

Connections between Noisy Fitness and Selection Strength

Vincent R. Ragusa^{1,2} and Clifford Bohm^{1,3}

¹Michigan State University, East Lansing, MI 48823

²ragusavi@msu.edu

³cliff@msu.edu

Abstract

Selection can be described as a filtering process which changes a population over time with regard to the result of some evaluation (i.e. a fitness function). We are interested understanding the relationship between different parameters for altering selection strength and rates of adaptation. In this work we perform a detailed assay exploring the relationship between population size, noisy phenotype evaluation, and tournament size, and their effects on rates of genomic change. We run our model on nearly 4,500 different scenarios.

We observe evolution on a smooth fitness landscape as well as nine deceptive landscapes using our model. We show that for the smooth landscape it is always best to have strong selection with noise-free fitness and a large population. For deceptive landscapes, there is an optimum configuration of tournament size and noise that balances exploration and exploitation. Population size, on the other hand, always increases genomic change when larger, because it not only increases selection strength but also maximizes mutational inflow and standing variation. We see that while these parameters for selection strength have similar effects, they each behave in unique ways. Finally, we suggest that evaluation noise is a better proxy for selection strength than population size.

Introduction

Selection is perhaps the most central concept in evolutionary theory. Selection can be described as a filtering process which changes a population over time with regard to the result of some evaluation (i.e. a fitness function). The filter can range from fully permissive, where the result of the evaluation is not considered, to maximally restrictive, where only the organism(s) with the best evaluation pass through. These two extreme cases are referred to as drift and elitism respectively. All other 'strengths of selection' fall somewhere in between drift and elitism.

Background

We are interested understanding the relationship between parameters of evolution and rates of adaptation. Here, we limit our investigation to the relationship between selection strength and rates of adaptation in the context of different fitness landscapes.

In evolutionary theory, the concept of a “fitness landscape” is the relationship between genotypes and fitness (WRIGHT, 1932). A position in a fitness landscape surrounded by regions of lower fitness is called a “local optimum”. If a population is positioned on a local optimum, then it must cross a region of lower fitness — called a “fitness valley” — in order to escape the local optima and potentially ascend to higher fitness.

Although a population does not always require a loss in fitness to cross a fitness valley (Iwasa et al., 2004) deleterious mutations are often necessary to explain valley-crossing events. For example, Covert et al. (2013) disallow deleterious mutations and observe that the adaptability of the system is hindered due to the inability to valley-cross. Similarly, it has been shown that elitism hinders adaptation on rugged fitness landscapes (Oliveto et al., 2018).

In this work we consider three specific evolutionary parameters that each effect selection strength: **population size**, **noisy evaluation**, and **tournament size** (a parameter specific to our model).

Population Size Population size relates to selection strength in that selection in larger populations results in less sampling error because of the greater number of sampling events when choosing parents. Small populations experience drift more readily and so experience less selection strength, while large populations are very sensitive to small fluctuations in fitness and experience strong selection.

Adjusting the population size in an evolutionary model, therefore, is a rather simple way to adjust the selection strength. While this convention is ubiquitous in evolutionary modeling, we believe it is somewhat flawed. Our concern with this approach is that altering population size changes more than selection strength. Increasing the population size increases mutation inflow and the standing variation of genotypes in the population. The increase in standing variation is a natural consequence of the diffusion of genotypes due to mutational inflow (FELLER, 1951; Kimura, 1964).

Noisy Evaluation Some recent work has shown that noisy fitness, for example due to environmental noise, can reduce

the strength of selection. For example, Wang and Zhang (2011) and Melbinger and Vergassola (2015) both see an increase in drift-like behavior when fitness is noisy and the selection algorithm has difficulty comparing organisms accurately. Both publications describe the change in behavior as being similar to a reduction in the effective population size.

A third study by Van Egeren et al. (2018) demonstrates how noisy fitness can also benefit adaptation. By creating new opportunities for deleterious mutations to hitchhike, the stochastic increases to fitness actually help populations cross fitness valleys that would have otherwise been impossible to cross. They also see an increased capability for valley crossing when the noise is increased. It is worth noting that these three studies do not conflict; The drift-like behavior, which disrupts the effectiveness of selection, is precisely how the deleterious hitchhikers survive to cross valleys.

