Bi-parental Reproduction may Enhance Species Sustainability by Conserving Shared Parental Traits more Faithfully than Mono-parental Reproduction

Assaf Marron[1], Smadar Szekely[1], Irun R. Cohen[2], and David Harel[1]
[1] *Department of Computer Science and Applied Mathematics,*
[2] *Department of Immunology and Regenerative Biology,*
Weizmann Institute of Science, Rehovot, 76100, Israel
Email: *<firstname>.<lastname>@weizmann.ac.il*

Keywords:
Sexual reproduction, Asexual reproduction, Bi-parental, Uniparental, Mono-parental, Unisexual, , Autoencoding, Probabilistic Sampling

Abstract:
Recognized effects of sexual reproduction and other forms of bi-parental reproduction in species sustainment and evolution include the increasing of diversity, accelerating adaptation, constraining the accumulation of deleterious mutations, and the homogenization of species genotype. Nevertheless, many questions remain open with regard to the evolution of bi-parental reproduction. In this paper we contribute an initial exploration of finer details of the homogenization effect that we believe deserve focused analysis and discussion. Specifically, rudimentary mathematical analyses suggest the perspective that bi-parental reproduction enhances the retention in the offspring of shared, i.e., common, properties of the parent generation, as compared with mono-parental, clonal reproduction. We argue that this is an intrinsic effect of merging the encodings of parents' traits, regardless of physical, chemical, biological and social aspects of the reproduction process and of the trait at hand. As the survival of offspring often depends on their ability to step in and actively fill the voids left by failing, dying or absent members of the species in interaction networks that sustain the species, cross-generation retention of common species properties helps sustain the species. This side-effect of sexual reproduction may have also contributed to the very evolution and pervasiveness of sexual reproduction in nature.

1 Introduction

Natural reproduction is of course about transferring parental features to offspring. It is also well accepted that bi-parental reproduction yields resemblance of offspring in certain aspects to one parent or the other, or to both, or to having a mixture of traits from both parents. In this paper we explore a finer effect of sexual reproduction which is not addressed explicitly in discussions of evolution of sexual reproduction like those cited below, namely, that bi-parental reproduction conserves traits that are *shared* by both parents, and that may be common to the entire parent generation within the species.

While there are multiple theories for the sustainment and evolutionary effects of bi-parental reproduction, both sexual (with two or more mating types) and unisexual (with one mating type, as is common, for example in fungi (Heitman, 2015)), over mono-parental, clonal, asexual reproduction, the underlying question is still considered open (MacPherson et al., 2023; MacPherson et al., 2021; Livnat and Papadimitriou, 2016; Billiard et al., 2012). and others.

The most commonly cited consequences of bi-parental reproduction include: (i) the introduction of diversity, which then leads to increased adaptivity to changing conditions, and enhanced competitiveness and the ability to co-evolve with pathogens, parasites, or symbionts; for example the Red Queen hypothesis (Van Valen, 1973); (ii) the constraining the irreversible penetration of deleterious mutations, often termed Muller's ratchet (Kondrashov, 1988; Muller, 1932); (iii) preservation of common properties by facilitating the raising of reproduction barriers between divergent subgroups within a species (Barton et al., 2007, CH.22), DNA repair (Bernstein et al., 1981), (Barton et al., 2007, Ch.12 p.335) that is part of meiosis, sexual selection which promotes reproduction of individuals with certain preferred properties, and interference with specialization and speciation of variants that find and prefer new niches (Barton et al., 2007, Ch.16). Indeed, the genome homogenization effect of bi-parental reproduction is often taken for granted when a population is panmictic, i.e., every organism of the species can mate with every other of the same species, and confined within a geographical region (Fitzpatrick et al., 2008, Abstract); it is also often implicit from discussions of gene flow in structured populations (Barton et al., 2007, Ch.16 p.441.), or in pondering the puzzle of sympatric speciation, i.e., why do panmictic populations in a given location sometime give rise to diversity and speciation (Fitzpatrick et al., 2008; Mayr, 1982).

