The Annals of Applied Probability
1992, Vol. 2, No. 1, 229–261

THE $3x + 1$ PROBLEM: TWO STOCHASTIC MODELS

BY J. C. LAGARIAS AND A. WEISS

AT & T Bell Laboratories

The $3x + 1$ problem concerns the behavior under iteration of the function $T: \mathbb{Z}^+ \to \mathbb{Z}^+$ defined by $T(n) = n/2$ if n is even and $T(n) = (3n + 1)/2$ if n is odd. The $3x + 1$ conjecture asserts that for each $n \geq 1$ some k exists with $T^{(k)}(n) = 1$; let $\sigma_\infty(n)$ equal the minimal such k if one exists and $+\infty$ otherwise. The behavior of $\sigma_\infty(n)$ is irregular and seems to defy simple description. This paper describes two kinds of stochastic models that mimic some of its features. The first is a random walk that imitates the behavior of $T \pmod{2^j}$; the second is a family of branching random walks that imitate the behavior of $T^{-1} \pmod{3^j}$. For these models we prove analogues of the conjecture that $\limsup_{n \to \infty}(\sigma_\infty(n)/\log(n)) = \gamma$ for a finite constant γ. Both models produce the same constant $\gamma_0 \doteq 41.677647$. Predictions of the stochastic models agree with empirical data for the $3x + 1$ problem up to 10^{11}. The paper also studies how many n have $\sigma_\infty(n) = k$ as $k \to \infty$ and estimates how fast $t(n) = \max(T^{(k)}(n)\colon k \geq 0)$ grows as $n \to \infty$.

1. Introduction. The $3x + 1$ problem concerns the behavior of iterates of the $3x + 1$ *function*

$$T(n) := \begin{cases} \dfrac{n}{2}, & \text{if } n \equiv 0 \pmod 2, \\ \dfrac{3n+1}{2}, & \text{if } n \equiv 1 \pmod 2, \end{cases} \tag{1.1}$$

defined on the positive integers $\mathbb{Z}^+$. The $3x + 1$ *conjecture* asserts that for each $n \in \mathbb{Z}^+$ there exists some iterate k such that $T^{(k)}(n) = 1$. This conjecture derives its appeal from the contrast between the simple form of the $3x + 1$ function and the apparently complicated behavior of the *trajectories* $\{T^{(k)}(n)\colon k = 0, 1, 2, \dots\}$ for different integers n. At present the $3x + 1$ conjecture appears to be an intractably difficult problem; Lagarias (1985) surveys known results. It has been numerically verified for all $n \leq 2 \cdot 10^{12}$ and for many larger numbers.

The *total stopping time* $\sigma_\infty(n)$ of $n \in \mathbb{Z}^+$ is the least $k \geq 0$ such that $T^{(k)}(n) = 1$ if such k exists and $\sigma_\infty(n) = \infty$ otherwise. Define also the *maximum excursion* $t(n) = \max_{k \geq 0}(T^{(k)}(n))$, where $t(n) = \infty$ if $T^{(k)}(n)$ is unbounded. This paper is motivated by the questions:

1. How fast does $\sigma_\infty(n)$ grow as $n \to \infty$, in the worst case?
2. How many n have $\sigma_\infty(n) = k$ as $k \to \infty$?
3. How fast does $t(n)$ grow as a function of n, as $n \to \infty$?

Received October 1990; revised June 1991.
AMS 1980 *subject classifications*. Primary 11A99, 60F10; secondary 26A18, 60J85.
Key words and phrases. $3x + 1$, large deviations, branching random walk.

These questions can be made more precise. Heuristic arguments suggest that $\sigma_\infty(n)$ should be of order $O(\log n)$ in some "average" sense (see Section 2). The first question is reformulated as the problem:

1$'$. Determine the $3x + 1$ *stopping constant*

$$\gamma := \limsup_{n\to\infty}\left(\frac{\sigma_\infty(n)}{\log n} : \sigma_\infty(n) < \infty\right). \tag{1.2}$$

This problem seems to be at least as hard as settling the $3x + 1$ conjecture. It is, however, logically possible that the $3x + 1$ conjecture is true and yet $\gamma = \infty$. To reformulate the second question, define $N_k = \#\{n: \sigma_\infty(n) = k\}$.

2$'$. Determine the $3x + 1$ *growth constant*

$$\delta := \limsup_{k\to\infty}\frac{\log N_k}{k}. \tag{1.3}$$

The third question is reformulated as:

3$'$. Determine the $3x + 1$ *maximum excursion constant*

$$\rho := \limsup_{n\to\infty}\frac{\log t(n)}{\log n}.$$

A necessary condition for ρ to be finite is that there are no *divergent trajectories*, that is, $t(n) = \infty$ never occurs.

One can easily obtain lower bounds for γ and ρ and both upper and lower bounds for δ. Since $T(n) \ge n/2$ it is immediate that $\sigma_\infty(n) \ge \log n/\log 2$, and using the oft-discovered fact that $T^{(k)}(2^k - 1) = 3^k - 1$, one obtains

$$\gamma \ge \frac{\log 2 + \log 3}{(\log 2)^2} \doteq 3.72931. \tag{1.4}$$

This seems to be the strongest rigorously proved result about the constant γ; it is much weaker than the conjectured "average case" size of $(\sigma_\infty(n)/\log n)$, which is $(\frac{1}{2}\log\frac{4}{3})^{-1} \doteq 6.95212$; see Section 2. The relation $T^{(k)}(2^k - 1) = 3^k - 1$ also implies that

$$\rho \ge \frac{\log 3}{\log 2} \doteq 1.58496. \tag{1.5}$$

Turning to bounds for δ, since an integer n has at most two preimages under T, one has $N_k \le 2^k$ and hence

$$\delta \le \log 2 \doteq 0.69315.$$

In Section 3 we show that

$$0.17328 \doteq \tfrac{1}{4}\log 2 \le \delta \le \tfrac{1}{2}\log 3 \doteq 0.54931. \tag{1.6}$$

Both of these bounds can be further improved, but apparently not sufficiently to determine δ.

Several authors have proposed stochastic models to imitate the pseudorandom behavior of successive $3x + 1$ function iterates. These include random walk models to estimate the average behavior of $\sigma_\infty(n)$, which were studied by Crandall (1978), Rawsthorne (1985) and Wagon (1985), and Markov chain models to estimate the amount of time trajectories $\{T^{(k)}(n): k \geq 0\}$ spend in various residue classes (mod M), which were studied by Matthews and Watts (1984, 1985) and Leigh (1986). In this paper we formulate and study two types of stochastic processes modelling the behavior of $3x + 1$ function iterates, which possess analogues of the constants γ, δ and ρ. Such stochastic models prove nothing about the behavior of iterates of the $3x + 1$ function, of course. However, they make various predictions that can be checked against empirical data for the $3x + 1$ function, and this is done in Section 5.

The first of these models, described in Section 2, is a set of independent random walks modelling the forward evolution of the process based on the behavior of iterates (mod 2^j). It produces constants γ_{RW} and ρ_{RW} defined in Theorems 2.1 and 2.3, respectively. The underlying random walk is similar to those of Crandall (1978) and Wagon (1985). The second of these models, described in Section 3, is a family of multitype branching random walks that model the backward evolution of the process (mod 3^j) using the multivalued function T^{-1} on $\mathbb{Z}^+$. It produces constants γ_{BP} and δ_{BP} defined in Section 3. (Here BP = branching process.)

The constant γ_{RW} is determined using basic facts from large deviation theory for random walks while γ_{BP} is calculable using results of Biggins (1976) for multitype branching random walks. In Section 4 we show that

$$\gamma_{\mathrm{RW}} = \gamma_{\mathrm{BP}} \doteq 41.677647\ldots. \tag{1.7}$$

The equality of these two constants, which we discovered numerically and then proved, initially surprised us. Both constants depend on "tails" of distributions, and as such are sensitive to model assumptions. (For example, a simple random walk having equal step sizes using a biased coin with bias adjusted to give the same drift as the model in Section 2 yields a constant different from γ_{RW}.) Also these two models are based on different aspects of the $3x + 1$ function, the behavior of T (mod 2^j) and of T^{-1} (mod 3^j), respectively. The equality $\gamma_{\mathrm{RW}} = \gamma_{\mathrm{BP}}$ is a theorem of probability theory expressing a kind of "duality" between certain repeated random walks and branching random walks; it is proved in Section 4.

The predictions of these two stochastic models lead us to formulate analogous conjectures for the $3x + 1$ function. Based on (1.7) we propose the following.

CONJECTURE 1. $\gamma = \gamma_{\mathrm{RW}} = \gamma_{\mathrm{BP}}$.

This conjecture seems as hard to decide as the $3x + 1$ conjecture itself. In Sections 5 and 6 we show that the available empirical evidence seems consistent with this conjecture.

The constant $\delta_{\text{BP}} = \log \frac{4}{3} \doteq 0.28768$ is calculable using the basic theory of multitype Galton–Watson processes (Corollary 3.1). This result plus much other evidence leads us to propose:

CONJECTURE 2. *The limit of* $\log N_k / k$ *as* $k \to \infty$ *exists and is* $\log \frac{4}{3}$.

This conjecture is almost certainly true. It is related to the problem of obtaining lower bounds of the form x^c for the quantity $\#\{n \le x: \sigma_\infty(n) < \infty\}$, which we hope to return to elsewhere.

Using large deviation theory for random walks we show that $\rho_{\text{RW}} = 2$ (Theorem 2.3).

CONJECTURE 3. *The* $3x + 1$ *maximal excursion constant* $\rho = 2$.

This conjecture agrees well with the empirical evidence for the $3x + 1$ function given in Table 3, but seems unprovable at present.

The final Section 6 describes greedy algorithms to find large values of $\gamma_\infty(n) := (\sigma_\infty(n))/\log n$ by backward search using T^{-1}, as well as probabilistic models of these greedy algorithms that use the branching random walk of Section 3. The predictions of suitable greedy branching random walks are in excellent agreement with empirical data for T^{-1}. Using a greedy algorithm to find large values of $\sigma_\infty(n)$ was proposed by Vyssotsky (1987), who made numerical experiments. The desire to explain his empirical data motivated this paper.

In order to keep the results accessible to the widest possible audience interested in the $3x + 1$ problem, we have used elementary arguments whenever possible, at the expense of occasionally not obtaining the sharpest available error bounds. For example, Section 2 uses the ballot theorem and Chernoff bounds rather than martingale arguments and the full machinery of large deviation theory.

2. Random walk model. One of the first observations made about the $3x + 1$ function was that the parity of its successive iterates behaves like independent flips of a fair coin. Given $n \in \mathbb{N}$ define the kth *parity bit* $b_k(n) \in \{0, 1\}$ by

$$b_k(n) \equiv T^{(k)}(n) \pmod 2,$$

and the kth *parity vector* $\mathbf{v}^{(k)}(n)$ by

$$\mathbf{v}^{(k)}(n) = (b_0(n), \ldots, b_{k-1}(n)).$$

The basic result is due independently to Terras (1976) and Everett (1977).

PROPOSITION 2.1. *The kth parity vector $\mathbf{v}^{(k)}(n)$ is a periodic function of n with period 2^k. Each vector $\mathbf{v} \in \{0, 1\}^k$ occurs as the parity vector $\mathbf{v}^{(k)}(n)$ for exactly one n with $1 \le n \le 2^k$.*

Now the value $T(n)$ is either $n/2$ or approximately $3n/2$ according to the parity of n, and after k iterations one has the approximation

$$T^{(k)}(n) \approx 3^{b_0(n)+\cdots+b_{k-1}(n)}2^{-k}n. \tag{2.1}$$

Thus $\{\log T^{(k)}(n): k \ge 1\}$ can be approximated by a walk on the real line which starts at $\log n$ and which takes a jth step of $\log \frac{1}{2}$ if the parity of $T^{(j-1)}(n)$ is even and $\log \frac{3}{2}$ if the parity of $T^{(j-1)}(n)$ is odd. The approximation (2.1) becomes less and less accurate as k increases, but it is reasonably accurate for the first $\log n$ steps. Proposition 2.1 states that if an integer n is drawn uniformly from the interval $[1, 2^k]$, then its parity vector has the uniform distribution on $\{0, 1\}^k$, so that, aside from its starting point, this walk then behaves like a simple random walk (with unequal size steps) for the first k steps.

