

How often does the ratchet click? Facts, heuristics, asymptotics.

BY A. ETHERIDGE*, P. PFAFFELHUBER† AND A. WAKOLBINGER

*University of Oxford, Ludwig-Maximilians University Munich
and Goethe-University Frankfurt*

17. July 2007

Abstract

The evolutionary force of recombination is lacking in asexually reproducing populations. As a consequence, the population can suffer an irreversible accumulation of deleterious mutations, a phenomenon known as *Muller's ratchet*. We formulate discrete and continuous time versions of Muller's ratchet. Inspired by Haigh's (1978) analysis of a dynamical system which arises in the limit of large populations, we identify the parameter $\gamma = N\lambda/(Ns \cdot \log(N\lambda))$ as most important for the speed of accumulation of deleterious mutations. Here N is population size, s is the selection coefficient and λ is the deleterious mutation rate. For large parts of the parameter range, measuring time in units of size N , deleterious mutations accumulate according to a power law in $N\lambda$ with exponent γ if $\gamma \geq 0.5$. For $\gamma < 0.5$ mutations cannot accumulate. We obtain diffusion approximations for three different parameter regimes, depending on the speed of the ratchet. Our approximations shed new light on analyses of Stephan et al. (1993) and Gordo & Charlesworth (2000). The heuristics leading to the approximations are supported by simulations.

1 Introduction

Muller's ratchet is a mechanism that has been suggested as an explanation for the evolution of sex (Maynard Smith 1978). The idea is simple; in an asexually reproducing population chromosomes are passed down as indivisible blocks and so the number of deleterious mutations accumulated along any ancestral line in the population can only increase. When everyone in the current 'best' class has accumulated at least one additional bad mutation then the minimum mutational load in the population increases: the ratchet clicks. In a sexually reproducing

*Research supported in part by EPSRC GR/T19537/01.

†Travel support from DFG, Bilateral Research Group FOR 498.

AMS 2000 subject classification. 92D15, 60J70 (Primary) 60K35, 60H35, 60H10 (Secondary).

Keywords and phrases. Muller's ratchet, selection, mutation, Fleming-Viot process, Wright-Fisher model, diffusion approximation

population this is no longer the case; because of recombination between parental genomes a parent carrying a high mutational load can have offspring with *fewer* deleterious mutations. The high cost of sexual reproduction is thus offset by the benefits of inhibiting the ratchet. Equally the ratchet provides a possible explanation for the degeneration of Y chromosomes in sexual organisms (e.g. Charlesworth 1978, 1996; Rice 1994; Charlesworth & Charlesworth 1997, 1998). However, in order to assess its real biological importance one should establish under what circumstances Muller’s ratchet will have an evolutionary effect. In particular, *how many generations will it take for an asexually reproducing population to lose its current best class?* In other words, what is the rate of the ratchet?

In spite of the substantial literature devoted to the ratchet (see Loewe 2006 for an extensive bibliography), even in the simplest mathematical models a closed form expression for the rate remains elusive. Instead various approximations have been proposed which fit well with simulations for particular parameter regimes. The analysis presented here unifies these approximations into a single framework and provides a more detailed mathematical understanding of their regions of validity.

The simplest mathematical model of the ratchet was formulated in the pioneering work of Haigh (1978). Consider an asexual population of constant size N . The population evolves according to classical Wright-Fisher dynamics. Thus each of the N individuals in the $(t+1)$ st generation independently chooses a parent from the individuals in the t th generation. The probability that an individual which has accumulated k mutations is selected is proportional to its *relative* fitness, $(1-s)^k$. The number of mutations carried by the offspring is then $k+J$ where J is an (independent) Poisson random variable with mean λ .

Haigh identifies $n_0 = Ne^{-\lambda/s}$ as an approximation (at large times) for the total number of individuals carrying the least number of mutations and finds numerical evidence of a linear relationship between n_0 and the average time between clicks of the ratchet, at least for ‘intermediate’ values of n_0 (which he quantifies as $n_0 > 1$ and less than 25, say). On the other hand, for increasing values of n_0 Stephan et al. (1993) note the increasing importance of s for the rate of the ratchet. The simulations of Gordo & Charlesworth (2000) also suggest that for n_0 fixed at a large value the ratchet can run at very different rates. They focus on parameter ranges that may be the most relevant to the problem of the degeneration of large non-recombining portions of chromosomes.

In our approach we use diffusion approximations to identify another parameter as being an important factor in determining the rate of the ratchet. We define

$$\gamma := \frac{N\lambda}{Ns \log(N\lambda)}. \quad (1.1)$$

Notice that $n_0 = N(N\lambda)^{-\gamma}$. In these parameters one can reinterpret Haigh’s empirical results as saying that if we measure time in units of size N , then the rate of the ratchet follows a power law in $N\lambda$. In fact our main observation in this note is that for a substantial portion of parameter space (which we shall

quantify a little more precisely later) we have the following

Rule of Thumb. *The rate of the ratchet is of the order $N^{\gamma-1}\lambda^\gamma$ for $\gamma \in (1/2, 1)$, whereas it is exponentially slow in $(N\lambda)^{1-\gamma}$ for $\gamma < 1/2$.*

There are two novelties here. First, the abrupt change in behaviour at $\gamma = 1/2$ and second the power law interpretation of the rate for $\gamma \in (1/2, 1)$. As an appetiser, Figure 1 illustrates that this behaviour really is reflected in simulated data; see also §6.

The rule of thumb breaks down for two scenarios: first, if $\gamma > 1$ then for large $N\lambda$ we have $n_0 < 1$ and so our arguments, which are based on diffusion approximations for the size of the best class, will break down. This parameter regime, which leads to very frequent clicks of the ratchet, was studied by Gessler (1995). Second, if $N\lambda$ is *too* large then we see the transition from exponentially rare clicks to frequent clicks takes place at larger values of γ .

The rest of this note is laid out as follows. In §2 we review the work of Haigh (1978). Whereas Haigh’s work focuses on discrete time dynamics, in §3 we write down instead a (continuous time) Fleming-Viot diffusion approximation to the model whose behaviour captures the dynamics of large populations when s and λ are small. We then pass in §4 to the infinite population (deterministic) limit. This system can be solved exactly and by, in the spirit of Haigh, using this deterministic system to estimate the behaviour of the ‘bulk’ of the population we obtain in §5 a one dimensional diffusion which approximates the size of the best class in our Fleming-Viot system. The drift in this one-dimensional diffusion will take one of three forms depending upon whether the ratchet is clicking rarely, at a moderate rate or frequently per N generations (but always rarely per generation). Performing a scaling of this diffusion allows us to predict the relationship between the parameters $N\lambda$ and Ns of the biological model and the value of γ at which we can expect to see the phase transition from rare clicking to power law behaviour of the rate of the ratchet. In §6 we compare our predictions to simulated data, and in §7 we discuss the connection between our findings and previous work of Stephan et al. (1993), Stephan & Kim (2002) and Gordo & Charlesworth (2000).