Here we show that in a population whose clonal reproduction is imperfect and may induce some change, as is the case with many kinds of DNA mutations (Barton et al., 2007, Ch.12), horizontal gene transfer (Keeling and Palmer, 2008) incorporation of viral genetic material (Koonin, 2010), etc., bi-parental reproduction that employs merging of encodings of inherited properties, has an underlying mathematical conservation effect that is independent of biological specifics, and is more general than the homogenization mechanisms listed above. Specifically, bi-parental reproduction conserves the properties that are common within the parents' generation better than mono-parental reproduction. This, of course, then extends to conservation of ancestral properties across multiple generations.

We support our claim that the effects of bi-parental reproduction are intrinsic, by initially showing that they exist in an abstracted setting: we deal with only a single property, and the mechanism does not attempt to model in detail meiosis, genetic recombination, epistasis, syngamy, the role of mating types, the role of the environment, and other complex aspects of reproduction in nature. While all these factors are of course important and should be part of realistic models of reproduction in nature, the particular aspects of bi-parental reproduction discussed here deserve a special focus beyond the published theories, mechanisms and effects.

1.1 Current theories do not imply conservation of shared traits.

Below we list some distinctions between offspring's retention of traits that are common to their two parents, from presently published theories and observations around sexual reproduction.

1. Consider at first the prominent effect of sexual reproduction, which is the enhancing of species diversity. This capability relies on the fact that individuals differ from each other in the first place. By contrast, the effect discussed here focuses on the traits that do not change
2. Plain merging of encodings is not the same as repair: It does not involve a preferred template, and does not call for assessment of whether a structure is broken or intact.
3. Conservation of shared properties is not about elimination of traits that interfere with fitness. In the present model, entity survival is not affected by whether its traits are deleterious or beneficial to its ability to reproduce, or its rate of reproduction.
4. There is no preferential attitude to properties, in contrast, for example, to how sexual selection contributes to species homogenization.
5. The present analysis is different from the homogenization effect of reproductive barrier raised between divergent subspecies, as the group we analyze is the entire population where every individual can and may mate with every other individual in the species.
6. The way sexual reproduction and associated gene flow constrains specialization and fitting into new niches by diluting new beneficial traits may be a special case of the mechanism described here. Bi-parental reproduction slows down property drift regardless of how useful the new properties may be in new niches. We emphasize that interference with speciation comes from retention of shared traits at the expense of diverse ones. Hypothetically, the dilution of over-specialized properties could occur by further randomizing then, and preventing them from becoming common and shared, without slowing down the drifting away of the properties that are shared and common.

7. Most importantly, homogenization of the next generation as induced by sexual reproduction, and discussed extensively in the literature, does not necessarily imply that a shared property of a homogeneous new generation is also closer to the shared property of the immediately preceding parent generation. For example, if the parent generation is comprised of clusters of different trait values, a homogeneous child generation may still have genotype and phenotype properties that are not in any of these clusters.

In the next section, we briefly discuss some of the motivation for the particular principles applied in the ensuing model. In Section 3 we use a synthetic model with encodings consisting of sets of real numbers to prove mathematically the inevitability of conservation of parental properties as a general effect of bi-parental reproduction derived from basic probability and the central limit theorem. In Section 5 we illustrate the effect with examples drawn from simulation runs that reproduce diverse objects including abstract sets of points, images of printed text, and images of flowers.

In summary, given that survival of organisms, species and entire ecosystems depends on organisms *fitting in* into a rich interaction network, and given that all living individuals are destined to eventually die, our interim results show that intrinsic properties of bi-parental reproduction likely help sustain each species by enabling new generations to carry out the interaction roles performed by their parents and by several preceding generations. Further studies are needed to generalize and expand the model, and to confirm that the conservation of shared traits by bi-parental reproduction is indeed a distinct natural phenomenon.