Tournament Size Tournament selection is a selection algorithm commonly used in computational models of evolution. In regards to our work it has several benefits, but the most relevant is that it has an adjustable strength of selection. The strength of selection is adjusted by the ‘tournament size’ parameter which controls how many organisms in the population compete with each other for each opportunity to reproduce. The size parameter can be set between drift and elitism, which makes it an ideal dial of selection strength.

Summary of our work

In this work we perform a detailed assay on the relationship between population size, noisy phenotype evaluation, and tournament size, and their effects on rates of genomic change. We run our model on nearly 4,500 different scenarios.

We observe evolution on a smooth fitness landscape as well as nine deceptive landscapes using our model. We show that for the smooth landscape it is always best to have strong selection with noise-free fitness and a large population. For deceptive landscapes, there is an optimum configuration of tournament size and noise that balances exploration and exploitation. Population size, on the other hand, always increases genomic change when larger, because it not only increases selection strength but also maximizes mutational inflow and standing variation. We see that while these parameters for selection strength have similar effects, they each behave in unique ways. Finally, we suggest that evaluation noise is a better proxy for selection strength than the other two methods.

Methods

For this work we implemented our evolutionary simulation using the MABE software (Bohm et al., 2017). In order to understand the relationship between noise, tournament size

and, population size, with rates of adaptation, we ran an array of nearly 4,500 scenarios. Each scenario combines a tournament size with a score noise and a population size. Each combination of parameters is run on 10 fitness functions, one non-deceptive landscape and nine deceptive landscapes. All nine deceptive fitness functions contain fitness valleys which must be crossed in order to reach higher score; The functions differ in the depth and width of their fitness valleys. We used tournament selection for all scenarios.

Organism Definition

Our evolution model is an agent-based simulation where each generation digital organisms are evaluated and selected for reproduction. Unlike more complicated digital organisms (Ofria and Wilke, 2004) we have simplified the concept considerably in order to remove as many confounding interactions from our model as possible.

In this work the organism has a genome that is represented by a vector of 500 integer values, $g = [g_1, g_2, \dots, g_{500}]$. This genome is used to give each organism a score which is used during selection and is inherited during reproduction. Each site in the genome may mutate during reproduction with probability $\mu = 0.0005$ per site. When a mutation does occur at some site g_i in the genome vector, the value is mutated following the rule $g'_i = g_i \pm 1$ where the offset $+1$ or -1 is decided randomly with equal probability.

Fitness Functions

An organism’s score is determined based on the organism’s genome vector g . First a sum of all sites in the genome, x , is computed:

$$x = \sum_{i=1}^{500} g_i \quad (1)$$

Note that since all g_i are integers, the sum x is also an integer.

The sum x is then passed through a function which maps specific values of x to a score.

We use ten fitness functions in this work. The first is a simple hill-climbing function $w(x) = x$. This function allows us to observe the dynamics of the system with no deception (i.e. no fitness valleys).

The other nine functions all include deceptive fitness valleys. There is a default fitness function, and eight alternate functions which modify the topology of the default function in order to explore how valley depth and width change the outcomes of the simulation.

All of the deceptive fitness functions are periodic, meaning after crossing any fitness valley there will be another exactly like it, indefinitely so. However, the score at each of the local optima is 5 higher than the score of the previous (increasing with x) which incentivizes the population to

continuously climb to ever higher peaks, similar to the simple hill-climbing function.

Due to the repeating nature of these functions, they can be defined by just one valley. Each function begins with a plateau one mutation across creating two positions of equal fitness. Following that, as x increases the score drops until finally jumping up to a height greater than the starting position. The difference in score between the starting position and the lowest point in the valley we call “valley depth”. The length, in x , of the valley, including the plateau, we call “valley width”. Valley depth and width are varied between the nine deceptive functions. The default function has a width and depth of 6. Shallow versions of the function have depth 3 while deep versions have depth 9. Similarly, narrow versions of the function have width 3 while wide versions have width 12. Figure 1 provides a visual representation of each of the nine deceptive fitness functions.

