) maximal lottery of the skew-comparison matrix. Since the number of

[17]An extended version of this paper gives a more complete account of the related literature (Brandl and Brandt, 2021).

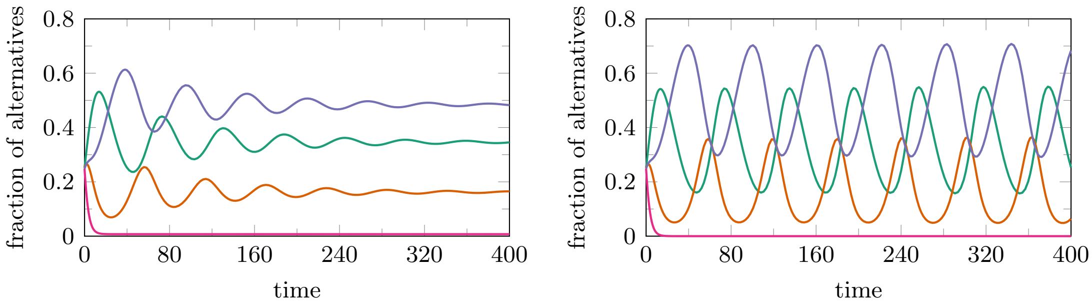


Figure 3: The continuous deterministic process $y(t)$ solving Equation (4) with $r = 0.01$ on the left and $r = 0$ on the right. For strictly positive r, $y(t)$ converges to a zero of $f^{(r)}$ (see Theorem 2). For $r = 0$, it approaches an orbit of constant entropy relative to a zero of $f^{(r)}$. The underlying skew-symmetric matrix $\tilde{M}$ is given by $\tilde{M}_{12} = \tilde{M}_{14} = \tilde{M}_{24} = \tilde{M}_{34} = {}^1\!/_3$, $\tilde{M}_{13} = -{}^1\!/_9$, and $\tilde{M}_{23} = {}^2\!/_9$.

balls in the urn increases, convergence is generally very slow.

Following earlier work by Allesina and Levine (2011), Grilli et al. (2017) consider a dynamic process in population biology to explain the stable coexistence of multiple species. Based on Laslier and Laslier's findings, Grilli et al. adapt the replicator equation to interactions of triples of individuals. In contrast to Laslier and Laslier, they keep the number of individuals constant and do not require the comparison matrix to be binary. They show that with a continuum of individuals, this process converges to an equilibrium strategy of the skew-comparison matrix.

Without mutations (i.e., $r = 0$), the deterministic process described by the differential equation (4) does not in general converge, but only approaches an orbit of constant entropy relative to a zero of the fitness function $f^{(0)}$. When introducing mutations, the limiting behavior of the process changes qualitatively (see Figure 3). As Theorem 2 shows, it then converges to a zero of the fitness function. To the best of our knowledge, this is a new observation.

Theorem 2. *Let $f^{(r)}$ and y be defined as in Equation* (4). *If $r > 0$, $f^{(r)}$ has a unique zero $p^{(r)}$ and $y(t)$ converges to $p^{(r)}$ as $t \to \infty$. Moreover, if r goes to 0, then $p^{(r)}$ converges to $ML(R)$ in Hausdorff distance.*

Table 1 summarizes the key differences between the above mentioned results and ours. In comparison, the main contribution of our work is that we are able to show for a discrete (rather than continuous) process based on stochastic (rather than deterministic) interactions between pairs (rather than triples) that the distribution in the urn is close to a maximal lottery most of the time (rather than convergence of the temporal average). Methodologically, the approach we take to cope with the discrete process is related to that of Benaïm and Weibull (2003), who study more general population processes in n-player games.[18]

We believe that Theorem 2 as well as Theorem 1 and Corollary 1 are of relevance to the natural sciences. In particular, a discrete model may describe the mentioned natural phenomena more accurately than continuous ones. As Corollary 1 shows, the expectation of the discrete process with a large number of individuals is a good approximation of the continuous process. Furthermore, the observation that convergence is only guaranteed if mutations occur with small probability and the number of individuals is large enough seems noteworthy.

Acknowledgments

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grants BR 2312/11-1, BR 2312/11-2, BR 2312/12-1, and BR 5969/1-1. The authors thank Stefano Allesina, Stergios Athanasoglou, Vincent Conitzer, Javier Esparza, Erwin Frey, Drew Fudenberg, Philipp Geiger, Umberto Grandi, Matthias Greger, Sean Horan, Johannes Knebel, Jean-François Laslier, Patrick Lederer, Hervé Moulin, Noam Nisan, Robert Schapire, Omer Tamuz, Nicolas Vieille, Jörgen Weibull, Peyton Young, and the participants of the COMSOC video seminar (March 2021), the International Conference on "New Directions in Social Choice" (St. Petersburg, July 2021), the Hausdorff Center for Mathematics Symposium (Bonn, August 2021), the seminar of the Center of Economic Research of ETH Zürich (Zurich, November 2021), the Nobel Symposium "One Hundred Years of Game Theory" (Stockholm, December 2021), the Microeconomics Research Seminar at the University of Hamburg (Hamburg, April 2022), the Economics Seminar at the University of Milano-Bicocca (Milan, May 2022), the Hi!Paris Symposium on Artificial Intelligence and the Social Sciences (Paris, June 2022), the 16th Meeting of the Society of Social Choice and Welfare (Mexico City, June 2022), and the Economics Seminar at Bielefeld University (Bielefeld, June 2022) for stimulating discussions and encouraging feedback.

[18] In Benaïm and Weibull's model, each player has a population of N individuals who play pure strategies. In each round, one individual of one player can update their strategy based on the distributions of pure strategies of all players. An update rule induces a deterministic process described by a differential equation similar to (4) below. They show that if N is large, the distributions of strategies among the individuals of each role in this stochastic process approximate the deterministic process described by the differential equation. Our setting corresponds to a symmetric two-player zero-sum game and an update rule based on the comparison matrix $\tilde{M}$. The special properties of this instance allow us to make more precise statements about the behavior of the deterministic process, and, thus, of the stochastic process for large N. In particular, we show that the deterministic process converges and that its limit approximates a maximal lottery.

APPENDIX: Proofs

As guidance for the reader, we outline the main steps in the proof of Theorem 1. Fix any $\delta, \tau > 0$.

In Appendix A, we consider, for any $p \in \Delta^{(N)}$, the expected value of $N\left(X^{(N,r)}(k+1,p) - X^{(N,r)}(k,p)\right)$, which, conditional on $X^{(N,r)}(k,p)$, is independent of k since $X^{(N,r)}$ is a time-homogeneous Markov process. Moreover, it is independent of N since the probability of replacing a ball of type j by one of type i is independent of N. Hence, these expected values induce a continuous function $f^{(r)} \colon \Delta \to \mathbb{R}^d$. We consider $g^{(r)} \colon \Delta \to \mathbb{R}^d$ with $g^{(r)}(p) = p + \frac{1}{2} f^{(r)}(p)$ and show that it maps to Δ. If $r > 0$, $g^{(r)}$ has a unique fixed-point $p^{(r)}$ (a zero of $f^{(r)}$), which is close to some lottery in $ML(R)$ for any small enough r. We choose r_0 so that $p^{(r)}$ is no more than $\frac{\delta}{2}$ away from $ML(R)$ for all $0 < r \leq r_0$. Fixing such an r, let $p^* \in ML(R)$ be a maximal lottery which is within δ of $p^{(r)}$.