2 The discrete ratchet – Haigh’s approach

The population dynamics described in the introduction can be reformulated mathematically as follows. Let N be a fixed natural number (the population size), $\lambda > 0$ (the mutation parameter) and $s \in (0, 1)$ (the selection coefficient). The population is described by a stochastic process taking values in $\mathcal{P}(\mathbb{N}_0)$, the simplex of probability weights on \mathbb{N}_0 . Suppose that $\mathbf{x}(t) = (x_k(t))_{k=0,1,\dots} \in \mathcal{P}(\mathbb{N}_0)$ is the vector of *type frequencies* (or *frequency profile*) in the t th generation (so for example $Nx_k(t)$ individuals in the population carry exactly k mutations). Let H be an \mathbb{N}_0 -valued random variable with $\mathbf{P}[H = k]$ proportional to $(1 - s)^k x_k(t)$, let J be a Poisson(λ)-random variable independent of

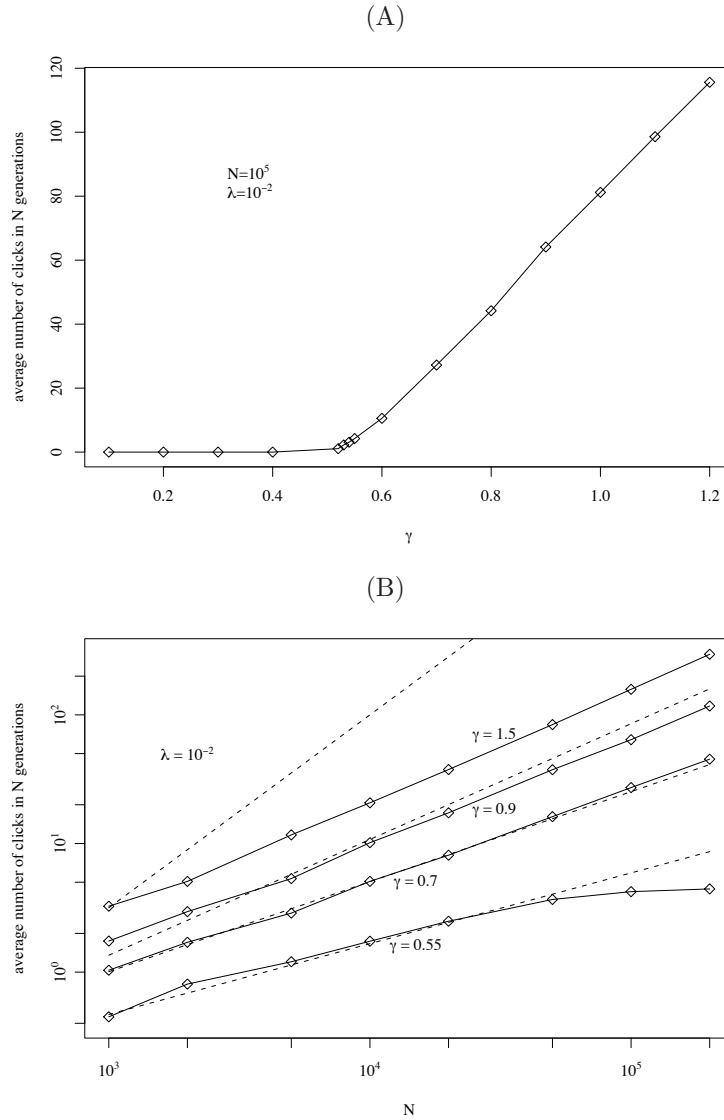


Figure 1: (A) We plot the rate of the ratchet against γ , where time is measured in units of N generations. As predicted by our rule of thumb we see a sharp change of behaviour around $\gamma = 0.5$. (B) We see the power-law behaviour of the rate for various values of γ . The solid lines are given by simulation of a Wright-Fisher model. The dashed lines fit the prediction that the time between clicks is a constant times $N(N\lambda)^{-\gamma}$ for $\gamma > 0.5$. Note that this breaks down for $\gamma > 1$.

H and let K_1, \dots, K_N be independent copies of $H + J$. Then the random type frequencies in the next generation are

$$X_k(t+1) = \frac{1}{N} \#\{i : K_i = k\}. \quad (2.1)$$

We shall refer to this as the *ratchet dynamics in discrete time*.

First consider what happens as $N \rightarrow \infty$. By the law of large numbers, (2.1) results in the deterministic dynamics

$$\mathbf{x}(t+1) := \mathbf{E}_{\mathbf{x}(t)}[\mathbf{X}(t+1)]. \quad (2.2)$$

An important property of the dynamics (2.2) is that vectors of Poisson weights are mapped to vectors of Poisson weights. To see this, note that when $N \rightarrow \infty$ the right hand side of (2.2) is just the *law* of the random variable $H + J$. If $\mathbf{x}(t) = \text{Poisson}(\alpha)$, the Poisson distribution with mean α , then H is $\text{Poisson}(\alpha(1-s))$ -distributed and consequently $\mathbf{x}(t+1)$ is the law of a $\text{Poisson}(\alpha(1-s) + \lambda)$ random variable. We shall see in §4 that the same is true of the continuous time analogue of (2.2) and indeed we show there that for every initial condition with $x_0 > 0$ the solution to the continuous time equation converges to the stationary point

$$\pi := \text{Poisson}(\theta)$$

as $t \rightarrow \infty$, where

$$\theta := \frac{\lambda}{s}.$$

Haigh’s analysis of the finite population model focusses on the number of individuals in the best class. Let us write k^* for the number of mutations carried by individuals in this class. In a finite population, k^* will increase with time, but the profile of frequencies *relative to* k^* , $\{X_{k+k^*}\}$, forms a recurrent Markov chain. We set

$$\mathbf{Y} := (Y_k)_{k=0,1,\dots} := (X_{k^*+k})_{k=0,1,\dots}$$

and observe that since fitness is always measured relative to the mean fitness in the population, between clicks of the ratchet the equation for the dynamics of \mathbf{Y} is precisely the same as that for \mathbf{X} when $X_0 > 0$. Suppose that after t generations there are $n_0(t) = Ny_0(t)$ individuals in the best class. Then the probability of sampling a parent from this class *and* not acquiring any additional mutations is

$$p_0(t) := (y_0(t)/W(t)) e^{-\lambda},$$

where

$$W(t) = \sum_{i=0}^{\infty} y_i(t)(1-s)^i. \quad (2.3)$$

Thus, given $y_0(t)$, the size of the best class in the next generation has a binomial distribution with N trials and success probability $p_0(t)$, and so the evolution of

the best class is determined by $W(t)$, the mean fitness of the population. We shall see this property of Y_0 reflected in the diffusion approximation of §3.

A principal assumption of Haigh's analysis is that immediately before a click, the individuals of the current best class have all been distributed upon the other classes, in proportion to their Poisson weights. Thus immediately after a click he takes the type frequencies (relative to the new best class) to be

$$\tilde{\pi} := \frac{1}{1 - \pi_0}(\pi_1, \pi_2, \dots) = \frac{1}{1 - e^{-\theta}} \left(\theta e^{-\theta}, \frac{\theta^2}{2} e^{-\theta}, \dots \right). \quad (2.4)$$

The time until the next click is then subdivided into two phases. During the first phase the deterministic dynamical system decays exponentially fast towards its Poisson equilibrium, swamping the randomness arising from the finite population size. At the time that he proposes for the end of the first phase the size of the best class is approximately $1.6\pi_0$. The mean fitness of the population has decreased by an amount which can also be readily estimated and, combining this with a Poisson approximation to the binomial distribution, Haigh proposes that (at least initially) during the second (longer) phase the size of the best class should be approximated by a Galton-Watson branching process with a Poisson offspring distribution.

Haigh's original proposal was that since the mean fitness of the population (and consequently the mean number of offspring in the Galton Watson process) changes only slowly during the second phase it should be taken to be constant throughout that phase. Later refinements have modified Haigh's approach in two key ways. First they have worked with a diffusion approximation so that the Galton-Watson process is replaced by a Feller diffusion and second, instead of taking a constant drift, they look for a good approximation of the mean fitness *given the size of the best class*, resulting in a Feller diffusion with *logistic* growth. Our aim in the rest of this paper is to unify these approximations in a single mathematical framework and discuss them in the light of simulations. A crucial building block will be the following extension of (2.4), which we call the *Poisson profile approximation* (or PPA) of \mathbf{Y} based on Y_0 :

$$\Pi(Y_0) := \left(Y_0, \frac{1 - Y_0}{1 - \pi_0}(\pi_1, \pi_2, \dots) \right). \quad (2.5)$$

As a first step, we now turn to a diffusion approximation for the full ratchet dynamics (2.1).

