# Synthetic Speciation Through Meiotic Inheritance

William Booker - Version 1.01

November 2016

## 1 Abstract

The field of artificial intelligence owes much to our understanding of biology. Many of the most promising machine learning techniques, such as neural nets and genetic algorithms, were inspired by biological precedent. Recently, this trend continues through research into a hitherto under-appreciated biological phenomenon: speciation.

Speciation, the process by which a single group of organisms diverges into a number of independent groups, is fundamental to evolution in the natural world. Understanding its finer points is critical to understanding life on this planet and predicting how it may continue to change and develop over time. Additionally, speciation shows promise in the realm of artificial intelligence, as a mechanism by which to narrow in on optimal solutions to multimodal problems.

Using computer simulations, we can now model large biological populations over the course of centuries, giving us unprecedented insight into this elusive process.However, no simulation is perfect; any real-world phenomenon must be simplified before it can be modeled, and these simplifications have the potential to reduce the accuracy of the simulation. This paper seeks to identify and address one such oversimplification in the paper Sexual Selection, Resource Distribution, and Population Size in Synthetic Sympatric Speciation by Mark Woehrer and to demonstrate how the use of non-destructive inheritance is critical to modeling biological populations. Furthermore, this paper seeks to address how biologically inspired inheritance methods could prove useful for inducing speciation and maintaining genetic diversity in genetic algorithms.

## 2 Background

In evolutionary biology, there are two distinct types of speciation: allopatric speciation, where isolated populations acquire new traits independently, and sympatric speciation, where a single population diverges over time. While both appear in the natural world, this research pertains primarily to sympatric speciation, as it is less understood, and has many potential uses in the realm of machine learning. Specifically, sympatric speciation shows promise in improving the quality of genetic algorithms by allowing it to converge around several

local optima in the solution space before deciding on the best one.While there are mechanisms for inducing a sort of allopatric speciation, such as niching [4], population topologies [6], and island models [8], these approaches require implicit assumptions about the solution space which must be explicitly tailored to each problem. Ideally, we would like the algorithm to induce speciation within populations independently, without the need for human input. To solve this problem, we again turn to the natural world for inspiration, by mimicking how sympatric speciation occurs in biological populations. Recent work in this area has focused around the possibility that sexual selection may lie at the root of sympatric speciation.

In Woehrer's Speciation paper, the researchers hypothesized that sexual selection may be key to inducing sympatric speciation in a population. The idea is that sexual selection may reproductively isolate sub-groups within the population, which could then evolve different traits in a similar manner to allotropic speciation. To test this hypothesis, the researchers set up a 2x2 factorial study using a simulated population of finches. In the study, the researches created two different distributions of seeds, bimodal and uniform, and observed the differences in population dynamics between finches that utilized sexual selection and those that did not.

While the researches successfully demonstrated that sexual selection was an effective method to induce sympatric speciation in a population, some of the results in their control study seemed to contradict our understanding of biological speciation. Specifically, when a population without sexual selection encountered a bimodal seed distribution, the population never expanded to cover both possible niches. This result goes against our understanding of biology, as a species is highly incentivized to take advantage of all available niches in its environment. The convergence exhibited in the bimodal seed random mating scenario is known as bottlenecking [5], and typically only occurs in small, declining populations [5]. To see bottlenecking occurring in a large, healthy population suggests that there may be a lurking oversimplification in the simulation.

## 3 Hypothesis

The cause for the population convergence in the bimodal seed random mating simulation was due to the oversimplification of genetic inheritance. Averaging the beak sizes of the parents inhibits a population's ability to maintain genetic diversity, which leads to population collapse in multi-niche environments.

## 4 Methods

This research consisted of five experiments: three primary experiments and two supplemental experiments. The three primary experiments are designed to test our hypothesis that the introduction of meiotic inheritance radically changes the behavior of the system. The two supplemental experiments are designed to

help us better understand why those changes arise.

## 4.1   Experiment 1 – Replicating Results

The initial simulation is designed to replicate the simulation from the Speciation paper. The details of this simulation are summarized in the following paragraphs. For a more in-depth analysis on the reasoning behind the choices made, please consult the original paper.

The simulation consists of an island 100x100 units in size, filled with seeds and inhabited by a population of finches. Each finch has a known beak size, age, gender, energy-level, and location. Each seed has a known size, energy-level and location.

