Wagner's Canalization Model

Emilia Huerta-Sanchez* and Rick Durrett†

September 22, 2006

*Supported by a Cornell-Sloan fellowship.

†Partially supported by NSF grants from the probability program (0202935) and from a joint DMS/NIGMS initiative to support research in mathematical biology (0201037).

Abstract

Wagner (1996) and Siegal and Bergman (2002) have studied a simple model of the evolution of a network of N genes, in order to explain the observed phenomenon that systems evolve to be robust. These authors primarily considered the case $N = 10$ and used simulations to reach their conclusions. Here we investigate this model in more detail, considering systems of different sizes with and without recombination, and with selection for convergence instead of to a specified limit. For the simpler evolutionary model lacking recombination, we analyze the system as a neutral network. This allows us to describe the equilibrium distribution networks within genotype space. Our results show that, given a sufficiently large population size, the qualitative observation that systems evolve to be robust, is itself robust, as it does not depend on the details of the model. In simple terms, robust systems have more viable offspring, so the evolution of robustness is merely selection for increased fecundity, an observation that is well known in the theory of neutral networks.

1 Introduction

The idea that wild-type genotypes are mutationally robust, i.e., are buffered against the effect of mutation goes back to Waddington (1942), an effect he called canalization. Waddington argued that for a well adapted trait almost all mutations with an effect are deleterious. For this reason, any modifier that reduces the effect of mutations and thus keeps the trait closer to optimum should be selected. Following the classical work of Schmalhausen (1949), Waddington (1957) and their contemporaries, research on robustness experienced a decline in the 1970s and 1980s. However, in the early 1990s as powerful molecular techniques to track and manipulate genotypes became routine, there was a renewed interest in the issue of genetic robustness.

Scharloo (1991) describes the results of a number of early experiments with *Drosophila*. More recently Rutherford and Lindquist (1998) and Queitsch, Sangster, and Lindquist (2002) have examined the role of the heat shock protein Hsp90. All these experiments show increased phenotypic variance in populations that carry a major mutation or are exposed to environmental stress and this is interpreted as evidence for the existence of an evolutionary buffer capable of hiding and then releasing genetic variation. See Stearns (2002) and Nivens (2004) for more discussion of the experiments.

The evolution of mechanisms underlying the buffering of the phenotype against genetic and environmental influences has received much theoretical attention, yet many issues remain unresolved. For discussions see Meiklejohn and Hartl (2002), Gibson and Wagner (2002), or de Visser et al. (2003). There have been a number of different approaches to modeling, see

e.g., Hermisson, Hansen, and Wagner (2003), Hermisson and Wagner (2004). This paper follows the approach of Wagner (1996) and Siegal and Bergman (2002) who investigated an interacting network of N genes described by an $N \times N$ matrix where $1 =$ on and $-1 =$ off . Masel (2004) later investigated a variant in which $1 =$ on and $0 =$ off.

In Section 2 we will describe the original version of Wagner's model and Masel's modification of the model. Section 3 introduces three population dynamics: one due to Wagner (1996), one to Siegal and Bergman (2002), and an intermediate model which is a hybrid of the two. Section 4 is devoted to a detailed study of the 2×2 case and a use of Markov chain theory to derive a result for the asymptotic behavior in an infinite population for a general N. In Section 5, we investigate systems with $N = 4, 7$, and 10 genes by simulation. In Section 6, we discuss the conclusions we have reached based on our analytical results and simulations.

2 The network model

Following Wagner (1996) we consider a finite population of M randomly mating individuals, each of which has an interacting network of N genes. These interactions are represented by an $N \times N$ matrix W, whose elements w_{ij} indicate the effect on gene i of the product of gene j, which may involve activation ($w_{ij} > 0$) or repression ($w_{ij} < 0$). Changing expression levels are modeled by the set of difference equations:

$$S_i(t+1) = \sigma \left[\sum_{j=1}^{N} w_{ij} S_j(t) \right] \tag{1}$$

where σ is the sign function; $\sigma(x) = 1$ if $x > 0$, $\sigma(x) = -1$ if $x < 0$, and $\sigma(0) = 0$. $h_i(t) = \sum_{j=1}^{N} w_{ij} S_j(t)$ represents the sum of all regulatory effects of all of the network genes on gene i. Ignoring the possibility that the sum is exactly zero, which in our model will have probability 0, $S_i(t)$ only takes the values -1(not expressed) or 1(expressed). Siegal and Bergman (2002) replace σ in (1) by a sigmoidal function:

$$f(x) = \frac{2}{1 + e^{-ax}} - 1 \tag{2}$$

They choose $a = 100$ so $f(x)$ is close to 1 when $x \notin [-.03, .03]$. Since this is only a small perturbation of the sign function, we will simplify the analysis by using Wagner's sign function, which results in a dynamical system on the finite set $\{-1, 1\}^N$. In the version of the model used by Masel (2004), $S_i(t)$ takes values 1 (expressed) and 0 (not expressed) where $\sigma(x) = 1$ if $x \geq 0$ and $\sigma(x) = 0$ if $x < 0$. This is also a dynamical system on the finite set $\{0, 1\}^N$. In the 0, 1 map, $S_i(t) = (0, \ldots, 0)$ can never be a fixed point. The 0, 1 mapping is more realistic biologically since genes that are off, have no effect on anything while in the -1,1 formulation if gene i is off it has a positive effect on gene j with $W_{i,j} < 0$. As we will see later the 0,1 map has many more fixed points due to turning off various subsets of genes. Note that by definition the all off state is not a fixed point.

2.1 Random networks

In a moment we will introduce the population dynamics into the system. However, in order to assess how selection changes the collection of possible gene networks, we will first

consider properties of randomly chosen 4×4 networks. For convenience we will rewrite the state space $\{-1, 1\}^4$ as $\{0, 1, \ldots 15\}$ by replacing -1 by 0 and regarding the sequence as a binary number. For example, $(1, -1, -1, 1)$ becomes $1001 = 9$. We generated w_{ij} that are independent and identically distributed uniform random variables on $[-1, 1]$ and then flipped the sign of each row (if necessary) to make $(1, 1, 1, 1) = 15$ a fixed point. We then iterated the system starting from $(1, 1, -1, -1) = 12$ and $(1, 1, 0, 0) = 12$ when using the 0, 1 model. This choice is somewhat arbitrary, but when $(1, 1, 1, 1)$ is a fixed point the qualitative behavior of the system only depends on the number of -1's in the initial state, so there are only three interesting choices: 8, 12, and 14. For each matrix we generated, we iterated until a fixed point was reached or the system settled into a periodic orbit. Then we counted the number of times each state was the limiting fixed point, how long it took to reach $(1, 1, 1, 1)$ when it was the limit, the probability of reaching $(1, 1, 1, 1)$ in $0 - 15$ time steps, and the lengths of the periodic orbits obtained. The results are presented in Figure 1.