These observations motivate modelling the $3x + 1$ iteration process using independent random walks for each value of n separately. For each n, we model $\log T^{(k)}(n)$ by a random walk $Z^*(n, k)$:

$$Z^*(n, k+1) = \begin{cases} Z^*(n,k) - \log 2, & \text{with probability } \frac{1}{2}, \\ Z^*(n,k) + \log \frac{3}{2}, & \text{with probability } \frac{1}{2}, \end{cases}$$

$$Z^*(n, 0) = \log n.$$

For this model we define $\omega_\infty(n)$ as the level crossing time to the set $\{Z \le 0\}$. The process $Z^*(n, k)$, as a function of k, has drift $-\frac{1}{2}\log\frac{4}{3}$, so $\omega_\infty(n) < \infty$ a.s. (i.e., the $3x + 1$ conjecture is true for this model). Furthermore,

$$E(\omega_\infty(n)) = \left(\tfrac{1}{2}\log\tfrac{4}{3}\right)^{-1}\log n \approx 6.95212\log n. \tag{2.2}$$

Now we formalize the model we just described. Actually we use an equivalent random walk process more convenient for analysis: $Z(n, k) = \log n - Z^*(n, k)$, which starts at 0 for all n, and has positive drift.

Independent random walks process. Given an i.i.d. set $\{X(n, k): n \ge 1, k \ge 0\}$ with

$$P[X(\cdot,\cdot) = \log 2] = \tfrac{1}{2},$$

$$P[X(\cdot,\cdot) = \log \tfrac{2}{3}] = \tfrac{1}{2},$$

define $Z(n, 0) = 0$ and

$$Z(n, k) = \sum_{i=1}^{k} X(n, i).$$

From this model we extract two sets of random variables analogous to $\sigma_\infty(n)$ and $t(n)$. Define for $n \geq 1$,

$$\begin{aligned}\sigma_\infty(\omega_n) &= \min_{k \geq 1}\{k\colon Z^*(n,k) \leq 0\}\\ &= \min_{k\geq 1}\{k\colon Z(n,k) \geq \log n\}\\ t(\omega_n) &= \sup_{k \geq 0}\{\exp(Z^*(n,k))\}\\ &= \sup_{k\geq 0}\{n\exp(-Z(n,k))\}.\end{aligned}$$

These quantities are set equal to $+\infty$ if they are otherwise undefined; however, they are each finite almost surely.

The distribution of large values of $(\sigma_\infty(\omega_n))/\log n$ can be analyzed using large deviation theory.

THEOREM 2.1. *The independent random walks process has*

$$\limsup_{n\to\infty}\frac{\sigma_\infty(\omega_n)}{\log n} = \gamma_{\mathrm{RW}} \quad a.s., \tag{2.3}$$

where $\gamma = \gamma_{\mathrm{RW}}$ *is the unique solution with* $\gamma > (\frac{1}{2}\log\frac{4}{3})^{-1}$ *of the equation*

$$\gamma g\left(\frac{1}{\gamma}\right) = 1, \tag{2.4}$$

where

$$g(a) := \sup_{\theta\in\mathbb{R}}\left[a\theta - \log\left(\left(\tfrac{1}{2}\right)\left(2^\theta + \left(\tfrac{2}{3}\right)^\theta\right)\right)\right]. \tag{2.5}$$

Numerically one finds that $\gamma_{\mathrm{RW}} \doteq 41.677647$.

PROOF. To prove (2.3) it suffices to show that for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty}\operatorname{Prob}[\sigma_\infty(\omega_n) > (\gamma+\varepsilon)\log n] < \infty, \tag{2.6}$$

$$\sum_{n=1}^{\infty}\operatorname{Prob}[\sigma_\infty(\omega_n) > (\gamma-\varepsilon)\log n] = \infty. \tag{2.7}$$

To establish these, we begin by noting that

$$\operatorname{Prob}[\sigma_\infty(\omega_n) > \beta\log n] = \operatorname{Prob}[Z(n,k) < \log n\colon 0 \leq k \leq \beta\log n]. \tag{2.8}$$

We use the bounds of Chernoff (1952), Theorem 1.

CHERNOFF'S THEOREM. *Let* $S_m = \sum_{i=1}^m X_i$ *be the sum of m i.i.d. random variables with distribution X. Set*

$$g(a) = \sup_{\theta\in\mathbb{R}}\{\theta a - \log(E[\exp(\theta X)])\}.$$

Then if $a \geq E(X)$,

$$\text{Prob}[S_m \geq ma] \leq \exp(-g(a)m). \tag{2.9}$$

Furthermore if $a \geq E(X)$ *and* $\varepsilon > 0$, *then*

$$\lim_{m\to\infty} \frac{\exp(-(g(a)+\varepsilon)m)}{\text{Prob}[S_m \geq ma]} = 0. \tag{2.10}$$

We will use the following sharpening of the second part of Chernoff's theorem, which is easily derived from the ballot theorem.

Lemma 2.1. *If* $a \geq E(X)$ *and* $\varepsilon > 0$, *then*

$$\lim_{m\to\infty} \frac{\exp(-(g(a)+\varepsilon)m)}{\text{Prob}[S_j \geq ja \text{ for } 1 \leq j \leq m]} = 0. \tag{2.11}$$

Proof. For any sequence of m i.i.d. random variables and any real a one has

$$\text{Prob}[S_j \geq ja \text{ for } 1 \leq j \leq m] \geq \frac{1}{m} \text{Prob}[S_m \geq ma].$$

This is a consequence of the observation that at least one of the m cyclic permutations $(X_i, X_{i+1}, \dots, X_m, X_1, \dots, X_{i-1})$ of any set of m real numbers with $\sum_{i=1}^m X_i = m\beta$ has jth partial sum at least $j\beta$ for $1 \leq j \leq m$. Indeed start at that i which minimizes $\{S_i - i\beta: 1 \leq i \leq m\}$ [see Takàcs (1977), Chapter 1]. Now (2.11) follows from the inequality above, using (2.10) with $\frac{1}{2}\varepsilon$ replacing ε.

Before continuing the proof we note that stronger versions of the Chernoff bounds exist, one of which asserts that

$$\text{Prob}[S_m \geq ma] = \exp(-g(a)m - \tfrac{1}{2}\log(m) + O(1));$$

see Ney (1984), Equation (2). Using this bound one can prove that (2.7) also holds for $\varepsilon = 0$.

The random walk $Z(n,k)$ has steps that are either $\log 2$ or $\log(\frac{2}{3})$. A single step has positive drift $E[X] = \frac{1}{2}\log\frac{4}{3}$, and *moment generating function* (Laplace transform)

$$M(\theta) := E[\exp(\theta X)] = \tfrac{1}{2}\left(2^\theta + \left(\tfrac{2}{3}\right)^\theta\right). \tag{2.12}$$

We define the quantity γ_{RW} by the condition $\gamma_{\text{RW}} > E[X]^{-1}$ and

$$\gamma g\left(\frac{1}{\gamma}\right) = 1. \tag{2.13}$$

The existence of a unique solution to (2.13) is proved in a standard way.

LEMMA 2.2. *The function*

$$g(a) := \sup_{\theta \in \mathbb{R}} \Big(\theta a - \log\Big(\tfrac{1}{2}\big(2^{\theta} + (\tfrac{2}{3})^{\theta}\big)\Big)\Big) \tag{2.14}$$

is finite, nonnegative and strictly convex for $\log(\frac{2}{3}) < a < \log 2$. *One has* $g(\frac{1}{2}\log\frac{4}{3}) = 0$.

PROOF. The function $\log M(\theta)$ in (2.12) is strictly convex on $\mathbb{R}$ since $d^2M/d\theta^2 > 0$ on $\mathbb{R}$. Because $g(a)$ is the conjugate convex function (Legendre transform) of $\log M(\theta)$, it is strictly convex where it is finite; see, for example, Roberts and Varberg [(1973), page 34]. By choosing $\theta = 0$ one sees that $g(a) \geq 0$ everywhere. Now a calculation shows that

$$\frac{d}{d\theta}(-\log M(\theta)) = -\log 2 + \log 3\left(\frac{3^{-\theta}}{1 + 3^{-\theta}}\right), \tag{2.15}$$

which is a monotone increasing function of θ with limit $\log\frac{2}{3}$ as $\theta \to -\infty$ and $\log 2$ as $\theta \to \infty$. Consequently equation $(d/d\theta)(a\theta - \log M(\theta)) = 0$ has a unique solution $\theta(a)$ for $\log\frac{2}{3} < a < \log 2$, which achieves the maximum on the right side of (2.14). One finds that $\theta(\frac{1}{2}\log\frac{4}{3}) = 0$ by substituting $\theta = 0$ in (2.15), which shows that $g(\frac{1}{2}\log\frac{4}{3}) = 0$.

Lemma 2.1 shows that $g(0) > 0$ and that $g(a)$ is strictly decreasing on the interval $[0, \frac{1}{2}\log\frac{4}{3}]$. Hence $\gamma g(1/\gamma)$ is a strictly increasing function on $[(\frac{1}{2}\log\frac{4}{3})^{-1}, \infty)$ with range $[0, \infty)$, so (2.13) has a unique solution $\gamma > (\frac{1}{2}\log\frac{4}{3})^{-1}$.

To continue proving Theorem 2.1, we apply Chernoff's theorem (2.9) with $\gamma = \gamma_{\mathrm{RW}}$ to obtain

$$\begin{aligned} \operatorname{Prob}[\sigma_\infty(\omega_n) > (\gamma + \varepsilon)\log n] &\leq \operatorname{Prob}[Z(n, (\gamma + \varepsilon)\log n) < \log n] \\ &\leq \exp\big(-g((\gamma + \varepsilon)^{-1})(\gamma + \varepsilon)\log n\big) \\ &\leq \exp(-(1 + \varepsilon')\log n) = n^{-1-\varepsilon'}, \end{aligned} \tag{2.16}$$

for some $\varepsilon' > 0$ depending on ε, since $\gamma g(1/\gamma)$ is strictly increasing. This proves (2.6). Since the random walks ω_n are independent, for different n, using the Borel–Cantelli lemma and letting $\varepsilon \to 0$ gives

$$\limsup_{n \to \infty} \frac{\sigma_\infty(\omega_n)}{\log n} \leq \gamma_{\mathrm{RW}} \quad \text{a.s.}$$

To establish (2.7) for $\gamma = \gamma_{\mathrm{RW}}$, we use Lemma 2.1 to obtain for $0 < \varepsilon_1 < g(\gamma - \varepsilon)$ that

$$\begin{aligned} &\operatorname{Prob}[\sigma_\infty(\omega_n) > (\gamma + \varepsilon)\log n] \\ &\quad\geq \operatorname{Prob}\left[Z(n, k) < \frac{k}{\gamma - \varepsilon} : 0 \leq k \leq (\gamma - \varepsilon)\log n\right] \\ &\quad\geq \exp\Big(-\big(g((\gamma - \varepsilon)^{-1}) + \varepsilon_1\big)(\gamma - \varepsilon)\log n\Big), \end{aligned} \tag{2.17}$$

for all $n > n_0(\varepsilon, \varepsilon_1)$. Choosing $\varepsilon_1 > 0$ small enough, one has

$$\mathrm{Prob}[\sigma_\infty(\omega_m) > (\gamma - \varepsilon)\log n] \ge \exp(-(1-\varepsilon')\log n) = n^{-1+\varepsilon'},$$

for some $\varepsilon' > 0$ and all sufficiently large n, which proves (2.7). Using the Borel–Cantelli lemma and letting $\varepsilon \to 0$ shows

$$\limsup_{n\to\infty} \frac{\sigma_\infty(\omega_n)}{\log n} \ge \gamma_{\mathrm{RW}} \quad \text{a.s.} \qquad \square$$

We next give a density result on the occurrence of large deviations.

THEOREM 2.2. *For the independent random walks process, with*

$$\left(\tfrac{1}{2}\log\tfrac{4}{3}\right)^{-1} < \alpha < \gamma_{\mathrm{RW}}$$

one has

$$E\left[\#\left\{n \le x: \frac{\sigma_\infty(\omega_n)}{\log n} \ge \alpha\right\}\right] \le \left(1 - \alpha g\left(\frac{1}{\alpha}\right)\right)^{-1} x^{1-\alpha g(1/\alpha)}. \tag{2.18}$$

For any $\varepsilon > 0$ one has

$$E\left[\#\left\{n \le x: \frac{\sigma_\infty(\omega_n)}{\log n} \ge \alpha\right\}\right] \ge x^{1-\alpha g(1/\alpha)-\varepsilon} \tag{2.19}$$

for all sufficiently large $x \ge x_0(\varepsilon)$.

