The Annals of Probability
1988, Vol. 16, No. 3, 1229–1241

PHASE TRANSITION IN REINFORCED RANDOM WALK AND RWRE ON TREES[1]

BY ROBIN PEMANTLE

Massachusetts Institute of Technology

A random walk on an infinite tree is given a particular kind of positive feedback so edges already traversed are more likely to be traversed in the future. Using exchangeability theory, the process is shown to be equivalent to a random walk in a random environment (RWRE), that is to say, a mixture of Markov chains. Criteria are given to determine whether a RWRE is transient or recurrent. These criteria apply to show that the reinforced random walk can vary from transient to recurrent, depending on the value of an adjustable parameter measuring the strength of the feedback. The value of the parameter at the phase transition is calculated.

1. Introduction. The idea of a reinforced random walk is due to Coppersmith and Diaconis. Imagine a person getting acquainted with a new town. She walks about the area near the hotel somewhat randomly, but tends to traverse the same blocks over and over as they become familiar. To model this, Coppersmith and Diaconis (1987) have defined the following process which they call reinforced random walk. A random walk is taken on the vertices of an undirected graph, beginning at a specified vertex. Initially all the edges are given weight 1, but whenever an edge is traversed the weight of that edge is increased by a fixed parameter Δ. To choose the next move from a particular vertex, an edge leading out from the vertex is chosen, with the probabilities for the various edges being proportional to their weights. So for example, if after one step the walk has reached a vertex with k neighbors, it will return to the starting point on the next step with probability $(1 + \Delta)/(k + \Delta)$. For a more formal description of reinforced random walk, see Coppersmith and Diaconis (1987).

This paper studies the case where the graph is an infinite tree (acyclic graph). The starting point of the investigation is a well known result from exchangeability theory. We apply this to the sequence of edges chosen each time the walk is at a fixed vertex. In Coppersmith and Diaconis (1987) such an approach via de Finetti's theorem is used to analyze a reinforced random walk on finite graphs. For $\Delta = 1$, they calculate the distribution of the random limiting fraction of the time the walk spends on each edge. The calculation involves the homology of the graph. Since the graphs considered in this paper are acyclic, the full force of these results is not needed; we use only the single result known as Pólya's urn.

In Section 3 Pólya's urn is used to construct a random walk in a random environment (RWRE) that is equivalent to the original reinforced random walk. In Sections 4 and 5 we study RWRE using Chernoff's (1952) theory of large

Received August 1986; revised September 1987.

[1]Research supported in part by a grant from the National Science Foundation.

AMS 1980 *subject classifications.* Primary 60J15; secondary 60J80.

Key words and phrases. Reinforced random walk, Pólya urn, mixture of Markov chains, random walk on trees.

deviations. Section 4 contains a sufficient criterion for a.s. transience of a certain class of RWRE. Section 5 gives a sufficient condition (via the stationary measure) for positive recurrence of the RWRE. The only cases remaining unsettled are transitional points where certain equalities hold. We apply these results in Section 6 to the RWRE from Section 3; surprisingly, the calculations for this class of RWRE reduce to a few lines. Thus the recurrence or transience of reinforced random walk on an infinite binary tree can be established except for one value of Δ.

2. Main results. Say the reinforced random walk is recurrent iff the probability of return to the root is 1. At the end of Section 3 it will be clear that the usual equivalences hold: The walk is recurrent iff it returns to the root infinitely often a.s., and it is transient iff it returns to the root finitely often a.s.

The mean recurrence time is always infinite. To see this for $\Delta \geq 1$, let ν_1 be one of k vertices adjacent to the root and let ν_2 be adjacent to ν_1 and distinct from the root. Then the probability of going from ν_1 to ν_2 at least $n + 1$ times before returning to the root is at least

$$\begin{aligned} &(1/k)(1/(2+\Delta))((1+2\Delta)/(2+3\Delta)) \times \cdots \\ &\times((1+2n\Delta)/(2+(2n-1)\Delta)) \\ &\geq 1/k(2+(2n-1)\Delta). \end{aligned} \tag{2.1}$$

The sum of these diverges, so the mean recurrence time is infinite. For $\Delta < 1$ the mean recurrence time is also infinite but we need Lemma 2 to see this.

Nevertheless, we can define a notion of mixed positive recurrence. In Section 3 the walk is decomposed into a mixture of Markov chains, the mixture being necessarily unique in the recurrent case. Call the walk "mixed positive recurrent" whenever the Markov chain is a.s. positive recurrent under the mixing measure.

THEOREM 1. *For a reinforced random walk on an infinite binary tree, there exists* $\Delta_0 \approx 4.29$ *so that*

$$\text{for } \Delta < \Delta_0 \text{ the walk is transient;} \tag{2.2}$$

$$\text{for } \Delta > \Delta_0 \text{ the walk is mixed positive recurrent.} \tag{2.3}$$

It should be noted that the author does not know whether increasing Δ always makes a reinforced random walk more recurrent in any quantitative sense. It seems reasonable to conjecture that the probability of return to the root is monotone in Δ.