3 The Fleming-Viot diffusion

For large N and small λ and s , the following stochastic dynamics on $\mathcal{P}(\mathbb{N}_0)$ in continuous time captures the conditional expectation and variance of the

discrete dynamics (2.1):

$$dX_k = \left(\sum_j s(j-k)X_jX_k + \lambda(X_{k-1} - X_k) \right) dt + \sum_{j \neq k} \sqrt{\frac{1}{N}X_jX_k} dW_{jk}, \quad (3.1)$$

$$k = 0, 1, 2, \dots,$$

where $X_{-1} := 0$, and $(W_{jk})_{j>k}$ is an array of independent standard Wiener processes with $W_{kj} := -W_{jk}$. This is, of course, just the infinite-dimensional version of the standard multi-dimensional Wright-Fisher diffusion. Existence of a process solving (3.1) can be established using a diffusion limit of the discrete dynamics of §2. The coefficient $s(j-k)$ is the fitness difference between type k and type j , $\lambda(X_{k-1} - X_k)$ is the flow into and out of class k due to mutation and the diffusion coefficients $\frac{1}{N}X_jX_k$ reflect the covariances due to multinomial sampling.

Remark 3.1. Often when one passes to a Fleming-Viot diffusion approximation, one measures time in units of size N and correspondingly the parameters s and λ appear as Ns and $N\lambda$. Here we have *not* rescaled time, hence the factor of $\frac{1}{N}$ in the noise and the unscaled parameters s and λ in the equations. \square

Writing

$$M_1(\mathbf{X}) := \sum_j jX_j$$

for the first moment of \mathbf{X} , (3.1) translates into

$$dX_k = \left(s(M_1(\mathbf{X}) - k) - \lambda X_k + \lambda X_{k-1} \right) dt + \sum_{j \neq k} \sqrt{\frac{1}{N}X_jX_k} dW_{jk} \quad (3.2)$$

In exactly the same way as in our discrete stochastic system, writing k^* for the number of mutations carried by individuals in the fittest class, one would like to think of the population as a travelling wave with profile $Y_k = X_{k+k^*}$. Notice in particular that

$$dY_0 = (sM_1(\mathbf{Y}) - \lambda)Y_0dt + \sqrt{\frac{1}{N}Y_0(1-Y_0)}dW_0,$$

where W_0 is a standard Wiener process. Thus, just as in Haigh's setting, the frequency of the best class is determined by the mean fitness of the population.

Substituting into (3.1) one can obtain a stochastic equation for M_1 ,

$$dM_1 = (\lambda - sM_2)dt + dG$$

where $M_2 = \sum_j (j - M_1)^2 X_j$ and the martingale G has quadratic variation

$$d\langle G \rangle = \frac{1}{N}M_2dt.$$

Thus the speed of the wave is determined by the variance of the profile. Similarly,

$$dM_2 = \left(-\frac{1}{N}M_2 + (\lambda - sM_3)\right)dt + dH$$

where $M_3 = \sum_j (j - M_1)^3 X_j$ and the martingale H has quadratic variation $d\langle H \rangle = \frac{1}{N}(M_4 - M_2)dt$ with M_4 denoting the fourth centred moment and so on.

These equations for the centred moments are entirely analogous to those obtained by Higgs & Woodcock (1995) except that in the Fleming-Viot setting they are exact. As pointed out there, Bürger (1991) obtained similar equations to study the evolution of polygenic traits. The difficulty in using these equations to study the rate of the ratchet is of course that they are not closed: the equation for M_k involves M_{k+1} and so on. Moreover, there is no obvious approximating closed system. By contrast, the infinite population limit, in which the noise is absent, turns out to have a closed solution.

4 The infinite population limit

The continuous time analogue of (2.2) is the deterministic dynamical system

$$dx_k = \left((s(M_1(\mathbf{x}) - k) - \lambda)x_k + \lambda x_{k-1} \right) dt, \quad k = 0, 1, 2, \dots \quad (4.1)$$

where $x_{-1} = 0$, obtained by letting $N \rightarrow \infty$ in our Fleming-Viot diffusion (3.2). Our goal in this section is to solve this system of equations. Note that Maia et al. (2003) have obtained a complete solution of the corresponding discrete system following (2.2).

As we shall see in Proposition 4.1, the stationary points of the system are exactly the same as for (2.2), that is $\mathbf{x} = \pi$ and all its right shifts $(\pi_{k-k^*})_{k=0,1,\dots}$, $k^* = 0, 1, 2, \dots$. Since the Poisson distribution can be characterised as the only distribution on \mathbb{N}_0 with all cumulants equal, it is natural to transform (4.1) into a system of equations for the cumulants, $\kappa_k, k = 1, 2, \dots$, of the vector \mathbf{x} . The cumulants are defined by the relation

$$\log \sum_{k=0}^{\infty} x_k e^{-\xi k} = \sum_{k=1}^{\infty} \kappa_k \frac{(-\xi)^k}{k!}. \quad (4.2)$$

We assume $x_0 > 0$ and set

$$\kappa_0 := -\log x_0. \quad (4.3)$$

Proposition 4.1. *For $\kappa_k, k = 0, 1, 2, \dots$ as in (4.2) and (4.3) the system (4.1) is equivalent to*

$$\dot{\kappa}_k = -s\kappa_{k+1} + \lambda, \quad k = 0, 1, 2, \dots$$

Setting $\underline{\kappa} := (\kappa_0, \kappa_1, \dots)$ this system is solved by

$$\underline{\kappa} = B\underline{\kappa}(0)^\top + \frac{\lambda}{s}(1 - e^{-st})\mathbf{1}, \quad B = (b_{ij})_{i,j=0,1,\dots}, \quad b_{ij} = \begin{cases} \frac{(-st)^{j-i}}{(j-i)!} & j \geq i \\ 0 & \text{otherwise.} \end{cases} \quad (4.4)$$

In particular,

$$x_0(t) = e^{-\kappa_0(t)} = x_0(0) \frac{\exp\left(-\frac{\lambda}{s}(1 - e^{-st})\right)}{\left(\sum_{k=0}^{\infty} x_k(0)e^{-stk}\right)} \quad (4.5)$$

and

$$\kappa_1(t) = \sum_{k=0}^{\infty} kx_k(t) = -\frac{\partial}{\partial \xi} \log \sum_{k=0}^{\infty} x_k(0)e^{-\xi k} \Bigg|_{\xi=st} + \frac{\lambda}{s}(1 - e^{-st}). \quad (4.6)$$

Remark 4.2. If $\mathbf{x}(0)$ is a Poisson(μ) distribution then substituting into (4.4) we see that $\mathbf{x}(t)$ is a Poisson distribution with parameter $\lambda/s + e^{-st}(\mu - \lambda/s)$. In other words, just as for the discrete dynamical system considered by Haigh, vectors of Poisson weights are mapped to vectors of Poisson weights. In particular $\pi := \text{Poisson}(\lambda/s)$ is once again a stationary point of the system. Moreover, this proposition shows that for any vector $\mathbf{x}(0)$ with $x_0(0) > 0$, the solution converges to this stationary point. The corresponding convergence result in the discrete case is established in Maia et al. (2003). More generally, if k^* is the smallest value of k for which $x_k(0) > 0$ then the solution will converge to $(\pi_{k-k^*})_{k=0,1,2,\dots}$.