PROOF. Using Chernoff's theorem, one derives similarly to (2.16) that for any fixed n,

$$\mathrm{Prob}[\sigma_\infty(\omega_n) > \alpha\log n] \le \exp\left(-\alpha g\left(\frac{1}{\alpha}\right)\log n\right).$$

Hence

$$E\left[\#\left\{n \le x: \frac{\sigma_\infty(\omega_n)}{\log n} \ge \alpha\right\}\right] = \sum_{n=1}^{x} \mathrm{Prob}[\sigma_\infty(\omega_n) > \alpha\log n] \le \sum_{n=1}^{x} n^{-\alpha g(1/\alpha)},$$

from which (2.18) follows.

One derives similarly to (2.17) that for any $\varepsilon' > 0$,

$$\mathrm{Prob}[\sigma_\infty(\omega_n) > \alpha\log n] \ge \mathrm{Prob}\left[Z(n,k) < \frac{k}{\beta}: 0 \le k \le \alpha\log n\right] \ge \exp\left(-\left(g\left(\frac{1}{\alpha}\right) + \varepsilon'\right)\alpha\log n\right)$$

holds for all sufficiently large $x \ge x(\varepsilon')$. Choosing ε' small enough, summing over all $n \le x$ gives the lower bound (2.19). $\square$

TABLE 1
Large deviation density $\alpha g(1/\alpha)$ and density ratio $r(\alpha)$ for event $\omega_n \geq \alpha \log n$

Density α	$\alpha g(1/\alpha)$	**Density ratio $r(\alpha)$**
6.952119	0.000000	0.500000
10	0.031883	0.539906
15	0.148531	0.570247
20	0.293284	0.585418
25	0.449402	0.594520
30	0.611249	0.600588
35	0.776375	0.604923
40	0.943568	0.608173
41.677647	1.000000	0.609090

Some specific values of these densities are given in Table 1. This table also gives values of the *density ratio*

$$r(\alpha) = \frac{1/\alpha + \log 2}{\log 3},$$

which is the ratio of the number of steps of size $\log(2/3)$ to the total number of steps in a random walk having $\sigma_\infty(\omega_n) = \alpha \log n$. [This ratio corresponds to the fraction of iterates $T^{(k)}(n) \equiv 1 \pmod 2$ occurring in a trajectory having $\sigma_\infty(n) = \alpha \log n$.]

Large deviation theory also predicts that "most" trajectories of the random walk with $\gamma(\omega_n)$ near the constant γ_{RW} when plotted logarithmically will appear roughly linear with slope $(\gamma_{\mathrm{RW}})^{-1} \doteq 0.02399$ for their entire length (see Figure 3 in Section 5). A precise version of this assertion can be formulated following Wentzell (1976) or Azencott and Ruget (1977).

Now we analyze the behavior of large values of $(\log t(\omega_n))/\log n$.

THEOREM 2.3. *The independent random walks process has*

$$\limsup_{n \to \infty} \frac{\log t(\omega_n)}{\log n} = \rho_{\mathrm{RW}} \quad a.s. \tag{2.20}$$

where $\rho_{\mathrm{RW}} = 2$.

PROOF. This is proved similarly to Theorem 2.1. Here we indicate only how the constant ρ_{RW} is determined. We will use Chernoff's theorem applied to the original random walk, which has moment generating function

$$\overline{M}(\theta) := \log \tfrac{1}{2}\left(\left(\tfrac{1}{2}\right)^{\theta} + \left(\tfrac{3}{2}\right)^{\theta}\right) = M(-\theta). \tag{2.21}$$

Now $\rho_{\mathrm{RW}} = 1 + \eta$ for that value of $\eta > 0$ for which

$$\mathrm{Prob}[S_m \geq \eta \log n \text{ for some } m \geq 0] = \exp(-(1 + o(1))\log n), \tag{2.22}$$

as $n \to \infty$. These trajectories of a random walk having negative drift that maximize the probability of reaching a given positive height H are approximately straight lines with a constant slope a, and Chernoff's theorem gives

$$\text{Prob}[S_m > am] = \exp(-\bar{g}(a)(1+o(1))m)$$

as $m \to \infty$, where

$$\bar{g}(a) = \sup_{\theta \in \mathbb{R}} \left(\theta a - \log \overline{M}(\theta)\right). \tag{2.23}$$

[Here $\bar{g}(a) = g(-a)$ for the function $g(a)$ of Lemma 2.2.] The trajectory takes H/a steps to attain height H, and to maximize $\exp(-((\bar{g}(a))/a)H)$ we must minimize $(\bar{g}(a))/a$. If $\theta = \bar{\theta}(a)$ attains the supremum on the right side of (2.23), then

$$\begin{aligned} \frac{d}{da}\bar{g}(a) &= \bar{\theta}(a) + \frac{\partial}{\partial\theta}\left(\theta a - \log \overline{M}(\theta)\right)\Big|_{\theta=\bar{\theta}} \\ &= \bar{\theta}(a). \end{aligned} \tag{2.24}$$

To minimize $(\bar{g}(a))/a$ we find the stationary point

$$0 = \frac{d}{da}\left(\frac{\bar{g}(a)}{a}\right) = \frac{1}{a^2}\left(\bar{g}(a) - a\frac{d\bar{g}(a)}{da}\right).$$

Any stationary value a^* satisfies

$$0 = \bar{g}(a^*) - a^*\frac{d\bar{g}(a)}{da} = \log M(\bar{\theta}(a^*)), \tag{2.25}$$

using (2.23) and (2.24). The strict convexity of $\log M(\theta)$ guarantees that $\bar{\theta}(a^*)$ is unique, and by inspection of (2.21) we see that $\bar{\theta}(a^*) = 1$. Furthermore (2.25) gives

$$\frac{\bar{g}(a^*)}{a^*} = \frac{d\bar{g}(a)}{da} = \bar{\theta}(a^*) = 1,$$

and this is easily checked to be a minimum for $(\bar{g}(a))/a$. Taking $H = \log n$, one obtains $\eta = 1$ and $\rho_{\text{RW}} = 2$. One can also verify that the optimal slope is

$$a^* = \tfrac{3}{4}\log 3 - \log 2 \doteq 0.1308. \qquad \square$$

It can be shown that nearly all trajectories with $t(\omega_n) > n^{2-\varepsilon}$ will appear roughly linear with slope $\beta_1 = \frac{3}{4}\log 3 - \log 2 \doteq 0.1308$ for their first $\beta_1^{-1}\log n \doteq 7.6445 \log n$ steps, and then roughly linear with slope $\beta_2 = (\frac{1}{2}\log\frac{3}{4})^{-1} = -0.1453$ for their next $\beta_1^{-1}(2\log n)$ steps, and thus have $\gamma(\omega_n) \cong \beta_1^{-1} + \beta_2^{-1} \doteq 21.5487$ (see Figure 3 in Section 5).

We also estimate the number of trajectories ω_n having $t(\omega_n) > n^{2-\alpha}$ for $0 \le \alpha < 1$.

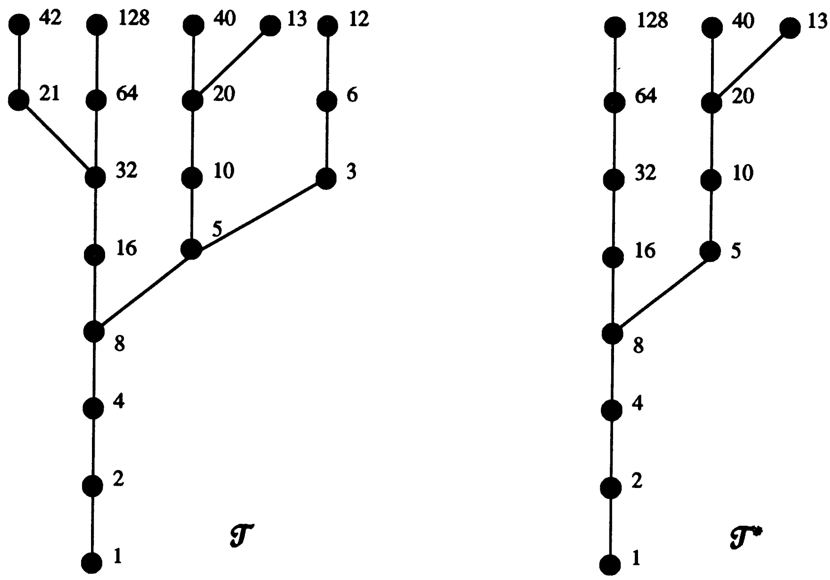

FIG. 1. $3x + 1$ *tree* $\mathcal{T}$ *and pruned tree* $\mathcal{T}^*$ *for depth* ≤ 7.

THEOREM 2.4. *For the independent random walks process*

$$E\left[\#\left\{n \leq x: \frac{\log t(\omega_n)}{\log n} \geq 2 - \alpha\right\}\right] = x^{\alpha(1+o(1))},$$

as $x \to \infty$, *almost surely.*

PROOF. This is proved similarly to Theorem 2.2. We omit the details. □

3. Branching process models. The backward evolution of the $3x + 1$ function from 1 can be represented by an (infinite) rooted tree $\mathcal{T}$ with root node labelled 1, whose nodes at level k from the root are labelled by those integers n having $\sigma_\infty(n) = k$, with edges from n to $T(n)$. The root node makes up level zero; see Figure 1. Associated to this tree are the quantities $N(k)$, which counts the number of vertices at level k, and $H(k)$, which equals the minimum label of a node at level k, that is,

$$N(k) = \#\{n: \sigma_\infty(n) = k\},$$

$$H(k) = \min\{n: \sigma_\infty(n) = k\}.$$

It is easy to see that the $3x + 1$ growth constant

$$\gamma = \limsup_{k \to \infty} \left\{\frac{k}{\log H(k)}\right\}. \tag{3.1}$$

The tree $\mathscr{T}$ can be recursively constructed from its unique node at level 2 (having the label 4), using the multivalued operator T^{-1} on the domain $\mathbb{Z}^+$, which is given by

$$T^{-1}(n) = \begin{cases} \{2n\}, & \text{if } n \equiv 0, 1 \pmod 3, \\ \left\{2n, \dfrac{2n-1}{3}\right\}, & \text{if } n \equiv 2 \pmod 3. \end{cases}$$

The elements at level $k + 1$ in this tree are produced from level k using T^{-1}.

More generally one can apply T^{-1} starting from any $m \geq 1$ to obtain an infinite rooted tree $\mathscr{T}(m)$, where T^{-1} applied to the labelled nodes at depth k gives labelled nodes at depth $k + 1$. If the root node m is in a cycle, then T^{-1} "unwraps" the cycle so that the same label appears repeatedly at different levels of the tree. If m is not in a cycle, then all node labels in the tree are distinct. [We regard the tree $\mathscr{T}$ as corresponding to $\mathscr{T}(4)$, rather than $\mathscr{T}(1)$. Note $\mathscr{T}(4)$ is obtained from $\mathscr{T}$ by deleting its first two levels.] The labelled nodes in level k of the tree $\mathscr{T}(m)$ are the kth generation of descendants of m; they have labels $\{n\colon T^{(k)}(n) = m\}$. Let $\mathscr{T}_k(m)$ denote the finite labelled tree consisting of the first k levels of $\mathscr{T}(m)$ and let $\overline{\mathscr{T}}_k(m)$ denote the tree with all node labels 0 and 1 obtained from $\mathscr{T}_k(m)$ by replacing each node label n with its parity bit $b(n) \equiv n \pmod 2$.

The pattern of branching of the tree $\mathscr{T}(m)$ at a node n is determined by $n \pmod 3$ according to the formula (3.1) for T^{-1}. Similarly the branching pattern to depth k of such a tree is entirely determined by $m \pmod{3^k}$; if $m_1 = m_2 \pmod{3^k}$, then $\overline{\mathscr{T}}_k(m_1)$ is identical to $\overline{\mathscr{T}}_k(m_2)$.

The trees $\mathscr{T}(m)$ have very different branching behavior depending on whether $m \equiv 0 \pmod 3$ or $m \not\equiv 0 \pmod 3$. If $m \equiv 0 \pmod 3$, the tree $\mathscr{T}(m)$ never branches, and has nodes $\{2^j m\colon j \geq 0\}$. We will show below that if $m \not\equiv 0 \pmod 3$, then the number of nodes $N(k; m)$ at level k of the tree $\mathscr{T}(m)$ grows exponentially in k.