The simulation begins by generating 400 finches with a roughly equal number of males and females. This initial population has a mean beak size of 5.5 and a standard deviation of 0.5 (Gaussian distribution). These finches have their energy level set to zero and are randomly spread around the island. The island then undergoes an annual cycle consisting of a dry season and a mating season. This cycle repeats 1000 times during the simulation.

The dry season lasts 100 days, and begins by spreading 5000 seeds randomly around the island. Each day, each bird has a chance to forage for food. The foraging is performed in a random order. Each bird searches a 10x10 plot of the island and consumes the first seed it finds that falls within one unit of its beak size. This search costs the finch 0.1 units of energy. Any finch that has less than zero units of energy at the end of the day is declared dead and removed from the population.

The simulation accounts for two different distributions of seed sizes: bimodal and uniform. The bimodal distribution consists of two Gaussian distributions of 2500 seeds each; these distributions are centered at three and eight units, each with a standard deviation of 0.5 units. The uniform distribution consists of a single uniform distribution of 5000 seeds ranging from one to ten units in size. In both scenarios, a seed contains between zero and two units of energy (uniform distribution).

After the dry season ends, the remaining finches participate in that year's mating season. During the mating season, each female chooses a single male to mate with. The simulation accounts for two types of mating: random mating, where females will choose any male from the population, and assorted mating, where females will only pick males within one unit of its beak size. Each male is only allowed to mate with five females per year. When presented with multiple possible mates, the female will choose between them at random. Mating produces a single offspring at the location of the mother. The child's beak size is an average of the beak sizes of the parents plus a small Gaussian-random mutation (Mean 0, Standard deviation 0.2).

## 4.2  Experiment 2 – Introduction of Genetic Structure

The objective of this experiment is to replace the simplistic inheritance structure with a more robust genetic model while still maintaining the averaging process. This step serves as a control study to demonstrate that the changes in the third experiment are a result of the introduction of meiotic inheritance rather than the introduction of the new genetic structure. This genetic structure is designed to mimic genes in the natural world as closely as possible.

Each finch is equipped with two sets of genes, one from each parent. Each gene contains an array of one-hundred sub-genes, each represented by a floating point value. The phenotype for a gene is determined by taking the sum of all of the sub-genes. In order to mimic the Speciation paper's initial population, the sub-genes in the initial population are randomly chosen from a uniform distribution with a mean of 0.055 and a range of 0.245 [2]. A single gene was used to determine the beak size of the individual. This choice reflects how a single gene, Bmp4, is responsible for much of the variation in beak sizes in real Darwinian finches [1].

Our experiments use both a complete dominance and an incomplete dominance model. In the complete dominance model, the larger of the two phenotypes is chosen for the finch's beak size. In the incomplete dominance model, the two genes are averaged to determine the beak size. Note that for the purposes of inheritance this choice is arbitrary, as neither choice affects the actual genetic material that is passed on to the child.

In order to maintain the averaging process, when mating, the sub-genes of each parent are averaged to determine the sub-genes for the child. To account for the mutation, each sub-gene has a 10% chance to randomly change after the averaging process. When a sub-gene is selected for mutation, we add a random value taken from a uniform distribution with a range of 0.310 and a mean of 0.

## 4.3  Experiment 3 – Introduction of Meiotic Inheritance

The setup for the third experiment is nearly identical to that of the second: the only change is the addition of meiotic inheritance. During mating, rather than average the sub-genes of the parent to create a new gene, the child receives a randomly selected gene from each of its parents. The mutation rate and range remain the same as in the previous step.

## 4.4  Experiments 4-5 Notes

In the fourth and fifth experiments, we made two major changes when compared to the initial three experiments. First, rather than testing all population conditions, we only looked at the meiotic population conditions and one non-meiotic population condition in the BSRM scenario. Second, we reduced the initial population size from 400 down to 40 to help smooth out the initial population collapse and subsequent rebound that appeared in our first three experiments.

## 4.5   Experiment 4 – Removing Mutation

The fourth experiment is designed to illustrate the convergent tendency of the averaging inheritance method by removing the noise caused by mutation. We achieved this effect by reducing the range of the running mutation rate from 0.310 to 0.00. We also quadrupled the initial mutation rate to 0.980 in order to give the starting population sufficient diversity to reach both niches.