Proof of Proposition 4.1. Using (4.1) we have

$$\begin{aligned} \frac{d}{dt} \log \left(\frac{1}{x_0} \sum_{k=0}^{\infty} x_k e^{-\xi k} \right) &= \frac{x_0}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} \left(-\frac{\sum_{k=0}^{\infty} x_k e^{-\xi k}}{x_0^2} \dot{x}_0 + \frac{1}{x_0} \sum_{k=0}^{\infty} \dot{x}_k e^{-\xi k} \right) \\ &= -s \sum_{j=0}^{\infty} jx_j + \lambda - s \frac{\sum_{k=0}^{\infty} kx_k e^{-\xi k}}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} + s \sum_{j=0}^{\infty} jx_j \\ &\quad + \frac{\lambda}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} \left(e^{-\xi} \sum_{k=1}^{\infty} x_{k-1} e^{-\xi(k-1)} - \sum_{k=0}^{\infty} x_k e^{-\xi k} \right) \\ &= s \frac{d}{d\xi} \log \left(\sum_{k=0}^{\infty} x_k e^{-\xi k} \right) + \lambda e^{-\xi}. \end{aligned}$$

Thus, (4.2) gives

$$\frac{d}{dt} \sum_{k=0}^{\infty} \kappa_k \frac{(-\xi)^k}{k!} = -s \sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} + \lambda e^{-\xi}.$$

Comparing coefficients in the last equation we obtain

$$\dot{\kappa}_k = -s\kappa_{k+1} + \lambda, \quad k = 0, 1, \dots$$

This linear system can readily be solved. We write

$$D := (\delta_{i+1,j})_{i,j=0,1,2,\dots}, \quad \underline{1} = (1, 1, \dots),$$

so that

$$\dot{\underline{\kappa}}^\top = -sD\underline{\kappa}^\top + \lambda\underline{1}^\top. \quad (4.7)$$

Since

$$(e^{-Dst})_{ij} = \begin{cases} \frac{(-st)^{j-i}}{(j-i)!} & j \geq i \\ 0 & \text{otherwise,} \end{cases}$$

the linear system (4.7) is solved by

$$\underline{\kappa}(t)^\top = e^{-Dst}\underline{\kappa}(0)^\top + \lambda \int_0^t e^{-Dsu}\underline{1}du = e^{-Dst}\underline{\kappa}(0)^\top + \frac{\lambda}{s}(1 - e^{-st})\underline{1}^\top.$$

□

Remark 4.3. With the initial condition $\mathbf{x}(0) := \tilde{\pi}$ given by (2.4), equations (4.5) and (4.6) become

$$x_0(t) = e^{-\theta} \frac{\theta e^{-st}}{1 - e^{-\theta e^{-st}}} \quad (4.8)$$

and

$$\kappa_1(t) = \theta - 1 + \frac{\theta e^{-st}}{e^{\theta e^{-st}} - 1}. \quad (4.9)$$

At time

$$\tau := \frac{\log \theta}{s}, \quad (4.10)$$

we have $x_0(\tau) = e^{-\theta} \frac{1}{1 - e^{-1}} \approx 1.6\pi_0$. Comparing with §2 we see that in our continuous time setting τ is precisely the counterpart of the time proposed by Haigh as the end of ‘phase one’.

In §5 our prediction for $M_1(\mathbf{Y})$ given Y_0 will require the value of $M_1(\mathbf{y}(\tau))$ for \mathbf{y} solving (4.1) when started from a Poisson profile approximation. This is the purpose of the next proposition.

Proposition 4.4. *For $y_0 \in (0, 1)$, let $\mathbf{y}(t)$ be the solution of (4.1) with the initial state $\mathbf{y}(0) := \Pi(y_0)$ defined in (2.5), and let τ be Haigh’s relaxation time defined in (4.10). Then for $A \geq 0$ with $\eta := \theta^{1-A}$*

$$M_1(\mathbf{y}(A\tau)) = \theta + \frac{\eta}{e^\eta - 1} \left(1 - \frac{y_0(A\tau)}{\pi_0} \right).$$

Proof. Since

$$\sum_{k=0}^{\infty} \pi_k e^{-\xi k} = \exp(-\theta(1 - e^{-\xi})) = \pi_0^{1 - e^{-\xi}},$$

we have

$$\sum_{k=0}^{\infty} y_k e^{-\xi k} \Big|_{\xi=sA\tau} = y_0 + \frac{1 - y_0}{1 - \pi_0} \pi_0 (e^{\theta e^{-\xi}} - 1) \Big|_{\xi=sA\tau} = y_0 + \frac{1 - y_0}{1 - \pi_0} \pi_0 (e^\eta - 1)$$

and

$$-\frac{\partial}{\partial \xi} \sum_{k=0}^{\infty} y_k e^{-\xi k} \Big|_{\xi=sA\tau} = \frac{1 - y_0}{1 - \pi_0} \pi_0^{1 - e^{-\xi}} \theta e^{-\xi} \Big|_{\xi=sA\tau} = \frac{1 - y_0}{1 - \pi_0} \pi_0 e^\eta \eta.$$

Using the solution (4.5) and (4.6) and $y_0(0) = y_0$

$$\begin{aligned} y_0(A\tau) &= y_0 \frac{\pi_0 e^\eta}{y_0 + \frac{1 - y_0}{1 - \pi_0} \pi_0 (e^\eta - 1)} \\ &= y_0 \frac{\pi_0 e^\eta (1 - \pi_0)}{y_0 (1 - \pi_0 e^\eta) + \pi_0 (e^\eta - 1)}, \end{aligned} \quad (4.11)$$

$$\begin{aligned} M_1(\mathbf{y}(A\tau)) &= \frac{\frac{1 - y_0}{1 - \pi_0} \pi_0 e^\eta \eta}{y_0 + \frac{1 - y_0}{1 - \pi_0} \pi_0 (e^\eta - 1)} + \theta - \eta \\ &= \theta + \eta \frac{\pi_0 - y_0}{y_0 (1 - \pi_0 e^\eta) + \pi_0 (e^\eta - 1)}. \end{aligned} \quad (4.12)$$

From (4.11),

$$y_0 = \frac{y_0(A\tau) \pi_0 (e^\eta - 1)}{\pi_0 e^\eta (1 - \pi_0) - y_0(A\tau) (1 - \pi_0 e^\eta)}$$

and thus

$$\begin{aligned} \pi_0 - y_0 &= \frac{\pi_0 e^\eta (\pi_0 - y_0(A\tau)) (1 - \pi_0)}{\pi_0 e^\eta (1 - \pi_0) - y_0(A\tau) (1 - \pi_0 e^\eta)}, \\ y_0 (1 - \pi_0 e^\eta) + \pi_0 (e^\eta - 1) &= \pi_0 (e^\eta - 1) \frac{\pi_0 e^\eta (1 - \pi_0)}{\pi_0 e^\eta (1 - \pi_0) - y_0(A\tau) (1 - \pi_0 e^\eta)}. \end{aligned}$$

Plugging the last two equations into (4.12) we find

$$M_1(\mathbf{y}(A\tau)) = \theta + \frac{\eta}{e^\eta - 1} \left(1 - \frac{y_0(A\tau)}{\pi_0} \right).$$

□

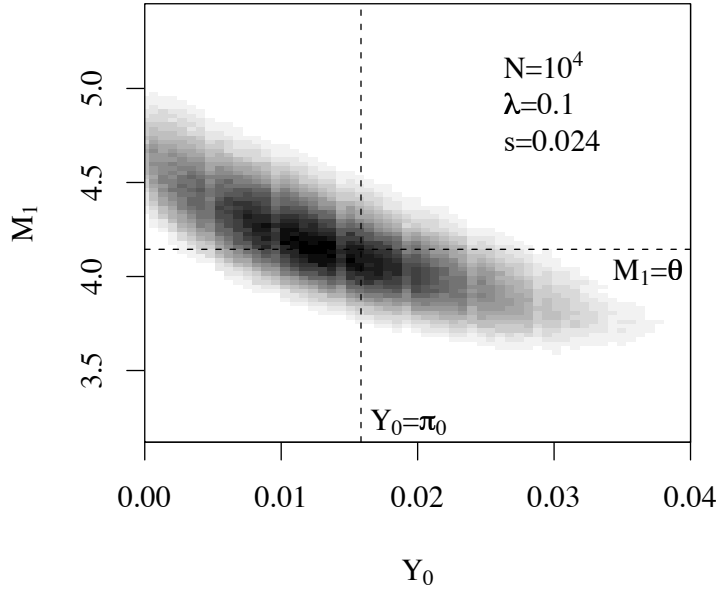


Figure 2: Using simulations (see also §6) we plot (Y_0, M_1) . There is a good fit to a linear relationship between Y_0 and M_1 . Note that $\gamma = 0.6$ in the figure.