2 Modeling Rationale

Before delving into the details of the mathematical model, in this section we offer some background and context for various choices applied in the model and its analysis. The reader may choose to skip this section, and return to it as needed.

2.1 Codes, interaction codes, and autoencoding

The present work is part of a larger research project. In (Cohen and Marron, 2023; Cohen and Marron, 2020) and related papers we have outlined how sustainment of a species depends on interaction networks in which it participates. In such networks, the role of each species is naturally and automatically encoded in what we term *species interaction code*, manifested in the genes, physiology and environment of the individuals of the species. We aim to extend our modeling to address not only species properties, but inter- and intra-species interactions, and research how sexual reproduction may help encode and conserve species interaction codes over multiple generations.

As part of our work on natural autoencoding we have experimented with autoencoding of basic structures like images of straight line segments (Bayer et al., 2023), which hypothetically, can be encoded in just four numbers: the two pairs of coordinates of the endpoints, or a pair of coordinates, an angle and a length. While not dealing with autoencoding per-se, the present model further reduces this two dimensional line encoding problem to a one dimensional continuous range of real numbers which can be encoded in just two numbers, like the value of its boundaries, or its center point and its length. For the formal proof we used a normal distribution with a given mean and variance, rather than a continuous segment.

In this spirit, we focus on the passing down of the encoding of a trait, not the trait itself, even though the phenotype is what in the end sustains the organism. For example, we assume that if the property of interest of an individual is its average height as an adult, or its expected life span, or its varying colors in various seasons, then this property must somehow be encoded before it can be recreated. The encoding itself does not have to be compact or minimalistic; for example, it may involve extensive redundancy, to overcome decoding errors, or may be captured physically, as is the case in a human-made diagram or graph, which may be recorded in many pixels.

This perspective connects the present work to quantitative genetics (Barton et al., 2007, Ch.14;16;17;19;26),(Kang, 2020),(Fisher, 1918). While *blending inheritance* theory, where parents traits are mixed to form the offspring features, is long considered obsolete, quantitative genetics studies the many quantitative phenotypic traits, like lifespan, weight, strength, or speed, which are affected by vast genetic information, commonly represented in Quantitative Trait Loci (QTL) which may include multiple genes, regulatory sequences, epigenetic effects, etc. The phenotype aggregation of such polygenic features involves among others operations that resemble adding and averaging the various genetic factors within the organism. Furthermore, the way offspring QTL information is determined from the merging of parents' QTL is known to be highly complex (Cui et al., 2004; Niu et al., 2022), and cannot yet be reduced to a small set of rules; the model makes a modest attempt to reflect this complexity.

Throughout this paper, we use the term *an encoding* to refer to an object or structure that encodes, or represents, another object or concept. Thus, an encoding is synonymous here with the term *a code*, which we presently reserve for the broader concept of *species interaction code.*

While the model does not aim at mimicking meiosis and creation of gametes, the model includes an elaborate step of recreating a parent's encoding that will then serve as a seed for subsequent reproduction steps; this is instead of just directly copying the parent's encoding.

2.2 Information aggregation by set operations

While the relation to quantitative phenotype traits allows some measure of numerical calculations on traits, we wanted the determination of offspring trait encodings to be carried out largely by set operations (like set union) on parent encodings, relying less on pure numerical operations on the two sources.

The present setting is amenable to extensions using more elaborate methods for deriving child traits from parents' traits.

Note that pure averaging of a single property represented by a random variable would be analogous to combining two one-dimensional random walks, with the same step size and equal probabilities for stepping in either direction. Such random walks can be readily shown to have a mean position which is the average of the two origins, and, more importantly, its variance is half the (equal) variance of the two original random walks. This is aligned with the known fact that bi-parental reproduction reduces the variance within the offspring generation.