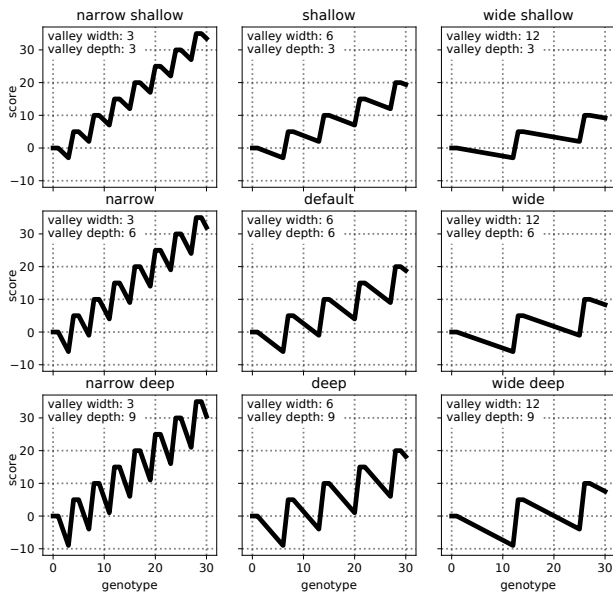


Figure 1: A visual representation of the 9 deceptive fitness landscapes we use throughout this work. The functions are arranged in a 3x3 grid such that the depth of the fitness valley increases from top to bottom and the length (in mutational distance) of each valley increases from left to right. The names and dimensions of each fitness function are displayed above each plot.

Real-Valued Tournament Selection

In this work, we use tournament selection. The tournament selection algorithm is typically defined as shown in Algorithm 1 where T , the integer variable that determines the tournament size ($T \in \mathbf{Z}$), is restricted such that $T \geq 1$. Here, we wish to consider very weak selection strength. In order to achieve this we extended the tournament algorithm

to include tournament sizes that are between 2 and 1. Therefore, we define Algorithm 2 as a means of defining and interpreting real-valued tournament sizes.

In Algorithm 2, T is still restricted such that $T \geq 1$, but is now a real value ($T \in \mathbf{R}$). Algorithm 2 amends Algorithm 1 such that a tournament of size $\lfloor T \rfloor$ is always conducted but an extra challenger is drawn with probability $T - \lfloor T \rfloor$ resulting in an average tournament size of T . Note that every tournament conducted by Algorithm 2 is composed entirely of integer-sized tournaments.

The idea of fractional tournament size is, as far as we are aware, a new concept. It may be challenging to conceive of the meaning of a tournament size of 1.1. The way to think of $T = 1.1$ is to see it as tournament 1 (random selection) 90% of the time and tournament 2 10% of the time. So, given a population of size 100, this would mean that 90% of the next generation would be the result of random selection and only 10% would be the result of tournament 2.

Algorithm 1: Standard Tournament Selection

Result: A digital organism destined to reproduce
count = 1;
best = chooseRandomlyFrom(population);
while count < T **do**
 challenger = chooseRandomlyFrom(population);
 if challenger.fitness > best.fitness **then**
 best = challenger;
 end
end
return best;

Algorithm 2: Real-Valued Tournament Selection

Result: A digital organism destined to reproduce
count = 1;
best = chooseRandomlyFrom(population);
whole = floor(T);
remainder = T - whole;
if randomFloat(0,1) \leq remainder **then**
 whole = whole + 1;
end
while count < whole **do**
 challenger = chooseRandomlyFrom(population);
 if challenger.fitness > best.fitness **then**
 best = challenger;
 end
end
return best;

Applying Noise to Score

In some experiments we investigate the effects of noisy evaluation on rates of adaptation. Whenever we add noise to the

score of an organism as part of our experiments, it is applied as the last step before selection. N is the variable that controls the level of noise. Once an organism's score has been determined by the relevant fitness function, a random value in the range $[-N, N]$ is chosen with uniform probability and then added to the organism's score. This creates a level of uncertainty; Organisms whose scores are closer together than $2N$ may sometimes have their competitive advantages reversed by the noise applied to their score.

Model Justification

The genetic model, selection algorithm, and fitness functions we use in this work were chosen after some consideration. We are primarily interested in the dynamics surrounding valley-crossing events and how those dynamics are affected by the parameters that we vary.

Our genomic model removes the influence of many genetic phenomena such as sexual recombination or crossover, allele dominance, gene epistasis, horizontal gene transfer, and gene translation.