5 One dimensional diffusion approximations

Recall from §3 that in our Fleming-Viot model the frequency Y_0 of the best class follows

$$dY_0 = \left(sM_1(\mathbf{Y}) - \lambda \right) Y_0 dt + \sqrt{\frac{1}{N} Y_0 (1 - Y_0)} dW_0, \quad (5.1)$$

where W_0 is a standard Wiener process. The system of equations (3.2) is too complex for us to be able to find an explicit expression for $M_1(\mathbf{Y})$, which depends on the whole vector \mathbf{Y} of class sizes. Instead we seek a good approximation of M_1 given Y_0 . Substituting this into equation (5.1) will then yield a one-dimensional diffusion which we use as an approximation for the size of the best class. Of course this assumption of a functional dependence between Y_0 and M_1 is a weakness of the one-dimensional diffusion approximation, but simulations show that there is a substantial correlation between Y_0 and M_1 , see, for example, Figure 2.

To understand our approach to finding a map $Y_0 \mapsto M_1$, recall as a first

step Haigh's approximation that immediately after a click of the ratchet the profile has the form (2.4). The reasoning is as follows. Deviations of \mathbf{Y} from a Poisson profile can only be due to the randomness arising from resampling in a finite population. Since resampling has no tendency to increase or decrease the frequency of a given class, the average profile immediately after a click of the ratchet is approximated by the state where π_0 is distributed evenly over all other classes according to their equilibrium frequencies. During his short 'phase one', Haigh then allows this profile to 'relax' through the action of the discrete dynamical system (2.2) and it is the mean fitness in the population after this short relaxation time which determines the behaviour of the best class during 'phase two'.

A natural next step in extending this argument is to suppose that also *in between* click times the resampling distributes the mass $\pi_0 - Y_0$ evenly on all other classes. In other words, given Y_0 , approximate the state of the system by $\mathbf{Y} = \Pi(Y_0)$ given by (2.5).

Of course in reality the dynamical system interacts with the resampling as it tries to restore the system to its Poisson equilibrium. If this restoring force is strong, just as in Haigh's approach one estimated mean fitness during phase two from the 'relaxed' profile, so here one should approximate the mean fitness M_1 not from the PPA, but from states which arise by evolving the PPA using the dynamical system for a certain amount of time. We call the resulting states *relaxed Poisson Profile approximations* or RPPA. There are three different parameter regimes with which we shall be concerned. Each corresponds to a different value of η in the functional relationship

$$M_1 = \theta + \frac{\eta}{e^\eta - 1} \left(1 - \frac{Y_0}{\pi_0}\right). \quad (5.2)$$

of Proposition 4.4. These can be distinguished as follows:

$$A \text{ small,} \quad \eta \approx \theta, \quad M_1 \approx \frac{\theta}{1 - \pi_0} (1 - Y_0), \quad (5.3a)$$

$$A = 1, \quad \eta = 1, \quad M_1 \approx \theta + 0.58 \left(1 - \frac{Y_0}{\pi_0}\right), \quad (5.3b)$$

$$A \text{ large,} \quad \eta \approx 0, \quad M_1 \approx \theta + \left(1 - \frac{Y_0}{\pi_0}\right) \quad (5.3c)$$

The resulting maps $Y_0 \mapsto M_1$ are plotted in Figure 3. Observe that for consistency, M_1 has to increase, on average, by 1 during one click of the ratchet.

Finally, before we can apply our one-dimensional diffusion approximation we must choose a starting value for Y_0 following equation (5.1). For A large, the system is already close to its new equilibrium at the time of a click and so we take $Y_0 = \pi_0$.

For $A = 1$, at the time of the click we observe a state which has relaxed for time τ from a state of the form $\tilde{\pi}$ from (2.4). We computed in Remark 4.3 that such a state comes with $Y_0 = 1.6\pi_0$.

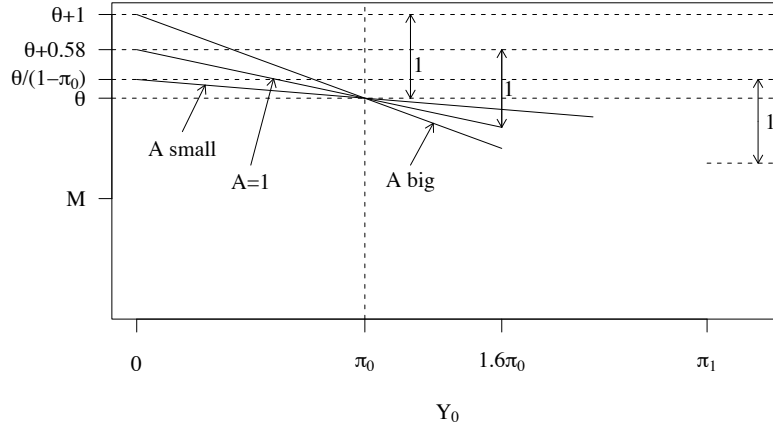


Figure 3: Since simulations show a strong correlation between the first moment M_1 and Y_0 , we use (5.3a)-(5.3c) to predict M_1 from Y_0 depending on the model parameters.

For small values of A , observe that the profile of the population immediately after a click is approximately $\tilde{\pi}$ from (2.4). Since $\tilde{\pi}$ is not a state of the form $\Pi(y_0)$ the arguments that led to Proposition 4.4 do not apply. Instead we follow Haigh in dividing the time between clicks into two phases. Consider first ‘phase one’. Recall from the dynamical system that

$$dY_0 = (sM_1 - \lambda)Y_0 dt.$$

We write

$$(sM_1 - \lambda)Y_0 = c (\pi_0 - Y_0),$$

where c (like Y_0 and M_1) depends on $r = \theta e^{-st}$. Starting in $\mathbf{Y}(0) = \tilde{\pi}$ we have from (4.8) and (4.9)

$$Y_0(t) = e^{-\theta} \frac{\theta e^{-st}}{1 - \exp(-\theta e^{-st})} = e^{-\theta} \frac{r}{1 - e^{-r}},$$

$$M_1(t) = \theta - 1 + \frac{\theta e^{-st}}{\exp(\theta e^{-st}) - 1} = \theta - 1 + \frac{r}{e^r - 1}.$$

We compute

$$\frac{c}{s} = \frac{(M_1 - \theta)Y_0}{\pi_0 - Y_0} = \frac{1 - \frac{r}{e^r - 1}}{1 - \frac{1 - e^{-r}}{r}} = \frac{r(1 - e^{-r}) - r^2 e^{-r}}{r(1 - e^{-r}) - (1 - e^{-r})^2}.$$

It can be checked that this expression lies between 1 and 1.25 for all $r > 0$ which suggests that the size of the best class in the initial phase after a click is reasonably described by the dynamics

$$dY_0 = s(\pi_0 - Y_0)dt + \sqrt{\frac{1}{N}Y_0}dW_0 \quad (5.4)$$

started from $\frac{\pi_1}{(1-\pi_0)}$. We allow Y_0 to evolve according to equation (5.4) until it reaches $1.6\pi_0$, say, and then use our estimate of M_1 from equation (5.3) to estimate the evolution of Y_0 during the (longer) ‘phase two’.

We assume that states of the ratchet are RPPAs, i.e., Poisson profile approximations (2.5) which are relaxed for time $A\tau$, where $\tau = \frac{1}{s} \log \theta$, which leads to the functional relationship (5.2). Consequently, we suggest that (5.1) is approximated by the ‘mean reversion’ dynamics

$$dY_0 = s \frac{\eta}{e^\eta - 1} \left(1 - \frac{Y_0}{\pi_0}\right) Y_0 dt + \sqrt{\frac{1}{N}Y_0} dW_0, \quad (5.5)$$

with $\eta = \theta^{1-A}$, where we have used a Feller noise instead of the Wright-Fisher term in (5.1). In other words, Y_0 is a *Feller branching diffusion with logistic growth*.