Appendix B studies the following differential equation with $p \in \Delta$, $t \in \mathbb{R}_{\geq 0}$, and $y(\cdot, p) \colon \mathbb{R}_{\geq 0} \to \Delta$.

$$\begin{aligned} \frac{d}{dt} y(t,p) &= f^{(r)}(y(t,p)) \\ y(0,p) &= p \end{aligned} \tag{5}$$

A solution to (5) is a *deterministic* process that can be interpreted as the *stochastic* process we consider with a continuum of balls. We show that the unique solution $y^{(r)}(\cdot, p)$ of (5) converges to $p^{(r)}$ for any initial state $p \in \Delta$ as t goes to infinity and the convergence is uniform in p. This is done by showing that the entropy of $p^{(r)}$ relative to $y^{(r)}(t,p)$ decreases monotonically at a rate proportional to the square of the distance between $p^{(r)}$ and $y^{(r)}(t,p)$.

Appendix C relates the discrete-time stochastic process $X^{(N,r)}$ to the continuous-time deterministic process $y^{(r)}$. To this end, we extend the former to the real time axis by letting $\bar{X}^{(N,r)}(t,p) = X^{(N,r)}(k,p)$ for $t \in [\frac{k-1}{N}, \frac{k}{N})$. Given any $T > 0$, one can show that with probability close to 1, $\bar{X}^{(N,r)}$ *approximately* satisfies the integral equation corresponding to (5) for t between 0 and T and uniformly in $p \in \Delta^{(N)}$ if N is large. Using Grönwall's inequality, we show that with probability close to 1,

$\bar{X}^{(N,r)}(t,p)$ and $y^{(r)}(t,p)$ are close to each other for all t from 0 to T.[19] However, for t larger than T, they may (and almost surely will) be arbitrarily far apart.

To deal with this, we partition the time axis into consecutive intervals of length T and synchronize the deterministic process with the stochastic process at the beginning of each interval. More precisely, since $y^{(r)}(t,p)$ converges to $p^{(r)}$ as t goes to infinity uniformly in p, we can find $T > 0$ such that $y^{(r)}(t,p)$ is no more than $\frac{\delta}{4}$ away from $p^{(r)}$ for all but possibly a $1 - \frac{\tau}{2}$ fraction of the interval $[0,T]$ for all p. Moreover, we can choose N large enough so that with probability at least $1 - \frac{\tau}{2}$, the distance between $\bar{X}^{(N,r)}$ and $y^{(r)}$ is less than $\frac{\delta}{4}$ for all t in an interval of length T provided both processes start at the same point at the beginning of the interval. We chop up the time axis into intervals $[0,T]$, $[T,2T]$, ... On the interval $[(k-1)T, kT]$, we compare $\bar{X}^{(N,r)}(t,p)$ to $y^{(r)}(t-(k-1)T, \bar{x}_{k-1})$, where $\bar{x}_{k-1} = X^{(N,r)}((k-1)T, p)$. That is, we reset $y^{(r)}$ to the position of $\bar{X}^{(N,r)}$ at the beginning of the interval. In those intervals where the distance between both processes is never more than $\frac{\delta}{4}$, $\bar{X}^{(N,r)}$ is no more than $\frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2}$ away from $p^{(r)}$ for all but a $\frac{\tau}{2}$ fraction of the interval. By the choice of N, the union of those intervals is almost surely at least a $1 - \frac{\tau}{2}$ fraction of the time axis. Summing over all intervals, this is enough to conclude that $\bar{X}^{(N,r)}$ is no more than $\frac{\delta}{2}$ away from $p^{(r)}$ at least a $1-\tau$ fraction of the time. Since $p^{(r)}$ is no more than $\frac{\delta}{2}$ away from p^*, we can get the same conclusion with δ in place of $\frac{\delta}{2}$ and p^* in place of $p^{(r)}$. Translating this statement back to $X^{(N,r)}$ gives the first part of Theorem 1.

A. A Continuous Vector Field Induced by the Markov Chain

In this section, we define a continuous mapping from Δ to Δ based on the *expected* urn distribution in the subsequent round for each state of the Markov chain. We then show that this mapping admits a unique fixed-point corresponding to an approximate maximal lottery.

[19] In the language of functional analysis, this step corresponds to an approximation of an operator semi-group. Consider the operators $\Gamma(t)$ on probability measures on Δ induced by mapping $p \in \Delta$ to $y^{(r)}(t,p)$. Then $\{\Gamma(t) \colon t \geq 0\}$ is an operator semi-group (that is, $\Gamma(s+t) = \Gamma(s)\Gamma(t)$). On $\Delta(\Delta^{(N)})$, we approximate $\Gamma(t)$ by $(P^{(N,r)})^{Nt}$.

Recall that $\{X^{(N,r)}(n, p_0) \colon n \in \mathbb{N}_0\}$ is a discrete-time, time-homogeneous Markov chain with state space $\Delta^{(N)}$ and transition probability matrix

$$P^{(N,r)}(p, p') = \begin{cases} 2(1-r)p_i p_j M(i,j) + \frac{r}{d} p_j & \text{if } i \neq j \\ 2(1-r)\sum_{k=1}^d p_k^2 + \frac{r}{d} & \text{if } i = j \end{cases}$$

for $p \in \Delta^{(N)}$ and $p' = p + \frac{e_i}{N} - \frac{e_j}{N}$ for $i, j \in [d] = \{1, \ldots, d\}$ with $p' \in \Delta^{(N)}$. All other transition probabilities are 0. If $r > 0$, it is irreducible and aperiodic and thus admits a unique stationary distribution in $\Delta(\Delta^{(N)})$, a probability distribution over urn distributions. We omit writing the initial state p_0 whenever it is convenient.

For $i \in [d]$, we calculate the expected change in the ith component of $X^{(N,r)}$ times N given that $X^{(N,r)}$ is in state $p \in \Delta^{(N)}$.

$$\begin{aligned}
&N\,\mathbb{E}\left(X_i^{(N,r)}(n+1) - X_i^{(N,r)}(n) \mid X^{(N,r)}(n) = p\right) \\
&= N \sum_{p' \in \Delta^{(N)}} (p'_i - p_i) P^{(N,r)}(p, p') \\
&= 2(1-r) \sum_{j \neq i} p_i p_j \left(M(i,j) - M(j,i)\right) + \frac{r}{d} \sum_{j \neq i} (p_j - p_i) \\
&= 2(1-r) p_i \sum_{j \neq i} \tilde{M}(i,j) p_j + \frac{r}{d} \left(1 - p_i - (d-1)p_i\right) \\
&= 2(1-r) p_i (\tilde{M}p)_i + r\left(\frac{1}{d} - p_i\right)
\end{aligned}$$

For the last equality, recall that $\tilde{M}(i,i) = 0$ since $\tilde{M}$ is skew-symmetric.

Based on this, we define the continuous function $f^{(r)} \colon \Delta \to \mathbb{R}^d$ with

$$f_i^{(r)}(p) = 2(1-r) p_i (\tilde{M}p)_i + r\left(\frac{1}{d} - p_i\right). \tag{6}$$

Let $g^{(r)} \colon \Delta \to \Delta$ with $g^{(r)}(p) = p + \frac{1}{2} f(p)$ for $p \in \Delta$. We show that $g^{(r)}$ is well-defined (that is, indeed maps to Δ) and has a fixed-point. If $r > 0$, this fixed-point is unique and we denote it by $p^{(r)}$. As r goes to 0, $p^{(r)}$ converges to the set of maximal lotteries for the profile R that induces $\tilde{M}$. We note that if $r = 0$, $g^{(r)}$ has a unique fixed-point if and only if there is a unique maximal lottery.