In more generality, we can allow the tree itself to be random as long as everything is sufficiently i.i.d. in the following sense. Let each vertex ν have $M(\nu)$ children with $M(\nu)$ i.i.d. and bounded and $\mathbb{E}(M) = \lambda > 1$. So for binary trees $M(\nu) \equiv \lambda = 2$. Let the transition probabilities from ν to its neighbors $\nu_1, \nu_2, \ldots, \nu_M$ be denoted by the vector $\mathbf{p}$ with p_0 being the probability of transition to the parent of ν and the conditional distribution of $\mathbf{p}$ given M being symmetric in coordinates $p_1, \dots, p_M$. Let

$$\phi(\nu) = \frac{\text{prob(transition from parent of } \nu \text{ to } \nu)}{\text{prob(transition from parent of } \nu \text{ to grandparent of } \nu)}. \tag{2.4}$$

So ϕ is the same on the class of all children of the same node and is i.i.d. on such equivalence classes except at the root and any of its children. (As a harmless fiction we sometimes pretend these exceptions do not exist.) Let

$$m(r) = \inf\{\exp(-rt)\mathbb{E}(\phi^t): t \in \mathbb{R}\} \tag{2.5}$$

be the rate function for $\ln(\phi)$ as in (4.7). Assume that $\mathbb{E}(\ln(\phi))$ exists, possibly $\pm\infty$. The following theorem collects all results on RWRE.

THEOREM 2. *Conditional on the tree being infinite*:

(2.6) *If* $\mathbb{E}(\ln(\phi)) \geq 0$ *then the walk is a.s. transient.*

(2.7) *If* $\mathbb{E}(\ln(\phi)) < 0$ *and* $\sup\{\lambda r m(\ln(r)): 0 < r \leq 1\} < 1$ *then the walk is a.s. positive recurrent.*

(2.8) *If* $\mathbb{E}(\ln(\phi)) < 0$ *and* $\sup\{\lambda r m(\ln(r)): 0 < r \leq 1\} > 1$ *then the walk is a.s. transient.*

(2.9) *If* $\mathbb{E}(\phi) < 1/\lambda$ *then the walk is a.s. positive recurrent.*

(2.10) *If* $1 \leq \mathbb{E}(\phi) \leq \infty$ *then the suprema in* (2.7) *and* (2.8) *need only be evaluated at* $r = 1$.

Boundedness of M is not really needed except in (2.7); even here it may be replaced by a weaker condition. (2.9) is always included in (2.7) but is given for ease of calculation.

3. Reduction to RWRE. In this section we study reinforced random walk in order to prove the equivalence in Lemma 2. Our notation for trees is as follows. The set of vertices or nodes is a finite or countable set $\mathbb{T}$. The starting node, or root, is denoted ρ. Every node ν other than ρ is adjacent to a parent par(ν), which is closer to the root, and zero or more children, denoted $c1(\nu)$, $c2(\nu)$, etc. We will write $\nu_1 \leq \nu_2$ for ν_1 an ancestor of ν_2. A branch of length $n \leq \infty$ is a sequence of nodes of length n beginning with ρ where each is the parent of the next. $\mathbb{T}_n$ denotes the set of nodes at distance n from ρ.

Fix a single node ν. It has parent ν_0 and children ν_i for some possibly empty set of i. Edges e_i connect ν to ν_i. When the reinforced random walk first reaches ν, the edge weights must be $1 + \Delta$ for e_0 and 1 for each other edge. If the walk returns later to ν it must do so along the same edge by which it left. So the weight of one edge will increase by 2Δ while the others remain fixed. As long as the walk keeps returning to ν, the weights of e_i increment in this fashion. It is easy to see that the sequence of edges by which the walk leaves ν is an exchangeable sequence stopped at a random time. In fact it is equivalent to Pólya's urn, a version of which was first introduced by Eggenberger and Pólya (1923).

Pólya's urn contains balls of different colors. At each turn a ball is drawn and replaced along with n extra balls of the same color. Of course the probability of choosing a color is just the fraction of balls in the urn of that color. The mathematics still makes sense (although the mechanism does not) if n is allowed to be nonintegral. Letting $n = 2\Delta$ and the initial numbers of each color equal $1 + \Delta$ for color 0 and 1 for each other color, gives the sequence of edges chosen at each visit to ν. The following results can be obtained from Feller (1957, Volume 2, Chapter VII, Section 4) using rational approximations.

LEMMA 1 (Pólya's urn). *Let the urn begin with w_i balls of colour i, $1 \le i \le k$. Then the sequence of draws is distributed as a mixture of sequences of i.i.d. draws with the common probability of choosing color i being a random variable p_i. The vector* $\mathbf{p}$ *ranges over the unit simplex and has the Dirichlet distribution with parameters $w_1/n, \dots, w_k/n$. In particular if $W = w_1 + \cdots + w_k$ then the density of p_i on $(0,1)$ is given by*

$$[\Gamma(W)/\Gamma(w_i/n)\Gamma(W - w_i/n)]\, x^{(w_i/n - 1)}(1 - x)^{(W - w_i/n - 1)}. \tag{3.1}$$

A consequence of the acyclicity of the graph is that the sequence of edges chosen from ν is (except for a random stopping time) independent of what happens on edges not incident to ν. So we can model the reinforced random walk by independent Pólya's urns at each node, making the decisions about where to go from that node. The urns can be replaced in turn, according to the lemma, by random values $p_0(\nu), p_1(\nu), \dots, p_{M(\nu)}(\nu)$ chosen with the specified Dirichlet distribution independently at every node. Conditional upon these choices, the walk is a Markov chain with transition probabilities $\text{prob}(\nu \to \text{par}(\nu)) = p_0(\nu)$ and $\text{prob}(\nu \to ci(\nu)) = p_i(\nu)$ for $1 \le i \le M(\nu)$.