Using the three regimes from (5.3), we have the approximations

$$A \text{ small,} \quad dY_0 = \lambda(\pi_0 - Y_0)Y_0 dt + \sqrt{\frac{1}{N}Y_0}dW, \quad (5.6a)$$

$$A = 1, \quad dY_0 = 0.58s \left(1 - \frac{Y_0}{\pi_0}\right) Y_0 dt + \sqrt{\frac{1}{N}Y_0}dW_0, \quad (5.6b)$$

$$A \text{ large,} \quad dY_0 = s \left(1 - \frac{Y_0}{\pi_0}\right) Y_0 dt + \sqrt{\frac{1}{N}Y_0}dW_0, \quad (5.6c)$$

(where in the first equation we have used that $\frac{1}{1-\pi_0} \approx 1 + \pi_0$ and that $Y_0\pi_0$ is negligible).

An equation similar to (5.6b) was found (by different means) by Stephan et al (1993) and further discussed in Gordo & Charlesworth (2000). Stephan and Kim (2002) analyse whether a prefactor of 0.5 or 0.6 in (5.6b) fits better with simulated data. We discuss the relationship with these papers in detail in §7.

The expected time to extinction of a diffusion following (5.5) is readily obtained from a Green function calculation similar to that in Lambert (2005). We refrain from doing this here, but instead use a scaling argument to identify parameter ranges for which the ratchet clicks and to give evidence for the rule of thumb formulated in the introduction.

Consider the rescaling

$$Z(t) = \frac{1}{\pi_0} Y_0(N\pi_0 t).$$

For A small equation (5.6a) becomes

$$\begin{aligned} dZ &= N\lambda\pi_0^2(1-Z)Zdt + \sqrt{Z}dW \\ &= (N\lambda)^{1-2\gamma}(1-Z)Zdt + \sqrt{Z}dW. \end{aligned} \quad (5.7)$$

For $A = 1$ on the other hand we obtain from (5.6b)

$$\begin{aligned} dZ &= 0.58N s\pi_0(1-Z)Zdt + \sqrt{Z}dW \\ &= 0.58\frac{1}{\gamma \log(N\lambda)}(N\lambda)^{1-\gamma}(1-Z)Zdt + \sqrt{Z}dW. \end{aligned} \quad (5.8)$$

For A large we obtain from (5.6c) the same equation without the factor of 0.58.

From this rescaling we see that the equation that applies for small A , i.e. (5.7), is strongly mean reverting for $\gamma < 1/2$. Recall that the choice of small A is appropriate when the ratchet is clicking frequently and so this indicates that frequent clicking simply will not happen for $\gamma < 1/2$. To indicate the boundary between rare and moderate clicking, equation (5.8) is much more relevant than equation (5.7). At first sight, equation (5.8) looks strongly mean reverting for all $\gamma < 1$, which would seem to suggest that the ratchet will click only exponentially slowly in $(N\lambda)^{1-\gamma}$. However, the closer γ is to one, the larger the value of $N\lambda$ we must take for this asymptotic regime to provide a good approximation. For example, in the table below we describe parameter combinations for which the coefficient in front of the mean reversion term in equation (5.8) is at least five. We see that for $\gamma < 1/2$ this coefficient is large for most of the reasonable values of $N\lambda$, whereas for $\gamma > 1/2$ it is rather small over a large range of $N\lambda$.

γ	0.3	0.4	0.5	0.55	0.6	0.7	0.8	0.9
$N\lambda \geq$	20	10^2	$9 \cdot 10^2$	$4 \cdot 10^3$	$2 \cdot 10^4$	$4 \cdot 10^6$	$2 \cdot 10^{11}$	$8 \cdot 10^{26}$

Thus, for example, if $\gamma = 0.7$ we require $N\lambda$ to be of the order of 10^6 in order for the strong mean reversion of equation (5.8) to be evident. This is not a value of $N\lambda$ which will be observed in practice. Indeed, as a ‘rule of thumb’, for biologically realistic parameter values, we should expect the transition from no clicks to a moderate rate of clicks to take place at around $\gamma = 0.5$.

6 Simulations

We have argued that the one-dimensional diffusions (5.6) approximate the frequency in the best class and from this deduced the *rule of thumb* from §1. In this section we use simulations to test the validity of our arguments.

For a population following the dynamics (2.1), the $(t + 1)$ st generation is formed by multinomial sampling of N individuals with weights

$$p_k(t) = \sum_{j=0}^k \frac{x_{k-j}(t)(1-s)^{k-j}}{W(t)} e^{-\lambda} \frac{\lambda^j}{j!}, \quad (6.1)$$

where $W(t)$ is the average fitness in the t th generation from (2.3) and it is this Wright-Fisher model which was implemented in the simulations.

To supplement the numerical results of Figure 1 we provide simulation results for the average time between clicks (where time is measured in units of N generations) for fixed N and γ and varying λ ; see Figure 4. Note that, for fixed γ in equation (1.1), s is increasing with λ . We carry out simulations using a population size of $N = 10^5$ and λ varying from 10^{-4} to 1. For $\gamma = 0.5$ we observe that the power law behaviour breaks down already for $N\lambda = 10^3$ and the diffusion (5.6b) predicts the clicking of the ratchet sufficiently well. For increasing γ , the power law breaks down only for larger values of $N\lambda$. For $\gamma = 0.7$, in our simulations we only observe the power law behaviour but conjecture that for larger values of $N\lambda$ the power law would break down; compare with the table above.

For a finer analysis of which of the equations (5.6) works best, we study the resulting Green functions numerically; see Figure 5. In particular, we record the relative time spent in some dY_0 in simulations and compare this quantity to the numerically integrated, normalised Green functions given through the diffusions (5.6a) and (5.6b). (We do not consider (5.6c) because it only gives an approximation if the ratchet clicks rarely.) We see in (A) that for $\gamma = 0.5$ not only does (5.6b) produce better estimates for the average time between clicks (Figure 4) but also for the time spent around some point y_0 . However, for $\gamma = 0.9$, clicks are more frequent and we expect (5.6a) to provide a better approximation. Indeed, although both (5.6a) and (5.6b) predict the power law behaviour, as (B) shows, the first equation produces better estimates for the relative amount of time spent in some dY_0 .

To support our claim that the states observed are relaxed Poisson Profile Approximations we use a phase-plane analysis; see Figure 6. At any point in time of a simulation, values for Y_0 and M_1 can be observed. The resulting plots indicate that we can distinguish the three parameter regimes introduced in §5. In the case of rapid clicking of the ratchet (so that the states we observe have not relaxed a lot and thus are approximately of the form $\Pi(y_0)$) we see in (A), (B) that the system is driven by the restoring force to $M_1 < \theta$. The reason is that M_1 is small at click times and these are frequent. However, the slope of the line relating Y_0 and M_1 is low, as predicted by (5.3a). (We used $(1 - Y_0)/(1 - \pi_0) \approx 1 - Y_0 + \pi_0$ in the plot here.) For the case $A = 1$ the system spends some time near $Y_0 = 0$ and thus the ratchet clicks, but not frequently. So, the dynamical system restores states partly to equilibrium and we see that the slope given in (5.3b) gives the most reasonable prediction in (C), (D), (E). For rare clicking, i.e. A large, the dynamical system has even more time and (F) shows that the slope is as predicted by (5.3c).