## 4.6   Experiment 5 – Modifying The Initial Population

The fifth experiment is a response to an unexpected result in our third experiment, where complete dominance exhibited far higher speciation rates than incomplete dominance. In this experiment, we changed the size of the initial population to see the effects on the relative extinction rates between incomplete dominance and complete dominance. For the first test, we widened the initial population by quadrupling the mutation rate to 0.980. In the second test, we reduced the initial population by halving the mutation rate to 0.1225, and moving the mean to 0.80. This change centers the population in the middle of the upper niche.

# 5   Results

## 5.1   Experiments 1-2 – Replication and Gene Structure

## 5.2   Experiment 3 – Introduction of Meiotic Inheritance

## 5.3   Experiment 4 – Removing Mutation

## 5.4   Experiment 5 – Modifying The Initial Population

# 6   Analysis

In our first experiment, we were largely successful in replicating the results from the Speciation paper. In the BSAM, BSRM, and USAM cases, the population distributions, the withs of the population bands, and the number of extinction events lined up extremely well with what was described in the paper. In the USRM scenario, the population distribution and the withs of the bands were correct, but there were significantly more extinction events than expected. We suspect that this anomaly might be due to a minor difference between our implementation and that of the initial experiment that was not covered in the original paper. Furthermore, it's worth noting that we experienced another anomaly in the USRM scenario in the second experiment, suggesting that the extinction rates in this scenario might be unusually sensitive to minor changes in experimental setup. However, barring this anomaly, we are confident that we have successfully replicated the Speciation paper's results.

In the second experiment, we introduced the new genetic structure but kept the averaging method the same, then compared the results against our findings from the first experiment. As expected, the population trees and extinction rates closely matched our original results. Again, the only major variation was in the extinction rates for the complete dominance structure in the USRM scenario. We suspect that this variation may have been the result of the slight skewing of the phenotypes of the initial population, caused by the switch from incomplete dominance to complete dominance. Nevertheless, this experiment demonstrates that the new genetic structure had little impact on the behavior of the populations.

In the third experiment, we used the same genetic structure as the second experiment with the addition of meiotic inheritance. This change radically altered the result of the simulation. Rather than collapsing into narrow bands, each population rapidly expanded to cover the entire range of seed sizes. These extremely thick population bands indicate that the meiotic populations exhibited far more genetic diversity than the non-meiotic populations. Furthermore, unlike the non-meiotic populations, the meiotic populations exhibited evidence of speciation in the BSRM scenario. This observation contradicts the results from the initial Speciation paper, where the researchers demonstrated that sexual selection was requisite for sympatric speciation. The key to this discrepancy lies in the mechanism by which genes are passed on to the next generation.

In the original experiment, we averaged the beak-sizes of the parents to determine the beak-size of the child. While seemingly sensible, this approach destroys important genetic diversity in the population. Barring mutation, any child produced using this method will have a beak-size between the beak-sizes of its parents. When scaled up, this effect causes each generation to cover a smaller range of beak-sizes than the generation that proceeded it. Thus, over time, the species begins to coalesce around a single beak-size: population convergence. As the inheritance method pushes the population inwards, the high mutation rate pushes the population outwards, eventually reaching equilibrium at two-unit thick population bands. This equilibrium is why the population bands are the same with regardless of niche size: the inheritance method is the force pushing the population inwards, not natural selection. This effect also explains why we don't see speciation in the BSRM scenario for non-meiotic populations. Because the gap between the local maxima at three and eight are more than two units apart, there is no stable population that can contain both, forcing one niche to be abandoned.

Contrast this tendency with non-destructive meiotic inheritance, in which genetic information is passed unaltered from parent to child. By keeping the genes the same between generations, meiotic inheritance protects the diversity of the population. There is no guarantee that a child's beak-size will fall between that of its parents, so future populations can reassert genetic diversity that may not appear in the current distribution of phenotypes. By introducing meiotic inheritance in the third experiment, we removed the pressure towards convergence, allowing the population to diverge to cover the entire range of seed sizes. This change is why the population bands from the third experiment are so

much wider than those of the first two: by removing the convergent inheritance method, the only factor restraining the population is the size of its niche. Furthermore, because meiotic populations have no tendency for convergence, and are thus stable at an arbitrary range, they can exploit both population niches in the BSRM scenario.