Lemma 1. *For $r > 0$, $g^{(r)}$ has a unique fixed-point $p^{(r)}$. Moreover, for every $\delta > 0$, there is r_0 so that $p^{(r)} \in B_\delta(ML(R))$ for all $r \leq r_0$.*

Proof. We verify that $g^{(r)}$ maps to Δ. For all $p \in \Delta$, $\sum_{i\in[d]} f_i^{(r)}(p) = 2(1-r)p^\intercal \tilde{M} p + r\left(1 - \sum_{i\in[d]} p_i\right) = 0$ since $\tilde{M}$ is skew-symmetric and $p \in \Delta$. Moreover,

$$f_i^{(r)}(p) = 2(1-r)p_i \underbrace{(\tilde{M}p)_i}_{\geq -1} + r\left(\frac{1}{d} - p_i\right) \geq -2p_i.$$

Thus,

$$g_i^{(r)}(p) \geq p_i + \frac{1}{2}(-2p_i) \geq 0.$$

It follows that $g^{(r)}$ maps to Δ. Moreover, $g^{(r)}$ is continuous since $f^{(r)}$ is continuous. Hence, by Brouwer's Theorem, $g^{(r)}$ has a fixed point $p^{(r)}$.

Now let $r > 0$. Then, for all $p \in \Delta$ with $f^{(r)}(p) = 0$, we have for all $i \in [d]$, $p_i > 0$ since $p_i = 0$ implies $f_i^{(r)}(p) = r\frac{1}{d} > 0$. Hence, we can rewrite $f^{(r)}(p) = 0$ as follows: for all $i \in [d]$,

$$2(1-r)(\tilde{M}p)_i = r\left(1 - \frac{1}{p_i d}\right) \tag{7}$$

To show that $f^{(r)}$ has a unique zero, assume that $f^{(r)}(p) = f^{(r)}(q) = 0$ for $p, q \in \Delta$. We have

$$\begin{aligned}
0 &= 2(1-r)\left(p^\intercal \tilde{M} q + q^\intercal \tilde{M} p\right) \\
&= 2(1-r)\sum_{i\in[d]} p_i\left(\tilde{M}q\right)_i + q_i\left(\tilde{M}p\right)_i \\
&\overset{(7)}{=} r\sum_{i\in[d]} p_i\left(1 - \frac{1}{q_i d}\right) + q_i\left(1 - \frac{1}{p_i d}\right) \\
&= \frac{r}{d}\sum_{i\in[d]} \frac{p_i q_i - p_i}{q_i} + \frac{p_i q_i - q_i}{p_i} \\
&= -\frac{r}{d}\sum_{i\in[d]} \frac{(p_i - q_i)^2}{p_i q_i} \leq -\frac{r}{d}|p-q|_2^2
\end{aligned}$$

where the first equality uses the skew-symmetry of $\tilde{M}$ (hence, $p^\intercal \tilde{M} q = -q^\intercal \tilde{M} p$), the third equality follows from (7) and the fact that p and q are zeros of $f^{(r)}$, and the last two are algebra. ($|\cdot|_2$ denotes the L^2-norm.) This sequence of equalities implies that $p = q$. Hence, $p^{(r)}$ is the unique zero of $f^{(r)}$ for $r > 0$. Since every fixed-point of $g^{(r)}$ is a zero of $f^{(r)}$, $g^{(r)}$ has a unique fixed-point.

For the last statement, let $\delta > 0$. By (7), for all $r > 0$ and $i \in [d]$,

$$\left(\tilde{M} p^{(r)}\right)_i = \frac{r}{2(1-r)} \left(1 - \frac{1}{p_i^{(r)} d}\right) \leq \frac{r}{2(1-r)}. \tag{8}$$

Suppose for every $r_0 > 0$, there is $r < r_0$ so that $p^{(r)} \notin B_\delta(ML(R))$. Then we can find a sequence (r_n) going to 0 so that $p^{(r_n)} \notin B_\delta(ML(R))$ for all n. By passing to a subsequence, we may assume that $p^{(r_n)} \to p \notin B_\delta(ML(R))$. But from (8) it follows that $\tilde{M} p \leq 0$ so that $p \in ML(R)$, which is a contradiction.

□

B. Properties of the Deterministic Process

In this section, we study a deterministic version of the stochastic process described by the Markov chain. We thus have a continuum of balls and continuous time, and show that this process converges to the unique fixed-point identified in the previous section.

Function $f^{(r)}$ defined in Equation (6) gives rise to a (first-order ordinary) differential equation for continuously differentiable functions from $[0, \infty)$ to Δ, that is, functions in $\mathcal{C}^1([0, \infty), \Delta)$. For $y \in \mathcal{C}^1([0, \infty), \Delta)$ and $p_0 \in \Delta$, consider

$$\begin{aligned} \frac{d}{dt} y(t) &= f^{(r)}(y(t)) \\ y(0) &= p_0 \end{aligned} \tag{9}$$

We show that (9) has a unique global solution $y^{(r)}$ for all $r > 0$ and $p_0 \in \Delta$. Moreover, this solution converges to the zero $p^{(r)}$ of $f^{(r)}$ as t goes to infinity. Since r remains fixed throughout this section, we frequently omit the superscript (r).

The proof that (9) has a unique local solution with values in $\mathbb{R}^d$ is standard. Only the fact that the solution does not leave the domain Δ of f and can thus be extended to a global solution requires attention.

Lemma 2. *For every $p_0 \in \Delta$, (9) has a unique solution $y \in \mathcal{C}^1([0,\infty), \Delta)$ with $y(0) = p_0$.*

Proof. Note that f is Lipschitz-continuous in a neighborhood of Δ. It follows from the Picard-Lindelöf Theorem that for any $t_0 \in [0,\infty)$ and $p \in \Delta$, the system

$$\begin{aligned} \frac{d}{dt} y(t) &= f(y(t)) \\ y(t_0) &= p \end{aligned} \tag{10}$$

has a unique local solution, that is, a solution $y \in \mathcal{C}^1((t_0 - \varepsilon, t_0 + \varepsilon), \mathbb{R}^d)$.

We observe that y maps to Δ. First, by the same arguments as in the proof of Lemma 1, we have

$$\frac{d}{dt} \sum_{i \in [d]} y_i(t) = \sum_{i \in [d]} f_i(y(t)) = 0$$

whenever $y(t) \in \Delta$. Second, if $y_i(t) = 0$, then $\frac{d}{dt} y_i(t) = f_i(y(t)) > 0$. Hence, $y(t) \in \Delta$ for all $t \in (t_0 - \varepsilon, t_0 + \varepsilon)$. Since $t_0 \in [0,\infty)$ was arbitrary, it follows that y can be uniquely extended to a global solution in $\mathcal{C}^1([0,\infty), \Delta)$. □

Denote by $y^{(r)}(t, p_0) \in \mathcal{C}^1([0,\infty), \Delta)$ the unique solution to (9) with $y^{(r)}(0, p_0) = p_0$. We will sometimes suppress the argument p_0 when it is clear from the context.

We want to show that if $r > 0$, $y^{(r)}(t, p_0)$ converges to the zero $p^{(r)}$ of $f^{(r)}$ as t goes to infinity. Moreover, the convergence is uniform in p_0. The proof of this fact in Lemma 4 uses the relative entropy (aka the Kullback–Leibler Divergence) of $p, q \in \Delta$, which is defined as

$$D(p \mid q) = \sum_{i \in [d]} p_i \log\left(\frac{p_i}{q_i}\right).$$

Moreover, the following lower bound on the relative entropy will be helpful (see, e.g., Cover and Thomas, 2006, Lemma 11.6.1).