To see the effect of the structure of the mapping on the statistics, we generated a random mapping ϕ of $\{0, 1, \ldots 15\}$ into itself with $\phi(15) = 15$, started from 12, and iterated ϕ and recorded the outcomes as before. The results from the random map are much different from the Wagner map and from the 0, 1 map. For the random map, 15 and 12 are more likely to be the limiting fixed points, but all other limits occur with almost equal frequency, and the time to reach 15 is much longer for the random map. For the 0, 1 map (the black bars) 15 is a fixed point most of the time and 0 is never a fixed point.

One reason for the difference in qualitative behavior is that the Wagner map has an

antisymmetric property: $\sigma(W(-s(t))) = -\sigma(Ws(t)) = -s(t+1)$. To take this into account we generated random maps of $\{0, 1, \ldots 15\}$ into itself with $\phi(15) = 15$ and $\phi(15 - x) = 15 - \phi(x)$. This reduced the discrepancy between the random map and the Wagner map. Antisymmetry dictates that 3 can never be the limiting fixed point since we start at 12, but 4, 8, 13, and 14 more frequently occur as limits in the Wagner system compared to the antisymmetric one. The two antisymmetric systems have a similar distribution of the time to reach 15. This is due in part to the fact that the states $0 - 7$ are the mirror images of states $8 - 15$ for these systems so if a fixed point is to be reached at all it must happen before 8 steps. For the 0, 1 map the time to reach 15 is almost always 1 or 2, and 15 is reached on average faster than with the Wagner map, see Fig. 1(b). After evolving the 4×4 matrices as explained in a later section, the probability of reaching the fixed point 15 in fewer time steps increases. Both the Wagner and the random antisymmetric system share a lack of long period orbits of odd length. Indeed, a periodic orbit of length 9 is impossible under an antisymmetric map because its reflection would be a second disjoint periodic orbit. Periodic orbits of length 7 are not ruled out by antisymmetry but do not occur for the Wagner map. The size of periodic orbits in the 0, 1 map is with high probability 2 or 3, while sizes 4 to 10 are less likely.

3 Population Dynamics

For simplicity, we will describe the simulations for the 4×4 case. We consider three different evolutionary simulations. For all three, we assume the Wagner mapping for the network

dynamics and for comparison in one of the scenarios we apply the 0, 1 map.

In the first simulation, which is similar to the approach of Wagner (1996), we have a fixed optimum phenotype, which without loss of generality we can suppose is $S^{\text{opt}} = (1, 1, 1, 1)$. To generate the founding population, we generate $10,000$ random matrices with entries that are independent and uniformly distributed on $[-1, 1]$, and flip the sign of each row (if necessary) so that we have $15 = (1, 1, 1, 1)$ as a fixed point. To see if this matrix will be included in the initial population we iterate starting from $12 = (1, 1, -1, -1)$ and see if there is convergence to the fixed point $15 = (1, 1, 1, 1)$. The rest is like a Wright-Fisher model: To create the $(n + 1)$th generation, we randomly pick an individual (matrix) from the nth generation. In each row with probability $1/4$ we mutate one randomly chosen entry, and we replace that entry by a uniform$(-1, 1)$ random number. After mutation, we check if the individual is developmentally stable and if its fixed point is $S^{\text{opt}} = 15$. If so, then it is included in the next generation. We continue this process until we have 10,000 individuals for the $(n + 1)$th generation.

The model of Siegal and Bergman (2002) differs from the first simulation in two important ways: they have selection for convergence but not to a predefined limit and reproduction occurs with recombination. In our version of their situation, to generate the founding population, we generate a random matrix with entries that are independent and uniformly distributed on $[-1, 1]$, and that starting at $15 = (1, 1, 1, 1)$, reaches a fixed point. Again this starting point can be chosen without loss of generality. Once this matrix has been found, we clone it 10,000 times. To create the $(n + 1)$th generation, we randomly pick two indi-

viduals (matrices) from the nth generation. To simulate recombination, we randomly pick rows from these two matrices (with equal probability) to create a child. Once the child has been made, we mutate at most one entry in each row with probability $.1/4 = 1/40$ (we use the same mutation rate as Siegal and Bergman (2002), .1 per matrix), and we replace that entry by a uniform$(-1, 1)$ random number. After mutation, we check if the individual is developmentally stable (i.e., if the iteration still reaches a fixed point). If it does, then it is included in the next generation. We continue this process until we have the desired number of individuals for the $(n + 1)$th generation. This simulation was done using both Wagner map and the 0, 1 map. When we use the 0, 1 map, the only differences are the mapping and the state space. For the 0, 1 map, the sign function is 0 when $(WS)_i$ (the sum of all regulatory effects on gene i) is less than zero, and the sign function is 1 when $(WS)_i$ is greater or equal to zero. The state space (phenotypic states) are now vectors whose entries are ones and zeros instead of ones and negative ones.

To interpolate between the first two scenarios, we will also consider a system with selection for convergence, but not to a predefined limit, and reproduction occurs without recombination.

The simulations above have some minor differences from the simulations by Wagner (1996) and Siegal and Bergman (2002). One significant difference is that in Siegal and Bergman (2002), instead of using the sign function to determine the next state, they use a sigmoidal function (2). Since the state space is continuous, they need a criterion to determine when a fixed point has been reached. They average the distances between $S(\theta)$ and $\overline{S(t)}$ for

$\theta = t - \tau, .., t$, where $\overline{S(t)}$ is the average over the last $\tau = 10$ time steps. When this quantity is less than $\epsilon = 10^{-4}$, then a fixed point is considered to have been reached. Individuals that do not reach an equilibrium are assigned fitness zero.

4 Markov chain model

To gain further insight into the structure of the Wagner map, we will look at the simplest nontrivial case. In the 2×2 we can explicitly draw out what the trajectories on the $\{-1, 1\}^2$ state space will look like. Figure 2 has all the possible types of systems and how many of each type occur (by permuting the positions of orbits and fixed points while still keeping the system antisymmetric). The antisymmetry is geometrically an invariance through the center of the square whose corners represent the states.