In fact the nodes $m \equiv 0 \pmod 3$ in the tree $\mathscr{T}$ do not have significant effect on $N(k)$, $H(k)$ and γ, and we obtain a simplification by eliminating them. Let $\mathscr{T}^*$ be the *pruned tree* obtained from $\mathscr{T}$ by deleting all vertices with labels $n \equiv 0 \pmod 3$ in $\mathscr{T}$. Since all descendants of $n \equiv 0 \pmod 3$ are themselves $0 \pmod 3$, $\mathscr{T}^*$ is a tree (see Figure 1). The tree $\mathscr{T}^*$ may be recursively constructed from its unique node at level two (labelled 4) using the operator $(T^*)^{-1}$ on the domain $\mathbb{Z}^+$, given by

$$(T^*)^{-1}(n) = \begin{cases} \{2n\}, & \text{if } n \equiv 1, 4, 5 \text{ or } 7 \pmod 9, \\ \left\{2n, \dfrac{2n-1}{3}\right\}, & \text{if } n \equiv 2, 8 \pmod 9. \end{cases} \tag{3.2}$$

The analogues of the quantities $M(k)$ and $N(k)$ for $\mathscr{T}^*$ are

$$N^*(k) = \#\{n\colon \sigma_\infty(n) = k \text{ and } n \not\equiv 0 \pmod 3\},$$

$$H^*(k) = \min\{n\colon \sigma_\infty(n) = k \text{ and } n \not\equiv 0 \pmod 3\}.$$

The following lemma shows that there is essentially no loss in generality in considering $\mathscr{T}^*$ instead of $\mathscr{T}$.

LEMMA 3.1. *For all* $k \geq 1$,

$$\frac{1}{k}N(k) \leq N^*(k) \leq N(k), \tag{3.3}$$

$$H(k) \leq H^*(k) \leq 4H(k). \tag{3.4}$$

Consequently

$$\gamma = \limsup_{k\to\infty} \left\{ \frac{k}{\log H^*(k)} \right\}. \tag{3.5}$$

PROOF. The upper bound in (3.3) and lower bound in (3.4) are immediate. To prove the lower bound in (3.3), let $\tilde{N}(k) = \{n\colon \sigma_\infty(n) = k \text{ and } n \equiv 3 \pmod 6\}$. Since $T(n) = (3n+1)/2 \not\equiv 0 \pmod 3$ for such n, one has

$$\tilde{N}(k) \leq N^*(k-1).$$

Next

$$N(k) = N^*(k) + \sum_{j=1}^{k} \tilde{N}(j),$$

since any $n \equiv 0 \pmod 6$ with $\sigma_\infty(n) = k$ is the unique descendant of some $m \equiv 3 \pmod 6$ at a lower level. Hence

$$N(k) \leq \sum_{j=1}^{k} N^*(k) = kN^*(k).$$

To prove the upper bound in (3.4) suppose $H(k) = n \equiv 0 \pmod 3$. Set $n = 2^j\tilde{n}$ with $\tilde{n} = 1 \pmod 2$. Then $T^{(j+1)}(n) = (3\tilde{n}+1)/2$ and $m^* = 2^j(3\tilde{n}+1)$ also has $\sigma_\infty(m^*) = k$ so

$$H^*(k) \leq m \leq 2^j(4\tilde{n}) = 4H(k).$$

Finally (3.5) follows from (3.1) and the upper bound in (3.4). □

One constructs pruned trees $\mathscr{T}^*(m)$, $\mathscr{T}_k^*(m)$ and $\bar{\mathscr{T}}_k^*(m)$ in exact parallel with the construction of the trees $\mathscr{T}(m)$, $\mathscr{T}_k(m)$ and $\bar{\mathscr{T}}_k(m)$, using the operator $(T^*)^{-1}$ in place of T^{-1}. The trees $\mathscr{T}^*(m)$, $\mathscr{T}_k^*(m)$ are obtained from the trees $\mathscr{T}(m)$, $\mathscr{T}_k(m)$ by deleting all nodes with labels $n \equiv 0 \pmod 3$; they are nonempty only for $m \not\equiv 0 \pmod 3$.

The branching pattern of a tree $\mathscr{T}^*(m)$ at a node n is specified by $n \pmod 9$ according to the definition of $(T^*)^{-1}$. Similarly its branching structure to depth k is determined by $m \pmod{3^{k+1}}$; if $m_1 \equiv m_2 \pmod{3^{k+1}}$, then $\bar{\mathscr{T}}_k^*(m_1)$ is identical to $\bar{\mathscr{T}}_k^*(m_2)$.

For any $m \not\equiv 0 \pmod 3$ the number $N^*(k, m)$ of depth k nodes in $\mathscr{T}^*(m)$ satisfies

$$N^*(k, m) \geq 2^{[k/4]}.$$

This follows because the recursion (3.2) shows that any path in $\mathscr{T}^*(m)$ must encounter a branch at least once in each four levels, because if $n \not\equiv 0 \pmod 3$, then at least one of $\{n, 2n, 4n, 8n\}$ is congruent to 2 or 8 (mod 9). As a consequence

$$N(k) \geq N^*(k) = N^*(k-2, 4) \geq 2^{[(k-2)/4]}. \tag{3.6}$$

In similar fashion we obtain for $m \not\equiv 0 \pmod 3$ the upper bound

$$N^*(2k, m) \leq 2^{k/2}$$

because every tree $\mathscr{T}_2^*(m')$ of depth 2 has at most three branches. Consequently, using Lemma 3.1

$$N(k) \leq kN^*(k) \leq k3^{(k+1)/2}.$$

This bound and (3.6) imply that

$$\tfrac{1}{4}\log 2 \leq \delta \leq \tfrac{1}{2}\log 3. \tag{3.7}$$

The actual growth rate of $N(k)$ is probably $\log \frac{4}{3}$. Evidence for this is provided by the following result showing that the average size of $N^*(k, m)$ as m varies is $(\frac{4}{3})^k$.

THEOREM 3.1. *The pruned trees $\overline{\mathscr{T}}_k(m)$ have*

$$\sum_{\substack{m \pmod{3^{k+1}} \\ m \not\equiv 0 \pmod 3}} N^*(k, m) = 2 \cdot 4^k. \tag{3.8}$$

Hence if a residue class $m \pmod{3^{k+1}}$ with $m \not\equiv 0 \pmod 3$ is picked with the uniform distribution, the expected number of leaves in $\overline{\mathscr{T}}_k^*(m)$ is $(\frac{4}{3})^k$.

PROOF. View an edge from level j to $j + 1$ in a tree $\overline{\mathscr{T}}_k^*(m)$ as being labelled 0 or 1 by the parity bit of its node n at level j. To each leaf n of a tree $\overline{\mathscr{T}}_k^*(m)$ we assign the parity sequence of its edges $\mathbf{v}(k, n) = (b(n), b(T(n)), \ldots, b(T^{(k-1)}(n)))$ to the root m. No two leaves in a fixed tree $\overline{\mathscr{T}}_k^*(m)$ have the same parity sequence. We ask: How many leaves over all these trees have a given parity sequence $\nu = (\nu_1, \ldots, \nu_k) \in \{0, 1\}^k$? Answer: It equals the number of different residue classes $\pmod{3^{k+1}}$ possible for $T^{(k)}(n)$, for those $n \pmod{2^k}$ having parity sequence ν. Now

$$2^k T^{(k)}(n) = 3^{\nu_1 + \cdots + \nu_k} n + \sum_{i=1}^{k-1} \nu_i 3^{\nu_1 + \cdots + \nu_i} 2^i,$$

and reduced $\pmod{3^k}$ this equality shows that $n \pmod{3^{k+1-(\nu_1 + \cdots + \nu_k)}}$ determines $T^{(k)}(n) \pmod{3^{k+1}}$. Then the condition $n \not\equiv 0 \pmod 3$ implies that there are exactly $2 \cdot 3^{k-(\nu_1 + \cdots + \nu_k)}$ choices of such leaves. Now the number of

$\nu \in \{0,1\}^k$ with $\nu_1 + \cdots + \nu_k = j$ is $\binom{k}{j}$, so the total number of leaves is

$$\sum_{j=0}^{k} \binom{k}{j} 2 \cdot 3^{k-j} = 2(1+3)^k,$$

which is (3.8). Since there are $2 \cdot 3^k$ residue classes $m \pmod{3^{k+1}}$ with $m \equiv 0 \pmod 3$, the expected number of leaves in a uniform draw of $\overline{\mathscr{T}}_k^*(m)$ is $(\frac{4}{3})^k$. □

One can prove similarly that if $N(k;m)$ denotes the number of leaves in the tree $\mathscr{T}_k(m)$, then

$$\sum_{m \pmod{3^k}} N(k;m) = 4^k,$$

so the expected number of leaves in a randomly drawn tree $\mathscr{T}_k(m)$ is also $(\frac{4}{3})^k$. But if $m \equiv 0 \pmod 3$ it has one leaf, while if $m \not\equiv 0 \pmod 3$ it has the expected number $\frac{3}{2}(\frac{4}{3})^k$ leaves.

Now we describe a family of branching processes called *multitype branching random walks* that mimic the behavior of all trees $\mathscr{T}^*(m)$ for $m \not\equiv 0 \pmod 3$, and in particular $\mathscr{T}^*$. A *multitype branching random walk* describes the evolution over time of a population consisting of a finite number p of types of individuals sitting on the real line $\mathbb{R}$, starting from a single individual located at 0. Each individual of type i produces offspring distributed at *locations* on $\mathbb{R}$ (measured from its position) described by a (multitype) *point process* $\mathscr{P}_i$, and does so independently of all other individuals. The point process $\mathscr{P}_i$ produces a total of $\mathbf{n} := (n_1, \dots, n_p)$ offspring of the various types with probability distribution $\{P_i(\mathbf{n}) \colon \mathbf{n} \in \mathbb{N}^p\}$ in which each of these $m := n_1 + n_2 + \cdots + n_p$ individuals have locations $(l_1, \dots, l_m)$ drawn from a distribution on $\mathbb{R}^m$ depending on $\mathbf{n}$. The locations of the offspring of one individual may be correlated in this model; see Biggins (1976). [In the special case that all locations are nonnegative, *locations* are called *birth times*, and individuals are viewed as being born and living forever. Such processes are called *Crump–Mode processes*, after Crump and Mode (1968, 1969).]

Associated to any branching random walk is a simpler branching process that assigns to each individual the number $\mathbf{n}$ of progeny of each type and gives all individuals unit lifetimes; this process is a *multitype Galton–Watson process*; see, for example, Athreya and Ney (1972).

Any realization ω of a multitype Galton–Watson process starting from a single individual may be represented as a rooted tree with edges indicating the "offspring of" relation, and nodes are labelled by type of individual. Those nodes at depth k from the root form the kth *generation of descendants* of the root individual. Any realization of a multitype branching random walk may be represented by a similar tree, where in addition each edge of the tree is assigned a label giving the location of the offspring corresponding to that edge.

The family of multitype branching random walks $\{\mathscr{B}(3^j) \colon j = 0, 1, 2, \dots\}$ that we study models the behavior of the multivalued function $(T^*)^{-1} \pmod{3^j}$.

Branching random walk $\mathscr{B}[1]$. There is one type of individual. With probability $\frac{2}{3}$ an individual has a single offspring located at a position $\log 2$ from its progenitor, and with probability $\frac{1}{3}$ it has two offspring located at positions shifted $\log 2$ and $\log(2/3)$ from their progenitor.

Branching random walk $\mathscr{B}[3^j]$ $(j \geq 1)$. There are $p = 2 \cdot 3^{j-1}$ types of individuals, indexed by residue classes $m \pmod{3^j}$ with $m \not\equiv 0 \pmod 3$. The distribution of progeny of an individual of type $m \pmod{3^j}$ is determined as follows: Regard $m \pmod{3^{j+1}}$ with probability $\frac{1}{3}$ each as being one of the three residue classes $\tilde{m} \pmod{3^{j+1}}$ with $m \equiv \tilde{m} \pmod{3^j}$. The tree $\mathscr{T}_1(\tilde{m})$ given by $(T^*)^{-1}(\tilde{m})$ has either one or two progeny, and their labels are determined $\pmod{3^j}$. An offspring $2\tilde{m} \pmod{3^j}$ has location shifted $\log 2$ from that of its progenitor and an offspring $(2\tilde{m} - 1)/3 \pmod{3^j}$ has location shifted $\log(2/3)$ from that of its progenitor.

