

review articles

DOI:10.1145/2934662

Looking at the mysteries of evolution from a computer science point of view yields some unexpected insights.

BY ADI LIVNAT AND CHRISTOS PAPADIMITRIOU

Sex as an Algorithm The Theory of Evolution Under the Lens of Computation

LOOK AROUND YOU, and you will be stunned by the work of evolution. According to Nobel Laureate Jacques Monod, a strange thing about evolution is that all educated persons think they understand it fairly well, and yet very few—if any, one may grumble—actually do. Understanding evolution is essential: “Nothing in biology makes sense except in the light of evolution,” famously said the eminent 20th century biologist Theodosius Dobzhansky. And evolution is closer to home than black holes and other mysteries of science—it feels almost like your family history.

More so than most scientific fields, the theory of evolution has a sharp beginning: the publication

of Charles Darwin's *The Origin of Species* in 1859.⁸ But of course nothing is that simple: during the first half of the 19th century, several scientists were convinced the diversity of life we see around us must be the result of evolution (see the sidebar on Charles Babbage and the accompanying figure). Darwin's immense contribution lies in three things: His identification of natural selection as the engine of evolution; his articulation of the common descent hypothesis, stating that different species came from common ancestors, and further implying that all life came from a common source; and the unparalleled force of argument with which he empowered his theory. But of course *The Origin* was far from the last word on the subject: Darwin knew nothing about genetics, and had no clue about the role of sex in evolution, among several important gaps. On the ultimate reason for sex, for instance, he wrote, “the whole subject is as yet hidden in darkness.”

Mathematics has informed the theory early on. Mendel discovered genetics by appreciating mathematical patterns in the ratios of sibling pea plants exhibiting different characteristics and model building. When his laws were rediscovered 40 years later, their discrete nature was misunderstood as being at loggerheads with Darwin's

» key insights

- Sexual reproduction is nearly ubiquitous in nature. Yet, despite a century of intense research, its evolutionary role and origin is still a mystery.
- Recent research at the interface of evolution and CS has revealed that evolution under sex possesses a surprising and multifaceted computational nature: It can be seen as a coordination game between genes played according to the powerful Multiplicative Weights Update Algorithm; or as a randomized algorithm for deciding whether genetic variants perform well across all possible genetic combinations; it allows mutation to process and transmit information from transient genetic combinations to future generations; and much more.
- Computational models and considerations are becoming an indispensable tool for unlocking the secrets of evolution.



40X

SEX

continuous perception of traits. A deep scientific crisis raged for two decades, and was eventually resolved with the help of mathematics: discrete alleles (see the glossary) can result, through cumulative contributions, in continuous phenotypic traits. During the 1920s and the next two decades, R.A. Fisher, J.B.S. Haldane, and S. Wright developed mathematical equations for predicting how the genetic composition of a population changes over the generations. This mathematical theory of population genetics is introduced briefly in the sidebar “The Equation that Reconciled Darwin and Mendel.” It is key to what is called the “modern evolutionary synthesis”—the 20th-century view of evolution, because it proposed one way of unifying Darwinism and Mendelism.

During the near-century since then, the study of evolution has flourished into a mature, comprehensive, and prestigious scientific discipline, while over the past two decades it has been inundated by a deluge of molecular data, a vast scientific gold mine that informs—and often challenges—its tenets. And yet, despite the towering accomplishments of modern evolutionary biology, there are several important questions that are beyond our current understanding:

► **What is the role of sex in evolution?**

Reproduction with recombination is almost ubiquitous in life (even bacteria exchange genetic material), while obligate asexual species appear to be rare evolutionary dead ends. And yet there is no agreement among the experts as to what makes sex so advantageous.

► **What exactly is evolution optimizing, if anything?**

With all the evolution-inspired optimization heuristics coded by computer scientists (as we discuss next), this question—also very much contemplated by biologists over the decades—comes to the fore.

► **The paradox of variation.**

Genetic variation in humans, and in many other species, is much higher than the theory predicted due to selection; and assuming the variation is neutral has its problems too.

► **Are mutations completely random?**

We know they are far from uniformly distributed in the genome, but can they be the results of elaborate genetic mechanisms?

Not only is sex essentially universal, but it seems to be very much center stage in life, the basis of a fantastic variety of behavior and structure.

As we recount in this article, recent joint research by computer scientists and biologists, bringing ideas and concepts from computation into biology, has made quite unexpected progress in these questions. Additional background and literature in the online appendix is available in the ACM Digital Library (dl.acm.org) under Source Material.

Evolution and Computer Science

Over the past 70 years, computer scientists, starting with von Neumann,³⁷ have been inspired and intrigued by evolution. During the 1950s, computer scientists working in optimization developed local search heuristics: Start with a random solution and repeat the following step: If there is a “mutation” of this solution that is better than the current one, change to that, until a local optimum is discovered. By “mutation” (much more often the word “neighbor” is used), we mean a solution differing from the present one in a very small number of features; in the traveling salesman problem, for example, a mutation could change two, or three, edges of the tour to form a new tour. This process is repeated many times from random starts, a stratagem that can be seen as a sequential way of maintaining a population (see Papadimitriou and Steiglitz³² for a survey on the 1980s).