For a binary tree, we give a formal description of this; the formalisms for the general tree are equally routine. Let $\{A(\nu), B(\nu): \nu \in \mathbb{T}\}$ be independent random variables with $A(\rho) = 0$, the density of A for $\nu \neq \rho$ given by

$$[\Gamma((3 + \Delta)/2\Delta)/\Gamma((1 + \Delta)/2\Delta)\Gamma(1/\Delta)]\, x^{(1-\Delta)/2\Delta}(1 - x)^{1/\Delta - 1} \tag{3.2}$$

and the density of B given by

$$[\Gamma(1/\Delta)/\Gamma(1/2\Delta)\Gamma(1/2\Delta)]\, x^{1/2\Delta - 1}(1 - x)^{1/2\Delta - 1}. \tag{3.3}$$

Then the vector $(A(\nu), (1 - A(\nu))B(\nu), (1 - A(\nu))(1 - B(\nu)))$ has the Dirichlet distribution with parameters $(1 + \Delta)/2\Delta, 1/2\Delta, 1/2\Delta$. Let Z_i be i.i.d. uniform on $(0,1)$ for $i = 1, 2, \dots,$ and let the random variable

$$\prod_{\nu \in \mathbb{T}} A(\nu) \times \prod_{\nu \in \mathbb{T}} B(\nu) \times \prod_{i \in \mathbb{N}} Z_i$$

be defined on some space Ω. For each $\omega \in \Omega$, generate a sequence of nodes recursively by $\nu_1(\omega) = \rho$ and

$$\begin{aligned} \nu_{n+1}(\omega) &= \text{par}(\nu_n(\omega)), && \text{if } Z_n(\omega) < A(\nu_n(\omega)), \\ &= c1(\nu_n(\omega)), && \text{if } A(\nu_n(\omega)) \le Z_n(\omega) \\ & && \quad < A(\nu_n(\omega)) + (1 - A(\nu_n(\omega)))B(\nu_n(\omega)), \\ &= c2(\nu_n(\omega)), && \text{otherwise.} \end{aligned}$$

The following equivalence should now be clear.

LEMMA 2. *The distribution of the random sequence $\nu_1, \nu_2, \dots$ is the same as the distribution of sequences of nodes visited by a reinforced random walk.*

The process $\omega \to (\nu_1(\omega), \nu_2(\omega), \dots)$ is our RWRE. $A(\nu, \omega)$ and $B(\nu, \omega)$ are the random environment; conditional upon their values for all $\nu \in \mathbb{T}$, the Z_i determine a Markov random walk. (Unfortunately the distribution function for A does not vary pointwise monotonically with respect to Δ so it is difficult to compare the RWRE's for different values of Δ.)

We can now prove that the mean recurrence time is infinite even for $\Delta < 1$. Let R be the mean recurrence time given the values of A and B at all nodes so that the mean recurrence time is $\mathbb{E}(R)$. By looking at the subtree below the first node visited we get the equation

$$\mathbb{E}(R) = 2 + \mathbb{E}((1 - A)/A)\mathbb{E}(R).$$

But for $0 \le \Delta < 1$, we get $1 \le \mathbb{E}((1 - A)/A) < \infty$ by a calculation similar to (6.2), so $\mathbb{E}(R)$ must be infinite.

Altering the values of A and B for finitely many ν will not affect whether an environment is recurrent. So recurrence is a tail property of the i.i.d. pairs $(A(\nu), B(\nu))$ given the number of children of each node. It follows from the 0–1 law for tails that, in the case of binary trees, the environment must be a.s. transient or a.s. recurrent. In other words, the mixing measure for the RWRE does not mix recurrence and transience, so the claim at the beginning of Section 2 is established. For general trees one must first condition on the tree being infinite. We use an argument due to Harry Kesten (personal communication). The process is recurrent iff the process restricted to each subtree with root ν for $\nu \in \mathbb{T}_1$ is recurrent. So the recurrence probability is a fixed point of the offspring generating function. If it is less than 1 it is bounded by the extinction probability and hence equal to the extinction probability. So conditional upon nonextinction, the recurrence probability is either 1 or 0.

4. Transience of RWRE. In this section and the next we discuss a general RWRE on a binary tree with transition probabilities i.i.d. as described in Section 2. In particular, if we let

$$C(c1(\nu)) = (1 - A(\nu))B(\nu), \tag{4.1}$$

$$C(c2(\nu)) = (1 - A(\nu))(1 - B(\nu)), \tag{4.2}$$

$$\phi(\nu) = C(\nu)/A(\mathrm{par}(\nu)), \tag{4.3}$$

then this agrees with the definition of ϕ in (2.4). The main results of this section are the transience criteria for RWRE, (2.6) and (2.8). We restate them here.

THEOREM 3. *Suppose that for some $r \in (0, 1]$*

$$\lambda m(\ln(r)) > 1. \tag{4.4}$$

Then the RWRE is transient.

The proof breaks into three pieces: Lemma 3, which is a transience criterion for a single environment, Chernoff's identification of the rate function for large deviations and a lemma on branching processes, providing the hypotheses for Lemma 3.