We observed these effects more clearly in our fourth experiment, where we removed the ability for the population to mutate over time. Without the divergent pressure from mutation, the averaging method immediately coalesced around a single beak-size. As predicted, each generation exhibited less diversity than the generation that proceeded it, eventually forcing the population to abandon one of the niches within a dozen generations. However, with meiotic inheritance, no such trend appeared. While some genes did die out, the handful of genes that survived were stable for the duration of the simulation. Some phenotypes even temporarily disappeared from the population, only to reemerge in later generations. These observations line up well with what we observe in real-world biological populations. Like the meiotic simulations, biological populations tend to exhibit a countable number of discrete phenotypes, and demonstrate high levels of diversity despite extremely low mutation rates [1]. While simulated non-meiotic populations may look normal due to high mutation rates, this experiment demonstrates that they are actually governed by vastly different forces than biological populations, making them unreliable for predicting biological behavior.

However, while we have illustrated why meiotic populations can exhibit speciation in BSRM scenarios, we have not addressed why complete dominance exhibited far higher rates of speciation than incomplete dominance. This effect is caused by an unrelated property of genetic inheritance: gene making. Gene masking is a mechanism by which a population can bridge the evolutionary valley between the two optimums via genetic drift. In the case of complete dominance, only one gene is actually expressed, thus allowing the non-dominant gene to move through the low-fitness valley without reducing the organism's evolutionary fitness. Eventually, these recessive genes can wander into the second niche, and, when matched, they produce an individual capable of exploiting these untapped resources. With incomplete dominance, both genes are expressed, so the population is unable to bridge the gap on its own. By masking the effect of one of the genes, complete dominance acts as a sort of multiplier on the mutation rate, allowing beneficial mutations to develop over the course of several generations rather than within a single individual.

Our fifth experiment was designed to test this analysis. Assuming that the change in speciation rates was due to the introduction of an improved mutation mechanism, we would expect that introducing more mutation in the initial population would reduce its impact, while reducing the initial mutation would increase it. When we ran the experiment, we saw exactly this result. When the initial population was widened, incomplete dominance performed almost identically to complete dominance, with thirty-nine speciation events to forty-three. However, when the initial population was narrowed and confined to a single niche, there were no speciation events for incomplete dominance and twelve for

complete dominance. Note that the averaging method did not produce any speciation events in either condition. These results suggest that both incomplete dominance and complete dominance are equally effective in protecting genetic diversity in a population; complete dominance only performed better in the third experiment because it added a method to bridge the low-fitness valley.

# 7    Conclusion

This research demonstrates the paramount importance of inheritance methods in biological simulations and genetic algorithms. A slight change in a population's inheritance mechanism can radically alter the population dynamics, due to the potential introduction of strong evolutionary pressures. Obviously, for biological simulations, the inheritance process should mimic meiotic inheritance as closely as possible to avoid potential inaccuracy. While the applications for machine learning are less clear, we postulate that meiotic inheritance may have many advantages over associative mating for inducing speciation in genetic algorithms.

First, meiotic speciation has an advantage in that it makes very few assumptions about the nature of the solution space. As demonstrated, destructive inheritance has a tendency to collapse into narrow population bands that are only sustained by high mutation rates. If the mutation rate is too low, these narrow population bands run the risk of getting stuck local maxima before they are able to identify the optimal solution. This observation significantly reduces the generalizability of the sexual selection technique, as the mutation rate will likely have to be hand-crafted to optimize for a given solution space. However, meiotic inheritance avoids this problem entirely by preventing population collapse during runtime. By keeping the population as wide as the niches will allow, meiotic inheritance is better protected against premature optimization in most solution spaces.

Second, meiotic speciation has the potential to radically speed up the mating process, as random mating can be performed in O(n) time while sexual selection may take up to $O(n^2)$ time. Unlike with destructive inheritance, where sexual selection was requisite for speciation, in meiotic inheritance, sexual selection is unnecessary, and in some cases, even counter effective [2]. While it did not have a major impact on our simulation, if an algorithm wanted to have a high chance of finding the optimal solution by seeding a large portion of the problem space, this difference may radically improve performance.