This basic idea of local search was enhanced in the 1980s by a thermodynamic metaphor,²⁰ to help the algorithm escape local optima and barriers: *Simulated annealing*, as this variant of local search is called, allows the adoption of even a disadvantageous mutation, albeit with a probability decreasing with the disadvantage, and with time. A further variant called *go with the winners*¹ is closer to evolution in that it keeps a population of solutions, teleporting the individuals that are stuck at local optima to the more promising spots. Notice that all these heuristics are inspired by asexual evolution (no recombination between solutions happens); heuristics of this genre have been used successfully in many realms of practice, and there are several practically important hard problems, such as graph partitioning, for which such heuristics are competitive with the best known.

During the 1970s, a different family of heuristics called genetic algorithms was proposed by John Holland.¹⁴ A population of solutions evolves through mutations and recombination. Recombination is much more difficult to apply to optimization problems, because it presupposes a genetic code, mapping the features of the problem to recombinable genes (to see the difficulty, think of the traveling salesman problem: tours can mutate, but they cannot recombine). The evolutionary fitness of each individual solution in the population (that is, the number of children it will have) is proportional to the quantity being maximized in the underlying problem. After many generations, the population will presumably include some excellent solutions. Holland's idea had instant appeal and immense following, and by now there is a vast bibliography on the subject (see, for example, Mitchell²⁸ and Goldberg¹²). The terms *evolutionary algorithms* and *evolutionary computation* are often used as rough synonyms of "genetic algorithms," but often they describe more general concepts, such as the very interesting algorithmic work—also categorized as research in artificial life—whose purpose is not to find good solutions to a practical optimization problem, but instead to understand evolution in nature by exposing novel, complex evolutionary phenomena in silico, for an example, see Jong.¹⁶

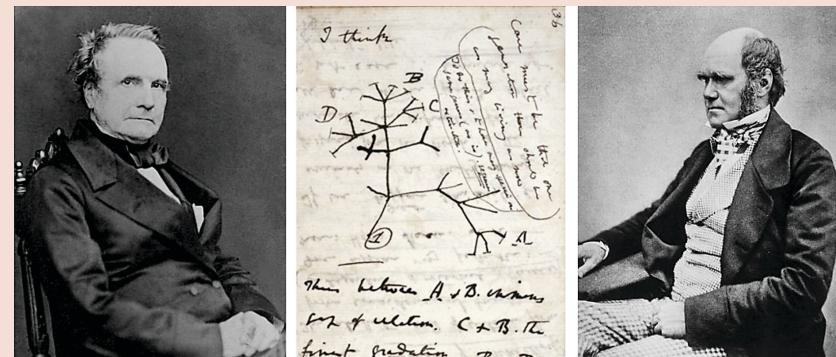
Coming back to the genetic algorithms, several successes in solving actual optimization problems have been reported in the literature, but the general impression seems to prevail that genetic algorithms are far less successful in practice than local search and simulating annealing. If this is true, it is quite remarkable—a great scientific mystery—because in nature the exact opposite happens: Recombination is successful and ubiquitous, while obligate asexual reproduction is extremely rare and struggling.

The authors started collaborating on a computational understanding of evolution in 2006, precisely in order to investigate this mystery, and we recount our findings in this article. At exactly the same year, Leslie Valiant first wrote on his theory of evolvability, another attempt at understanding evolution in computational terms.³⁶ Valiant sees evolution as an approximation of an

Charles Babbage on Evolution

Charles Babbage, the grandfather of computer science, was undoubtedly the first to think of evolution in algorithmic terms. In 1837—22 years before the publication of *The Origin of Species* and the same year that Darwin jotted down a sketch of the tree of life in his notebook (see the figure below)—Babbage wrote the Ninth Bridgewater Treatise.⁴ The treatise was ninth in a series of philosophical essays on the subject of the deity, in which he refers extensively to his Difference Engine and how it can be programmed to go on for days calculating unimaginably large integers, with no further intervention by the programmer. Babbage used this image to argue that a benevolent creator can design not necessarily species, but an algorithm (he did not use the word) for creating species.

Portraits of Charles Babbage (left) and Charles Darwin (right). In the middle is Darwin's sketch showing the principle of common descent leading to a branching pattern.



Glossary

Allele: Variant of a gene. Some are known to affect our susceptibility to diseases.

Diploidy: The state of having two instances of each chromosome, and thus each gene, in a cell. Organisms with one instance of each chromosome and gene are called haploid.

Fitness: The ability of an individual to survive and reproduce in its natural environment, manifested in its expected number of surviving offspring.

Gene: A unit of heredity and a region of the DNA that encodes a functional product. It is thought that humans have more than 20,000 of these. However, now that coding is known to be far more complex than originally thought, it is no longer clear how to define these units and their boundaries.

Genotype: The whole or a part of the genetic make-up of an individual or that is common to a group of individuals.

Heterozygous: An individual having two different alleles of a certain gene.

Mutation: A change in the hereditary material.

Phenotype: The characteristics of an individual other than its genetic code.

Recombination: In sexual reproduction, taken broadly, this term means that the genetic material of an offspring is a bricolage of the genetic material of its parents, due to both the independent assortment of chromosomes during the halving of chromosome numbers and the crossing over of chromosomes—the shuffling of segments between two corresponding chromosomes (in humans, typically 2–3 per chromosome)—that occur in the generation of gametes. More generally, DNA recombination in the sense of exchange of genetic material between two DNA molecules or between segments of the same DNA molecule is an important part of mutational mechanisms.