LEMMA 3. *Let* $k \in \mathbb{N}$, $M \in \mathbb{R}^+$, $r \in (0,1]$ *and* $\delta > 0$ *be fixed constants. Suppose a nonempty set of nodes* $S \in \mathbb{T}$ *can be found such that if* S_i *denotes* $S \cap \mathbb{T}_{ik}$, *the nodes of* S *at distance* ik *from the root. Then*

$$[\nu \in S \text{ and } \nu_0 < \nu] \Rightarrow \nu \in S; \tag{4.5}$$

$$\nu_0 \in S_i \Rightarrow \operatorname{card}\{\nu \in S_{i+1}: \nu_0 < \nu\} \geq r^k; \tag{4.6}$$

$$\begin{gathered} \text{for each branch segment } \nu_0 < \nu_1 < \cdots < \nu_k \\ \text{with } \nu_0 \in S_i \text{ and } \nu_k \in S_{i+1}, \\ \sum_{1 \leq i \leq k} \ln(\phi(\nu_i)) \geq k \ln(r) + \delta; \end{gathered} \tag{4.7}$$

$$\phi(\nu)^{-1} \leq M \quad \text{for all } \nu \in S. \tag{4.8}$$

Then the environment is transient.

To see the intuition behind this lemma, suppose $r = 1$. Then (4.5) and (4.6) say that S contains at least one infinite branch. By (4.7) and (4.8), the lim inf average of $\ln(\phi)$ along initial segments of any branch in S is at least δ/k. For any such branch, the Markov chain gotten by considering only moves along that branch will be transient; this is because the sequence $\varepsilon_i = \prod_{\nu < \nu_i} \phi(\nu)^{-1}$ is summable, which is the standard test for transience in the one-dimensional case. But any environment containing a transient subtree is transient either because it wanders to infinity on the subtree or because it fails to return to the subtree infinitely often.

PROOF OF LEMMA 3. We find the appropriate martingale to generalize the standard test to the case $r < 1$. Define a function $s: \mathbb{T} \setminus \{\rho\} \to [0,1]$ by $s(\nu) = 0$ for $\nu \notin S$ and by

$$s(\nu) = \operatorname{card}\{\nu' \in S_{i+1}: \nu \leq \nu'\} / \operatorname{card}\{\nu' \in S_{i+1}: \operatorname{par}(\nu) \leq \nu'\}$$

for $\nu \in \mathbb{T}_j$ with $ik < j \leq (i+1)k$. Clearly, the sum over i of $s(ci(\nu))$ is 1 for any $\nu \in S$. See Figure 1 for an example of the function s. Now define $t: \mathbb{T} \to \mathbb{R}^+$ by $t(\rho) = 1$ and for $\nu \neq \rho$ by $t(\nu) = (s(\nu)\phi(\nu)^{-1} t(\operatorname{par}(\nu))$. Define $u: \mathbb{T} \to \mathbb{R}^+$ by $u(\nu) = \sum_{\nu' \leq \nu} t(\nu')$.

I claim that $u(\nu_i)$ is a bounded martingale for any $\nu_1 \neq \rho$, where $\nu_1, \nu_2, \ldots$ is a random walk on the given environment, stopped if it reaches ρ. To see it is a martingale, just calculate

$$\begin{aligned} \mathbb{E}(u(\nu_{i+1}) | \nu_i) &= A(\nu_i) u(\operatorname{par}(\nu_i)) + \sum_j C(cj(\nu_i)) u(cj(\nu_i)) \\ &= u(\nu_i) + A(\nu_i) \left[-t(\nu_i) + \sum_j \phi(cj(\nu_i)) t(cj(\nu_i)) \right] \\ &= u(\nu_i) + A(\nu_i)/t(\nu_i) \left[-1 + \sum_j s(cj(\nu_i)) \right] = u(\nu_i) \end{aligned} \tag{4.9}$$

for $\nu_i \in S$. For $\nu_i \notin S$ the result is true because $t(\nu_i) = 0$.

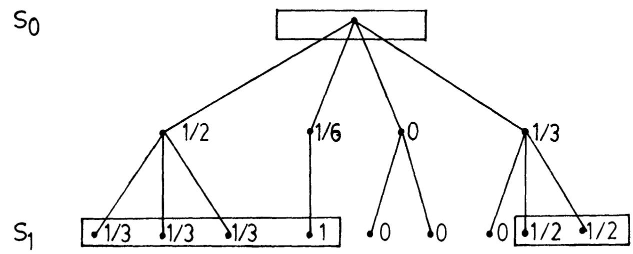

FIG. 1.

For boundedness, first consider the case $\nu \in S_i$. Find $\nu_0 \in S_{i-1}$ with $\nu_0 < \nu$. Then $\prod_{\nu_0 < \nu' \le \nu} s(\nu')$ is a telescoping product and is at most r^k by (4.6). Also $\prod_{\nu_0 \le \nu' \le \nu} \phi(\nu)^{-1} \le e^{-\delta} r^{-k}$ by (4.7). Then $t(\nu) \le t(\nu_0)e^{-\delta}$ and by induction $t(\nu) \le e^{-i\delta}$. Now for any $\nu \in S \cap \mathbb{T}_n$, apply (4.8) to see that $t(\nu)$ decreases at least geometrically in n. Therefore u is bounded on S. But for $\nu \notin S$, $u(\nu) = u(\text{par}(\nu))$ so u is bounded on all of $\mathbb{T}$.