These locations are chosen to approximate the logarithm of the growth in size of inverse iterates of the $3x + 1$ function.

Let ω denote a realization of the branching random walk $\mathscr{B}[3^j]$ that starts from a single individual denoted $\omega_{0,1}$ located at 0, and let $\mathscr{T}(\omega)$ denote its tree of progeny. The number of individuals of the various types in the kth generation of ω is represented by the vector

$$\mathbf{M}_k(\omega) := \left(N_k^{(1)}(\omega), \ldots, N_k^{(p)}(\omega)\right).$$

We denote the *total number of individuals* (of all types) *in the* kth *generation* by $N_k(\omega) := \sum_p N_k^{(p)}(\omega)$. The collection of progeny of the kth generation is denoted $\{\omega_{k,i}\colon\ 1 \leq i \leq N_k(\omega)\}$, ordered as explained below. $L(\omega_{k,i})$ is the *location* of $\omega_{k,i}$ on $\mathbb{R}$. Let $\omega_{k,i}[l]$ denote the ancestor of $\omega_{k,i}$ in the lth generation (for $0 \leq l \leq k$), so that $\omega_{k,i}[0] = \omega_{0,1}$. Assign to $\omega_{k,i}$ the *parity* $b(\omega_{k,j}) = 0$ or 1 according as the quantity $L(\omega_{k,i}) - L(\omega_{k,i}[k-1])$ is $\log 2$ or $\log \frac{2}{3}$. Now assign to $\omega_{k,i}$ the *parity vector*

$$\mathbf{v}(\omega_{k,i}) = \left(b(\omega_{k,i}[1]), \ldots, b(\omega_{k,i}[k])\right).$$

For fixed k no two $\omega_{k,i}$ have the same parity vector, and we order them so that the vectors $\mathbf{v}(\omega_{k,j})$ are in increasing lexicographic order. $T(\omega_{k,i}) \pmod{3^j}$ denotes the *type* of the individual $\omega_{k,i}$.

We also define quantities directly analogous to the $3x + 1$ process. The *density ratio* $r(\omega_{k,i})$ of $\omega_{k,i}$ is

$$r(\omega_{k,i}) = \frac{1}{k} \sum_{l=1}^{k} b(\omega_{k,l}). \tag{3.9}$$

The *size*

$$S(\omega_{k,i}) = \exp(L(\omega_{k,i}))$$

is analogous to a node label in the $3x + 1$ tree $\mathscr{T}$. The density ratio $r(\omega_{k,i})$

and location $L(\omega_{k,i})$ are related by

$$L(\omega_{k,i}) = k(\log 2 - r(\omega_{k,i})\log 3). \tag{3.10}$$

The evolution of the process $\mathscr{B}[3^j]$ starting from a single individual of type 4 (mod 3^j) provides an analogue of the tree $\mathscr{T}^*$. We study the behavior as $k \to \infty$ of analogues of $3x + 1$ quantities. The *first birth* among member of the kth generation,

$$L_k^*(\omega) := \min\{L(\omega_{k,i}): 1 \le i \le N_k(\omega)\},$$

and the *asymptotic first birth* for ω in $\mathscr{B}[3^j]$,

$$\beta^{(j)}(\omega) := \limsup_{k\to\infty} \frac{1}{k}(L_k^*(\omega)).$$

The analogue of the $3x + 1$ stopping constant γ [given by (3.1)] for the process $\mathscr{B}[3^j]$ is the quantity

$$\begin{aligned}\gamma_{\mathrm{BP}}^{(j)}(\omega) &:= \limsup_{k\to\infty} \frac{k}{L_k^*(\omega)} \\ &= \left(\beta^{(j)}(\omega)\right)^{-1}.\end{aligned} \tag{3.11}$$

The analogue of the $3x + 1$ growth constant δ for $\mathscr{B}[3^j]$ is

$$\delta_{\mathrm{BP}}^{(j)}(\omega) = \limsup_{k\to\infty} \frac{\log N_k(\omega)}{k}.$$

The transition matrices for $\mathscr{B}[1]$, $\mathscr{B}[3]$ and $\mathscr{B}[9]$ are given in Table 2. The rows give the individual's type and the columns represent the type of offspring. Separate transition matrices $\mathbf{E}_j^{(b)}$ are given for the two possible birth times of offspring, corresponding to parity $b = 0$ or 1. Using classical results in branching processes it is easy to show that for the process $\mathscr{B}[3^j]$ the counts of individuals $\mathbf{M}_k(\omega)$ grow geometrically like $(\frac{4}{3})^k$, as follows.

THEOREM 3.2. *For all $j \ge 0$, the branching process $\mathscr{B}[3^j]$ starting from a single individual of type m (mod 3^j) has counts of individuals $\mathbf{M}_k(\omega)$ satisfying*

$$\lim_{k\to\infty} \mathbf{M}_k(\omega)\left(\tfrac{3}{4}\right)^k = \mathbf{e}W_m \quad a.s., \tag{3.12}$$

where $\mathbf{e} = (1, 1, \ldots, 1)$ is the uniform distribution on the individual types and W_m is a scalar-valued random variable with

$$\mathrm{Prob}\{a < W_m < b\} = \int_a^b w_m(x)\,dx,$$

where $w_m(x)$ is strictly positive on $(0, \infty)$.

PROOF. Let $\mathbf{E}_j^{(b)}$ represent the transition matrices for the branching probabilities of offspring having parity $b = 0$ or 1 for the process $\mathscr{B}[3^j]$. The *mean*

TABLE 2
Branching probabilities for branching processes $\mathscr{B}[3^j]$, $j = 0, 1, 2$

(a) $\mathscr{B}[1]$

$H_0^{(0)} = [1]$ $\qquad$ $H_0^{(1)} = [1/3]$

(b) $\mathscr{B}[3]$

$H_1^{(0)}$:

	1	2
1	0	1
2	1	0

$H_1^{(1)}$:

	1	2
1	0	0
2	1/3	1/3

(c) $\mathscr{B}[9]$

$H_2^{(0)}$:

	1	4	7	2	5	8
1		**0**		1	0	0
4				0	0	1
7				0	1	0
2	0	1	0		**0**	
5	1	0	0			
8	0	0	1			

$H_2^{(1)}$:

	1	4	7	2	5	8
1		**0**			**0**	
4						
7						
2	1/3	1/3	1/3	0	0	0
5	0	0	0	0	0	0
8	0	0	0	1/3	1/3	1/3

matrix of the underlying multitype Galton–Watson process is

$$\mathbf{E}_j = \mathbf{E}_j^{(0)} + \mathbf{E}_j^{(1)};$$

that is, $\mathbf{E}_j[k, l]$ is the expected number of offspring of type k from an individual of type l.

CLAIM. For each j, $\mathbf{E}_j$ has maximal eigenvalue $\frac{4}{3}$, which is simple and has left eigenvector $\mathbf{e} = (1, 1, \ldots, 1)$.

PROOF OF CLAIM. The matrix $\mathbf{E}_j^{(0)}$ is a permutation matrix sending $m \to 2m \pmod{3^j}$, hence has a left eigenvector $\mathbf{e}$ with eigenvalue 1. The matrix $\mathbf{E}_j^{(1)}$ has all column sums equal to $\frac{1}{3}$ because the equation $m \equiv (2n - 1)/3 \pmod{3^k}$ has exactly one solution $n \pmod{3^{j+1}}$ for each $m \pmod{3^j}$ with $m \not\equiv 0 \pmod 3$. Hence it has $\mathbf{e}$ as left eigenvector with eigenvalue $\frac{1}{3}$ and $\mathbf{E}_j$ thus has the left eigenvector $\mathbf{e}$ with eigenvalue $\frac{4}{3}$.

It now suffices to show that $\mathbf{E}_j$ is *strictly positive* (also called *primitive*), that is, some power of $\mathbf{E}_j$ has all entries positive. For if so, then Perron–Frobenius theory guarantees that $\mathbf{E}_j$ has a simple real eigenvalue of maximum modulus, that this eigenvalue has a positive real eigenvector and that it is the only eigenvalue having a positive real eigenvector. Thus $\frac{4}{3}$ must be the simple maximal eigenvalue.

To show that $\mathbf{E}_j$ is strictly positive, one first uses the fact that the permutation in $\mathbf{E}_j^{(0)}$ is cyclic because 2 is a primitive root $(\bmod 3^j)$ for all j. Second, $\mathbf{E}_j^{(1)}$ has the fixed point -1 $(\bmod 3^j)$. Consequently $(\mathbf{E}_j^{(0)} + \mathbf{E}_j^{(1)})^L$ has all positive entries for $L = 4 \cdot 3^{j-1}$. □

This claim and its proof show that for all j the process $\mathscr{B}[3^j]$ is supercritical, positive regular and nonsingular [in the terminology of Athreya and Ney (1972)]. Now Theorem 3.2 follows directly from Athreya and Ney [(1972), Theorems 1 and 2, page 192]. □

COROLLARY 3.1. *For all $j \geq 0$ the branching process $\mathscr{B}[3^j]$ has*

$$\lim_{k \to \infty} \frac{\log N_k(\omega)}{k} = \delta_{\mathrm{BP}} \quad a.s.,$$

where $\delta_{\mathrm{BP}} = \log \frac{4}{3}$.

PROOF. This follows from Theorem 3.2 since the density $w_m(x)$ has no mass at 0. □

What is the expected location of members of the kth generation? By (3.9) this is determined by the expected density ratio. We prove a result suggesting that the expected density ratio is about $\frac{1}{4}$ for $\mathscr{B}[1]$.

THEOREM 3.3. *For the branching random walk $\mathscr{B}[1]$ let*

$$V_k(\omega) \cdot k \sum_{i=1}^{N_k(\omega)} r(\omega_{k,i}),$$

which is the sum of parities over the paths to all leaves in the tree $\mathscr{T}_k(\omega)$. Then

$$E[N_k] = \left(\tfrac{4}{3}\right)^k,$$

$$E[V_k] = \tfrac{1}{4} k E[N_k] = \tfrac{1}{4} k \left(\tfrac{4}{3}\right)^k.$$

PROOF. Look at the root node of the tree $\mathscr{T}_k(\omega)$. With probability $\frac{2}{3}$ it has a single edge of parity 0 attached to a tree $\mathscr{T}_{k-1}(\omega_1)$, while with probability $\frac{1}{3}$ it has two edges of parity 0 and 1, respectively, attached to two trees $\mathscr{T}_{k-1}(\omega_1)$ and $\mathscr{T}_{k-1}(\omega_2)$. This gives the recursions

$$E[N_k] = \tfrac{2}{3} E[N_{k-1}] + \tfrac{1}{3}(2E[N_{k-1}]),$$

$$E[V_k] = \tfrac{2}{3} E[V_{k-1}] + \tfrac{1}{3}(2E[V_{k-1}] + E[N_{k-1}]),$$

which when solved give $E[N_k] = (\frac{4}{3})^k$, $E[V_k] = \frac{1}{4} k (\frac{4}{3})^k$, using $E[N_0] = 1$, $E[V_0] = 0$. □

It can be shown for $\mathscr{B}[1]$ that all but an exponentially small fraction of $\{r(\omega_{k,i})\colon 1 \le i \le N_k(\omega)\}$ are in an interval $[\frac{1}{4} - \varepsilon, \frac{1}{4} + \varepsilon]$, by using a Chernoff bound for branching random walks; compare with Biggins (1977). (A similar result undoubtedly holds for all $\mathscr{B}[3^j]$.)

Kingman (1975) showed that any realization ω of a single-type Crump–Mode process almost surely has an asymptotic first birth which is a constant β depending on the process. Biggins [(1976), Section 6] extended this result to multitype branching random walks. Applying Biggins' method, we obtain the following result.