Lemma 3. *For all $p, q \in \Delta$,*

$$D(p \mid q) \geq \frac{1}{2\log 2} |p-q|^2 .$$

To ease notation, we write χ_S for the indicator function of a set $S \subset \mathbb{R}^d$ and $\bar{\chi}_S = 1 - \chi_S$ for the indicator function of the complement of S.

Lemma 4. *Let $r > 0$. Then,*

$$\lim_{t\to\infty} \sup\left\{\left|y^{(r)}(t, p_0) - p^{(r)}\right| : p_0 \in \Delta\right\} = 0.$$

Proof. Fix p_0 in the interior of Δ and write $y = y(\cdot, p_0)$. We show that the entropy of $p^{(r)}$ relative to $y(t)$ decreases at a rate of at least $\frac{r}{d\sqrt{d}}\left|p^{(r)} - y(t)\right|_2^2$.

$$\begin{aligned}
\frac{d}{dt}D(p^{(r)} \mid y(t)) &= \frac{d}{dt}\sum_{i\in[d]} p_i^{(r)} \log\left(\frac{p_i^{(r)}}{y_i(t)}\right) = -\sum_{i\in[d]} p_i^{(r)} \frac{\frac{d}{dt}y_i(t)}{y_i(t)} \\
&\overset{(i)}{=} -\sum_{i\in[d]} p_i^{(r)} \frac{f_i(y(t))}{y_i(t)} \\
&= -\sum_{i\in[d]} p_i^{(r)} \frac{2(1-r)y_i(t)(\tilde{M}y(t))_i + r\left(\frac{1}{d} - y_i(t)\right)}{y_i(t)} \\
&= -2(1-r)\sum_{i\in[d]} p_i^{(r)}(\tilde{M}y(t))_i - r\sum_{i\in[d]} p_i^{(r)}\left(\frac{1}{y_i(t)d} - 1\right) \\
&\overset{(ii)}{=} 2(1-r)\sum_{i\in[d]} y_i(t)(\tilde{M}p^{(r)})_i - r\left(\sum_{i\in[d]} \frac{p_i^{(r)}}{y_i(t)d} - 1\right) \\
&\overset{(iii)}{=} \sum_{i\in[d]} y_i(t) r\left(1 - \frac{1}{p_i^{(r)}d}\right) - r\left(\sum_{i\in[d]} \frac{p_i^{(r)}}{y_i(t)d} - 1\right) \\
&= r\left(2 - \frac{1}{d}\sum_{i\in[d]} \frac{y_i(t)}{p_i^{(r)}} + \frac{p_i^{(r)}}{y_i(t)}\right) \\
&\overset{(iv)}{=} -\frac{r}{d}\sum_{i\in[d]} \frac{(p_i^{(r)} - y_i(t))^2}{p_i^{(r)}y_i(t)} \leq -\frac{r}{d\sqrt{d}}\left|p^{(r)} - y(t)\right|^2
\end{aligned}$$

Here, (i) follows from the fact that y satisfies (9), (ii) uses the skew-symmetry of $\tilde{M}$ and $\sum_{i\in[d]} p_i^{(r)} = 1$, (iii) uses (7), and (iv) uses $a^2+b^2=(a+b)^2-2ab$ for any $a,b\in\mathbb{R}$.

For $t \geq t_0 \geq 0$, we have

$$0 \leq D(p^{(r)} \mid y(t)) = D(p^{(r)} \mid y(t_0)) + \int_{t_0}^{t} \frac{d}{ds} D(p^{(r)} \mid y(s)) ds \leq D(p^{(r)} \mid y(t_0)).$$

Combining this with the sequence of equalities above, we see that

$$0 \leq \frac{r}{d}\sum_{i=1}^{d}\int_{t_0}^{t}\frac{(p_i^{(r)}-y_i(s))^2}{p_i^{(r)}y_i(s)}ds = -\int_{t_0}^{t}\frac{d}{ds}D(p^{(r)} \mid y(s))ds \leq D(p^{(r)} \mid y(t_0)). \tag{11}$$

We want to prove that $y(t,p_0)$ converges to $p^{(r)}$ uniformly in p_0 as t goes to ∞. That is, for all $\varepsilon > 0$, there exists $T > 0$ such that for all $t \geq T$ and all $p_0 \in \Delta$, $|y(t,p_0) - p^{(r)}| < \varepsilon$. To this end, first note that if $y_i(t,p_0) < \frac{r}{4d}$, then

$$\frac{d}{dt}y_i(t,p_0) = 2(1-r)y_i(t,p_0)\underbrace{\left(\tilde{M}y(t,p_0)\right)_i}_{\geq -1} + r\left(\frac{1}{d} - y_i(t,p_0)\right) \geq -\frac{r}{2d} + \frac{r}{d} \geq \frac{r}{2d}.$$

Hence, for all $p_0 \in \Delta$, $i \in [d]$, and $t \geq 1$, $y_i(t,p_0) \geq \frac{r}{4d}$. We can thus upper bound $D(p^{(r)} \mid y(t,p_0))$ for all $p_0 \in \Delta$ and $t \geq 1$ by $C = \max_{p\in\Delta^r} D(p^{(r)} \mid p) < \infty$, where $\Delta^r = \{p \in \Delta \colon p_i \geq \frac{r}{4d} \text{ for all } i \in [d]\}$.

Now we prove uniform convergence in p_0. Let $\varepsilon > 0$. It follows from (11) with $t_0 = 1$ that given $\delta > 0$, for all $p_0 \in \Delta^r$,

$$\int_{t\geq 1} \bar{\chi}_{B_\delta(p^{(r)})}(y(t,p_0))dt \leq \frac{Cd\sqrt{d}}{r\delta^2}.$$

Hence, for every $p_0 \in \Delta^r$, we can find $t_0(p_0,\delta) \in [1, 1+\frac{Cd\sqrt{d}}{r\delta^2}]$ such that

$$\left|y(t_0(p_0,\delta),p_0) - p^{(r)}\right| < \delta. \tag{12}$$

Using the estimate $\log(x-\delta) \geq \log(x) - \frac{\delta}{x-\delta}$ for the last inequality, we find that

$$\begin{aligned} D(p^{(r)} \mid y(t_0(p_0,\delta))) &\leq \sum_{i\in[d]} \log\left(\frac{p_i^{(r)}}{p_i^{(r)}-\delta}\right) = \sum_{i\in[d]} \log\left(p_i^{(r)}\right) - \log\left(p_i^{(r)}-\delta\right) \\ &\leq \sum_{i\in[d]} \frac{\delta}{p_i^{(r)}-\delta} \leq \delta C' \end{aligned}$$

where $C' = 2d \max\{\frac{1}{p_i^{(r)}} : i \in [d]\}$ if $\delta \in (0, \frac{1}{2}\min\{p_i^{(r)} : i \in [d]\})$.

We use this bound and the fact that the relative entropy is non-increasing in t to show that $|y(t,p_0) - p^{(r)}| < \varepsilon$ for $t \geq t_0(p_0,\delta)$ for sufficiently small δ. By Lemma 3, we have for all $p \in \Delta$, $D(p^{(r)} \mid p) \geq \frac{1}{2\log 2}|p^{(r)} - p|^2$. Hence, $|p^{(r)} - y(t,p_0)| \leq \sqrt{2\log(2)\delta C'}$ for $t \geq t_0(p_0,\delta)$. Recalling that $t_0(p_0,\delta) \leq 1 + \frac{Cd\sqrt{d}}{r\delta^2} =: T$, we have for $\delta \in (0, \frac{\varepsilon^2}{2\log(2)C'})$ that $|p^{(r)} - y(t,p_0)| < \varepsilon$ for all $t \geq T$ and $p_0 \in \Delta$. Since ε was arbitrary, this proves uniform convergence. □

The next lemma states that for any $\delta > 0$, if the process $y^{(r)}$ starts sufficiently close to $p^{(r)}$, it will never get further than δ away from $p^{(r)}$.