We conclude by the bounded martingale theorem that $u(\nu_i)$ converges a.s. to a limit u with $\mathbb{E}(u) = \mathbb{E}(u(\nu_1)) > 1$. Then $\text{prob}(u = 1) < 1$ so the walk stays away from the root with nonzero probability. □

The case (2.6) is easily disposed of using Lemma 3 and the strong law of large numbers, so we assume for the remainder of this section that $\mathbb{E}(\ln(\phi)) < 0$. At this point we require Chernoff's estimates for the probabilities of large deviations.

THEOREM 4 [Chernoff (1952), Theorem 1 and Lemma 6]. *Let $S_n = X_1 + \cdots + X_n$, where the X_i are i.i.d. with common distribution function F. Define*

$$m(r, F) = \inf\{\exp(-rt)\mathbb{E}(\exp(tX_1)) : t \in \mathbb{R}\}. \tag{4.10}$$

Assume $r > \mathbb{E}(X_1) \ge -\infty$. Then

$$\text{prob}(S_n \ge nr) \le m(r, F)^n \tag{4.11}$$

and

$$\lim_{n\to\infty} m_1^{-n}\, \text{prob}(S_n \ge nr) = \infty \quad \text{for any } m_1 < m(r, F). \tag{4.12}$$

Furthermore, $m(r, F)$ is continuous in r and strictly decreasing between $C = \mathbb{E}(X_1)$ and $D = \text{essential sup}(X_1)$ with $m(C, F) = 1$ and $m(D, F) = \text{prob}(X_1) = D$.

We will apply this with $F = \Phi$, the distribution function for $\ln(\phi)$. With $m(r)$ denoting $m(r, \Phi)$, the notations in (4.10) and (2.5) agree. Roughly speaking, (4.12) tell us that under the condition (4.4), there are enough branches on which $\ln(\phi)$ averages more than $\ln(r)$ to make (4.5)–(4.8) possible. To make this into a proof we need some facts about branching processes.

The branching processes we will consider begin with a single ancestor. Each individual bears a random number of children which is i.i.d. and equals i with probability p_i. Let $f(x)$ be the generating function for the p_i's, with $f'(1) = M < \infty$, and assume $M > 1$ so the process is finite with some probability $b < 1$.

Lemma 4. *Pick any $K > 0$, $M_1 < M$. Then*

$$\lim_{n\to\infty} \text{prob}(\textit{size of the nth generation} < KM_1^n) = b.$$

Proof. See Harris [(1963), Theorem 8.1 and Remark 1, pages 13–14]. □

Say a branching process is d-infinite for $d \in \mathbb{N}$ if there is some nonempty subset of individuals such that each individual in the subset has at least d children in the subset. Say that a given individual has a d, n-subtree if $n = 0$ or the individual has at least d children each of whom has a $d, n-1$-subtree. Suppose that B is a branching process with generating function f and C is the process with generating function $f(r + (1-r)x)$, being identical to B except that births are aborted with probability r.

Lemma 5. *Suppose that for the process C, the probability of an individual having at least d children is at least $1 - r$. Then the process B is d-infinite with probability at least $1 - r$.*

Proof. We show by induction that the probability of any individual having a d, n-subtree is at least $1 - r$. The case $n = 0$ is trivial. Now assume it is true for some arbitrary n. Then the probability of an individual having a $d, n+1$-subtree is just the probability of having at least d children, provided that the ones who will not have a d, n-subtree are aborted. By the induction hypothesis, children will be aborted with probability at most r, so by the hypothesis of the lemma, the probability of having a $d, n+1$-subtree is at least $1 - r$.

Thus the probability of the initial ancestor having a d, n-subtree for all n is at least $1 - r$. Since each node has only finitely many children, the process will be d-infinite in these cases. □

For any branching process, B, let $B^{(k)}$ denote the process whose nth generation is the nkth generation of B, with the relation of parenthood in $B^{(k)}$ corresponding to ancestry in B.

Lemma 6. *For a branching process B, let f, b, M and M_1 be as in Lemma 4. Then there is some $k \in \mathbb{N}$ such that*

$$\text{prob}(B^{(k)} \textit{ is } \lfloor M_1^k \rfloor\textit{-infinite}) \geq (1-b)/2. \tag{4.13}$$

Proof. By Lemma 4 we can pick N large enough so that for all $i \geq N$,

$$\text{prob}(\text{size of } i\text{th generation of } B > 4M_1^i/(1-b)) > 3(1-b)/4. \tag{4.14}$$

By increasing N if necessary we can also assume that the following holds: Given a population of size at least $4M_1^N/(1-b)$, each member of which is killed independently with probability $(1+b)/2$,

$$\text{prob}\big(\text{at least } M_1^N \text{ of them survive}\big) > 3(1-b)/4. \tag{4.15}$$

Now let $k = N$ and apply Lemma 5 to $B^{(k)}$ with the probability of abortion equal to $(1+b)/2$. Then

$$\begin{aligned}
&\text{prob}\big(\text{having at least } M_1^N \text{ children in } B^{(N)} \text{ with abortion}\big)\\
&\quad \geq 1 - \text{prob}\big(\text{fewer than } 4M_1^N/(1-b) \text{ children in } B^{(N)} \text{ without abortion}\big)\\
&\qquad - \text{prob}\big(\text{from at least } 4M_1^N/(1-b) \text{ conceptions fewer than } M_1^N \text{ are born}\big)\\
&\quad \geq 1 - (1-b)/4 - (1-b)/4 = (1-b)/2 \quad \text{by (4.14) and (4.15)}.
\end{aligned}$$