Third, meiotic speciation has the ability to take advantage of gene masking, which we discovered may be a powerful tool for exploring new niches without having to increase the mutation rate. This ability is advantageous, as increasing mutation rates often result in the accidental abandonment of good solutions [7]. Meiotic inheritance is best equipped to utilize this technique, as the amount of genetic drift is not limited by the convergent tendency of the inheritance method.

Fourth, and perhaps most importantly, meiotic inheritance may prove useful

in situations where the population space is evolving over time, potentially even in response to the genetic algorithm. In this case, it's easy to imagine how premature optimization may trap a population in a niche from which it is unable to escape once it becomes unfavorable. Meiotic inheritance would be well equipped to deal with such a scenario, given the high level of diversity and the ease with which it exploits new niches. When coupled with the additional search range offered by gene masking, meiotic inheritance gains a major advantage over sexual selection in dynamic solution topographies.

Of course, there are some disadvantages to this technique. One issue is that our experiments use a dynamic population size, which, while biologically accurate, is uncommon in genetic algorithms. To expand the scope of this research, we would need to demonstrate an effective method by which to apply the benefits of meiotic inheritance onto a static population. Additionally, genetic algorithms that use an indirect encoding scheme often implicitly utilize a form of non-destructive inheritance, so there is little to be gained by altering these inheritance methods. Furthermore, in some applications convergent tendency is desired, as it can help to home in on a solution after the optimal peak has been identified. In fact, we hypothesize that the most effective inheritance method for solving multimodal problems likely begins with a non-destructive inheritance method and switches to a destructive inheritance method after identifying the major peaks. Thus, perhaps the most critical aspect of this research is that it demonstrates the importance of inheritance methods on populations, and illustrates the need for increased scrutiny of a rather under-appreciated aspect of genetic algorithms.

Finally, from a biological standpoint, it still remains to be seen how sexual selection impacts sympatric speciation in the natural world. While we have demonstrated that the *Speciation* paper's results cannot be applied to biological populations, we have not established whether or not sexual selection is a key component of biological sympatric speciation. The main reason that we hesitate to extend our results to biological populations is because the speciation that we observed in meiotic populations in the BSRM scenario is likely very different from the speciation exhibited in biological population. Specifically, for the purposes of this paper, we defined speciation as any set of visibly distinct populations exploiting separate niches. However, while useful for genetic algorithms, this definition is weaker than biological speciation, which typically requires that different species be reproductively isolated [3]. Thus, it could be argued that the distinct bands that arise in meiotic populations in the BSRM scenario do not actually encompass two different species, but rather a single species exploiting multiple niches. Further research is required to determine if sexual selection is required for this split population to diverge into separate species, or if there is another potential cause.

In summary, the findings of this paper are fourfold. One, we successfully replicated the results of the *Speciation* paper and confirmed their findings. Two, we demonstrated that meiotic inheritance radically changes the behavior of the simulation, thus providing evidence against the *Speciation* paper's biological conclusions. Three, we demonstrated that non-destructive inheritance allows for

speciation in multi-niche environments, which may have applications in genetic algorithms. And four, we demonstrated that gene masking gives populations the opportunity to explore the low-fitness valley, even when the mutation rate is smaller than the expanse.

# References

[1] Arhat Abzhanov. *Bmp4 and Morphological Variation of Beaks in Darwin's Finches.* Science, 2004.

[2] Appendix. *!!!APPENDIX!!!* Appendix, 2016.

[3] Niel A. Campbell, Brad Williamson, and Robin J. Heyden. *Biology Exploring Life.* Pearson, 2006.

[4] Jeffrey Horn, Nicholas Nafpliotis, and David E Goldberg. *A Niched Pareto Genetic Algorithm for Multiobjective Optimization.* IEEE, 1994.

[5] Citation Needed. *!!!Citation Needed!!!* Citation Needed, 2016.

[6] Jayshree A Sarma. *An Analysis of Decentralized and Spatially Distributed Genetic Algorithms.* Bombay University, 1998.

[7] William M. Spears. *Simple Subpopulation Schemes.* World Scientific, 1994.

[8] Darrell Whitley, Soraya Rana, and Robert B Heckendorn. *The Island Model Genetic Algorithm: On Separability Population Size and Convergence.* Colorado State University, 1998.