Our prediction that we observe profiles which are well approximated by a relaxed PPA applies especially well at click times. We check this numerically by observing the frequency of the (new) best class at click times; see Figure 7. For small γ the ratchet clicks rarely and the system has some time to relax to its

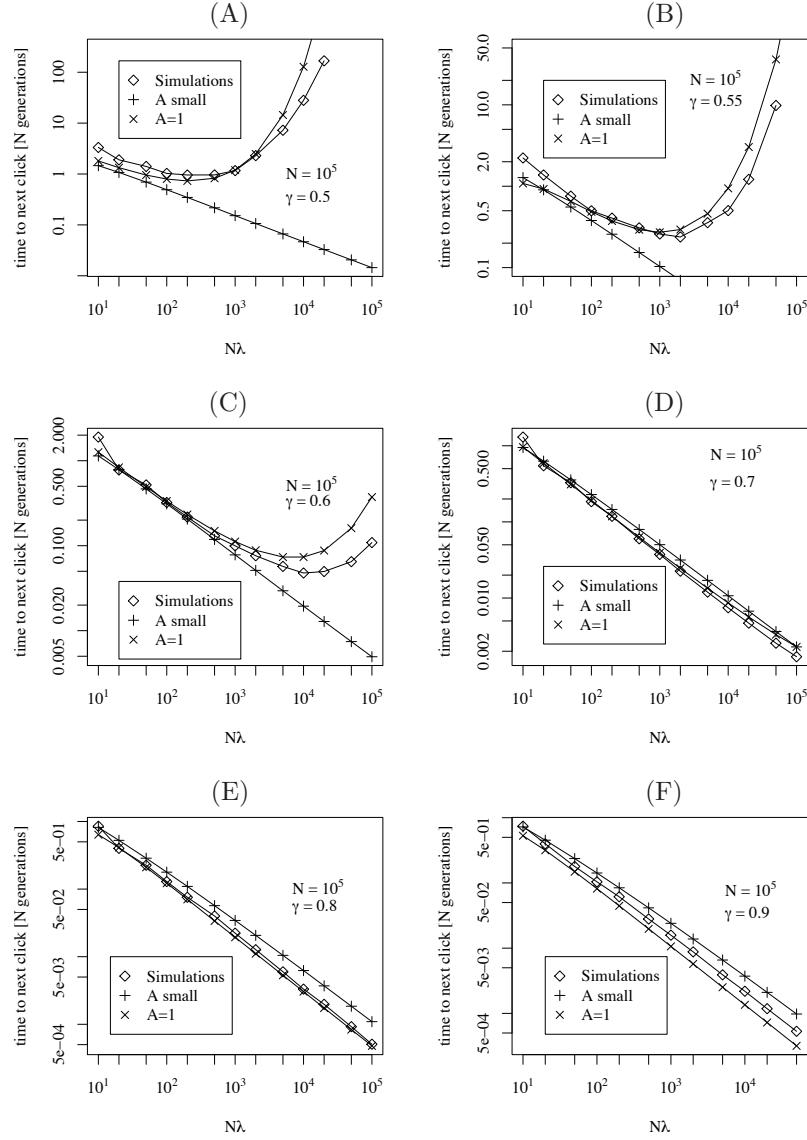


Figure 4: The power law behaviour of the rate of the ratchet with respect to γ is valid for a large portion of the parameter space. (A) For $\gamma = 0.5$ clicks become rare and the power law does not apply for $N\lambda > 10^3$. (B), (C) For $\gamma = 0.55$ and $\gamma = 0.6$, we have to explore a larger portion of the parameter space in order to see that the power law does not apply any more. (D), (E), (F) For $\gamma \geq 0.7$ we never observe a deviation from the power law. For every plot, we used $N = 10^5$ and simulations ran for $5 \cdot 10^6$ ($\gamma = 0.5$: $2 \cdot 10^7$) generations for each value of $N\lambda$.

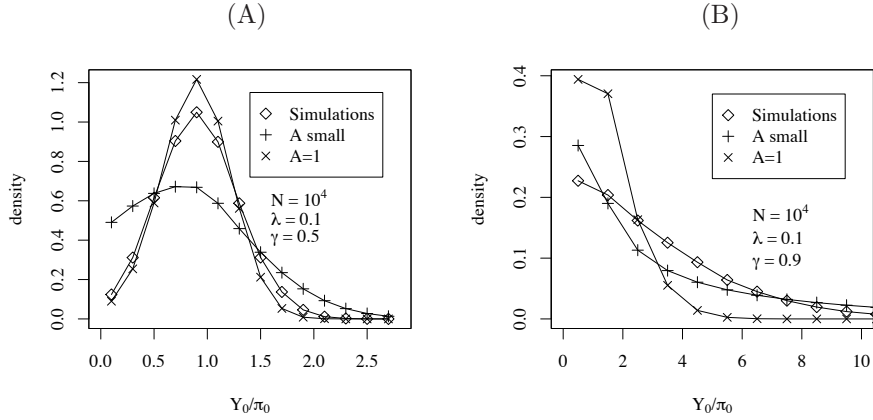


Figure 5: We compare the plots for the occupation density of Y_0 from the simulations with theoretical curves corresponding to the Green functions for the cases of small A and $A = 1$ in (5.6). (A) If clicks are rare, $A = 1$ produces better results than small A . (B) If clicks are frequent, the simulated densities of Y_0 are better approximated by small A . Every plot is based on the simulation of $5 \cdot 10^5$ generations.

new equilibrium even before the click of the ratchet. As a consequence, we see that the frequency of the (new) best class at the time of the click is already close to π_0 . However, if γ is large and the ratchet clicks frequently, the dynamical system has no time before the click to relax the system to the new equilibrium. Therefore, we observe that the frequency of the new best class is close to π_1 .

7 Discussion

Haigh (1978) was the first to attempt a rigorous mathematical analysis of the ideas of Muller (1964). However, in spite of the apparent simplicity of Haigh's mathematical formulation of the model, the exact rate of Muller's ratchet remains elusive. In this note, we have developed arguments in the spirit of Haigh (1978), Stephan et al. (1993) and Gordo & Charlesworth (2000) to give approximations for this rate.

Haigh gave the empirical formula

$$4N\pi_0 + 7 \log \theta + \frac{2}{s} - 20$$

for the average time between clicks of the ratchet (where time is measured in generations). A quantitative understanding of the rate was first obtained by Stephan et al. (1993) using diffusion approximations and later extended by Gordo & Charlesworth (2000). Both obtain the diffusion (5.6b) as the main equation giving a valid approximation for the frequency path of the best class.