The Equation that Reconciled Darwin and Mendel

After Fisher showed how a continuous trait can result from adding up the contributions of multiple discrete Mendelian factors, he, Wright, and Haldane established the classic population genetic framework for studying how a population of genotypes evolves from one generation to the next. A central idea is fitness, a mathematical articulation of Darwin's conception of evolutionary advantage, a useful summary of an organism's phenotype. Formally, this fitness of a genotype can be thought of as the number of surviving offspring an organism with this genotype will have, in expectation. For simplicity, imagine a haploid organism (one copy of each gene) in which fitness depends only on two genes, and these genes come in three alleles each. The fitness function of this species is perhaps captured by this 3×3 table and denoted w_{ij} : $i = 1, 2, 3; j = 1, 2, 3$:

| | $j=1$ | $j=2$ | $j=3$ |
|-------|-------|-------|-------|
| $i=1$ | 1.05 | 1.00 | 0.96 |
| $i=2$ | 1.01 | 1.03 | 1.02 |
| $i=3$ | 0.94 | 1.02 | 0.99 |

Suppose that at some point in time, in the t th generation of this species, the nine genotypes have these frequencies, denoted P_{ij}^t :

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.1111 | 0.1111 | 0.1111 |
| $i=2$ | 0.1111 | 0.1111 | 0.1111 |
| $i=3$ | 0.1111 | 0.1111 | 0.1111 |

Assume reproduction by recombination (sex); this means that a child inherits each allele from either of its parents; assume that it inherits it from either with equal probability, independently for the two genes (this holds, for example, if the two genes reside in different chromosomes). Then the genotype frequencies in the next generation, P_{ij}^{t+1} , to the fourth decimal place, will be as shown in this table:

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.1164 | 0.1109 | 0.1064 |
| $i=2$ | 0.1120 | 0.1142 | 0.1131 |
| $i=3$ | 0.1042 | 0.1131 | 0.1098 |

while the frequencies in 10 and 100 generations will be as follows:

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.1210 | 0.1226 | 0.0919 |
| $i=2$ | 0.1263 | 0.1469 | 0.1171 |
| $i=3$ | 0.0825 | 0.1085 | 0.0831 |

in 10 generations

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.0848 | 0.1409 | 0.0163 |
| $i=2$ | 0.2320 | 0.4424 | 0.0542 |
| $i=3$ | 0.0086 | 0.0185 | 0.0022 |

in 100 generations

The formula for going from the P_{ij}^t 's to the P_{ij}^{t+1} 's is the following:

$$P_{ij}^{t+1} = \frac{w_{ij}}{\bar{w}} \left(\frac{1}{2} P_{ij}^t + \frac{1}{2} \sum_l P_{il}^t \sum_k P_{kj}^t \right),$$

where \bar{w} is the sum of the numerators for all ij 's. Of course, this equation is based on certain crucial assumptions: It is assumed that generations do not overlap (as if all members of one generation procreate simultaneously just before death), that the population is so large that it can be thought of as being infinite, that mating is completely random and without separate sexes, and that expectations can be added and divided with impunity. However, for the sake of simplicity, these assumptions have been used very widely and successfully in evolutionary theory for many decades. So far we have been assuming sexual reproduction. If this species were asexual, the next generation, the 10th and 100th hence would be like this:

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.1164 | 0.1109 | 0.1064 |
| $i=2$ | 0.1120 | 0.1142 | 0.1131 |
| $i=3$ | 0.1042 | 0.1131 | 0.1098 |

next generation

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.1693 | 0.1039 | 0.0691 |
| $i=2$ | 0.1148 | 0.1397 | 0.1267 |
| $i=3$ | 0.0560 | 0.1267 | 0.0940 |

in 10 generations

| | $j=1$ | $j=2$ | $j=3$ |
|-------|--------|--------|--------|
| $i=1$ | 0.7767 | 0.0059 | 0.0001 |
| $i=2$ | 0.0160 | 0.1135 | 0.0428 |
| $i=3$ | 0.0000 | 0.0428 | 0.0022 |

in 100 generations

and the equation is much simpler:

$$P_{ij}^{t+1} = \frac{w_{ij}}{\bar{w}} P_{ij}^t,$$

with \bar{w} playing the same role. Naturally, much of evolution involves events not entering these equations at all: mutations creating new alleles, speciation events (the creation of a new branch in the tree of life with its own equations), and changes in the environment. These equations should be thought of as a mathematical description of a piece of the puzzle, still studied after nearly a century of research.

ideal fitness function by a polynomially large population of genotypes in polynomially many generations³⁵ through learning by mutations. Notably, there is no recombination (sex) in his theory, even though it can be added for a modest advantage.¹⁷ Several natural classes of functions are evolvable in this sense; in fact, functions susceptible to a limited weak form of learning called *statistical learning*.¹⁸ In the section “A Game between Genes,” we discuss another interesting connection between machine learning algorithms and evolution.

The Role of Sex

Sex is nearly universal in life: it occurs in animals and plants by the coming together of sperm and egg, in fungi by the fusion of hyphae, and even in bacteria.³⁴ Two bacterial cells can pair up, for example, and build a bridge between them through which genes are transferred. Many species engage in asexual reproduction, or in selfing, at some times, but also engage in sexual reproduction at other times, keeping their genotypes well shuffled. In contrast, species that do not exchange genes in any form or manner, called “obligate asexuals,” are extremely rare, inhabiting sparse, recent twigs of the tree of life, coming from sexual ancestral species that lost their sexuality, and heading toward eventual extinction without producing daughter species.²²

Not only is sex essentially universal, but it seems to be very much center stage in life, the basis of a fantastic variety of behavior and structure: from bacterial conjugation to the intense molecular machinery of meiosis (cell division producing gametes), from flower coloration to bird courtship dances, from stag fights all the way to the drama of human passion, much of life seems to revolve around sex. So why? What role might sex play in evolution?