The total number of systems that are possible is determined by a particular choice of sign of the quantities $w_{i1} \pm w_{i2}$ for $i = 1, 2$ because the form of Wagner's map means that these quantities determine what the mapping does to $(1, 1)$ and $(1, -1)$, which with the asymmetry property specifies the full dynamics. For example any system whose matrix entries satisfy $w_{11} + w_{12} > 0$, $w_{11} - w_{12} > 0$, $w_{21} + w_{22} > 0$ and $w_{21} - w_{22} > 0$ will map $(1, 1) \to (1, 1)$ and $(1, -1) \to (1, 1)$. This leads to a geometrical way to categorize such systems since the first 2 inequalities restrict to the intersection of two half planes in the (w_{11}, w_{12}) plane, see Figure 3 and the second two inequalities restrict to a similar region in the (w_{21}, w_{22}) plane. Using this counting procedure there are 16 possible systems in all because there are 4 distinct regions in the (w_{11}, w_{12}) plane times another 4 in the (w_{21}, w_{22}) plane. Suppose we have a network

with matrix W, and we regard its (w_{11}, w_{12}) entries and (w_{21}, w_{22}) entries as points in two planes (separate copies of R^2). As already mentioned, it is enough to know what the signs of $w_{i1} \pm w_{i2}$, $i = 1, 2$, are in order to specify the dynamics completely. Those sign combinations place the points representing each row in a particular sector of the (w_{i1}, w_{i2}) planes.

Now consider what happens when a mutation occurs on this network: mutations are independent so the effect on row 1 and row 2 of the matrix can be considered separately. With this assumption it is seen that if row i is affected by mutation, the point representing the row will be translated in the w_{i1} or w_{i2} direction, i.e. parallel to the coordinate axes. There are two possibilities at this stage: either the new point still lies within the same sector, and therefore the network keeps the same dynamics, or it leaves the sector which means it has switched to become a different dynamical system. For a particular example, consider a network in which $w_{i1} \pm w_{i2}$ are > 0. The sector that row 1 belongs to is shown in Figure 3 (and of course a similar picture exists for row 2). The possible effects of mutation are depicted by the arrows.

To simplify the situation we suppose that there are K possible values for each entry in the matrix and that mutations are uniform over the set of possible values. This way we can write down quite easily a transition probability that gives the effect of the mutation step on the first row of a matrix:

$$p(w_{11}, w_{12}; w_{11}, w_{12}) = 1 - 2\mu$$

$$p(w_{11}, w_{12}; w_{11}^*, w_{12}) = \mu/K$$

$$p(w_{11}, w_{12}; w_{11}, w_{12}^*) = \mu/K$$

where μ is the probability of mutation. To make the notation more compact we will write x for the current state (w_{11}, w_{12}) and y for the state after a possible mutation. Since a mutation can take us to a nonviable state we have $\sum_y p(x,y) < 1$, where the sum is over all states y in the optimum phenotype region (the neutral network). In an infinite population in which the initial state consisted of a fraction $\mu(x)$ of matrices with row 1 equal to x, the state after one time step would be

$$\frac{\sum_x \mu(x) p(x,y)}{\sum_{x,y} \mu(x) p(x,y)}$$

and the state after time n would be

$$\frac{\sum_x \mu(x) p^n(x,y)}{\sum_{x,y} \mu(x) p^n(x,y)} \tag{3}$$

where n is the nth power of the transition probability. Being a symmetric nonnegative matrix, see e.g., Seneta (1973), as $n \to \infty$

$$p^n(x,y) \sim \lambda^n u(x) u(y)$$

where λ is the largest eigenvalue and u is the associated eigenvector. Using this with (3) we see that asymptotically the population becomes distributed according to $u(y)/\sum_x u(x)$.

This result was derived earlier by van Nimwegen et al. (1999) page 9717. In that work, formulas are derived for the proportions of the population on nodes of the neutral network in equilibrium. This corresponds to the stationary distribution of our Markov Chain model. The equilibrium equations of that model are equivalent to ours where we identify the largest eigenvalue as the proportion of individuals remaining on the network after one time step (one generation).

The eigenvector does not depend on the mutation rate, only the convergence rate does. To see this let A be the matrix given by the transition probabilities above and note it can be written as $(\mu B + I)$ where I is the identity matrix and B is a matrix *independent* of μ. If u is an eigenvector of A with eigenvalue $\lambda(\mu)$ then $Au = (\mu B + I)u = \lambda(\mu)u$ so

$$Bu = \frac{\lambda(\mu) - 1}{\mu} u$$

If we change the mutation rate from μ to θ, then u will be an eigenvector with eigenvalue

$$\theta \frac{\lambda(\mu) - 1}{\mu} + 1$$

Thus the family of eigenvectors stays the same and all the matrices have the same dominant eigenvector. Again, this result can be found in van Nimwegen et al. (1999), see equation [6]. Here we have restricted our attention to $N = 2$ for simplicity. The analysis above can be extended to discretized $N \times N$ networks, or using the theory of Harris chains to our original continuous chain.

Figure 4*a* gives the dominant eigenvector calculated from the transition probability p. Note that, as one would expect, the distribution has smaller probabilities near the boundaries, since matrices lying near the boundaries of developmental stability are more sensitive to mutations than those further away. This feature that the evolved ensemble tends to have fewer members near the boundaries is a concrete visualization of what is described in De Visser et al. (2003), " A population that is mostly concentrated in the interior of wide parts of the plateau will be much less sensitive than a population that is distributed over the narrower parts, where mutations easily push individuals over the edge." In van Nimwegen et al. (1999), this corresponds to the highly connected region of the neutral space. The individuals in the interior of the triangular region are connected to many same phenotype neighbors by single mutation moves; ones at the narrow part of the triangular region nearest the origin have the fewest neighbors in the neutral network. Figure 4*b* gives the distribution of w_{11} and w_{12} after 200 generations for our first simulation (selection for $(1, 1)$ as an optimum and no recombination). The distributions have been normalized such that the total probability equals 1 and we average over 1000 runs. Figures 4*c* and 4*d* show what happens when we reduce the mutation rate by a factor of 10. Figure 4*c* is the distribution after 200 generations. Since the mutation rate is reduced, we do not see the concentration of probability in the interior after this time. However by running for 10 times longer (2000 generations), Figure 4*d*, we recover a distribution which is similar in the higher mutation rate case.