So (4.13) follows from Lemma 5. □

Proof of (2.8). Fix r with $\lambda r m(\ln(r)) > 1$. Fit in a few more constants:

$$m(\ln(r)) = (1+\delta_1)/\lambda r > (1+\delta_2)\lambda r > (1+\delta_3)/\lambda r > 1/\lambda r.$$

Apply (4.12) of Chernoff's theorem with $m_1 = (1+\delta_2)/\lambda r < m(\ln(r))$. Then for N sufficiently large and δ_0 sufficiently small,

$$\mathbb{E}\left(\text{card}\left\{\nu \in \mathbb{T}_N \colon \sum_{\nu' < \nu} \ln(\phi(\nu')) > N\ln(r) + \delta_0\right\}\right) > \lambda\big((1+\delta_2)/r\big)^N. \tag{4.16}$$

We can now pick M sufficiently large to amend this to

$$\begin{aligned}
&\mathbb{E}\left(\text{card}\left\{\nu \in \mathbb{T}_N \colon \sum_{\nu' < \nu} \ln(\phi(\nu')) > N\ln(r) + \delta_0\right.\right.\\
&\qquad\qquad \left.\left. \text{and } \phi(\nu')^{-1} < M \text{ for all } \nu \leq \nu\right\}\right)\\
&\quad > \lambda\big((1+\delta_2)/r\big)^N.
\end{aligned} \tag{4.17}$$

Now define a branching process B with ρ as its initial ancestor, whose individuals are elements of $\mathbb{T}_0, \mathbb{T}_N, \mathbb{T}_{2N}, \dots$ such that $\nu_0 \in \mathbb{T}_{iN}$ has $\nu \in \mathbb{T}_{(1+1)N}$ as a child iff $\nu_0 < \nu$ and $\sum_{\nu_0 \leq \nu' < \nu} \ln(\phi(\nu')) \geq N\ln(r) + \delta_0$ and $\phi(\nu')^{-1} < M$ for all $\nu_0 \leq \nu' < \nu$ and ν is the first child of $\text{par}(\nu)$ that qualifies under these conditions. By Lemma 6 there is a j such that $B^{(j)}$ is $\lfloor((1+\delta_3)/r)^{jN}\rfloor$-infinite with nonzero probability. In fact j can be chosen large enough so that the expression in greatest-integer brackets is at least $(1/r)^{jN}$. Now the criterion given by Lemma 3 applies with $k = jN$ to show that the probability of transience is nonzero. By the reasoning in Section 3, this means the probability of transience, given an infinite tree, is 1. □

5. Recurrence of RWRE. The main result in this section is a proof of (2.7).

(2.7) Suppose $\mathbb{E}(\ln(\phi)) < 0$ and $\sup\{\lambda r m(\ln(r)) \colon r \in (0,1]\} < 1$. Then the RWRE is a.s. positive recurrent.

To prove (2.7) we calculate a stationary distribution. Sufficient conditions for a measure μ to be stationary are that for every ν, i,

$$\mu(\nu)C(ci(\nu)) = \mu(ci(\nu))A(ci(\nu)). \tag{5.1}$$

If we let $\mu(\rho) = 1$ and for $\nu \neq \rho$ let

$$\mu(\nu) = A(\nu)^{-1} \prod_{\rho < \nu' \le \nu} \phi(\nu'), \tag{5.2}$$

then μ is stationary, satisfying (5.1). If $\mu(\mathbb{T}) < \infty$ then the walk is positive recurrent.

The statement (2.9) follows immediately, since in this case $\mathbb{E}(\mu(\mathbb{T}))$ is finite so $\mu(\mathbb{T})$ is a.s. finite.

Roughly speaking, the reason μ is finite under the hypotheses of (2.7) is that there are fewer than r^{-n} nodes of measure r^n for each $r < 1$. This must be formulated precisely and then integrated over $r \in [0,1]$. The methods are elementary, though in the case of Lemma 8 a more elegant argument ought to be possible.

Let $f(\nu) = A(\nu)\mu(\nu)$. Note that $f(\nu)$ depends only on transition probabilities of nodes strictly above ν.

Lemma 7. *Fix any $k \in (1,\infty)$ and $r \in (0,1]$. Assume $\mathbb{E}(\ln(\phi)) < \ln(r)$. Let $J_n = \{\nu \in \mathbb{T}_n\colon f(\nu) \ge r^{-n}\}$. Then*

$$\operatorname{prob}\big(\operatorname{card}(J_n) \ge (\lambda k m(\ln(r)))^n \ \textit{for infinitely many } n\big) = 0.$$

Proof. For each $\nu \in \mathbb{T}_n$, (4.11) gives $\operatorname{prob}(f(\nu) \ge r^n) \le m(\ln(r))^n$. So

$$\mathbb{E}(\operatorname{card}(J_n)) \le (\lambda m(\ln(r)))^n \tag{5.3}$$

and so

$$\mathbb{E}\big(\Sigma\operatorname{card}(J_n)/(\lambda k m(\ln(r)))^n\big) \text{ is finite.} \tag{5.4}$$

In the event that $\operatorname{card}(J_n) > (\lambda k m(\ln(r)))^n$ infinitely often, the sum in (5.4) would be infinite; the event therefore has probability 0. □