One common answer is that sex generates vast genetic diversity, and hence it must help evolution. But, just as sex puts together genetic combinations, it also breaks them down: a highly successful genotype will be absent from the next generation, as children inherit half their genes from each parent. To say the role of sex is to create particular, highly favorable genetic combinations is like watching a man catch fish only to toss them back

to sea, and concluding that he wants to bring food to his family’s table. (Incidentally, the designers of genetic algorithms are well aware of this downside of sex, and often allow the most successful individuals into the next generation, a stratagem known as “elitism,” which however cannot be easily imitated by nature.)

Evolutionary theorists have labored for about a century to find other explanations for the role of sex in evolution, but all 20th century explanations are valid only under specific conditions, contradicting the prevalence of sex in nature.^{9,a}

This is not a small problem. Imagine, for example, that even though much of the terrestrial world is green, we had no clue why leaves exist. That would have been a pretty big gap in our understanding of nature. Not knowing the role of sex is an even bigger gap, because far more life forms exchange genes than photosynthesize. It is no wonder the role of sex has been called “the queen of problems” in evolutionary biology.⁶

Since sex breaks down genetic combinations, it has been mainly thought in evolution that effective selection acts on individual alleles,³⁸ that is, each (non-neutral) allele is either beneficial or detrimental on its own. According to this line of thinking, two main forces drive allele frequencies: selection acting on alleles as independent actors (where alleles are often assumed to be making additive contributions to fitness), and random genetic drift (chance sampling effects on allele frequency, as discussed previously). The interaction between alleles, within and between loci—even though it has been of interest in population genetics from the start^{10,39}—has played a secondary role, often being treated as a mere correction to the above, under the term “epistasis.” A few years ago, while working with biologists Marcus Feldman and Jonathan Dushoff and computer scientist Nicholas Pipenger, we asked whether interactions between alleles could be crucial to the understanding of the role of sex in a yet unexplored manner.^{23–26} Based on the

standard equations used to describe how genotype frequencies change over generations (see the Darwin and Mendel sidebar), we demonstrated an important difference between sex and asex: In asexual evolution, the best combination of alleles always prevails. In the presence of sex, however, natural selection favors “mixable” alleles, those alleles that, even though they may not participate in any truly great genetic combinations, perform adequately across a wide variety of different combinations.^{23–26} To put it differently, in the hypothetical three-by-three fitness landscape in the sidebar, the winner of asexual evolution will be the largest entry of the fitness matrix (in this case, 1.05). In contrast, sexual evolution will favor, roughly speaking, those alleles (rows and columns) with larger average value; where “average” takes into account the prevalence of these genotypes in the population, as we will explore.^b

Sex as a Randomized Algorithm

One of the most central and striking themes of algorithms research in the past few decades has been the surprising power of randomization.²⁹ Paradoxically, evoking chance is often the safest and most purposeful and effective way of solving a computational problem. For one, it helps avoid the worst case, as in Quicksort. Second, sampling from a distribution D helps decide between competing hypotheses about D : Randomized algorithms for software testing, as well as for testing primality, or validity of polynomial identities, are like this.

Evolution under sex can be seen as an instance of a randomized algorithm of the latter type. Suppose we want to design a hypothetical evolutionary experiment for determining whether a new allele of a particular gene performs better than its alternative, across all genetic combinations. If the population is asexual, this could be done by inserting this mutation in the genome of one individual, and gauging the lineage thus founded to see if it thrives. This kind of sampling

a See the appendix available in the ACM Digital Library (dl.acm.org) under Source Material for a more extensive bibliography on this and other subjects.

b The original paper²⁴ refers to the unweighted average fitness as mixability, instead of the more natural average weighted by genotype frequency.

is very inefficient, because we sample from a small pool (the genotypes that happen to be available in the population), and must repeat the insertion many times—in many individuals. But if the population is sexual, then by inserting the mutation once, after $\log n$ generations, where n is the number of genes with which this particular allele interacts,³³ we will be sampling from all possible genetic combinations that could in principle be constructed. Sex enables evolution to sample quickly from the entire space of genetic combinations, in the distribution under which they appear in the population. What is more, evolution under sex not only decides among the competing hypotheses (which allele performs better), but also implements this decision (eventually, and with high probability, it will fix the winner).

Finally, for yet another take on explaining the ubiquity of sex in life, in terms and concepts familiar to computer scientists, view The Red-Blue Tree Theorem sidebar.

A Game between Genes

The search in population genetics for a quantity that is optimized by natural selection has a long history. Fisher wanted a theory for evolution with a mathematical law as clean and central as the second law of thermodynamics,¹⁰ while Wright pointed out that the frequency of an allele in a diploid locus changes in the direction that increases the population's mean fitness.⁴⁰ Later, investigators tried their hand again at looking for a Lyapunov function that will describe evolution, albeit with little success.

Our search for an analytical maximization principle involving mixability ended with a surprise: We did not answer the question “What is evolution optimizing?” but, perhaps more interestingly, we identified the quantity that each gene seems to be optimizing during evolution under sex. Together with Erick Chastain and Umesh Vazirani,⁷ we focused on the standard equations described in the Darwin Mendel sidebar, in a particular evolutionary regime known as *weak selection*.³⁰ Weak selection is the widely held assumption that fitness differences between genotypes are small. The fitness of a genotype g in this regime

is written $F_g = 1 + s \Delta_g$, where s is small and Δ_g is the *differential fitness* of the genotype, ranging in $[-1, 1]$.