This is expected given the independence of the dominant eigenvector from mutation rate shown above. There is one caveat with this result – it does not hold if the product of the

population size and the mutation rate, μ, is small. In that case the population becomes concentrated at one point, very few offspring are lost to mutation, and all viable points are visited with equal probability. For more discussion of this point, see the section on finite population effects in van Nimwegen et al (1999).

4.1 Recombination

Consider the $N = 2$ case with recombination and selection *only* for developmental stability. As described in 2×2 *networks*, there are sixteen dynamical systems possible depending on the signs of the sums and differences of rows in a given network matrix. The set of dynamical systems partitions into those that starting from $(1, 1)$ converge to a fixed point and those that end up in periodic orbits. The second case corresponds to being developmentally unstable, and we can ask whether moving away from these genotypes during simulated evolution naturally leads to a more robust genotype among the developmentally stable networks.

To examine this we have studied by simulation the fraction of a population of networks that have a given sign combination after evolution, where the initial state is chosen to be (1,1). In Table 1, $-+$ is short for $w_{i1} + w_{i2} < 0$ and $w_{i1} - w_{i2} > 0$ where $i = 1, 2$ is the row number. The left margin gives the sign combinations for row 1 and the top margin gives the sign combinations for row 2. So the combination (+-,-+) can be read as the regions that satisfy: $w_{11} + w_{12} > 0$, $w_{11} - w_{12} < 0$ and $w_{21} + w_{22} < 0$,$w_{21} - w_{22} > 0$. This combination has a zero in the table since starting from $(1, 1)$ the system moves into a periodic orbit and so will never be developmentally stable. The four nonzero entries that lie in the first

two rows and first two columns of the table have $(1,1)$ as a fixed point while the other two have $(1,-1)$ as their fixed point. As can be seen the combination $(++,+-)$ is the most robust followed by $(++,++)$ and $(+-,+-)$ since they make up the largest percentages of the evolved population, 54.6%, 14.9% and 16.4% respectively. Furthermore, these prevalent combinations all have path lengths of zero (starting from $(1,1)$). We can get a picture of why $(++,+-)$ is the most robust if we consider that after a recombination event, typically one row of the parent matrix will be swapped with a completely random one. Consider a matrix that has $++$ as its first row: during recombination, if its second row gets swapped (i.e. we move horizontally along Table 1), there is only a 1 in 4 chance the child will be developmentally unstable. (Compare with the other sign combinations for row 1 which lead to developmental instability more frequently). Similarly it is clear that the sign combination $+-$ for the second row is optimal. Therefore the most robust network to recombination events should have the combination $(++,+-)$ as is observed. We also observe that under selection pressure for developmental stability, a high proportion of the evolved networks have their initial state and limiting fixed point being equal i.e. they have a path length of zero.

For the 0, 1 map the regions are more complicated. For $N = 2$, the limiting fixed points starting from $(1,1)$ are $(1,1)$, $(1,0)$ and $(0,1)$. The regions where the individuals are viable are $w_{11} + w_{12} \geq 0$ if one wants $(1,1)$ to be a fixed point. The regions $w_{11} + w_{12} \geq 0$ and $w_{21} + w_{22} < 0$ and $w_{11} \geq 0$ and $w_{21} < 0$ in order to have $(1,0)$ as a fixed point. To have $(0,1)$ be a fixed point we need $w_{11} + w_{12} < 0$ and $w_{21} + w_{22} \geq 0$ and $w_{22} \geq 0$ and $w_{12} < 0$. After simulated evolution we have 79.1% of the matrices have $(1,1)$ as a fixed point, 10.4%

of the matrices have $(1, 0)$ as a fixed point, and 10.4% of the matrices have $(0, 1)$ as a fixed point.

5 Path length and probability of viability for $N = 4, 7, 10$

In higher dimensions it is impossible to visualize the distribution of even one row, so we instead investigate statistics associated with the mapping. We compute the mean path length at each time point for each run and then take the average over the number of runs. Path length is the number of steps that the individual takes to reach an equilibrium (fixed point), and a run simply means that a population of networks has been evolved for a fixed number of generations which is 200 for the first simulation, and 1500 for the other two, which have a mutation rate that is 10 times smaller.

In all three simulations, for all values of N, the path length gets smaller with time. Our result in Figure 5*a* with $N = 10$ is similar to the one in Wagner (1996). Bergman and Siegal (2002) have longer path lengths than we do in 5*c* because they have to wait at least ten iterations (their τ) to be able to compute $\overline{S(t)}$, which is an average of the expression levels in the time interval $(t - \tau, ..., t)$, and must wait longer for the average to be within 10^{-4} of its limit.

When there is selection for convergence but not to a particular optimum then in the absence of recombination path lengths are larger (see 5*b*), but are reduced dramatically by recombination. Note that the path length for $N = 4$ is close to zero, i.e., the starting point is often a fixed point. This phenomenon cannot be observed in the Bergman and Siegal

(2002) set-up due to their definition of convergence time. Figure 5*d* shows the results for the path length for the 0, 1 map. The initial path length is a lot smaller than for the Wagner map. This is expected since the total number of fixed points for the 0, 1 map is bigger. If we generate a million random matrices the percentage of fixed points reached starting from 15 is 49.8%, 40.1%, 32.3% for $N = 4, 7, 10$ respectively for the 0, 1 map. However, for the Wagner map the number of fixed points reached is 22.5%, 12.1%, 7.1% for $N = 4, 7, 10$.

The second measurement we take is the probability of being viable after mutation, which quantifies the degree of canalization. In the simulation with selection for a particular optimum, we perturb the evolved matrices (mutation of one single entry in each matrix) and check whether the matrix still has the desired fixed point. When there is selection for convergence but not to a particular optimum, we check whether the evolved matrices still reach a fixed point after being perturbed. Figure 6 shows that the probability of being viable after mutation in all simulations increases as the number of generations and the limit is larger for larger N. Thus as, Wagner (1996) and Siegal and Bergman (2002) have seen, canalization increases with complexity. (They keep N fixed at 10 and vary the fraction of nonzero entries.) For the 0, 1 map, the probability of being viable starts at a pretty high value, increases with generation number and then levels off.