Lemma 8. *Lemma 7 holds with μ in place of f.*

Proof. Let $G_n = \{\nu \in \mathbb{T}_n\colon \mu(\nu) \ge r^n\}$. Suppose to the contrary that for some $a > 0$,

$$\operatorname{prob}\big(\operatorname{card}(G_n) \ge (\lambda k m(\ln(r)))^n \text{ infinitely often}\big) = a. \tag{5.5}$$

By continuity of m we can choose r_1 and k_1 so that $r > r_1 > 0$, $k > k_1 > 1$, $m(\ln(r_1)) < 1$ and $k_1 m(\ln(r_1)) \ge km(\ln(r))$. Then (5.5) holds with r_1 and k_1 in place of r and k. Pick k_2 so that $k_1 > k_2 > 1$ and $k_2 m(\ln(r_1)) = b < 1$ for some b. By Lemma 7 with r_1 and k_2 in place of r and k, we can pick N_0 large enough so that

$$\text{prob}\Big(\text{card}\{\nu \in \mathbb{T}_n: f(\nu) \geq r_1^n\} \geq \big(\lambda k_2 m(\ln(r_1))\big)^n \text{ for some } n \geq N_0\Big) < a/2. \tag{5.6}$$

By picking a larger N we can assume that $\text{prob}(C(\nu) \leq (r_1/r)^n) < \delta/L$ for any fixed δ, where L is a bound for $M(\nu)$. [This is the only place that the boundedness of M is used. Any weaker condition still implying the truth of this lemma can be substituted in (2.7).] We fix δ small enough so that

(5.7) for any $n \geq N_0$ and any collection of individuals killed independently with probability δ, the probability of the fraction of survivors being at least b is greater than $1 - a/2$.

Now if the event in (5.5) occurs, pick the first $n \geq N$ for which $\text{card}(G_n) \geq (\lambda k_1 m(\ln(r_1)))^n$. For any $\nu_0 \in G_n$ and any child ν of ν_0, $f(\nu) > r^{n+1}$ unless $C(\nu) \leq r_1^{n+1}/r^n < (r_1/r)^n$. Thus the event in (5.6) will hold for $n + 1$ unless

$$\big(\lambda k_1 m(\ln(r_1))\big)^n \frac{\text{card}\{\nu: \text{par}(\nu) \in G_n \text{ and } C(\nu) \leq n(r1/r)\}}{\text{card}\{\nu: \text{par}(\nu) \in G_n\}} \leq \big(\lambda k_2 m(\ln(r_1))\big)^{n+1}. \tag{5.8}$$

But $(\lambda k_2 m(\ln(r_1)))^{n+1} < \lambda b(\lambda k_1 m(\ln(r_1)))^n$, so if (5.8) holds then

$$\text{card}\Big(\big\{\nu: \text{par}(\nu) \in G_n \text{ and } C(\nu) \geq (r_1/r)^n\big\}\Big)/\lambda\, \text{card}(G_n) \leq b. \tag{5.9}$$

While the event $\text{par}(\nu) \in G_n$ is not independent of $C(\nu)$, it is easy to see that $\text{prob}(C(\nu) < x)$ can only decrease when conditioned on $\text{par}(\nu) \in G_n$. [Given the values of $A(\nu')$ and $B(\nu')$ for $\rho \leq \nu' < \text{par}(\nu)$, the indicator function of the event $\text{par}(\nu) \in G_n$ is a decreasing function of $A(\text{par}(\nu))$. Since $A(\text{par}(\nu))$ is independent of this σ-field, $\text{prob}(A(\text{par}(\nu)) < x)$ increases for any x when conditioned on $\text{par}(\nu) \in G_n$. Then $\text{prob}(C(\nu) < x)$ must decrease since $C(\nu) = (1 - A(\text{par}(\nu)))Y$ where Y is independent from all the preceding variables.] If we think of a node in G_n as being killed if the value of C at any of its children is less than $(r_1/r)^n$, then we can apply (5.7) to show that the probability of (5.9) is less than $a/2$. Thus the event in (5.5), having probability at least a, entails the disjunction of events in (5.6) and (5.9), each having probability less than $a/2$, a contradiction. □

To finish proving (2.7) we use a compactness argument to estimate $\mu(\mathbb{T}_n)$. Let $\sup\{\lambda r m(\ln(r)): r < 1\} = 1 - \delta_1$ and pick δ_2 and δ_3 with $1 - \delta_1 < 1 - \delta_2 < 1 - \delta_3 < 1$. Let $\mathscr{I}$ be the collection of intervals $\{(g(x), x): x \in (0,1)\} \cup \{(g(1), 1]\}$ where g is any function such that

$$g(x) < x \quad \text{and} \quad \lambda x m(\ln(g(x))) < 1 - \delta_3 \quad \text{for } x \in (0,1]. \tag{5.10}$$

These cover $(0,1]$ so, by compactness, pick a finite subcover J of $\mathscr{I}$. Let $(a_1, b_1), \ldots, (a_k, b_k), (a_{k+1}, 1]$ be the elements of J written in ascending order of a_i. Since $\mathbb{E}(\ln(\phi)) < 0$ and $m(\mathbb{E}(\ln(\phi))) = 1$, we can assume without loss of generality that $\mathbb{E}(\ln(\phi)) < \ln(a_1) < \ln(1/\lambda)$. Apply Lemma 8 to each $(a, b) \in J$ with $r = a$ and $k = (1 - \delta_3)/(1 - \delta_2)$. Then there is a.s. some N such that for all $(a, b) \in J$ (including $a = 1$),

$$\begin{aligned} &\operatorname{card}\left(\{\nu \in \mathbb{T}_n : \mu(\nu) \geq a^n\}\right) \\ &\quad < \left[\lambda((1 - \delta_3)/(1 - \delta_2))m(\ln(a))\right]^n \quad \text{for } n \geq N. \end{aligned} \tag{5.11}$$