Working with the weak selection assumption, and after some algebraic manipulation, we noticed the equations of the evolution of a population under sex are mathematically equivalent to a novel process, which entails an entirely different way of looking at evolution:

- The process is a game between the genes of the organism. Recall that a game is a mathematical model of the interaction between several players, each player having a set of available actions or pure strategies, and a utility function: the objective the player is trying to maximize in the game, a mapping from choices of actions, one for each player, to a real number. The game here is repeated, that is, the same precise game is played over and over for many rounds, with each round corresponding to a generation of the population.

- The available actions of each gene/player correspond to the gene's alleles.

- At every generation, each gene chooses and plays a mixed strategy: it randomizes over its actions, that is to say, its alleles. The probability assigned to each allele in this mixed strategy corresponds to the frequency of the allele in the population during this generation.

- The intricacy of game theory is mostly due to conflicts between the objectives of the players. But in the game played by the genes there are no conflicts—this is not Dawkins's selfish gene metaphor: All players have the same utility function, which is the fitness of the genotype resulting from the players' choices. Games of identical utilities are called coordination games in game theory, and are the simplest possible kind. They are of interest only in cases in which the players are cognitively weak, or cannot communicate effectively.

- In a repeated game, the players must update their mixed strategies from one round to the next, based on the outcome of the previous rounds. Here lies the biggest surprise: The update rule used by the genes in this game is identical to a venerable learning algorithm, well known to computer scientists for its prowess in work-

ing surprisingly well in a multitude of difficult problems and contexts: the *multiplicative weights* update algorithm (MWU),² also known in machine learning as “no-regret learning” or “hedge” (see more in box “The Experts Problem” in the online appendix). MWU changes the frequencies x_i of the i -th allele of the gene as follows

$$x_i \leftarrow x_i (1 + sm_i), \quad (1)$$

where m_i is the expected differential fitness, positive or negative, of the i -th allele in the current gene pool. This quantity m_i is a measure of what we have called the *mixability* of allele i , its ability to form fit combinations with alleles of other genes in the current genetic mix. To summarize, at each generation in sexual evolution, each gene boosts the frequency of each of its alleles by a factor that increases with the mixability of this allele in the current generation. Naturally, the quantities resulting from the equation are normalized appropriately so as to add to one.

This is a completely new way of looking at evolution. And it is a productive view, because it gets more interesting: Let us look back at the update Equation 1 and ask once again the question: This choice of the new probabilities for the alleles by the gene, is it optimizing something? For once, the answer is very clean: Yes, the choice of allele frequencies by the gene shown in Equation 1 optimizes the following function, specific to this gene, of the allele frequencies:

$$\Phi(x) = \sum_i x_i M_i - \frac{1}{s} \sum_i x_i \ln x_i. \quad (2)$$

Here M_i denotes the cumulative relative fitness of allele i , that is, the sum of the m_i in Equation 1 over all generations up to and including $t-1$. It is easy to notice that Φ is a strictly concave function, and thus has one maximum, and this maximum can be checked by routine calculation to be exactly the new frequencies as updated in Equation 1! Now notice that the second term of Φ is plainly the entropy of the distribution x , a well-known measure of a distribution's diversity.

There is much that is unexpected and evocative here, but perhaps most surprising of all is that this radically new interpretation of evolution was lurking for almost a century so close to the surface of these well-trodden equa-

tions. That an algorithm as effective as MWU is involved in evolution under sex is also significant. It was pointed out in a commentary on our paper⁵ that the MWU is also present in asex. Indeed, asexual evolution can be trivially thought of as MWU helping nature select genotypes. However, our point is that in sexual evolution, the picture is far more sophisticated and organic, occurring deeper in the hierarchy of life: Individual genes interact, each “managing its investments in alleles” using MWU, in a context created by the other genes and, of course, by the environment.

The function Φ is a rare explicit optimization principle in evolution. The second term, and especially its large constant coefficient (recall that the selection strength s is small, and $|m_i| \leq 1$) suggests that attention to diversity is an important ingredient of this mechanism, a remark that may be relevant to the question of how genetic diversity is maintained. But there is a mystery in the non-diversity terms: the cumulative nature of the fitness coefficients M suggests that performance during any previous generation is as important as the current generation for the determination of the genetic make-up of the next generation. How can this be?

This surprising connection between evolution and algorithms, through game theory and machine learning, as well as the maximization principle Φ , are very recent, and their full interpretation is a work in progress.

The insights about sex as we have discussed shed some light on the mystery of genetic algorithms, as explained previously. There is a mismatch between heuristics and evolution. Heuristics should strive to create populations that contain outstanding individuals. In contrast, evolution under sex seems to excel at something markedly different: at creating a “good population.” So, it is small wonder that genetic algorithms are not the best heuristics around. On the other hand, these insights also suggest that genetic algorithms may be valuable when the robustness of solutions is sought, or when the true objective is unknown or uncertain or subject to change.