6 Discussion

We have examined a model of the evolution of a population of genetic networks that extends earlier work of Wagner (1996) and Siegal and Bergman (2002), who investigated particular

cases of the model by simulation. The aim of this investigation is obtain a more thorough understanding of the properties of the model, in order to better understand the conclusions that can be drawn from the observed behavior. As in previous work, we find that the networks evolve to be more robust. We find these conclusions in Wagner's system with no recombination and selection for convergence to a particular optimum (stabilizing selection), in Siegal and Bergman's system with recombination and selection for convergence but not to a particular limit (developmental stability), as well as in a hybrid model with no recombination and selection for convergence but not to a particular optimum.

There are differences in the degree of canalization under the three schemes. As Azevedo et al. (2006) have pointed out, sexual reproduction and the accompanying recombination enhances canalization. However, the qualitative behavior of our three models are similar indicating that the evolution of robustness, is itself a robust conclusion. We found also that the evolution of robustness did not depend on our choices of mapping for the network dynamics. The reason for the evolution of robustness is easy to understand. Robust systems are less sensitive to mutation and hence have a larger number of viable offspring. Thus the selection for robustness is simply selection for greater fecundity, as Siegal and Bergman (2002) remark on page 10530 in their results and discussion.

We also observe, as Wagner (1996) and Siegal and Bergman (2002) did previously, a reduction in the mean length of the path to equilibrium, since longer paths are more easily disturbed by mutation. A second by product of mutational robustness is the presence of negative epistatic interactions observed by Azevedo et al. (2006). The equilibrium distribu-

tion of the model concentrates on configurations that are less easily harmed by mutation, so mutants will on the average experience a greater damage by a second mutation.

While many of our conclusions are similar to earlier work, a few aspects of our work are new. To better understand the impact of the population dynamics on the properties of individual networks, we generated an ensemble of networks with either: a random dynamical mapping, a random antisymmetric dynamical mapping or random Wagner-type mapping. The distribution of the fixed points reached and the lengths of periodic orbits of these randomly generated maps provide a baseline with which to compare the properties of the evolved systems.

A second technical point is that we have returned to the original model of Wagner (1996) in order to have dynamical systems with state space $\{-1, 1\}^d$. These systems are trivial to analyze since they can only have fixed points or periodic orbits. In addition, they allow us to define the time, needed to reach the fixed point without averaging in time which artificially inflates the convergence time and obscures the patterns in Figure 5.

The simple nature of the dynamical systems allows us to recognize the model as a neutral network, i.e., a collection of mutually neutral genotypes which are connected by single mutational steps, a framework that has earlier been used in the study of RNA secondary structure, see e.g., van Nimwegen, Cruthchfield, and Huynen (1999) and Ancel and Fontana (2000). This viewpoint allows not only a convenient mental picture of the "genotypic landscape" on which the system evolves, but also enables us to use results from Markov chain theory to analyze the asymptotic behavior of the model in an infinite population, which is

determined by the dominant eigenvector.

References

Ancel, L.W., and W. Fontana. (2000) Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.* 288, 242-283

Azevedo, R., Lohaus, R., Srinivasan, S., Dang, K., and Burch. C. (2006) Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* 440, 87-90

Bergman, A., and Siegal, M.L. (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature.* 424, 549–552

De Visser, J. A. G. M., and 18 co-authors. (2003) Perspective: Evolution and Detection of Genetic Robustness. *Evolution* 57, 1959–1972

Gibson, G., and Wagner, G. (2000) Canalization in evolutionary genetics: a stabilizing theory? *BioEssays* 22, 372–380

Hermisson, J., Hansen, T.F., and Wagner, G.P. (2003) Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *American Naturalist.* 161, 708–734

Hermisson, J., and Wagner, G.P. (2003) The population genetic theory of hidden variation and genetic robustness. *Genetics.* 168, 2271–2284

Masel, J. (2004). Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. *J. Evol. Biol.* 17,

1106-1110.

Meiklejohn, C.D., and Hartl, D.L. (2002) A single mode of canalization. *Trends in Ecology and Evolution.* 17, 468–473

Niven, J.E. (2004) Channeling evolution: Canalization and the nervous system. *PLoS Biology.* 2, 22–24

Queitsch, C., Sangster, T.A., and Lindquist, S. (2002) Hsp90 as a capacitor of phenotypic variation. *Nature.* 417, 618–624

Rutherford, S. L., and Lindquist, S. (1998) Hsp90 as capacitor of morphological evolution. *Nature* 396, 336-342.

Scharloo, W. (1991) Canalization: genetic and developmental aspects. *Annu. Rev. Ecol. Syst.* 22, 65–93

Schmalhausen, I.I. (1949) *Factors of Evolution: the Theory of Stabilizing Selection.* Blakiston, Philadelphia. Reprint 1986, U. of Chicago Press.

Seneta, E. (1973) *Nonnegative Matrices.* John Wiley and Sons, New York

Siegal, M. L., and Bergman, A. (2002) Waddington's canalization revisited: Developmental stability and evolution. *Proc. Natl. Acad. Sci. USA* 99, 10528–10532

Stearns, S.C. (2002) Progress on canalization. *Proc. Nat. Acad. Sci.* 99, 10229–10230

van Nimwegen, E., J. P. Crutchfield, and M. Huynen. (1999) Neutral evolution of mutational robustness. *Proc.natl.Acad. Sci. U.S.A.* 96, 9716–9720

Waddington, C.H. (1942) The canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565