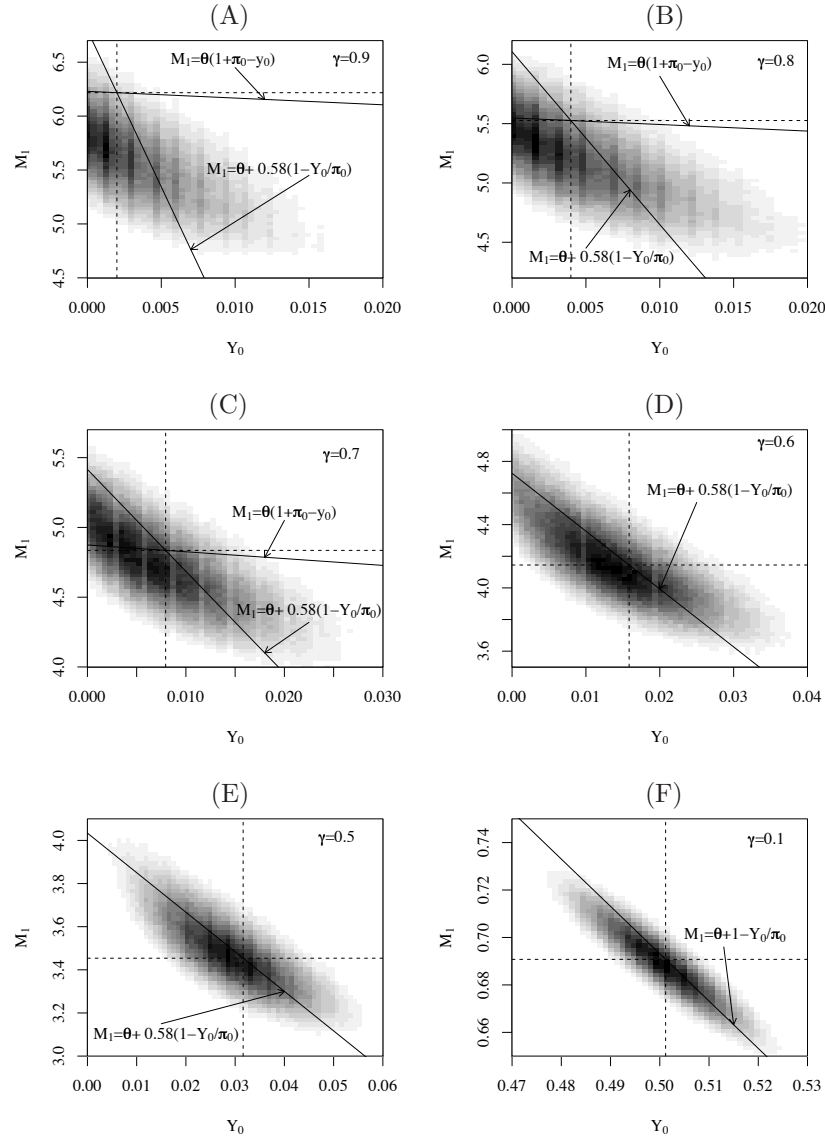


Figure 6: There are three regimes for the relationship between Y_0 and M_1 , as given in (5.3). If clicks are frequent, at least the slope of the relationship between Y_0 and M_1 fits roughly to equation (5.3a). If clicks occur reasonably often, (5.3b) gives a good approximation. If clicks are rare, (5.3c) gives a reasonable prediction. The plots show simulations for different values of γ . The dashed horizontal and vertical lines are $M_1 = \theta$ and $Y_0 = \pi_0$, respectively. For every plot we used $N = 10^4$, $\lambda = 0.1$ and simulations ran for 10^6 generations.

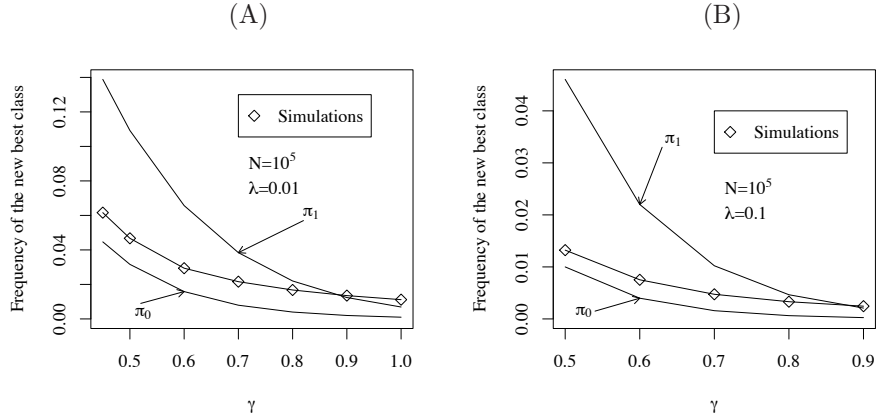


Figure 7: Our heuristic that observed states come from a RPPA applies in particular at click times. (A) For small γ , i.e., rare clicking, the frequency of the best class is already close to π_0 while (B) it is close to π_1 for larger γ , i.e., frequent clicking. Every plot is based on the simulation of $5 \cdot 10^6$ generations.

The reasoning leading to (5.6b) in these papers is twofold. Stephan et al. (1993) and Stephan & Kim (2002) argue that although fitness decreases by $se^{-\lambda}$ during one ‘cycle’ of the discrete ratchet model from §2 (in which the system advances from one Poisson equilibrium to the next), at the actual click time only a fraction of the fitness has been lost. They suggest $kse^{-\lambda}$ for $k = 0.5$ or $k = 0.6$ as the loss of fitness at click times. In other words they predict the functional relationship $M_1(Y_0)$ discussed in §5 by linear interpolation between $M_1(\pi_0) = \theta$ and $M_1(0) = \theta + ke^{-\lambda} \approx \theta + k$; compare with Figure 3. On the other hand, Gordo & Charlesworth (2000) use a calculation of Haigh which tells us that if the dynamical system (2.1) is started in $\tilde{\pi}$ from (2.4), then at the end of phase one (corresponding in the continuous setting, as we observed in Remark 4.3, to time $\frac{1}{s} \log \theta$) we have $sM_1 \approx 1 - e^{-\lambda}(1 + 0.42s)$. This leads to the approximation $M_1(1.6\pi_0) = \theta - 0.42$ and again interpolating linearly using $M_1(\pi_0) = \theta$ gives (5.6b).

Simulations show that (5.6b) provides a good approximation to the rate of the ratchet for a wide range of parameters; see e.g. Stephan and Kim (2002). The novelty in our work is that we derive (5.6b) explicitly from the dynamical system. In particular, we do not use a linear approximation, but instead derive a functional linear relationship in Proposition 4.4. In addition, we clarify the rôle of the two different phases suggested by Haigh. As simulations show, since phase one is fast, it is already complete at the time when phase two starts. Therefore, in practice, we observe states that are relaxed PPAs.

The drawback of our analysis is that we cannot give good arguments for the choice of $A = 1$ in (5.3b) and (5.6b). However, note that the choice of $A = 1$ is

essential to obtain the prefactor of 0.58 in (5.3b). E.g., if $\theta = 10$, the choice of $A = 0.5$ leads to a prefactor of 0.13 while $A = 2$ leads to the prefactor of 0.95, neither of which fits with simulated data; see Figure 6.

We obtain two more diffusion approximations, which are valid in the cases of frequent and rare clicking, respectively. In practice, both play little rôle in the prediction of the rate of the ratchet. For fast clicking, (5.6b) shows the same power law behaviour as (5.6a) and rare clicks are never observed in simulations.

Of course from a biological perspective our mathematical model is very naive. In particular, it is unnatural to suppose that each new mutation confers the same selective disadvantage and, indeed, not all mutations will be deleterious. Moreover, if one is to argue that Muller's ratchet explains the evolution of sex, then one has to quantify the effect of recombination. Such questions provide a rich, but challenging, mathematical playground.

Acknowledgements We have discussed Muller's ratchet with many different people. We are especially indebted to Ellen Baake, Nick Barton, Matthias Birkner, Charles Cuthbertson, Don Dawson, Wolfgang Stephan, Jay Taylor and Feng Yu.

References

- Bürger, R. (1991). Moments, cumulants and polygenic dynamics. *J. Math. Biol.*, 30:199–213.
- Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. USA*, 75(11):5618–5622.
- Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.*, 6:149–162.
- Charlesworth, B. and Charlesworth, D. (1997). Rapid fixation of deleterious alleles can be caused by Muller's ratchet. *Genet. Res.*, 70:63–73.
- Charlesworth, B. and Charlesworth, D. (1998). Some evolutionary consequences of deleterious mutations. *Genetica*, 102/103:2–19.
- Gessler, D. D. (1995). The constraints of finite size in asexual populations and the rate of the ratchet. *Genet. Res.*, 66(3):241–253.
- Gordo, I. and Charlesworth, B. (2000). The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics*, 154(3):1379–1387.
- Haigh, J. (1978). The accumulation of deleterious genes in a population—Muller's Ratchet. *Theor. Popul. Biol.*, 14(2):251–267.
- Higgs, P. G. and Woodcock, G. (1995). The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. *J. Math. Biol.*, 33:677–702.

- Lambert, A. (2005). The branching process with logistic growth. *Ann. Appl. Probab.*, 15(2):1506–1535.
- Loewe, L. (2006). Quantifying the genomic decay paradox due to Muller’s ratchet in human mitochondrial DNA. *Genet. Res.*, 87:133–159.
- Maia, L. P., Botelho, D. F., and Fontatari, J. F. (2003). Analytical solution of the evolution dynamics on a multiplicative fitness landscape. *J. Math. Biol.*, 47:453–456.
- Maynard Smith, J. (1978). *The Evolution of Sex*. Cambridge University Press.
- Rice, W. (1994). Degeneration of a non-recombining chromosome. *Science*, 263:230–232.
- Stephan, W., Chao, L., and Smale, J. (1993). The advance of Muller’s ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genet. Res.*, 61(3):225–231.
- Stephan, W. and Kim, Y. (2002). Recent applications of diffusion theory to population genetics. In *Modern Developments in Theoretical Population Genetics*, Oxford University Press, pages 72–93.