Our selection algorithm only considers the relative rank of scores in the population and therefore maintains a constant strength of selection (unlike fitness-proportional selection algorithms, which experience weakened selection as the average score increases).

Our fitness functions are designed to have an unlimited number of identical valley-crossing opportunities regardless of the population's position on the function. This allows us to characterise how well a particular configuration of the model crosses valleys. By looking at the final score of each run, we can determine how many valleys were crossed during the runtime and compare this with other configurations to easily compare rates of adaptation under different conditions.

System Configuration

Every experiment condition was replicated 100 times. Each replicate was run for 20,000 generations. Every generation, the parent organisms are fully replaced with the offspring they generated. Population size, tournament size, and score noise are varied between experiments and are noted in the results section.

Readers wishing to replicate the results from this paper are directed to the supplemental materials which include files and instructions for generating the data presented in this paper (see: <http://github.com/cliff-bohm/ALIFE-2021-Connections>).

Results

Smooth Landscapes

Figure 2 shows the final scores of runs on the smooth landscape. The smooth landscape is a simple hill climb function in which each increase in x awards the same increase in score. The hill-climbing nature of the fitness function

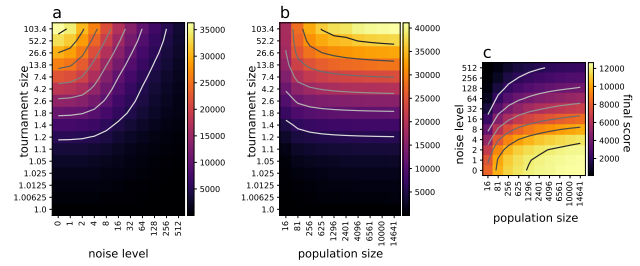


Figure 2: Heatmap showing the final scores of all runs on the smooth fitness function. Contour lines indicate lines of equal score. Every square represents the average score of 100 replicates. a) pairs tournament sizes with noise levels while fixing population size at 625. b) pairs tournament size with population sizes while fixing noise at 0. c) pairs noise level with population size while fixing tournament size at 2.

removes the explore-vs-exploit trade-off leaving pure exploitation as the optimal strategy. This is reflected in the data; Strong selection, large populations, and no score noise result in the highest final scores.

Deceptive Landscapes

Next we consider deceptive fitness landscapes. We use nine fitness functions that contain fitness valleys of different widths and depths. The functions are each labeled as deep/shallow or narrow/wide versions of a default valley shape. Unlike the simple hill-climbing function, these deceptive landscapes require a balance between exploration and exploitation.

Noise vs Population size Figure 3 shows the final outcomes of runs with varying population sizes and score noise amounts with tournament size fixed at 2. If altering the population size was fundamentally the same as changing the score noise we would expect to see a one-to-one mapping that relates pairs of parameters to the same rate of adaptation. Instead, we see that there is no one-to-one mapping between the two parameters.

Further, we see that the narrow fitness valleys are similar to the simple hill-climbing landscape. This is most likely due to the narrow valley being so short that tunneling across via mutation is possible. Under these circumstances low noise and large populations are advantageous. The large populations ensure that rare tunneling events happen more frequently, while the noise-free score help ensure that selection detects those individuals.

In the default and wide valley conditions tunneling is less likely, so the effects of purifying selection become an obstacle. Under these circumstances we see that score noise creates uncertainty for selection and deleterious mutations can persist long enough to assist the population in crossing the valley. When noise is too low, the population is stuck