Waddington, C.H. (1957) *The Strategy of the Genes.* MacMillan, New York

Wagner, A. (1996) Does evolutionary plasticity evolve? *Evolution* 50, 1008–1023

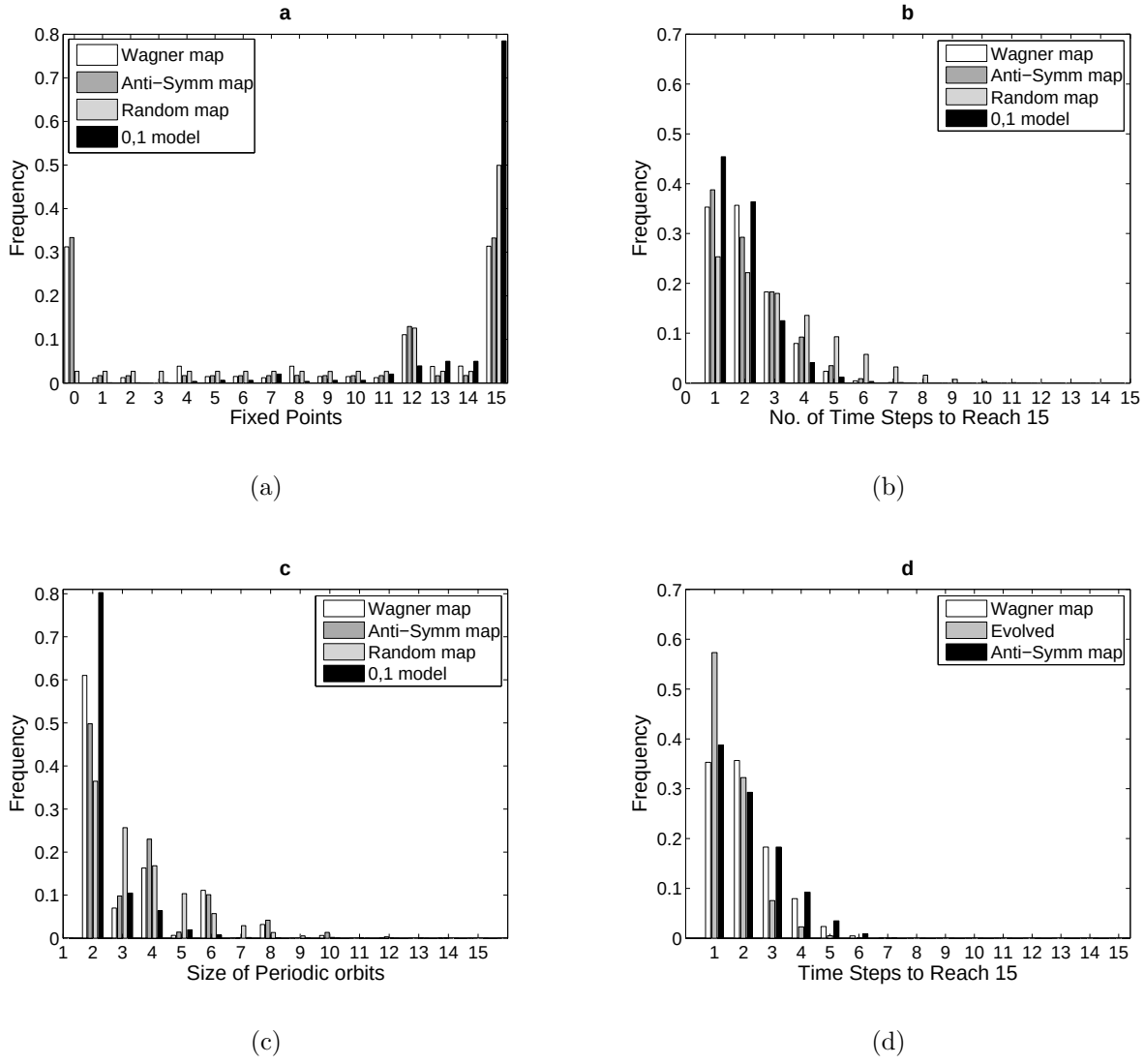

Figure 1: Comparing the Wagner map, the antisymmetric map, the random map, and for (a)-(c) we include the 0, 1 model. **a**. Frequency of fixed points, **b**. Frequency of time steps to reach 15, **c**. Frequency of periodic orbits (by size), **d.** Frequency of reaching 15 (at different time steps) evolved refers to networks that were evolved with no recombination and selection for an optimum. The initial state is 12.

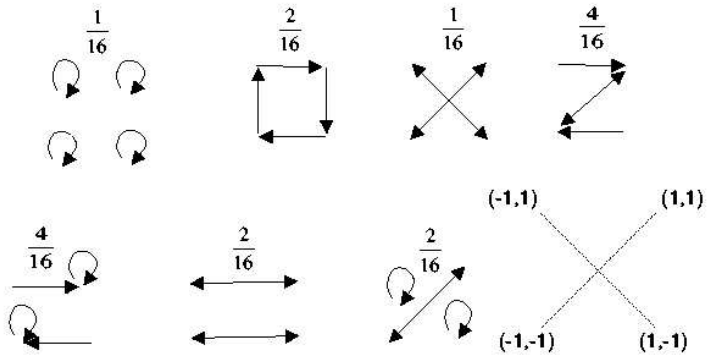

Figure 2: An enumeration of all the possible dynamical systems of the Wagner type for a 2 gene interacting network. If a network is generated completely randomly it will be of one of these forms with the probability given.

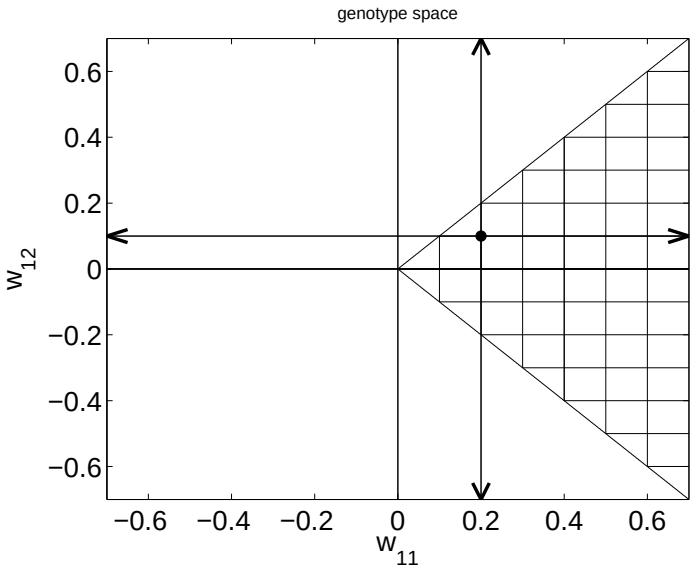

Figure 3: A particular discretization of the (w_{11}, w_{12}) where $w_{1i} \in \{-.7+0.1m | m = 0, .., 14\}$. Since the other points have fitness zero, we keep track of the states within the triangular region in the G matrix only. However transition probabilities are evaluated on the basis that a state can move to anywhere in the square moving in a horizontal or vertical direction after a mutation.