Lemma 5. *Let $r > 0$ and $\delta > 0$. Then, there is $\eta > 0$ such that*

$$\sup\left\{\left|y^{(r)}(t,p) - p^{(r)}\right| : t \geq 0,\, p \in B_\eta(p^{(r)})\right\} < \delta$$

Proof. Recall that $p_i^{(r)} > 0$ for all $i \in [d]$. By Lemma 3, if $p \notin B_\delta(p^{(r)})$, then $D(p^{(r)} \mid p) \geq \frac{1}{2\sqrt{2}}\delta^2 =: C$. Since $D(p^{(r)} \mid \cdot)$ is continuous on the interior of Δ and $D(p^{(r)} \mid p^{(r)}) = 0$, there is $\eta > 0$ such that $D(p^{(r)} \mid p) < C$ for all $p \in B_\eta(p^{(r)})$. In the proof of Lemma 4, we have seen that $D(p^{(r)} \mid y^{(r)}(t,p))$ is non-increasing in t. Hence, for $p \in B_\eta(p^{(r)})$, it follows that $|y^{(r)}(t,p) - p^{(r)}| < \delta$ for all $t \geq 0$. □

Summarizing Lemma 1, Lemma 2, and Lemma 4, we get the following theorem.

Theorem 2. *Let $f^{(r)}$ and y be defined as in Equation (4). If $r > 0$, $f^{(r)}$ has a unique zero $p^{(r)}$ and $y(t)$ converges to $p^{(r)}$ as $t \to \infty$. Moreover, if r goes to 0, then $p^{(r)}$ converges to $ML(R)$ in Hausdorff distance.*

Remark 6 (Drawing without replacement)**.** For the urn process with drawing without replacement, $f^{(r)}$ as derived in Appendix A would depend on N. The solution $y^{(r)}$ of the differential equation (9) and the unique zero $p^{(r)}$ of $f^{(r)}$ would thus also depend on N. The previous lemmas carry over to this case with the straightforward adaptions.

C. Properties of the Stochastic Process

In this section, we study the behavior of the Markov chain $X^{(N,r)}$ by exploring its connections to the deterministic process $y^{(r)}$.

We estimate the distance between $X^{(N,r)}$ and the set of maximal lotteries in several steps. First, we choose T_0 large enough so that $y^{(r)}(\cdot, p_0)$ is close to $p^{(r)}$ for all but a small fraction of the time interval $[0, T_0]$ for all initial states p_0. In Lemma 6, we show that if N is large enough, $X^{(N,r)}$ approximately solves (the integral equation equivalent to) the differential equation (9) with high probability on the interval $[0, T_0]$ for any initial state. From this we conclude in Lemma 7 that for large enough N, $X^{(N,r)}$ is close to $y^{(r)}$ with high probability on any interval of length T_0, provided both processes start with the same state at the beginning of that interval. Thus, $X^{(N,r)}$ is with high probability approximately equal to $p^{(r)}$ for all but a small fraction of rounds in any interval of length T_0. Now we chop up the time line into successive intervals of length T_0. In expectation, $X^{(N,r)}$ stays close to $y^{(r)}$ in a large fraction of these intervals. Using an adaption of the strong law of large numbers, we show in Lemma 9 that $X^{(N,r)}$ is *almost surely* close to $p^{(r)}$ for all but a small fractions of rounds. Lastly, since by Lemma 1, $p^{(r)}$ is close to a maximal lottery if r is small enough, Theorem 1 follows.

The integral equation equivalent to (9) is

$$\begin{aligned} y(t) - y(0) &= \int_0^t f^{(r)}(y(s))ds \\ y(0) &= p_0 \end{aligned} \tag{13}$$

We show that $X^{(N,r)}$ approximately satisfies (13) (with the integral replaced by a sum) for large N on bounded time intervals. Lemma 6 below states that for any time T and any $\delta > 0$, we can choose N large enough so that with high probability, $X^{(N,r)}(n, p_0)$

does not violate (13) by more than δ within the first NT rounds independently of the initial state $p_0 \in \Delta^{(N)}$. For the proof, we use the following proposition due to Kurtz (1970, Proposition 4.1). (The statement is adapted to our setting.)

Proposition 1 (Kurtz, 1970)**.** *Let $(z^{(N)})_{N\in\mathbb{N}}$ be a sequence of discrete-time Markov chains with states spaces $A^{(N)}$ and probability transition matrices $Q^{(N)}$. Suppose there exist sequences of positive number (α_N) and (ε_N),*

$$\lim_{N\to\infty} \alpha_N = \infty \qquad \text{and} \qquad \lim_{N\to\infty} \varepsilon_N = 0$$

such that

$$\sup_{N\in\mathbb{N}} \sup_{p\in A^{(N)}} \alpha_N \sum_{q\in A^{(N)}} |p-q| Q^{(N)}(p,q) < \infty \tag{14}$$

and

$$\lim_{N\to\infty} \sup_{p\in A^{(N)}} \alpha_N \sum_{q\in A^{(N)}, |p-q|>\varepsilon_N} |p-q| Q^{(N)}(p,q) = 0. \tag{15}$$

Let

$$G^{(N)}(p) = \alpha_N \sum_{q\in A^{(N)}} (q-p) Q^{(N)}(p,q).$$

Then, for every $\delta > 0$ and $T > 0$,

$$\lim_{N\to\infty} \sup_{p\in A^{(N)}} \mathbb{P}\left(\sup_{n\leq \alpha_n T} \left| z^{(N)}(n) - z^{(N)}(0) - \sum_{k=0}^{n-1} \frac{1}{\alpha_N} G^{(N)}(z^{(N)}(k)) \right| > \delta \mid z^{(N)}(0) = p \right) = 0.$$

The following lemma applies this result to $(X^{(N,r)})_{N\in\mathbb{N}}$ for a fixed r.

Lemma 6. *For every $T > 0$ and $\delta > 0$,*

$$\lim_{N\to\infty} \sup_{p\in\Delta^{(N)}} \mathbb{P}\left(\sup_{n\leq NT} \left| X^{(N,r)}(n,p) - X^{(N,r)}(0,p) - \sum_{k=0}^{n-1} f^{(r)}\left(X^{(N,r)}(k,p)\right) \right| \geq \delta \right) = 0 \tag{16}$$

Proof. Recall that $P^{(N,r)}$ is the transition probability matrix of $X^{(N,r)}$. We apply Proposition 1 with $z^{(N)} = X^{(N,r)}$, $A^{(N)} = \Delta^{(N)}$, $Q^{(N)} = P^{(N,r)}$, $\alpha_N = N$, and $\varepsilon_N = \frac{2}{N}$ and check (14) and (15):

$$\begin{aligned}&\sup_{N\in\mathbb{N}} \sup_{p\in\Delta^{(N)}} N \sum_{q\in\Delta^{(N)}} |p-q| P^{(N,r)}(Np, Nq)\\ =&\sup_{N\in\mathbb{N}} \sup_{p\in\Delta^{(N)}} N \sum_{i,j=1}^{d} \frac{1}{N} |e_i - e_j| P^{(N,r)}(Np, Np - e_i + e_j) \le 2\end{aligned}$$

and

$$\lim_{N\to\infty} \sup_{p\in\Delta^{(N)}} N \sum_{q\in\Delta^{(N)}:\, |p-q|>\frac{2}{N}} |p-q| P^{(N,r)}(Np, Nq) = 0$$

Recalling the definition of $f^{(r)}$ shows that $G^{(N)} = f^{(r)}$ for all N. Hence, (16) follows. □

Note that Lemma 6 does not use the full strength of Proposition 1 since $G^{(N)} = f^{(r)}$ is independent of N. Recall from Remark 6 that for the urn process without replacement, $f^{(r)}$ does depend on N. Hence, the additional flexibility of Proposition 1 is needed in that case.