THEOREM 3.4. *There is a constant β_{BP} such that for all $j \ge 0$ the branching process $\mathscr{B}[3^j]$ has asymptotic first birth*

$$\lim_{k\to\infty} \frac{1}{k} L_k^*(\omega) = \beta_{\mathrm{BP}} \quad a.s. \tag{3.13}$$

This constant $\beta_{\mathrm{BP}} \doteq 0.02399$ is determined uniquely by $\beta_{\mathrm{BP}} > 0$ and

$$\tilde{g}(\beta_{\mathrm{BP}}) = 0, \tag{3.14}$$

where

$$\tilde{g}(a) := -\sup_{\theta \le 0}\Big(a\theta - \log\big(2^{\theta} + \tfrac{1}{3}\big(\tfrac{2}{3}\big)^{\theta}\big)\Big). \tag{3.15}$$

This theorem shows there is a *branching process stopping constant* γ_{BP}, analogous to γ, which is defined by

$$\begin{aligned}\gamma_{\mathrm{BP}} &:= \limsup_{k\to\infty} \frac{k}{L_k^*(\omega)} \quad \text{a.s.}\\ &= (\beta_{\mathrm{BP}})^{-1}.\end{aligned} \tag{3.16}$$

PROOF. The proof of Theorem 3.2 showed that the process $\mathscr{B}[3^j]$ is supercritical and positive regular, hence the main result of Biggins [(1976), Section 6] applies and shows that (3.13) holds with a constant β_j depending on the process $\mathscr{B}[3^j]$.

To compute β_j, we follow Biggins [(1976), equation (2.5)] and associate to $\mathscr{B}[3^j]$ a *Laplace transform matrix* $\Phi^{(j)}(\theta)$ depending on a parameter $\theta \ge 0$, as follows. Let Z_n^1 denote the point process for births by an individual of type n and $E_m[\cdot]$ the expectation for individuals of type m. Then $\Phi^{(j)}(\theta)$ has entries

$$\begin{aligned}\Phi_{mn}^{(j)} &= E_m\left[\int_{-\infty}^{\infty} e^{-\theta t}\, dZ_n^1(t)\right]\\ &= \exp(-(\log 2)\theta)\mathbf{E}_j^{(0)} + \exp(-(\log 2 - \log 3)\theta)\big(\mathbf{E}_j^{(0)} + \mathbf{E}_j^{(1)}\big).\end{aligned}$$

The proof of Theorem 3.2 showed that $\mathbf{E}_j^{(0)}$ and $\mathbf{E}_j^{(1)}$ both have eigenvector $\mathbf{e} = (1, 1, \dots, 1)$ with eigenvalues $1, \frac{1}{3}$, respectively. It is then easy to show (as in the claim in Theorem 3.2) that for all $\theta \ge 0$, $\Phi(\theta)$ is a strictly positive matrix having a simple maximal eigenvalue

$$\phi(\theta) = \exp(-(\log 2)\theta) + \tfrac{1}{3}\exp(-(\log 2 - \log 3)\theta) = 2^{-\theta}(1 + \tfrac{1}{3}3^{\theta})$$

with associated eigenvector $\mathbf{e}$. The constant β_j is solely a function of $\phi(\theta)$, and since $\phi(\theta)$ is independent of j, β_j does not depend on j, so is a constant β_{BP}. Biggins defines

$$\mu(a) = \inf\{e^{\theta a}\phi(\theta) : \theta \geq 0\} \tag{3.17}$$

and shows that

$$\beta_{\mathrm{BP}} := \inf\{a : \mu(a) \geq 1\}. \tag{3.18}$$

On taking logarithms, we obtain

$$\begin{aligned} \tilde{g}(a) := \log \mu(a) &= \inf_{\theta \geq 0}\{a\theta + \log \phi(\theta)\} \\ &= -\sup_{\theta \geq 0}\{-a\theta - \log \phi(\theta)\} \\ &= -\sup_{\theta \leq 0}\{a\theta - \log \tilde{M}(\theta)\}, \end{aligned}$$

where

$$\tilde{M}(\theta) := \exp((\log 2)\theta) + \tfrac{1}{3}\exp((\log \tfrac{2}{3})\theta) \equiv \phi(-\theta) \tag{3.19}$$

is the moment generating function of the locations of the progeny of an individual of $\mathscr{B}[1]$. It is easy to check that $\tilde{g}(a)$ is continuous and strictly increasing on the interval $-\log \frac{3}{2} < a \leq \frac{1}{4}\log\frac{16}{3}$, and has constant value $\tilde{g}(a) = \log\frac{4}{3}$ for $\frac{1}{4}\log\frac{16}{3} \leq a < \infty$. On choosing $a = 0$ and $\theta = 1 + \varepsilon$ one sees that $\tilde{g}(0) < 0$; hence there is a unique value $\beta > 0$ with $\tilde{g}(\beta) = 0$. Then $\mu(\beta) = 1$; hence by (3.18), $\beta = \beta_{\mathrm{BP}}$. □

One can consider more complicated branching random walks $\mathscr{B}[3^j, d]$ that perfectly mimic the $3x + 1$ inverse map for trees of depth d. Given a residue class $m \pmod{3^j}$, choose uniformly with probability 3^{-d} a residue class $\tilde{m} \pmod{3^{j+d}}$ with $\tilde{m} \equiv m \pmod{3^j}$, and assign the tree $(T^*)^{-d}(\tilde{m})$, which has all its leaves (offspring) specified $(\bmod\, 3^j)$. Leaves are uniquely specified by a sequence $(x_1, x_2, \ldots, x_d)$ of parity labels read from the root of $(T^*)^{-d}(\tilde{m})$ and are assigned the locations $d\log 2 - (x_1 + \cdots + x_d)\log 3$. It is possible to prove that the Laplace transform matrix of such a process has largest eigenvalue $(\frac{4}{3})^d$, and the associated eigenvalue function $\phi_d(\theta) = [\phi(\theta)]^d$, so that the analogue of Theorem 3.4 holds for these processes as well.

The trees produced by branching processes $\mathscr{B}[3^j]$ differ from the trees produced by the $3x + 1$ function in that they have more variability. Any $3x + 1$ process tree $\mathscr{T}_d^*(m)$ is determined by the congruence class of $m \pmod{3^{d+1}}$ so that there are at most $2 \cdot 3^d$ distinct edge-labelled trees $\mathscr{T}_d^*(m)$ that are possible. By contrast, the total number of trees of depth d that can be realized by the branching random walk $\mathscr{B}[3^j]$ is at least $\exp(\exp(c_j d))$ as $d \to \infty$, for some positive constant c_j.

The "last birth" in generation k of the branching random walk $\mathscr{B}[3^j]$ is analogous to the logarithm of the quantity $H^+(k) = \max\{n\colon \sigma_\infty(n) = k\} = 2^k$ for the $3x + 1$ function. It is easy to check directly that this analogy is perfect: For each $\mathscr{B}[3^j]$ the "last birth" in the kth generation is at location $k \log 2$.

4. Relationships between random walk and branching process models. The independent random walks model represents enumerating $3x + 1$ starting values n in order of increasing n, while the branching random walks model represents enumerating $3x + 1$ starting values n in order of increasing $\sigma_\infty(n)$. These orderings are quite different. For the first ordering, the density ratio in the trajectory of a "random" n for the independent random walks process is about $\frac{1}{2}$, while for the branching random walk models $\mathscr{B}[3^j]$ with the nodes ordered by increasing depth k [corresponding to $\sigma_\infty(n)$] the expected density ratio in a trajectory to the root node is about $\frac{1}{4}$ (Theorem 3.3). Besides the difference in their ordering of instances, the branching process model incorporates dependencies between different individuals while the independent random walks model does not.

Nevertheless, the two models are related in a way that makes them exhibit the same asymptotic behavior $\gamma_{\mathrm{RW}} = \gamma_{\mathrm{BP}}$. This relationship is on the level of their moment generating functions. A step X of the independent random walks model has the moment generating function

$$M_{\mathrm{RW}}(\theta) := E[e^{\theta X}] = \tfrac{1}{2}\left(\tfrac{1}{2}\right)^{\theta} + \tfrac{1}{2}\left(\tfrac{3}{2}\right)^{\theta}, \tag{4.1}$$

while the point process Z^1 for birth locations in the branching random walk model $\mathscr{B}[1]$ has moment generating function

$$M_{\mathrm{BP}}(\theta) := E\left[e^{\theta Z^1}\right] = 1 \cdot 2^{\theta} + \tfrac{1}{3}\left(\tfrac{2}{3}\right)^{\theta}. \tag{4.2}$$

[The validity of (4.2) for $\mathscr{B}[3^j]$ depends on the fortuitous fact that $M_{\mathrm{BP}}(\theta)$ is the largest eigenvalue in the moment generating matrix $\Phi^{(j)}(\theta)$; see (3.19).] The *duality relation* is

$$M_{\mathrm{BP}}(\theta) \equiv M_{\mathrm{RW}}(-\theta - 1). \tag{4.3}$$

This is a duality in the sense that the map $\theta \to -1 - \theta$ is an involution. More generally, one may define two point processes to be *dual* in this sense if their moment generating functions are related by (4.3).

We use the duality relation to derive the following result.

THEOREM 4.1. *$\gamma_{\mathrm{RW}} = \gamma_{\mathrm{BP}}$.*

PROOF. The formulae for γ_{RW} and γ_{BP} derived in Sections 2 and 3 are expressed in terms of Legendre transforms of suitable moment generating functions. Theorem 2.1 asserts that $\gamma_{\mathrm{RW}} > (\frac{1}{2}\log\frac{4}{3})^{-1}$ satisfies

$$g\left(\frac{1}{\gamma_{\mathrm{RW}}}\right) = \frac{1}{\gamma_{\mathrm{RW}}},$$

where

$$g(a) := \sup_{a \in \mathbb{R}} (a\theta - \log M_{\mathrm{RW}}(-\theta)),$$

since $M(\theta) = M_{\mathrm{RW}}(-\theta)$. Theorem 3.4 asserts that $\gamma_{\mathrm{BP}} > 0$ is the unique positive solution of

$$\tilde{g}\left(\frac{1}{\gamma_{\mathrm{BP}}}\right) = 0,$$

where

$$\tilde{g}(a) := -\sup_{a \le 0} (a\theta - \log M_{\mathrm{BP}}(\theta)), \tag{4.4}$$

since $\tilde{M}(\theta) = M_{\mathrm{BP}}(\theta)$. On the range $0 < a < \frac{1}{4}\log\frac{16}{3}$ we can replace (4.4) by the Legendre transform

$$\tilde{g}(a) = -\sup_{a \in \mathbb{R}} (a\theta - \log M_{\mathrm{BP}}(\theta)), \tag{4.5}$$

using the expression (4.2) for $M_{\mathrm{BP}}(\theta)$ to verify this. The duality relation $M_{\mathrm{RW}}(\theta) = M_{\mathrm{BP}}(-\theta - 1)$ now gives

$$\tilde{g}(a) \equiv -g(a) + a, \tag{4.6}$$

for $0 < a < \frac{1}{4}\log\frac{16}{3}$. Since $1/\gamma_{\mathrm{RW}}$ and $1/\gamma_{\mathrm{BP}}$ fall in this interval and $\tilde{g}(a)$ is monotone increasing there, (4.6) implies that $1/\gamma_{\mathrm{RW}} = 1/\gamma_{\mathrm{BP}}$. □

Next we use the duality relation to show that when the individuals produced by a realization of the branching random walk model $\mathscr{B}[1]$ are ordered by increasing size, their distribution resembles that of the starting positions $\exp(Z^*(n, 0))$ of the independent random walks process, which consists of one individual at each of $n = 1, 2, 3, \dots$.

THEOREM 4.2. *For a realization* ω *of the process* $\mathscr{B}[1]$ *let*

$$I^*(t, \omega) := \#\{\omega_{k,j}: S(\omega_{k,j}) \le t\}$$

count the number of progeny of size $\le t$. *Then almost surely*

$$I^*(t, \omega) = t^{1+o(1)} \quad \text{as } t \to \infty. \tag{4.7}$$

PROOF. To estimate $I^*(t, \omega)$ we consider the sets

$$I_k^*(t, \omega) = \#\{\omega_{k,j}: k \text{ fixed and } S(\omega_{k,j}) \le t\}.$$

Set $a = k/\log t$. First observe that for $a \le \frac{1}{4}\log\frac{16}{3}$ one has the estimate

$$\begin{aligned} I_k^*(t, \omega) \le N_k(\omega) &\le \left(\tfrac{4}{3}\right)^{k(1+o(1))} \\ &\le t^{\alpha_c(1+o(1))} \quad \text{as } t \to \infty, \end{aligned} \tag{4.8}$$

where $\alpha_c = \frac{1}{4}(\log\frac{16}{3})(\log\frac{4}{3}) \doteq 0.120393$, by using Theorem 3.2. These individuals contribute negligibly to (4.7).