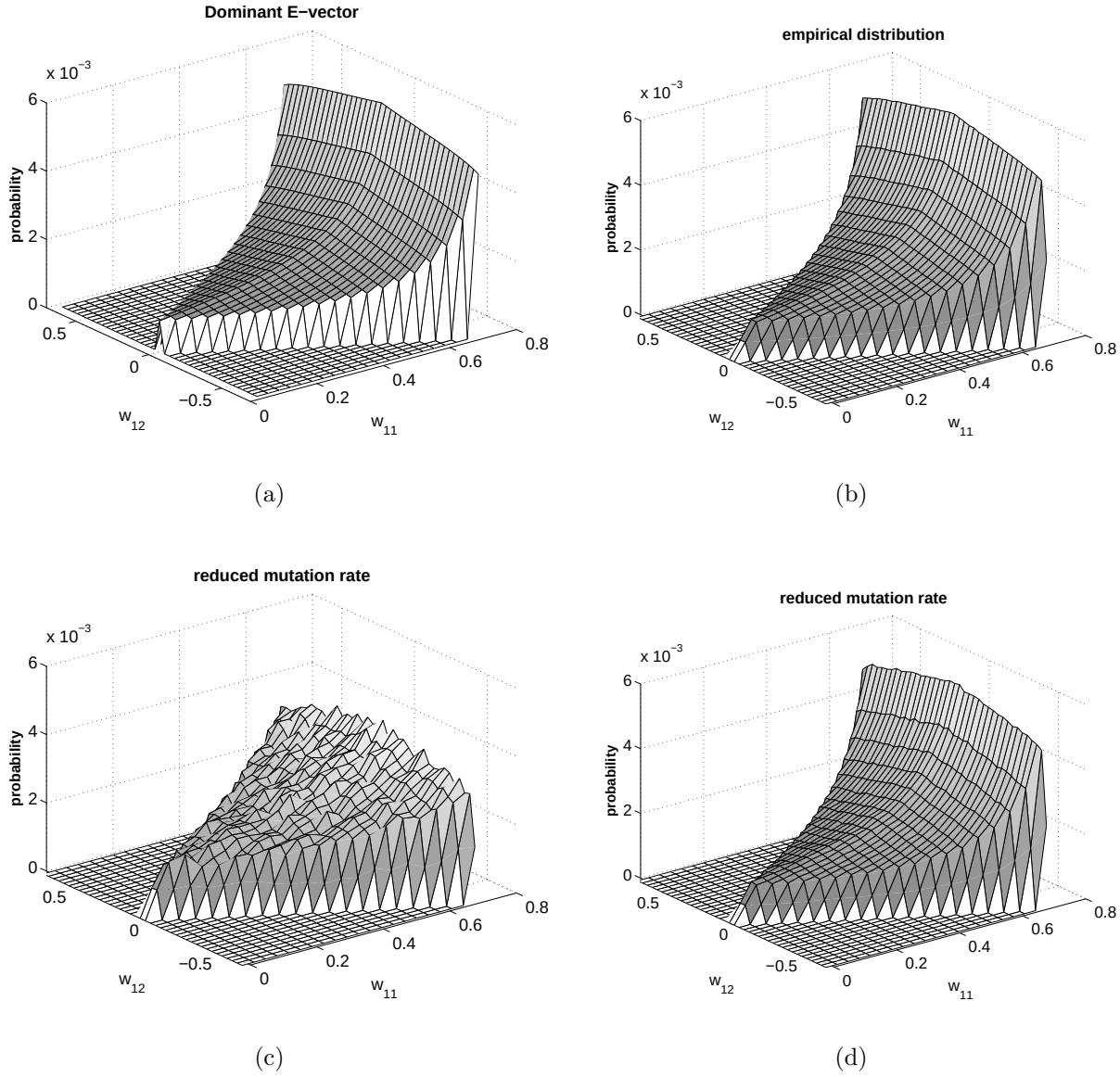

Figure 4: The distribution of w_{11} and w_{12} under different conditions. **a**. The stationary distribution given by the dominant eigenvector. **b**. The empirical distribution of w_{11} and w_{12} after 200 generations, **c** the empirical distribution with reduced mutation rate by a factor of ten at generations 200, **d**. same as (c) but at generation 2000. We see that the combinations with higher probability are away from the boundaries since individuals with those genotypes are more immune to mutations.

Table 1

Row1\Row2	++	+−	−+	−−
++	0.149260	0.546570	0	0.047830
+−	0.040460	0.163990	0	0
−+	0	0.051890	0	0
−−	0	0	0	0