Then letting $a_0 = 0$, $b_0 = a_1$, $a_{k+1} = 1$ and assuming $n \geq N$ we get that

$$\begin{aligned} \mu(\mathbb{T}_n) &\leq \sum_{0 \leq i \leq k+1} a_{i+1}^n \operatorname{card}\{\nu \in \mathbb{T}_n : \mu(\nu) \geq a_i\} \\ &\leq \lambda^n a_1^n + \sum_{1 \leq i \leq k} b_i^n \left[\lambda m(\ln(a_i))(1 - \delta_3)(1 - \delta_2)\right]^n \quad \text{[by (5.11)]} \\ &\leq (k+1)(1 - \delta_3)^n \quad \text{[by (5.10)]}, \end{aligned}$$

so $\mu(\mathbb{T})$ is finite.

6. Proof of Theorem 1 and further questions.

PROOF OF THEOREM 1. Applying (2.7) and (2.8) to equations (3.2) and (3.3), it remains only to calculate Δ_0. We first establish (2.10). Suppose $1 \leq \mathbb{E}(\ln(\phi)) \leq \infty$ and also that $m(0) < 1/\lambda$. Then for any $r \in (0, 1]$, (2.5) gives us

$$\lambda r m(\ln(r)) = \inf\{\lambda r^{1-t} \mathbb{E}(\phi^t) : t \in \mathbb{R}\}. \tag{6.1}$$

By assumption this is less than 1 for $r = 1$. The infimum for $r = 1$ must occur at some $t \leq 1$ since $\mathbb{E}(\phi) \geq 1$ implies $\mathbb{E}(\phi^t) > (\mathbb{E}(\phi))^t > \mathbb{E}(\phi)$ for $t \geq 1$ by Jensen's inequality. But for positive r and $t \leq 1$, (6.1) is increasing in r, so the supremum of (6.1) over $r \in (0, 1]$ must be less than 1, establishing (2.10).

Now we calculate $m(0)$ as a function of Δ. Unravelling the definitions gives

$$\begin{aligned} m(0) &= \inf\{\mathbb{E}[B(1 - A)/A]^t : t \in \mathbb{R}\} \\ &= \inf_t \left[\Gamma((3 + \Delta)/2\Delta)\Gamma(1/2\Delta)/\Gamma((1 + \Delta)/2\Delta)\Gamma(1/\Delta)\Gamma(1/2\Delta)\Gamma(1/2\Delta)\right] \\ &\quad \times \int\int [y(1 - x)/x]^t x^{(1-\Delta)/2\Delta}(1 - x)^{1/\Delta - 1} y^{1/2\Delta - 1}(1 - y)^{1/2\Delta - 1}\, dx\, dy \\ &= \inf_t \Gamma((1 + \Delta)/2\Delta - t)\Gamma(1/2\Delta + t)/\Gamma((1 + \Delta)/2\Delta)\Gamma(1/2\Delta). \end{aligned} \tag{6.2}$$

Since $\log(\Gamma)$ is concave, the minimum is reached when $t = \frac{1}{4}$, making the two factors in the numerator equal. For $\Delta \geq \Delta_0 \approx 4.29$, the expression (6.2) is less than $1/2$. Using (6.2) and (6.3) with $t = 1$ gives $\mathbb{E}(\phi) = 1/(1 + \Delta)$ for $\Delta > 1$ and $\mathbb{E}(\phi) = \infty$ for $\Delta \leq 1$. Also it is easy to see that $\mathbb{E}(\ln(\phi)) < 0$ for $\Delta > 1$ since for $\Delta = 1$ the distributions of A and $1 - A$ are identical. Then the conditions of (2.10) and (2.8) are satisfied for $\Delta > \Delta_0$, and (2.7) applies for $\Delta < \Delta_0$. □

Questions of reinforced random walk on other graphs are still wide open. Diaconis originally asked me about the d-dimensional integer lattice $\mathbb{Z}^d$. I believe it is not even known whether there is a $\Delta > 0$ for which the reinforced random walk on $\mathbb{Z}^2$ is recurrent!

Acknowledgments. I wish to thank the editors of this journal and especially Richard Durrett for their extremely careful criticisms and suggestions. In particular, my investigation of the stationary measure was at his suggestion. Thanks are also due Persi Diaconis for encouragement with the early stages of this work and to Kiwi hospitality which made it possible for me to continue discussing and revising this paper while on vacation in New Zealand.

REFERENCES

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

COPPERSMITH, D. and DIACONIS, P. (1987). Random walk with reinforcement. Unpublished.

EGGENBERGER, F. and PÓLYA, G. (1923). Über die Statistik verketter Vorgänge. *Z. Angew. Math. Mech.* **3** 279–289.

FELLER, W. (1957, 1966). *An Introduction to Probability Theory and Its Applications* **1**, **2**. Wiley, New York.

HARRIS, T. (1963). *Theory of Branching Processes.* Springer, Berlin.

DEPARTMENT OF MATHEMATICS
ROOM 2-342
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139