For the remaining range $\alpha > \frac{1}{4}\log\frac{16}{3}$, the quantity $I_k^*(t,\omega)$ measures a "tail event" which can be estimated using the Chernoff bound of Biggins [(1977), Theorem 1]. We obtain

$$I_k^*(k,\omega) = \left(\mu\left(\frac{1}{a}\right)\right)^{k(1+o(1))} \tag{4.9}$$
$$= t^{a\tilde{g}(1/a)(1+o(1))} \quad \text{as } t \to \infty$$

almost surely, where $\mu(a)$ is given by (3.17). Using the duality relation embodied in (4.6), one has for $a > \frac{1}{4}\log\frac{16}{3}$ that

$$a\tilde{g}\left(\frac{1}{a}\right) = a\left(\frac{1}{a} - g\left(\frac{1}{a}\right)\right) \le 1 \tag{4.10}$$

and equality holds exactly for $a = (\frac{1}{2}\log\frac{4}{3})^{-1}$ by Lemma 2.2. Then $k = (\frac{1}{2}\log\frac{4}{3})^{-1}\log t$ and one has

$$I^*(t,\omega) \ge I_k^*(t,\omega) \ge t^{1+o(1)} \quad \text{as } t \to \infty$$

almost surely, by (4.9). Finally one has

$$I^*(t,\omega) \le (60\log t)\max\{I_k^*(t,\omega) : 0 \le k \le 50\log t\}$$
$$\le t^{1+o(1)} \quad \text{as } t \to \infty$$

almost surely, using (4.8)–(4.10). [Actually one applies (4.7) for $a \le (\frac{1}{4}\log\frac{16}{3}) + 0.001$ and notes that (4.8) and (4.9) can then be made uniform in k as $t \to \infty$.] □

One can also check that the individuals near a given size produced by the branching process model $\mathscr{B}[1]$ have the density ratio $\frac{1}{2}$ expected for samples from the independent random walks model. Indeed the value $a = (\frac{1}{2}\log\frac{4}{3})^{-1}$ corresponds to the density ratio $r(a) = \frac{1}{2}$, and one can prove that the density ratio

$$r^*(t,\omega) := \frac{1}{I^*(t,\omega)}\sum_{k,j}\{r(\omega_{k,j}) : S(\omega_{k,j}) \le t\}$$

has $r^*(t,\omega) \to \frac{1}{2}$ almost surely as $t \to \infty$.

There are analogues of Theorems 4.1 and 4.2 for an arbitrary pair of dual branching random walks, provided that they satisfy some extra boundedness and sign conditions on their expected step sizes. The extra conditions are needed to complete those parts of the proof that go from (4.4) to (4.5).

5. Empirical results for the $3x + 1$ function. We compare predictions of the stochastic models of Sections 2 and 3 to the actual behavior of the $3x + 1$ function iterates using tables of the successive maximum values of $t(n)$ and $\sigma_\infty(n)$ for $n \le 10^{11}$ computed by Leavens (1989).

Table 3 gives the largest value of $(\log t(n))/\log n$ attained on each interval $10^k \le n \le 10^{k+1}$ for $1 \le k \le 10$. The data seem in excellent agreement with the random walks model's prediction that $\rho_{\mathrm{RW}} = 2$. Table 3 also gives data on

TABLE 3

Largest value of $\frac{\log t(n)}{\log n}$ *for* $10^k \le n \le 10^{k+1}$, $1 \le k \le 10$

k	n	$\frac{\log t(n)}{\log n}$	$\frac{\sigma_{\max}(n)}{\log n}$	$\gamma(n)$
1	27	2.560	13.65	21.24
2	703	1.791	7.32	16.48
3	9,663	1.790	2.83	12.86
4	77,671	1.819	3.46	13.14
5	704,511	1.788	2.75	11.59
6	6,631,675	1.976	5.86	23.05
7	80,049,391	1.903	5.06	19.84
8	319,804,831	2.099	4.65	19.10
9	8,528,817,511	1.909	5.20	20.03
10	77,566,362,559	1.897	6.86	19.02
Random walks model		2.000	7.645	21.55

$(\sigma_\infty(n))/\log n$ and $(\sigma_{\max}(n))/\log n$, where

$$\sigma_{\max}(n) := \min\{k : T^{(k)}(n) \ge T^{(l)}(n) \text{ for all } l \ge 0\}.$$

Recall that the random walks model predicts that $\sigma_\infty(n) \sim 21.55 \log n$ and $\sigma_{\max}(n) \sim 7.645 \log n$ (asymptotically) for those n having $t(n) = n^{2+o(1)}$ as $n \to \infty$. Furthermore the trajectories of such n plotted logarithmically should (asymptotically) approach two line segments of prescribed slopes. Figure 2 gives scaled logarithmic plots $(k/\log n, (\log T^{(k)}(n))/\log n)$ of the trajectories

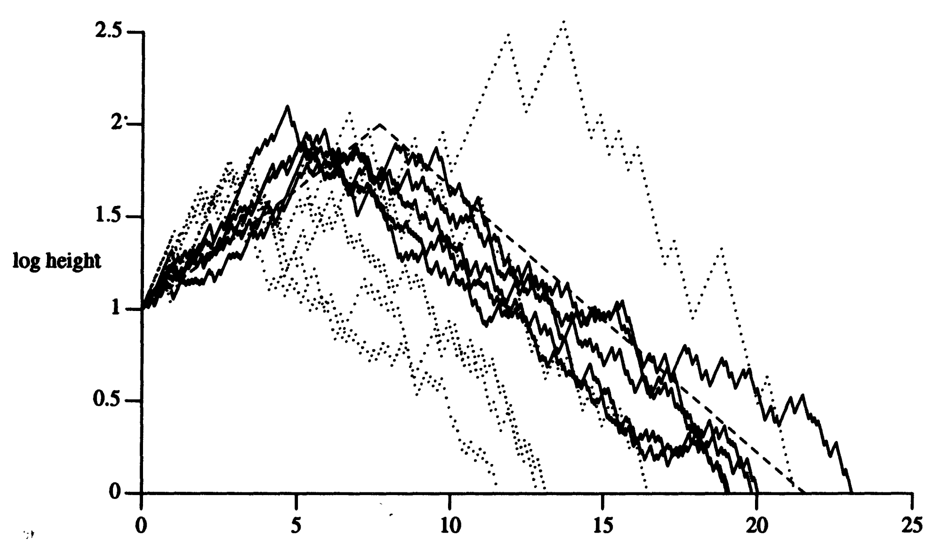

FIG. 2. *Scaled trajectories of* n_k *maximizing* $(\log t(n))/\log n$ *in* $10^k \le n \le 10^{k+1}$ *(dotted for* $1 \le k \le 5$; *solid for* $6 \le k \le 10$).

TABLE 4
Maximal value of $\gamma(n)$ in intervals $10^k < n \le 10^{k+1}$, $1 \le k \le 10$

k	n	$\sigma_\infty(n)$	$\gamma(n)$	$r(n)$
1	27	70	21.24	0.5857
2	703	108	16.48	0.5741
3	6,171	165	18.91	0.5818
4	52,527	214	19.68	0.5841
5	837,799	329	24.13	0.5927
6	8,400,511	429	26.91	0.5967
7	63,728,127	592	32.94	0.6030
8	127,456,254	593	31.77	0.6020
9	4,890,328,815	706	31.64	0.6020
10	13,371,194,527	755	32.38	0.6026
Random walks model			41.68	0.6091

for those n in Table 3; those trajectories for $1 \le k \le 5$ are dotted and for $6 \le k \le 10$ are continuous. The limiting trajectory predicted by the random walks model is indicated by dashed lines; it seems to be in good agreement with the data.

Table 4 gives the largest value of $\gamma(n) = (\sigma_\infty(n))/\log n$ attained on each interval $10^k < n \le 10^{k+1}$ for $1 \le k \le 10$. Excluding the small value $n_0 = 27$ there is a steady increase of the values of $\gamma(n)$ followed by a possible leveling off around a value of 32, which is somewhat less then Conjecture 1 predicts. Is this evidence in serious conflict with Conjecture 1? Large deviation theory predicts that "extremal" trajectories will exhibit a graph that appears roughly linear, with a slope -0.024. Figure 3 gives scaled logarithmic plots of the trajectories for those n in Table 4; those trajectories for $1 \le k \le 5$ are dotted and those for $6 \le k \le 10$ are solid. The limiting trajectory predicted by the random walks model is indicated by a dashed line.

We call the predicted extremal straight-line paths "train tracks" following Vyssotsky (1987). It is well known that for constant coefficient random walks the easiest way to do something improbable is to take a moderate pace, like a train climbing a mountain. In this train analogy, the trajectories going highest (as in Figure 2) are not the longest (Figure 3) because they expend their energy climbing and then must plunge down rapidly while a more conservative path is able to stay up longer. The actual extremal trajectories in Figure 3 seem to exhibit three regimes of behavior: an initial rise, a long "train track" portion and a final rapid plunge. This differs from the predicted "train track" in explicable ways. Since these trajectories are extremal, they must have an initial rise, since otherwise some trajectory with a lower starting point would be better. The final plunge comes from the nonrandom behavior of small ($\le 10^6$) numbers; none of these has $\sigma_\infty(n) > 24.15 \log n$. For these reasons we think that the data of Table 4 are consistent with Conjecture 1.

We examine "train track" trajectories further in the next section, using backwards iteration.

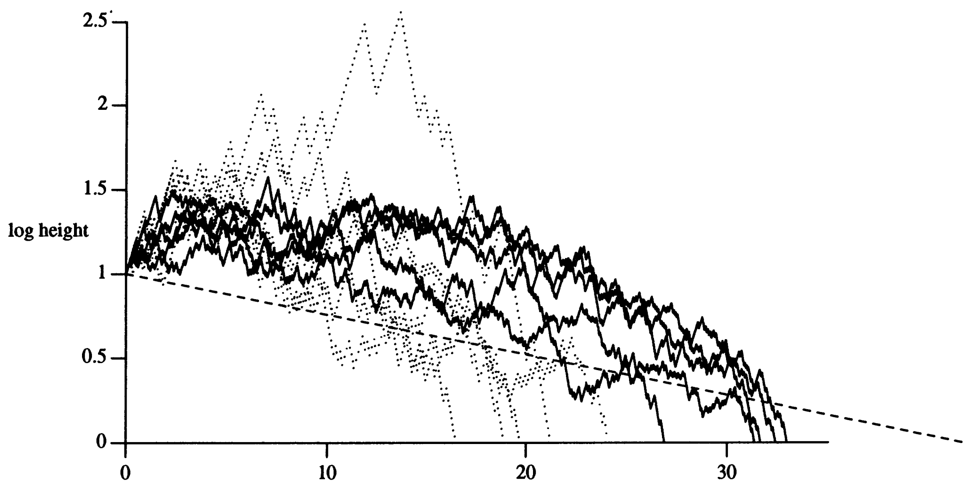

FIG. 3. *Scaled trajectories of n_k maximizing $\gamma(n)$ in $10^k \le n \le 10^{k+1}$ (dotted for $1 \le k \le 5$; solid for $6 \le k \le 10$).*

6. Greedy algorithms to find large stopping times. Vyssotsky (1987) suggested that one search for large n having large values of $(\sigma_\infty(n))/\log n$ by a backwards search using a "greedy" algorithm.

GREEDY ALGORITHM $\mathscr{G}_d^*$. *Recursively find $\{m_{kd}: k \ge 0\}$ by setting $m_0 = 4$ and choosing m_{kd} to be the smallest member of $(T^*)^{-d}(m_{(k-1)d})$; that is, the smallest leaf label in the tree $\mathscr{T}_d^*(m_{(k-1)d})$.*

The trajectories located by such an algorithm exhibit "train track" behavior with a characteristic slope depending on the depth d of the tree search.

We can model the behavior of this greedy algorithm by an analogous greedy algorithm for the branching processes of Section 3.

GREEDY ALGORITHM $\mathscr{G}_d[3^j]$. *Given a realization ω of the branching random walk $\mathscr{B}[3^j]$, find $\{\omega_{kd}^*: k \ge 0\}$ by taking ω_0^* to be the root and ω_{kd}^* to be that node at level kd of the tree $\mathscr{T}(\omega)$ of minimal height among all descendants of $\omega_{(k-1)d}^*$. (In case of ties, choose ω_{kd}^* to be first in the ordering $\omega_{kd,j}$ of nodes at level kd.)*

The performance of these algorithms can in principle be analyzed exactly, for any fixed values of d and j.