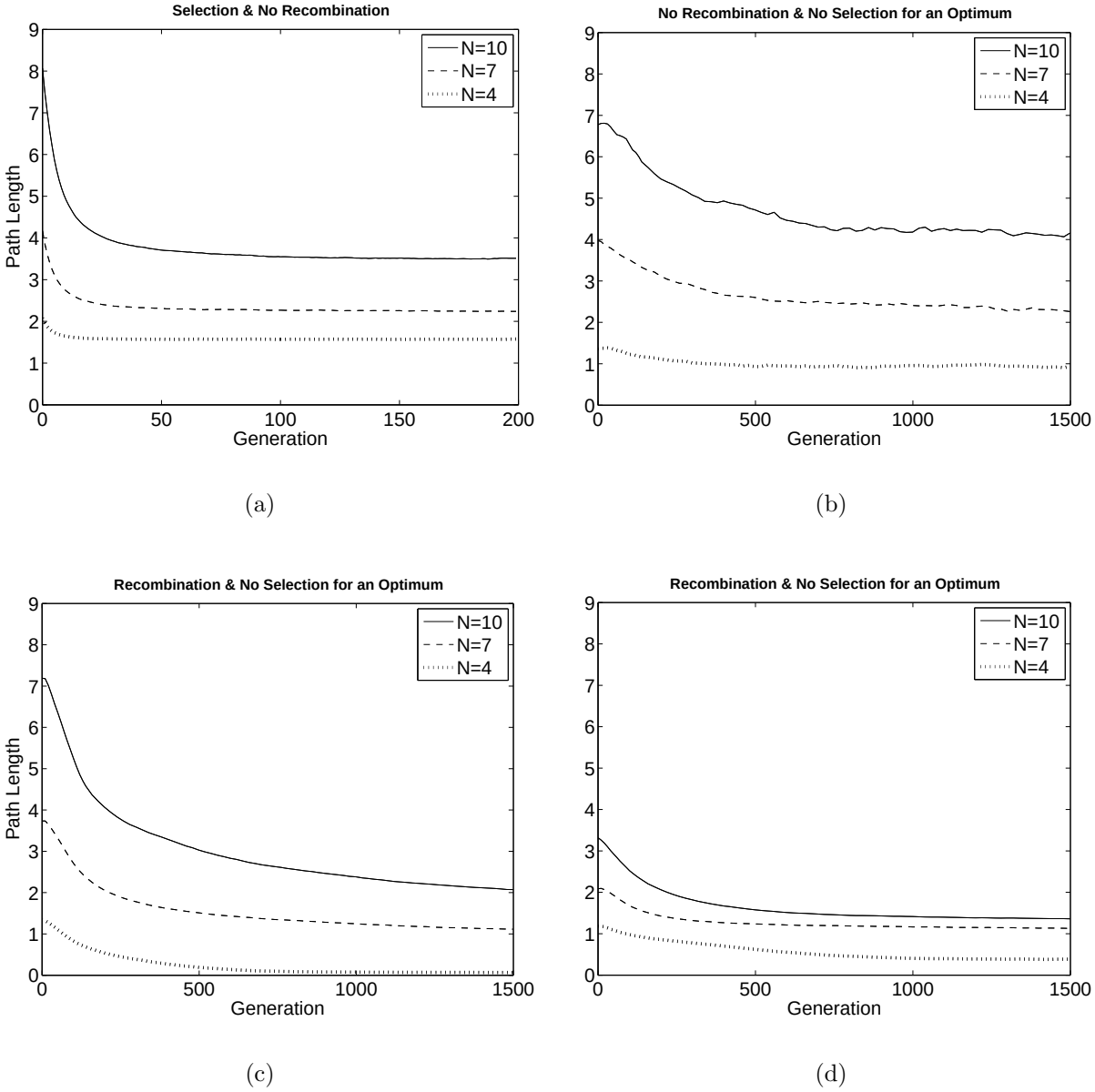

Figure 5: Path length defined as the number of steps that it takes to reach a fixed point. Population size is 10000 and number of runs is 200. (a) An asexual population with selection, the fixed point has to be the optimum state. (b) An asexual population with no selection for an optimum, (c) a sexual population with no selection for an optimum, and (d) As in (c) but we used the 0, 1 version of the model.

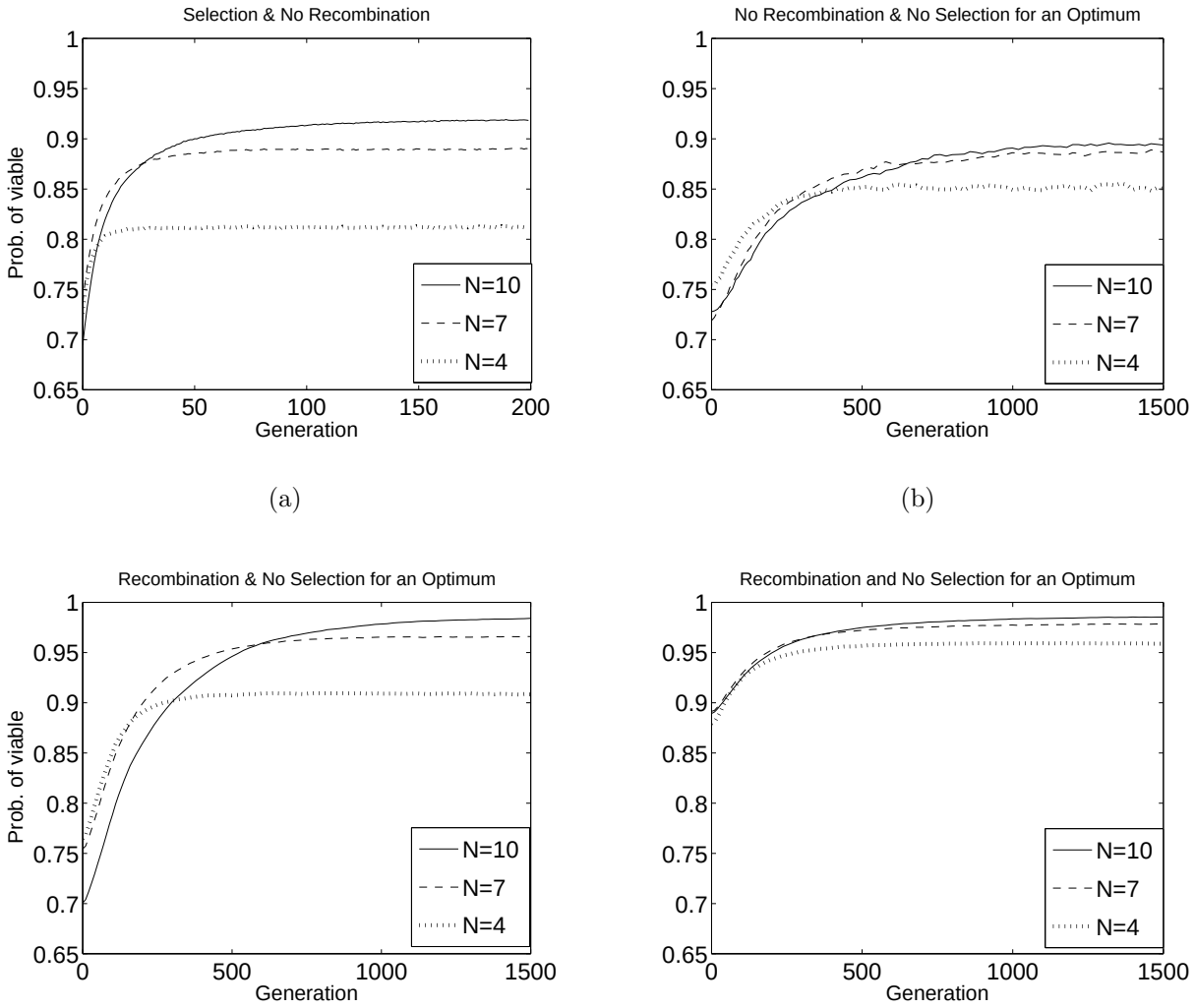

Figure 6: Probability of survival defined as being developmentally stable after one perturbation. Population size is 10000 and number of runs is 200. (a) An asexual population with selection, the fixed point has to be the optimum state. (b) An asexual population with no selection for an optimum, (c) a sexual population with no selection for an optimum, and (d) As in (c) but we used the 0, 1 version of the model.