Since we want to compare the discrete-time process $X^{(N,r)}$ to the continuous-time process $y^{(r)}$ solving (9), it is convenient to turn $X^{(N,r)}$ into a continuous-time process. To this end, let $\bar{X}^{(N,r)}(t,p) = X^{(N,r)}(\lfloor Nt \rfloor, p)$ for all $t \ge 0$ and $p \in \Delta$. (That is, time is scaled by $\frac{1}{N}$.) $\bar{X}^{(N,r)}$ is a right-continuous step function, which takes steps of length $\frac{1}{N}|e_i - e_j| = \frac{2}{N}$ and is constant on time intervals $[\frac{k}{N}, \frac{k+1}{N})$. Thus, as N grows, the steps become smaller and appear in shorter intervals. Lemma 6 shows that on any bounded time interval, $\bar{X}^{(N,r)}$ satisfies (13) up to some arbitrary error with high probability when N is large enough. That is, for every $T > 0$ and $\delta > 0$,

$$\lim_{N\to\infty} \sup_{p\in\Delta^{(N)}} \mathbb{P}\left(\sup_{t\le T} \left| \bar{X}^{(N,r)}(t,p) - \bar{X}^{(N,r)}(0,p) - \int_0^t f^{(r)}\left(\bar{X}^{(N,r)}(s,p) ds\right) \right| \ge \delta \right) = 0 \tag{17}$$

In Lemma 7, we show that this implies that the trajectories of $y^{(r)}(\cdot, p)$ and $\bar{X}^{(N,r)}(\cdot, p)$ stay close to each other with high probability on a given bounded time interval for any initial state p for large N. Importantly for later use, the bound on the probability is uniform in p.

Lemma 7. *For every $T > 0$ and $\delta > 0$,*

$$\lim_{N\to\infty} \sup_{p\in\Delta^{(N)}} \mathbb{P}\left(\sup_{t\leq T}\left|y^{(r)}(t,p) - \bar{X}^{(N,r)}(t,p)\right| \geq \delta\right) = 0 \tag{18}$$

Proof. First observe that since $f^{(r)}$ is continuously differentiable on the compact space Δ, there is $C \in \mathbb{R}_{\geq 0}$ such that $f^{(r)}$ is Lipschitz-continuous with constant C. Let $T > 0$, $\delta > 0$, and $p \in \Delta$. If $\sup_{t\leq T}|\bar{X}^{(N,r)}(t,p) - \bar{X}^{(N,r)}(0,p) - \int_0^t f^{(r)}(\bar{X}^{(N,r)}(s,p))ds| < \varepsilon$, then for all $t \in [0, T]$,

$$\begin{aligned}
\left|y^{(r)}(t,p) - \bar{X}^{(N,r)}(t,p)\right| &= \left|y^{(r)}(t,p) - y^{(r)}(0,p) - \bar{X}^{(N,r)}(t,p) + \bar{X}^{(N,r)}(0,p)\right| \\
&< \varepsilon + \int_0^t \left|f^{(r)}\left(y^{(r)}(s,p)\right) - f^{(r)}\left(\bar{X}^{(N,r)}(s,p)\right)\right| ds \\
&\leq \varepsilon + C\int_0^t \left|y^{(r)}(s,p) - \bar{X}^{(N,r)}(s,p)\right| ds
\end{aligned}$$

The first inequality follows from the assumption about $\bar{X}^{(N,r)}$ and the fact that $y^{(r)}$ satisfies (13). The second inequality uses the Lipschitz-continuity of $f^{(r)}$. We apply Grönwall's inequality to conclude that

$$\sup_{t\leq T}\left|y^{(r)}(t,p) - \bar{X}^{(N,r)}(t,p)\right| < \varepsilon e^{CT} < \delta$$

for $\varepsilon > 0$ small enough. Note that the choice of ε does not depend on $p = \bar{X}^{(N,r)}(0,p)$.

By (17), for every $\rho > 0$, we can find $N_0 \in \mathbb{N}$ such that for every $N \geq N_0$,

$$\sup_{p\in\Delta^{(N)}} \mathbb{P}\left(\sup_{t\leq T}\left|\bar{X}^{(N,r)}(t,p) - \bar{X}^{(N,r)}(0,p) - \int_0^t f^{(r)}\left(\bar{X}^{(N,r)}(s,p)\right) ds\right| \geq \varepsilon\right) < \rho$$

Hence, for all $N \geq N_0$,

$$\sup_{p \in \Delta^{(N)}} \mathbb{P}\left(\sup_{t \leq T} |y^{(r)}(t,p) - \bar{X}^{(N,r)}(t,p)| \geq \delta\right) < \rho$$

Since ρ was arbitrary, (18) follows. $\square$

The last tool, Lemma 8, is in essence a one-sided strong law of large numbers for indicator random variables. Instead of the usual assumption of i.i.d. random variables, it only assumes that the probability of each variable being 1 is conditionally upper bounded. The elegant proof was pointed out to us by an anonymous referee.

Lemma 8. *Let $\alpha \in [0,1]$. Let $\{Z_n : n \in \mathbb{N}_0\}$ be indicator random variables and, for $n \geq 1$, $S_n = \sum_{k=1}^{n} Z_k$. If $\mathbb{P}(Z_1 = 1) \leq \alpha$ and for all $n \geq 2$, $\mathbb{P}(Z_n = 1 \mid S_{n-1}) \leq \alpha$, then*

$$\mathbb{P}\left(\limsup_{n \to \infty} \frac{S_n}{n} > \alpha\right) = 0$$

Proof. The proof uses the moment generating functions of the random variables involved. Let Z be an indicator random variable with $\mathbb{P}(Z = 1) = \alpha$. Then, for all $n \geq 2$ and $t > 0$,

$$\mathbb{E}\left(e^{tS_n}\right) = \mathbb{E}\left(e^{tZ_n} e^{tS_{n-1}}\right) \leq \mathbb{E}\left(e^{tZ}\right) \mathbb{E}\left(e^{tS_{n-1}}\right),$$

where the inequality follows from the assumption that $\mathbb{P}(Z_n = 1 \mid S_{n-1}) \leq \alpha$ and Fubini's theorem. Repeated application of this inequality and the assumption $\mathbb{P}(Z_1 = 1) \leq \alpha$ give that for all $n \geq 1$ and $t > 0$,

$$\mathbb{E}\left(e^{tS_n}\right) \leq \mathbb{E}\left(e^{tZ}\right)^n.$$

Now let $\beta > \alpha$, $n \geq 1$, and $t > 0$. Then, by Markov's inequality,

$$\mathbb{P}(S_n \geq \beta n) = \mathbb{P}\left(e^{tS_n} \geq e^{t\beta n}\right) \leq e^{-t\beta n}\, \mathbb{E}\left(e^{tS_n}\right).$$

Combining the two preceding inequalities gives

$$\mathbb{P}\left(S_n \geq \beta n\right) \leq e^{-t\beta n}\,\mathbb{E}\left(e^{tZ}\right)^n = \mathbb{E}\left(e^{t(Z-\beta)}\right)^n.$$