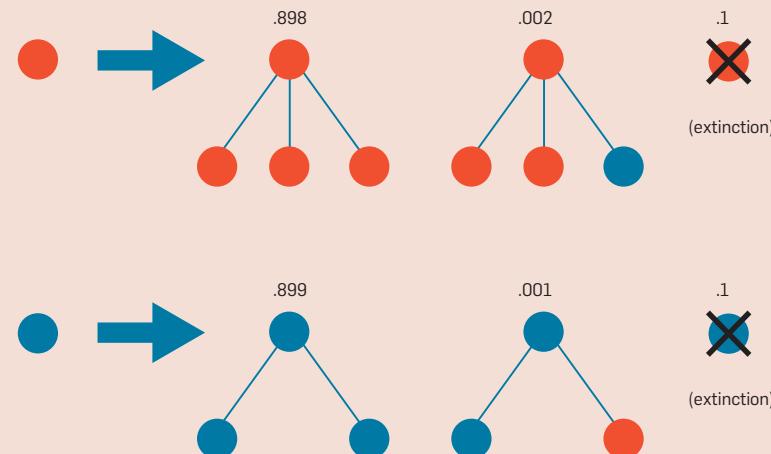
Are Mutations Random?

What is a mutation? Point mutations (changes in a single base such as an A turning into a C) are only part of the sto-

The Red-Blue Tree Theorem

A completely different take on the ubiquity of sex in evolution, based on unpublished work, uses concepts and images very familiar to computer scientists.

Imagine a tree with nodes of two kinds: red and blue. The tree starts from a root, which is blue, and is generated through a random process. At each step (depth of the tree), the nodes at this depth are replaced as shown here, each independently of the others according to the probabilities as shown below.



This is a very crude model for generating the tree of life (the probabilities here are indicative, and not supposed to reflect some biological reality). Red nodes are sexual species, and blue nodes are distinct asexual types (the term “species” does not perfectly apply to the latter). While there are reasons to believe that life started sexually,²² and not with a well-delineated species as exist today, we make the assumption that the tree begins with a blue node, which for our current purposes is conservative. According to the chosen parameters, sexuals and asexuals become extinct with equal probability. If extinction does not happen, sexual species branch out 50% more. With small probability, one of the new species can be of the other kind.

How would the leaves of the tree (that is, the life we see around us) look like, many steps down? After only 50 steps, the expected ratio of asexuals to sexuals is roughly 2:1000 (the probability of the top middle event).

Asexual nodes are rare! And, to a computer scientist, this stands to reason: We know that branching factors (in trees, computer virus propagation, recursive algorithms, among others) can dwarf in ultimate significance all other ingredients of a situation. Interestingly, this simple model immediately explains the main characteristics of the distribution of sexuality: It predicts that asexual forms are rare, sparsely distributed across the tree of life, and recently derived from sexual ancestors. Empirical evidence shows that all three points are true.²²

The precise relevant theorem is the following²³ (it is a consequence of well known properties of branching, see Athreya and Ney³):

Consider a continuous random process of generating a red-blue tree in which the rates of the events $B \rightarrow BB$, $B \rightarrow BR$, $B \rightarrow$ extinction, $R \rightarrow RR$, $R \rightarrow BR$ and $R \rightarrow$ extinction are, respectively, σ_A , β , ε_A , σ_S , α , and ε_S . Let $\delta_S = \sigma_S - \varepsilon_S$, $\delta_A = \sigma_A - \varepsilon_A$, and $\delta_S > \delta_A$. Suppose that $\alpha > 0$ and $\beta > 0$ be each small in comparison to $\delta_S - \delta_A$. Then the asymptotic ratio of asexual to sexual species at the leaves is $\approx \alpha / (\delta_S - \delta_A)$.

This argument suggests that, conceivably, a key advantage of sex might be its “branching factor” in the tree of life. Even if sex were a burden for species and organisms (which is at least a part of the truth), the mere fact that it promotes diversification and an exploration of new evolutionary niches might be enough for this strange trait to dominate the planet.

ry, as many important mutations are rearrangements of small stretches as well as large swaths of DNA: duplications, deletions, insertions, inversions, among others.¹³ For a long time it has been believed that mutations are the results of accidents such as radiation damage or replication error. But by now we have a

deluge of evidence pointing to involved biological mechanisms that bring about and affect mutations.²²

We know, for example, that the chance of a mutation varies from one region of the genome to another and is affected by both local and remote DNA.²² Nearly a quarter of all point mutations in humans

happen at a C base which comes before a G after that C is chemically modified (*methylated*);¹¹ methylation is known to be the result of complex enzymatic processes. As to rearrangement mutations, there are powerful agents of mutagenicity (*creation of mutations*) in the genome, such as transposable elements: DNA sequences prone to “jump” from one place of the genome to another, carrying other DNA sequences with them.¹³ A key step in mammalian pregnancy (*decidualization*), for instance, was the result of massive evolutionary rewiring of about 1,500 genes mediated in part by transposable elements.²⁷ Furthermore, the genetic sequences that participate in a rearrangement mutation are likely to be functionally related, since they are likely to be close to each other in 3D space and to bear sequence similarity, both of which allow interaction through recombination-based mutational mechanisms (see Livnat²² and references therein). Indeed, the same machinery that effects sexual recombination is also involved in mutations, and in fact produces different types of rearrangement mutations, depending on the genetic sequences that are present.¹³ Finally, different human populations undergo different kinds of mutations resulting in the same favorable effect, such as malaria resistance, suggesting that genetic differences between populations cause differences in mutation origination.²²