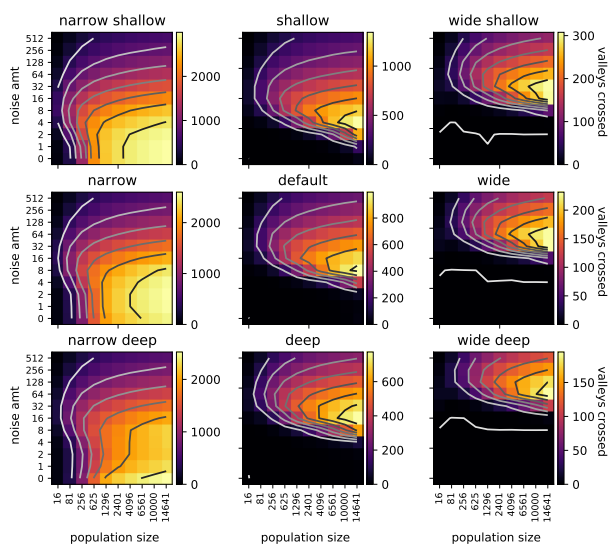


Figure 3: This figure shows the final outcomes of simulations that combine Population Size and Score noise parameters with tournament size $T = 2$. Every cell is the average of 100 replicates. The color intensity indicates the number of valleys crossed during the total runtime of 20,000 generations (final score divided by 5). The nine subplots each show the results of a different fitness function.

and we see large regions of low adaptation in Figure 3. We also see that when noise is too high adaptation begins to suffer due to the decreased ability to detect differences relative to the noise-free scores; As the noise begins to dominate, the system devolves into drift-like behavior. As in the narrow conditions, large populations result in fast adaptation, presumably due to increased standing variation (genetic diffusion) and increased mutational inflow. Apparently, these effects outweigh increases to selection strength because we see no point where population size becomes too big.

Tournament size vs Population size Figure 4 shows the final outcomes of runs with varying population sizes and tournament sizes with noise fixed at 0. As tournament size relates directly to selection strength, if altering the population size was fundamentally the same as changing the selection strength we would expect to see a one-to-one mapping that relates pairs of parameters to the same rate of adaptation. Instead, we see that there is no one-to-one mapping between the two parameters.

Further, we see that the narrow fitness valleys are similar to the simple hill-climbing landscape. This is most likely due to the narrow valley being so short that tunneling across via mutation is possible. Under these circumstances high selection strength and large populations are advantageous. The large populations ensure that rare tunneling events happen more frequently, while the strong selection ensures those

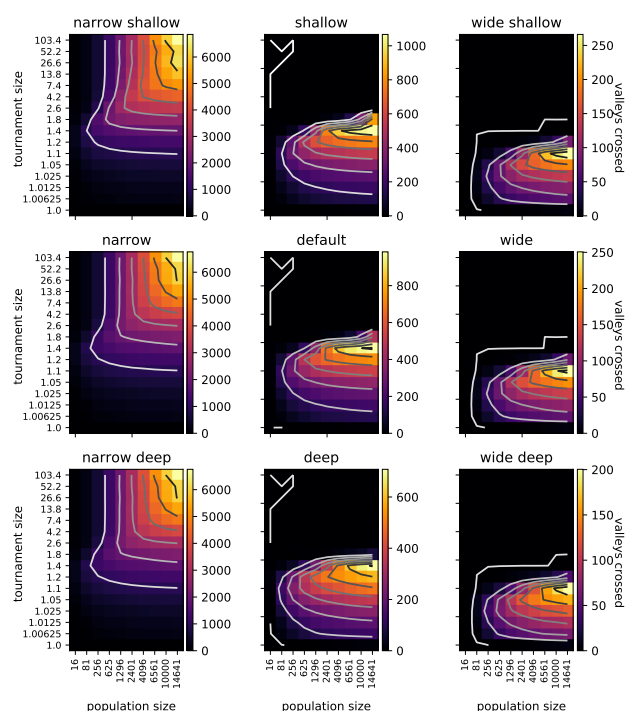


Figure 4: This figure shows the final outcomes of simulations that combine Tournament Size and population size parameters with noise $N = 0$. Every cell is the average of 100 replicates. The color intensity indicates the number of valleys crossed during the total runtime of 20,000 generations (final score divided by 5). The nine subplots each show the results of a different fitness function.

rare events go to fixation.

Once again, in the default and wide valley conditions tunneling is less likely, so the effects of purifying selection become an obstacle. Under these circumstances we see that only when tournament size is low can deleterious mutations persist long enough to assist the population in crossing the valley. When tournament size is too high, the population is purifying deleterious mutations too quickly to cross valleys and we see large regions of low adaptation in Figure 4. We also see that when tournament size is too low adaptation begins to suffer due to the decreased ability to discern differences in score. As tournament size approaches 1 the system tends towards drift-like behavior. Large populations continue to be advantageous here for the same reasons indicated in the Tournament size vs Population size results.