Since $\mathbb{E}\left(Z-\beta\right) < 0$, we have for sufficiently small $t > 0$ that $\mathbb{E}\left(e^{t(Z-\beta)}\right) < 1$. Hence, $\mathbb{P}\left(S_n \geq \beta n\right)$ decays exponentially in n, and so

$$\sum_{n\geq 1} \mathbb{P}\left(S_n \geq \beta n\right) < \infty.$$

The Borel-Cantelli lemma thus gives that $\limsup_n \frac{S_n}{n} \leq \beta$ almost surely. Since $\beta > \alpha$ was arbitrary, we conclude that $\limsup_n \frac{S_n}{n} \leq \alpha$ almost surely as claimed. $\square$

Putting together Lemma 4, Lemma 7, and Lemma 8, we show that $\bar{X}^{(N,r)}$ is almost surely close to $p^{(r)}$ most of the time for large enough N. More precisely, for $S \subset \Delta^{(N)}$, let

$$\bar{s}_t^{(N,r)}(S) = \frac{1}{t}\int_0^t \chi_S(\bar{X}^{(N,r)}(s))ds$$

be the fraction of time $\bar{X}^{(N,r)}$ spends in a state in S. For $\delta > 0$, we consider the fraction of time spent in $B_\delta(p^{(r)})$. If N is large enough, the limit of $\bar{s}_t^{(N,r)}(B_\delta(p^{(r)}))$ for t to infinity is almost surely close to 1.

Lemma 9. *Let $\delta, \tau > 0$. Then, for every $r > 0$, there is N_0 such that for all $N \geq N_0$ and $p_0 \in \Delta^{(N)}$,*

$$\mathbb{P}\left(\lim_{t\to\infty} \bar{s}_t^{(N,r)}(B_\delta(p^{(r)})) \geq 1-\tau\right) = 1 \tag{19}$$

Proof. Let $r > 0$. By Lemma 5, we can find $\eta > 0$ such that

$$\sup\left\{\left|y^{(r)}(t,p) - p^{(r)}\right| : t \geq 0,\, p \in B_\eta(p^{(r)})\right\} < \frac{\delta}{2}$$

By Lemma 4, we can find $T_1 > 0$ such that for all $T \geq T_1$,

$$\sup\left\{\left|y^{(r)}(T,p) - p^{(r)}\right| : p \in \Delta\right\} < \eta$$

Let $T_0 = \frac{2}{\tau}T_1$. Note that $y^{(r)}$ is time-invariant, that is, $y^{(r)}(t,p) = y^{(r)}(t-t_0, y^{(r)}(t_0,p))$ for all $t \geq t_0 \geq 0$. Combining these facts, it follows that for every $p \in \Delta$, the measure of $t \in [t_0, t_0 + T_0]$ for which $y^{(r)}(t,p)$ is in an $\frac{\delta}{2}$-ball around $p^{(r)}$ is at least $(1 - \frac{\tau}{2})T_0$. We may assume that T_0 is integral.

By Lemma 7, there is $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$,

$$\sup_{p\in\Delta^{(N)}} \mathbb{P}\left(\sup_{0\leq t\leq T_0} \left|y^{(r)}(t,p) - \bar{X}^{(N,r)}(t,p)\right| \geq \frac{\delta}{2}\right) < \frac{\tau}{2}.$$

Now fix $N \geq N_0$ and $p \in \Delta^{(N)}$. We upper bound the fraction of time $\bar{X}^{(N,r)}$ is further than δ away from $p^{(r)}$. To simplify notation, let $t_k = kT_0$ and $\bar{x}_k = \bar{X}^{(N,r)}(t_k,p)$. For $n \geq 1$, we calculate the expected number of intervals $[t_{k-1}, t_k]$, $1 \leq k \leq n$ so that $|\bar{x}(t,p) - p^{(r)}| \geq \delta$ for some $t \in [t_{k-1}, t_k]$. Let $Z_k^{(N,r)}$ be the indicator variable for the event that $\bar{X}^{(N,r)}(t,p)$ and $y^{(r)}(t - t_{k-1}, \bar{x}_{k-1})$ differ by at least $\frac{\delta}{2}$ on the time interval $[t_{k-1}, t_k]$ given that both start at the point $\bar{x}_{k-1}$ at time t_{k-1}. So $Z_k^{(N,r)}$ is 1 if $\sup_{t_{k-1}\leq t\leq t_k} |\bar{X}^{(N,r)}(t,p) - y^{(r)}(t - t_{k-1}, \bar{x}_{k-1})| \geq \frac{\delta}{2}$, and 0 otherwise. Notice that $\{Z_k^{(N,r)} \colon k \in \mathbb{N}_0\}$ satisfies the hypothesis of Lemma 8 with $\alpha = \frac{\tau}{2}$.[20]

If $Z_k^{(N,r)} = 0$, then

$$\begin{aligned}\int_{t_{k-1}}^{t_k} \bar{\chi}_{B_\delta(p^{(r)})}(\bar{X}^{(N,r)}(t,p))dt &\leq \int_{t_{k-1}}^{t_k} \bar{\chi}_{B_{\frac{\delta}{2}}(y^{(r)}(t-t_{k-1},\bar{x}_{k-1}))}(\bar{X}^{(N,r)}(t,p))dt \\ &\quad + \int_{t_{k-1}}^{t_k} \bar{\chi}_{B_{\frac{\delta}{2}}(p^{(r)})} y^{(r)}(t - t_{k-1}, \bar{x}_{k-1})dt \\ &\leq \frac{\tau}{2}T_0.\end{aligned}$$

It follows that

$$\frac{1}{nT_0}\sum_{k\in[n]} \int_{t_{k-1}}^{t_k} \bar{\chi}_{B_\delta(p^{(r)})}(\bar{X}^{(N,r)}(t,p))dt = \sum_{\substack{k\in[n] \\ Z_k^{(N,r)}=0}} \frac{1}{nT_0}\int_{t_{k-1}}^{t_k} \bar{\chi}_{B_\delta(p^{(r)})}(\bar{X}^{(N,r)}(t,p))dt$$

[20] While the probability that $Z_n^{(N,r)}$ equals 1 may depend on $Z_k^{(N,r)}$ for $k < n$, the bound of $\frac{\tau}{2}$ holds independently of the $Z_k^{(N,r)}$ since the bound obtained in Lemma 7 is uniform in the initial state p.

$$+ \sum_{\substack{k\in[n]\\ Z_k^{(N,r)}=1}} \frac{1}{nT_0}\int_{t_{k-1}}^{t_k} \bar{\chi}_{B_\delta(p^{(r)})}(\bar{X}^{(N,r)}(t,p))dt$$

$$\leq \frac{\tau}{2} + \frac{1}{n}\sum_{k=1}^{n} Z_k^{(N,r)}.$$

Applying Lemma 8 to $\{Z_k^{(N,r)} \colon k \in \mathbb{N}_0\}$ with $\alpha = \frac{\tau}{2}$ gives

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Z_k^{(N,r)} \geq \frac{\tau}{2}\right) = 0.$$

Hence, with the preceding inequality we get

$$\mathbb{P}\left(\lim_{t\to\infty} \bar{s}_t^{(N,r)}(B_\delta(p^{(r)}) \geq 1-\tau\right) = \mathbb{P}\left(\lim_{t\to\infty}\int_0^t \chi_{B_\delta(p^{(r)})}(\bar{X}^{(N,r)}) \geq 1-\tau\right) = 1$$

which is (19). $\square$

Theorem 1. *Let $\delta, \tau > 0$. Then, there is $r_0 > 0$ such that for all $0 < r \leq r_0$, there are $p^* \in ML(R)$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and $p_0 \in \Delta^{(N)}$, almost surely*

$$\lim_{n\to\infty} \frac{1}{n}\left|\left\{k \leq n \colon \left|X^{(N,r)}(k,p_0) - p^*\right| \leq \delta\right\}\right| \geq 1-\tau.$$