The idea that mutations may be non-accidental is still confronted with suspicion due to the legacy of Lamarck, who believed around 1800 that organisms can sense, through interaction with the environment, what is needed for an evolutionary improvement, and are able to make the correct heritable change. Since in the light of modern biology this seems impossible—to a computer scientist it sounds like reversing a one-way function, or a hash function—the accidental mutation notion prevailed. But in science one must not assume that the only relevant alternatives are the familiar, inside-the-box ones. Mutations are random—but it may be more productive to think of them as random in the same way that the outputs of randomized algorithms are random. Indeed, mutations are biological processes, and as such they must be affected by the interactions between genes. This new conception of heredity is exciting, because it creates

There is a mismatch between heuristics and evolution. Heuristics should strive to create populations that contain outstanding individuals. Evolution under sex seems to excel at something markedly different: creating a “good population.”

an image of evolution that is even more explicitly algorithmic. It also means that genes interacting in one organism can leave hereditary effects on the organism’s offspring.²² It no longer matters that a lucky genetic combination created by sex is doomed to vanish from the face of the earth (that the fisherman of our earlier metaphor throws the fish away): It may have achieved a lasting effect on the population through mutagenicity. Finally, the biological mechanisms affecting mutations may themselves evolve.

On the Preservation of Variation

If there is one idea that permeates all the various aspects of computational thinking about evolution, as explained in the past sections, it is this: Interactions between genes are crucial for understanding evolution. Gene interactions also come up in our recent work on the following question: How is variation within a species preserved?

Classic data indicates that, for a large variety of plants and animals taken together, the percentage of protein-coding loci that are polymorphic (in the sense that more than one protein variant exists that appears in more than 1% of the individuals in a population), and the percentage of such loci that are heterozygous in an individual, average around 30% and 7% respectively.³¹ Such a number is far greater than could be explained by traditional selection-based theories.²¹ Genetic variation is fed by mutations, and, according to the equations of population genetics, it is decreased by fixation, the eventual triumph of one allele and the extinction of all other alleles of the same gene. In much of the discourse on the subject, selection is assumed to act on individual genes, and therefore fitness is additive. The equations tell us that fixation will happen after $O(\frac{1}{\Delta})$ generations, where Δ is the difference in relative fitness between the most fit and the second most fit allele. Fixation looks rather speedy.

In the spring of 2014, the Simons institute brought to Berkeley 60 biologists and computer scientists to exchange ideas on evolution. It was during that time the authors, with Costis Daskalakis and Albert Wu, explored how our understanding of the speed of fixation would be affected if one takes into account gene interactions. We approached the subject through some decades-old work on the

complexity of local search;¹⁵ this theory examines how difficult it is for a local search process to reach a local optimum, and the conclusion has in many cases been: “pretty hard.” By applying this point of view to selection on interacting genes, we showed that there are n -gene systems, in which the fitness is the sum total of contributions of certain pairs of alleles—that is, the next step beyond selection on single alleles—for which fixation takes a number of generations proportional to 2^n to happen. A stronger result can also be obtained under the well-accepted complexity assumption that local search is intractable in general (see Johnson et al.¹⁵ for details). The implication is that, if gene interactions are taken into account, fixation may take much longer than in the regime of selection on individual genes.

Does this insight explain the mystery of variation? Not yet, because our analysis so far has been disregarding two other powerful forces in evolution, besides mutation and selection, acting on variation: the finiteness of the population, and heterozygosity (a diploid organism carrying two different alleles of a gene.).

First, finiteness. Because the number of individuals carrying the alleles in question is finite, say N , the number of individuals carrying each allele at each generation evolves as in a kind of a random walk within the confines of $[0, N]$, and, ignoring selection, this results in fixation after $O(N)$ generations.¹⁹ Second, diploidy introduces the possibility of overdominance, in which organisms with two different alleles of a gene are more fit than organisms with two copies of one allele or two copies of the other. In overdominance, the equations of selection point to stable variation, with both alleles enjoying stably high frequency in the population.

How these three effects, of finite population, of heterozygosity, and of selection acting on combinations of alleles across loci, interact with one another is an important subject for further research.

Epilogue

A computer scientist marvels at the brilliant ways in which evolution has achieved so much: Systems with remarkable resource efficiency, reliability and survivability, adaptability to exogenous circumstances, let alone ingenious and

pristine solutions to difficult problems such as communication, cooperation, vision, locomotion, and reasoning, among so many more. One is tempted to ask: *What algorithm could create all this in just 10^{12} steps?* The number 10^{12} —one trillion—comes up because this is believed to be the number of generations since the dawn of life $3.5 \cdot 10^9$ years ago (notice that most of our ancestors could not have lived for much more than a day). And it is not a huge number: cellphone processors do many more steps in an hour.

Over the past decade, computer scientists and evolutionary biologists working together have come up with new insights about central open problems surrounding evolution—including, rather surprisingly, a proposed answer to the “algorithm” question—by looking at evolution from a computational point of view. And, of course many more questions, inviting similar investigation, were opened up in the process. □

Additional background and literature appears in an online appendix available with this article in the ACM Digital Library (dl.acm.org) under Source Material.