Tournament size vs Noise Figure 5 shows the final outcomes of runs with varying tournament sizes and score noise amounts with the population size fixed at 625. If altering the tournament size was fundamentally the same as changing the score noise amount we would expect to see a one-to-one mapping that relates pairs of parameters to the same rate of

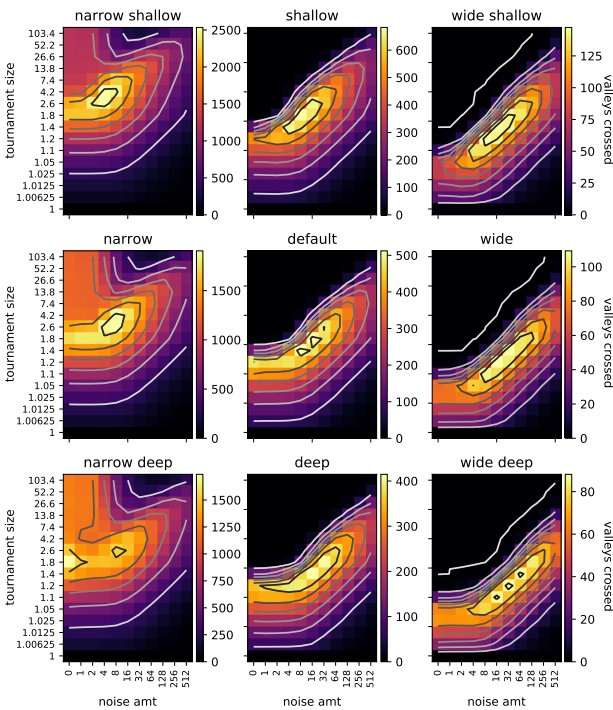


Figure 5: This figure shows the final outcomes of simulations that combine Tournament Size and Score noise parameters with Population size $P = 625$. Every cell is the average of 100 replicates. The color intensity indicates the number of valleys crossed during the total runtime of 20,000 generations (final score divided by 5). The nine subplots each show the results of a different fitness function.

adaptation. Unlike Figures 3 and 4, we see a very strong relationship between the two, though it is not one-to-one in all cases. There are, however, large areas within each plot where the relationship is approximately linear.

Here again, we see that the narrow fitness valleys are similar to the simple hill-climbing landscape. This, again, is most likely due to the narrow valley being so short that tunneling across via mutation is possible. Under these circumstances high selection strength and low noise are advantageous. The strong selection ensures rare high-fitness mutants go to fixation, while the noise-free score help ensure that selection detects those individuals.

As seen before, in the default and wide valley conditions tunneling is less likely, so the effects of purifying selection become an obstacle. Under these circumstances we see that all but the very low and very high tournament sizes are viable. The very low settings result in drift-like behavior, while the very high settings result in purifying selection that restricts valley crossing. Interestingly, the score noise seems to be capable of compensating for the increase in selection strength. As the noise increases, the optimal

tournament size increases too suggesting that although the score noise adds uncertainty, this can be overcome simply with more selection strength. However, this compensatory relationship begins to fail at larger noise strengths because once the score signal is dominated by the noise no amount of selection strength can compensate. As the score noise increases, the system will eventually tend towards drift-like behavior.

Discussion

In this work we show the relationships between all combinations of population size, noisy evaluation, and tournament size in the context of rate of adaptation.

One key observation is that there are relationships between each of the three parameters and selection strength. However, while the behavior of these parameters are each related to selection strength, the relationship is imperfect and there is no perfect one-to-one relationships between any of the parameters we have considered. This means that each of these parameters modify selection strength in different ways.

We began our discussion of deceptive landscapes by introducing the concept of fitness valley width and depth. In our simulations we examined valley depths between 3 and 9 and valley widths between 3 and 12. However, you could imagine setting the width or depth to zero. Doing so removes all deception from the fitness landscape. In other words, setting the width or depth to zero defines the smooth fitness landscape. With this in mind, we can see that all of the results sit on a continuum of deception ranging from none to high.