Moreover, there is $C > 0$ such that for all $n \in \mathbb{N}_0$,

$$\mathbb{P}\left(\left|X^{(N,r)}(n,p_0) - p^*\right| \leq \delta\right) \geq 1 - \tau - e^{-\lfloor Cn \rfloor}.$$

Proof. By Lemma 1, we can choose $r_0 > 0$ so that $p^{(r)} \in B_{\frac{\delta}{2}}(ML(R))$ for all $0 < r \leq r_0$. Let $0 < r \leq r_0$ and $p^* \in ML(R)$ so that $|p^{(r)} - p^*| \leq \frac{\delta}{2}$. Then, applying Lemma 9 to $\frac{\delta}{2}, \tau$, and r, we get $N_0 \in \mathbb{N}$ such that (19) holds (with $\frac{\delta}{2}$ in place of δ). Hence,

$$\mathbb{P}\left(\lim_{t\to\infty} \bar{s}_t^{(N,r)}(B_\delta(p^*)) \geq 1-\tau\right) = 1 \tag{20}$$

This is equivalent to the assertion in the first part of the theorem.

The second statement follows by recalling the standard fact that the distribution of an irreducible and aperiodic Markov chain converges exponentially to its stationary distribution in the total variation norm (Levin et al., 2009, Theorem 4.9). $\square$

References

S. Allesina and J. M. Levine. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences (PNAS)*, 108(14):5638–5642, 2011.

D. Angluin, J. Aspnes, Z. Diamadi, M. J. Fischer, and R. Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18(4): 235–253, 2006.

S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8:121–164, 2012.

K. J. Arrow. *Social Choice and Individual Values*. New Haven: Cowles Foundation, 1st edition, 1951. 2nd edition 1963.

J. Aspnes and E. Ruppert. An introduction to population protocols. In *Middleware for Network Eccentric and Mobile Applications*, chapter 5, pages 97–120. Springer-Verlag, 2009.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

J. P. Bailey and G. Piliouras. Multiplicative weights update in zero-sum games. In *Proceedings of the 19th ACM Conference on Economics and Computation (ACM-EC)*, pages 321–338, 2018.

S. Barberà. Majority and positional voting in a probabilistic framework. *Review of Economic Studies*, 46(2):379–389, 1979.

J. Bartholdi, III, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(3):157–165, 1989.

M. Benaïm and J. Weibull. Deterministic approximation of stochastic evolution in games. *Econometrica*, 71(3):873–903, 2003.

S. J. Brams and P. C. Fishburn. Approval voting. *The American Political Science Review*, 72(3):831–847, 1978.

S. J. Brams and P. C. Fishburn. Voting procedures. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 1, chapter 4. Elsevier, 2002.

F. Brandl and F. Brandt. Arrovian aggregation of convex preferences. *Econometrica*, 88(2):799–844, 2020.

F. Brandl and F. Brandt. A natural adaptive process for collective decision-making. Technical report, https://arxiv.org/abs/2103.14351, 2021.

F. Brandl, F. Brandt, and H. G. Seedig. Consistent probabilistic social choice. *Econometrica*, 84(5):1839–1880, 2016.

F. Brandl, F. Brandt, and C. Stricker. An analytical and experimental comparison of maximal lottery schemes. *Social Choice and Welfare*, 58(1):5–38, 2022.

F. Brandt and F. Fischer. Computing the minimal covering set. *Mathematical Social Sciences*, 56(2):254–268, 2008.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

F. Brandt, P. Lederer, and R. Romen. Relaxed notions of Condorcet-consistency and efficiency for strategyproof social decision schemes. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2022.

G. W. Brown. Iterative solutions of games by fictitious play. In *Activity Analysis of Production and Allocation*, pages 374–376. John Wiley and Sons, Inc., 1951.

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

O. Carbonell-Nicolau and E. A. Ok. Voting over income taxation. *Journal of Economic Theory*, 134(1):249–286, 2007.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons Inc., 2006.

M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

D. S. Felsenthal and M. Machover. After two centuries should Condorcet's voting procedure be implemented? *Behavioral Science*, 37(4):250–274, 1992.

P. C. Fishburn. Probabilistic social choice based on simple voting comparisons. *Review of Economic Studies*, 51(4):683–692, 1984.

D. P. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29(1–2):7–35, 1999.

Y. Freund and R. E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1–2):79–103, 1999.

D. Fudenberg and L. A. Imhof. Monotone imitation dynamics in large populations. *Journal of Economic Theory*, 140:229–245, 2008.

D. Fudenberg and D. M. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, 5:320–367, 1993.

W. V. Gehrlein and D. Lepelley. *Voting Paradoxes and Group Coherence*. Studies in Choice and Welfare. Springer-Verlag, 2011.

A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4): 587–601, 1973.

A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3):665–681, 1977.

A. Goel and D. T. Lee. Large-scale decision-making via small group interactions: The importance of triads. In *Proceedings of the 5th International Workshop on Computational Social Choice (COMSOC)*, 2014.

J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548:210–213, 2017.

S. Hart and A. Mas-Colell. A simple and adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

S. Hart and A. Mas-Colell. *Simple Adaptive Strategies*, volume 4 of *Economic Theory*. World Scientific Publishing Company, 2013.

J. C. Heckelman. Probabilistic Borda rule voting. *Social Choice and Welfare*, 21: 455–468, 2003.

L. N. Hoang. Strategy-proofness of the randomized Condorcet voting system. *Social Choice and Welfare*, 48(3):679–701, 2017.

J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998.

R. A. Holley and T. M. Ligget. Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, 3(4):643–663, 1975.

J. Knebel, M. F. Weber, T. Krüger, and E. Frey. Evolutionary games of condensates in coupled birth–death processes. *Nature Communications*, 6(6977), 2015.

T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970.

G. Laffond, J.-F. Laslier, and M. Le Breton. A theorem on symmetric two-player zero-sum games. *Journal of Economic Theory*, 72(2):426–431, 1997.

B. Laslier and J.-F. Laslier. Reinforcement learning from comparisons: Three alternatives is enough, two is not. *The Annals of Applied Probability*, 27(5):2907–2925, 2017.

J.-F. Laslier. Aggregation of preferences with a variable set of alternatives. *Social Choice and Welfare*, 17(2):269–282, 2000a.

J.-F. Laslier. Interpretation of electoral mixed strategies. *Social Choice and Welfare*, 17:283–292, 2000b.

J.-F. Laslier. And the loser is... plurality voting. In D. Felsenthal and M. Machover, editors, *Electoral Systems*, Studies in Choice and Welfare, pages 327–351. Springer, 2011.

D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

R. B. Myerson. Incentives to cultivate favored minorities under alternative electoral systems. *American Political Science Review*, 87(4):856–869, 1993.

R. L. Rivest and E. Shen. An optimal single-winner preferential voting system based on game theory. In *Proceedings of the 3rd International Workshop on Computational Social Choice (COMSOC)*, pages 399–410, 2010.

J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54(2): 296–301, 1951.

M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

P. Schuster and K. Sigmund. Replicator dynamics. *Journal of Theoretical Biology*, 100(3):533–538, 1983.

A. Slivkins. Introduction to multi-armed bandits. *Foundations and Trends in Machine Learning*, 12(1–2):1–286, 2019.

P. D. Taylor and L. B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1–2):145–156, 1978.

H. P. Young. An axiomatization of Borda's rule. *Journal of Economic Theory*, 9(1): 43–52, 1974.

H. P. Young and A. Levenglick. A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics*, 35(2):285–300, 1978.