References

1. Aldous, D. and Vazirani, U. ‘Go with the winners’ algorithms. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science* (1994), 492–501.
2. Arora, S., Hazan, E. and Kale, S. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing* 8, 1 (2012), 121–164.
3. Athreya, K. and Ney, P. *Branching Processes*. Springer, 1972.
4. Babbage, C. *The Ninth Bridgewater Treatise*. 2nd edn. John Murray, London, 1838.
5. Barton, N.H., Novak, S. and Paixão, T. Diverse forms of selection in evolution and computer science. In *Proceedings of the National Academy of Sciences* 111, 29 (2014), 10398–10399.
6. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. University of California Press, Berkeley, CA, 1982.
7. Chastain, E., Livnat, A., Papadimitriou, C. and Vazirani, U. Algorithms, games, and evolution. In *Proceedings of the National Academy of Sciences* 111, 29 (2014), 10620–10623.
8. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Murray, London, 1859.
9. Feldman, M.W., Otto, P. and Christiansen, F.B. Population genetic perspectives on the evolution of recombination. *Annual Review of Genetics* 30 (1997), 261–295.
10. Fisher, R.A. *The Genetical Theory of Natural Selection*. The Clarendon Press, Oxford, U.K., 1930.
11. Fryxell, K.J. and Moon, W.-J. CpG mutation rates in the human genome are highly dependent on local GC content. *Molecular Biology and Evolution* 22 (2005), 650–658.
12. Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, 1989.
13. Graur, D. and Li, W.-H. *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA, 2000.
14. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. U Michigan Press, 1975.
15. Johnson, D.S., Papadimitriou, C.H., and Yannakakis, M. How easy is local search? *J. Computer and System Sciences* 37, 1 (1988), 79–100.
16. Jong, K.A.D. *Evolutionary Computation: A Unified Approach*. MIT Press, Cambridge MA, 2006.
17. Kanade, V. Evolution with recombination. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, (2011), 837–846.
18. Kearns, M. Efficient noise-tolerant learning from statistical queries. *J. ACM* 45, 6 (1998), 983–1006.
19. Kimura, M. and Ohta, T. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61, 3 (1969), 763.
20. Kirkpatrick, S., Gelatt, Jr., C.D. and Vecchi, M.P. Optimization by simulated annealing. *Science* 220 (1983), 671–680.
21. Lewontin, R.C. and Hubby, J.L. A molecular approach to the study of genic heterozygosity in natural populations: amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54 (1966), 595–609.
22. Livnat, A. Interaction-based evolution: How natural selection and nonrandom mutation work together. *Biology Direct* 8, 1 (2013), 24.
23. Livnat, A., Feldman, M.W., Papadimitriou, C. and Pippenger, N. On the advantage to sexual species in diversification rates. Unpublished manuscript.
24. Livnat, A., Papadimitriou, C., Dushoff, J. and Feldman, M.W. A mixability theory for the role of sex in evolution. In *Proceedings of the National Academy of Sciences* 105, 50 (2008), 19803–19808.
25. Livnat, A., Papadimitriou, C. and Feldman, M.W. An analytical contrast between fitness maximization and selection for mixability. *J. Theoretical Biology* 273, 1 (2011), 232–234.
26. Livnat, A., Papadimitriou, C., Pippenger, N. and Feldman, M.W. Sex, mixability, and modularity. In *Proceedings of the National Academy of Sciences* 107, 4 (2010), 1452–1457.
27. Lynch, V.J., Leclerc, R.D., May, G. and Wagner, G.P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics* 43 (2011), 1154–1159.
28. Mitchell, M. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, 1996.
29. Motwani, R. and Raghavan, P. *Randomized Algorithms*. Cambridge University Press, 1995.
30. Nagylaki, T., Hofbauer, J. and Brunovský, P. Convergence of multilocus systems under weak epistasis or weak selection. *J. Mathematical Biology* 38, 2 (1999), 103–133.
31. Nevo, E., Beiles, A. and Ben-Shlomo, R. The evolutionary significance of genetic diversity: Ecological, demographic and life history correlates. *Lecture Notes in Biomathematics* 53 (1984), 13–213.
32. Papadimitriou, C. and Steiglitz, K. *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998.
33. Rabani, Y., Rabinovich, Y. and Sinclair, A. A computational view of population genetics. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, (1995), 83–92.
34. Stearns, S.C. and Hoekstra, R.F. *Evolution: An Introduction*. Oxford University Press, New York, 2005.
35. Valiant, L. *Probably Approximately Correct: Nature’s Algorithms for Learning and Prospering in a Complex World*. Basic Books, 2013.
36. Valiant, L.G. Evolvability. *J. ACM* 56, 1 (2009), 3.
37. Von Neumann, J. and A.W. Burks, A.W. Theory of self-reproducing automata. *IEEE Transactions on Neural Networks* 5, 1 (1966), 3–14.
38. Williams, G.C. *Adaptation and Natural Selection*, 8th edition. Princeton University Press, 1996.
39. Wright, S. Evolution in Mendelian populations. *Genetics* 16 (1931), 97–159.
40. Wright, S. The distribution of gene frequencies in populations. In *Proceedings of the National Academy of Sciences of the United States of America*, 23, 6 (1937), 307–320.

Adi Livnat (alivnat@univ.haifa.ac.il) is a Senior Lecturer in the Department of Evolutionary and Environmental Biology, and Institute of Evolution at the University of Haifa, Israel.

Christos Papadimitriou (christos@cs.berkeley.edu) is the C. Lester Hogan Professor in the Computer Science Division of the University of California at Berkeley.

Copyright held by authors. Publications rights licensed to ACM. \$15.00.



Watch the authors discuss their work in this exclusive *Communications* video. <http://cacm.acm.org/videos/sex-as-an-algorithm>

Copyright of Communications of the ACM is the property of Association for Computing Machinery and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.