THEOREM 6.1. *For the algorithm $\mathscr{G}_d[1]$ there are constants $\{\beta(d): d = 1, 2, \ldots\}$ such that*

$$\lim_{k \to \infty} \frac{1}{kd} L(\omega_{kd}^*) = \beta(d) \quad a.s. \tag{6.1}$$

One has

$$\lim_{d \to \infty} \beta(d) = \beta_{\mathrm{BP}}. \tag{6.2}$$

PROOF. We establish (6.1). Each step of the process $\{\omega_{kd}^*\}$ consists of independently drawing a depth d tree generated by $\mathscr{B}[1]$ with root at $\omega_{(k-1)d}^*$, then choosing ω_{kd}^* to be at the end of that branch from $\omega_{(k-1)d}^*$ having minimal value of $L(\omega_{kd}^*) - L(\omega_{(k-1)d}^*)$. The central limit theorem shows that $\beta(d)$ exists and is the expected value of the minimal location ratio for this random variable. The existence of the limit (6.2) and the assertion that its value is β_{BP} follow from Biggins [(1976), Theorem 4]. □

An equivalent form of Theorem 6.1 is that there are *limiting density ratios* $\{r(d)\colon d = 1, 2, \dots\}$ such that

$$\lim_{k \to \infty} r(\omega_{kd}^*) = r(d), \quad \text{a.s.}, \tag{6.3}$$

and $\lim_{r \to \infty} r(d) = r_{\mathrm{BP}} \doteq 0.60909$. The equivalence follows from the relation (3.10), which also shows that

$$\beta(d) = \log 2 - r(d) \log 3. \tag{6.4}$$

There is a simple recursion for computing the value of $r(d)$. Set

$$p_{d,j} = \mathrm{Prob}[\mathscr{T}_d(\omega) \text{ has no branch with } \geq j + 1 \text{ ones}].$$

TABLE 5
Limiting densities for probabilistic greedy algorithm for $\mathscr{G}_d[1]$

Depth d	Ones ratio $r(d)$	γ_d
1	0.33333	3.0586
2	0.33333	3.0586
3	0.33882	3.1161
4	0.34649	3.2001
5	0.35508	3.2999
10	0.39811	3.9095
20	0.45790	5.2606
30	0.49179	6.5420
40	0.51291	7.7126
50	0.52729	8.7825
100	0.56114	13.0438
200	0.58183	18.5376
400	0.59387	24.5634
600	0.59832	27.9213
800	0.60069	30.0124
1000	0.60216	31.6471
∞	0.60909	41.6776

Then

$$r(d) = \sum_{j=1}^{d} j(p_{d,j} - p_{d,j+1}) = \sum_{j=1}^{d} p_{d,j} \tag{6.5}$$

because $p_{d,d+1} = 0$. The definition of the process $\mathscr{B}[1]$ implies that the $p_{d,j}$ satisfy the recursion

$$dp_{d,j} = \tfrac{2}{3}(d-1)p_{d-1,j} + \tfrac{1}{3}(d-1)^2 p_{d-1,j}p_{d-1,j-1}. \tag{6.6}$$

Using $p_{0,0} = 1$ we can use this recurrence to evaluate all $p_{d,j}$ and hence $r(d)$ for small d. Table 5 gives values of $r(d)$ and the associated stopping time estimates $\gamma_d = \beta(d)^{-1}$ obtained in this fashion. (These values were computed using 100 digits precision, which was necessary to allow for the effects of roundoff error.)

The values $r(d)$ are apparently strictly increasing for $d \geq 3$, but we have no proof of this fact. Now $r(d)$ does have the subadditivity property

$$dr(d) \geq jr(j) + (d-j)r(d-j), \qquad 1 \leq j \leq d-1, \tag{6.7}$$

and one can use this and the data for $d \leq 68$ to show $r(d) > 1/2$ for all $d \geq 34$.

The predicted densities of the greedy algorithm $\mathscr{S}_d[1]$ in Table 5 may be compared with the actual $3x + 1$ data given in Table 4. The largest value n in Table 4 takes 755 iterates and has $\gamma(n) = 32.38$, while the "greedy" search to depth 800 using $\mathscr{S}_d[1]$ predicts a constant of 30.01.

TABLE 6
Greedy algorithm Markov chains for $d = 1$

	1	2
1	0	1
2	2/3	1/3

(a) $\mathscr{B}[3]$ *Markov chain*, $v = (\frac{2}{5}, \frac{3}{5})$

<table>
<tr><th></th><th>1</th><th>4</th><th>7</th><th>2</th><th>5</th><th>8</th></tr>
<tr><td>1</td><td colspan="3" rowspan="3">0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>7</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>2</td><td>1/3</td><td>1/3</td><td>1/3</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>5</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>8</td><td>0</td><td>0</td><td>0</td><td>1/3</td><td>1/3</td><td>1/3</td></tr>
</table>

(b) $\mathscr{B}[9]$ *Markov chain*, $v = (\frac{5}{21}, \frac{2}{21}, \frac{2}{21}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7})$

TABLE 7
Empirical densities for greedy algorithm $\mathscr{G}_1^*$

k	1	2	4	5	7	8	β
700	0.2379	0.2792	0.0997	0.1496	0.0912	0.1425	4.27
7000	0.2395	0.2863	0.0954	0.1418	0.0934	0.1435	4.52
Predicted density	0.2381 $\frac{5}{21}$	0.2857 $\frac{2}{7}$	0.0952 $\frac{2}{21}$	0.1429 $\frac{1}{7}$	0.0952 $\frac{2}{21}$	0.1429 $\frac{1}{7}$	4.498

The greedy algorithm $\mathscr{G}_d[1]$ does not, however, give a completely accurate model of the behavior of the algorithm $\mathscr{G}_d^*$. This is because when we choose sequences $\{\omega_{kd}^*\}$ having many parity 1 terms, the distribution of their values $(\text{mod } 3^j)$ cannot be uniformly distributed, which the model $\mathscr{B}[1]$ assumes.

In consequence we study the greedy algorithm $\mathscr{G}_1[3^j]$ that does a depth one optimization by always choosing a node ω_1^* of parity 1 if possible. This node choice creates a Markov chain in which all states are not equally probable. The resulting Markov chains for $j = 1$ and 2 are given in Table 6 along with the associated left eigenvector giving the steady state probabilities. The density ratio for $\mathscr{G}_1[3]$ is $r_1(1) = \frac{2}{5} \doteq 0.40000$ and for $\mathscr{G}_1[9]$ is $r_2(1) = \frac{3}{7} \doteq 0.42857$. Nothing new happens for $\mathscr{G}_1[3^j]$ for $j \geq 3$. The steady state vectors for the class $m \ (\text{mod } 3^j)$ are $3^{-(j-2)}$ times that for $m \ (\text{mod } 9)$ and $r_j(1) = \frac{3}{7}$ for all $j \geq 3$.

It appears that $\mathscr{G}_1[9]$ gives a qualitatively accurate model of the one-step greedy algorithm $\mathscr{G}_1^*$ for the $3x + 1$ problem. Table 7 gives statistics for 7000 iterations of $\mathscr{G}_1^*$ showing excellent agreement with this Markov chain.

One can construct more complicated Markov chains that simulate the behavior of the greedy algorithm $\mathscr{G}_d^*$. It appears likely that the "correct" model is the Markov chain $(\text{mod } 3^{d+1})$ arising from the algorithm that chooses the branch with the largest density ratio of a depth d tree generated by the process $\mathscr{B}[3^{d+1}, d]$ (defined at the end of Section 3). These models are so complicated that it may be easier to study $\mathscr{G}_d^*$ directly.

Table 8 presents empirical data on the densities of greedy algorithms $\mathscr{G}_d^*$ for $d = 5$, 10, 15, 20, 25 used to construct values m_{6000} whose 6000th iterate is 4. The values of $\gamma(m_{6000})$ in this table may be compared with the values for

TABLE 8
"Greedy" algorithm $\mathscr{G}_d^*$ applied to $m_0 = 4$

Search depth d	$\log(m_{6000})$	$\gamma(m_{6000})$
5	1116.28	5.38
10	1043.29	5.75
15	849.93	7.06
20	808.20	7.43
25	745.56	8.05

$\mathscr{S}_d[1]$ made in Table 5. Note that the depth d' needed in $\mathscr{S}_{d'}[1]$ to attain the same density as observed for $d = 25$ in $\mathscr{S}_d^*$ is about 45.

Vyssotsky (1987) used a "greedy" algorithm to find a number $n_0 \doteq 9.823 \times 10^{160}$, which has $\sigma_\infty(n_0) = 6000$, so that $(\sigma_\infty(n_0))/\log n_0 \doteq 16.186$. For the random walk model, Theorem 2.1 asserts that the probability a "random" n in $e^{370} \le n \le e^{371}$ has such a large value of α is $\le e^{-300}$. Thus backwards search is apparently much more efficient than "random" search for finding extreme values of $(\sigma_\infty(n))/\log n$.

Acknowledgments. We are indebted to V. S. Vyssotsky for describing his computations to us and to A. M. Odlyzko for warnings about possible numerical instability of computations using the recursion (6.5). We are indebted to a referee for simplification of some proofs in Section 2.

REFERENCES

ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York.

AZENCOTT, R. and RUGET, G. (1977). Melanges d'équations differentielles et grandes écarts à la loi des grandes nombres. *Z. Warsch. Verw. Gebiete* **38** 1–54.

BIGGINS, J. D. (1976). The first and last birth problems for a multitype age-dependent branching process. *Adv. in Appl. Probab.* **8** 446–459.

BIGGINS, J. D. (1977). Chernoff's theorem in the branching random walk. *J. Appl. Probab.* **14** 630–636.

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

CRANDALL, R. E. (1978). On the "$3x + 1$" problem. *Math. Comp.* **32** 1281–1292.

CRUMP, K. S. and MODE, C. J. (1968). A general age-dependent branching process I. *J. Math. Anal. Appl.* **24** 494–508.

CRUMP, K. S. and MODE, C. J. (1969a). A general age-dependent branching process II. *J. Math. Anal. Appl.* **25** 8–17.

CRUMP, K. S. and MODE, C. J. (1969b). An age-dependent process with correlations among sister cells. *J. Appl. Probab.* **6** 205–210.

EVERETT, C. J. (1977). Iteration of the number-theoretic function $f(2n) = n$, $f(2n + 1) = 3n + 2$. *Adv. in Math.* **25** 42–45.

KINGMAN, J. F. C. (1975). The first birth problem for an age-dependent branching process. *Ann. Probab.* **3** 790–801.

LAGARIAS, J. C. (1985). The $3x + 1$ problem and its generalizations. *Amer. Math. Monthly* **92** 3–23.

LEAVENS, G. T. (1989). A distributed search program for the $3x + 1$ problem. Technical Report TR89-22, Dept. Computer Science, Iowa State Univ.

LEIGH, G. M. (1986). A Markov process underlying the generalized Syracuse algorithm. *Acta Arith.* **46** 125–143.

MATTHEWS, K. R. and WATTS, A. M. (1984). A generalization of Hasse's generalization of the Syracuse algorithm. *Acta Arith.* **43** 167–175.

MATTHEWS, K. R. and WATTS, A. M. (1985). A Markov approach to the generalized Syracuse algorithm. *Acta Arith.* **45** 29–42.

NEY, P. (1984). Convexity and Large Deviations. *Ann. Probab.* **12** 903–906.

RAWSTHORNE, D. A. (1985). Imitation of an iteration. *Math. Mag.* **58** 172–176.

ROBERTS, A. W. and VARBERG, D. E. (1973). *Convex Functions*. Academic, New York.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

TAKÀCS, L. (1977). *Combinatorial Methods in the Theory of Stochastic Processes*. Krieger, Huntington, N.Y.

TERRAS, R. (1976). A stopping time problem on the positive integers. *Acta Arith.* **30** 241–252.
TERRAS, R. (1979). On the existence of a density. *Acta Arith.* **35** 101–102.
VYSSOTSKY, V. S. (1987). Private communication.
WAGON, S. (1985). The Collatz problem. *Math. Intelligencer* **7** (1) 72–76.
WENTZELL, A. D. (1976). Rough limit theorems on large deviations for Markov processes I, II. *Theory Probab. Appl.* **21** 227–242; 499–512.

AT & T BELL LABORATORIES
MATHEMATICAL SCIENCES RESEARCH CENTER
600 MOUNTAIN AVENUE
ROOM 2C-118
MURRAY HILL, NEW JERSEY 07974