In the data collected from the smooth function we see that for each pair of parameters the maximum rate of genomic change occurs where the parameters are tuned to maximum selection strength. The rates of change decrease when any parameter moves away from strong selection strength. The contour lines in Figure 2 show a trade-off between the values on each axis that indicates where strength of selection is roughly equal. As the functions become more deceptive, we see that the position of greatest change becomes uncoupled from the highest selection strength settings.

We see an interesting alignment between the areas of greatest change in Figure 5 and the contour lines in the smooth function plot (Figure 2.a). As deception increases the area of greatest change shifts to align with contour lines that are farther from the point of highest selection strength. This illustrates the trade-off between exploration and exploitation. As the functions become more deceptive, weaker selection is needed to allow more exploration across wider and deeper valleys. In addition, the shape of the regions in Figure 5 indicates an almost-compensatory trade-off between tournament size and noise amount which also aligns with the contour lines in the Figure 2.a.

The trade-off we identified in Figure 5 is not seen in Figures 3 and 4. If population size only effected selection strength we would expect to see that very large populations

would result in too much purifying selection to allow for exploration across the more deceptive valleys, but this is not the case. In these plots we still see that a large tournament size or low noise configurations result in reductions in the rate of adaptation, but larger populations always result in a faster rate of change.

Population size and selection strength

Increases to population size are known to also increase the selection pressure by improving the sampling of individuals in the population during selection. However, in the data presented we do not see an optimal population size; We observe that rate of adaptation increases with population size without seeing a reversal of this trend at high values. This seems to show that the trade-off between exploration and exploitation is not invoked when increasing the population size. Therefore, we conclude that there are other forces at work. Namely, we believe that the increase to mutational inflow and standing genetic variation that result from increasing the population size are responsible for the absence of this trade-off. Both of these forces increase the ability of the population to explore the fitness landscape. We believe these increases must be compensating for the increased selection strength that on its own would have caused exploration to be reduced due to purifying selection.

Proxies for selection strength

Based on our results, if someone wishes to design an experiment using an evolutionary model and wishes to vary selection strength in their model as an independent variable, they ought not to use population size as their proxy for selection strength. In the case where it is possible, it should be preferred to use a model of selection which comes with its own explicit parameter for selection strength, like tournament selection.

We find that, similar to tournament size, there is an optimum setting for noise when all other parameters are held constant. Lower than this level, the effects of selection may become too strong for the population to explore the fitness landscape effectively. Higher noise results in an unreliable signal for the selection algorithm to effectively exploit higher fitness individuals. In other words, adding noise to score creates the same trade-off as we see when we adjust the tournament size.

The relationship between tournament size and noise seems to be roughly one-to-one. This similarity suggests that evaluation noise is a rather good proxy for tournament size, which is itself a direct controller of selection strength. Ergo, noise is a rather good proxy for directly adjusting selection strength.

One benefit of using noise as an independent proxy for selection strength is that it is algorithm independent. Regardless of the selection algorithm in use, we can always apply noise to the scores it will operate on to reduce selection

strength.

Neutral Theory in the Context of Noisy Score

The way we have applied noise to our digital organisms' fitness is somewhat abstract. The noise could be interpreted as gene expression variance, or a stochastic environment where luck plays a role in survival and reproduction. In nature we might expect both of these sources of noise to co-occur with, possibly, even more. Therefore, natural organisms may experience a wide range of rates of adaptation even with all other factors usually attributed to controlling selection strength held constant.

This has large implications on the neutral theory of molecular evolution (Kimura, 1987). In brief, neutral theory posits that the majority of genetic variation is neutral variation. Since noisy fitness can change the effective strength of selection, noise can change what qualifies as neutral or not. What was once neutral may become detectable if the noise on fitness is reduced, while detectable traits may become neutral in periods of additional noise. This has far reaching consequences once you consider theories of local optima escape which require the discovery of neutral traits, for example extra dimensional bypass (Cariani, 2002). In the presence of noise, a pre-existing trait may become the neutral ridge that is used to bypass a fitness valley.

The effectiveness of small Tournaments

It came as a surprise that a large portion of the fast-evolving runs were using tournament sizes below $T = 2$. Seeing as how $T = 1.4$ means that more than half (60%) of the population is reproducing under drift-like conditions, it is fascinating that the population does not experience mutational meltdown. Perhaps $T = 1.4$ is like reducing the effective population size to 40% but also including a pool of additional organisms that might be shuffled in or out between generations. Interpreted this way is not at all surprising that $T = 1.4$ evolves. To the contrary, it would seem to have the benefit of maintaining diversity more effectively than $T \geq 2$.

The excitement over this should be tempered somewhat by acknowledging that the rates of change seen on the smooth function (Figure 2) for tournament sizes at or below 1.4 are extremely low. So, while these low tournament sizes help to maximize the rate of evolution by balancing exploration and exploitation, they cannot allow evolution to proceed faster than would be the case if the landscape was smooth and non-deceptive.

Even so, real-world fitness landscapes are high dimensional and likely highly deceptive. As a consequence, its interesting to see that very small tournament sizes can have observable long-term outcomes on evolution for deceptive landscapes. Systems experiencing near-drift behavior are still able to undergo directional selection when the time period of observation are long enough.

Future Work

One interesting consequence of the results presented here are the implications for multi-trait evolutionary systems. For example, consider a digital organism whose total fitness is given by summing together progress on both the shallow and narrow fitness functions. If the system is evolving under tournament selection without noise the optimum for the narrow function is $T = 1.8$ while for the shallow function it is $T = 1.2$. Further, if the shallow function is evolved at $T = 1.8$ it is virtually stuck. Therefore, the choice of tournament size could cause an organism on the multi-trait landscape to evolve on both functions or on just one. However, say we choose poorly ($T \geq 1.8$ which will halt progress on shallow), adding noise to the system, say $N = 8$, could move both traits into a fast-evolving scenario.

Our work has focused on results using the tournament selection algorithm. As a rank based algorithm, tournament selection is going to behave differently from more traditional fitness-proportional selection algorithms (such as roulette). Fitness proportional selection algorithms are known to have variable selection strength as a result of increasing organism scores. In preliminary work we attempted to apply a correction to roulette selection that would fix the diminishing selection strength problem, but ultimately we chose to go with tournament selection because no such correction was necessary. Future investigations should use a fitness proportional selection method and compare the results with tournament.

In our analysis of score noise, we find that there is often a single optimal value for this parameter, given the other parameters are already chosen. There is a potentially untapped engineering technique in the use of score noise to enhance the rates of adaptation of evolutionary optimization algorithms. The single optimum makes tuning this parameter rather simple; Simple enough to automate during an evolutionary search.

Acknowledgements

This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research. This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454.

References

- Bohm, C., G., N. C., and Hintze, A. (2017). Mabe (modular agent based evolver): A framework for digital evolution research. In *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*, page 76–83. MIT Press.
- Cariani, P. A. (2002). Extradimensional bypass. *Biosystems*, 64(1–3):47–53.
- Covert, A. W., Lenski, R. E., Wilke, C. O., and Ofria, C. (2013). Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences*, 110(34):E3171–E3178.

- FELLER, W. (1951). Diffusion processes in genetics. *Proc. Second Berkeley Symp. on Math. Stat. and Prob.*, page 227–246.
- Iwasa, Y., Michor, F., and Nowak, M. A. (2004). Stochastic tunnels in evolutionary dynamics. *Genetics*, 166(3):1571–1579.
- Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability*, 1(2):177–232.
- Kimura, M. (1987). Molecular evolutionary clock and the neutral theory. *Journal of Molecular Evolution*, 26(1):24–33.
- Melbinger, A. and Vergassola, M. (2015). The impact of environmental fluctuations on evolutionary fitness functions. *Scientific Reports*, 5(11):15211.
- Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life*, 10(2):191–229.
- Oliveto, P. S., Paixão, T., Pérez Heredia, J., Sudholt, D., and Trubenová, B. (2018). How to escape local optima in black box optimisation: When non-elitism outperforms elitism. *Algorithmica*, 80(5):1604–1633.
- Van Egeren, D., Madsen, T., and Michor, F. (2018). Fitness variation in isogenic populations leads to a novel evolutionary mechanism for crossing fitness valleys. *Communications Biology*, 1(11):1–9.
- Wang, Z. and Zhang, J. (2011). Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proceedings of the National Academy of Sciences*, 108(16):E67–E76.
- WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress of Genetics*, 1:356–366.