2.3 Minimalism, abstraction and accessible language

The model is not intended to be a full model of what happens in nature. In fact, we try to show that in situations of noisy inheritance, it is the very merging of parent encodings that enables the cross-generation conservation of ancestral traits. Examples of elements of the abstraction that we employed include: (i) The property of interest and its encoding both involve real numbers. Extending the arguments to pure symbolic entities, or concrete objects like molecules, is left for a future research; (ii) The inevitable variation in natural reproduction is modeled by replacing direct copying with sampling from a normal distribution; (iii) the bi-parental reproduction is unisexual, i.e., there are no distinct mating types.

Since here we aim to only show relations between certain values, rather than establishing those values, we adopt the "story proof" argumentation style, as defined and used in(Blitzstein and Hwang, 2015); this style relies on text of intuitive, yet well founded, logic statements to support the claims, and keeps mathematical formulas to a minimum. In addition, the occasional use of terms that are commonly applied to populations of living organisms is only for convenience.

3 Problem Formalization

In this section we define the bi-parental and mono-parental reproduction methods in the context of a specific model.

3.1 The formal model

Consider the following synthetic model of a population of entities, also referred to as individuals. The size of the population is kept fixed at a number n. The analysis does not include considerations of time. The population reproduces in methods discussed below. A set of n entities, referred to as the cohort of the next, or child, generation, appears instantly and synchronously.

In the present discussion it does not matter whether the parent generation persists or disappears as long as the entities in the parent generation do not reproduce after the appearance of the child generation. For simplicity, we assume that the entire cohort of each parent generation disappears immediately and synchronously upon the appearance of the cohort of the child generation.

One reason for this choice is to avoid dealing with inter-generation mating in a population in which diversity among co-existing distant generations is unbounded.

We label the individual entities in a cohort as $\{e_1, e_2, ..., e_n\}$;

All entities are associated with exactly one "phenotype" property P, which is manifested as a real number representing some quantity. Different entities may have the same or different values for P. For illustration, the abstract entities may be thought of as fixed-width sticks where P is their length, or as lumps of material where P is their weight. Let p denote any value of P, and, for example, p_i is the value of P in the entity e_i. As for the focus on a quantitative property, see (Barton et al., 2007, Ch.16 p.441). With each new generation, under the two reproduction methods, the p_i values and their distribution within each new cohort may change. In each cohort c we are first interested in the mean of the p_i values within the cohort

$$p_c = \frac{1}{n}\sum_{i=1}^{n} p_i \quad .$$

In each parent generation c, one can view p_c as representing a common property, or capability, which is manifested with some diversity in the living individuals. And, if the value of p in an individual is essential for survival within the species' ecosystem, then the distribution of the property P among the members of the next generation c' affects the likelihood that individuals from c' will be able to participate in ecological networks in which their parents from c no longer function.

The "genotype" encoding of P in each individual e_i is a set g_i of k of real values which are also referred to as the points in g_i. We allow the rare case of any two points in a set being equal to each other, without resorting to other mathematical terms like g_i being a multiset or a tuple. The integer k is a fixed parameter in the model, and it is large, say, $k \geq 100$; we also constrain k to be an even number.

The phenotype manifestation of the property P in e_i is the mean of the points in g_i. That is, given an individual e_i, $1 \leq i \leq n$, with genotype encoding $g_i = \{x_{i_1}, x_{i_2}, ..., x_{i_k}\}$,its phenotype value p_i is computed as follows:

$$p_i = \frac{1}{k}\sum_{\ell=1}^{k} x_{i_\ell} \quad .$$

3.2 Reproduction methods

We model and compare the effects of two different reproduction processes, termed *Mono-parental Reproduction* (MR, hereafter) and *Bi-parental Reproduction* (BR, hereafter).

Below we describe one reproduction step, termed *a transition* in each of the methods.

Mono-parental Reproduction

1. Denote the current population as being the set c and create an empty set c'.
2. Repeat n times to create n children:
   (a) With equal probability, pick a single individual $e_i \in c$, by picking a random integer $1 \leq i \leq n$; the selected individual will be the single parent of the child created in this iteration; in separate iterations, the same individual e_i may be selected again. Note: hereafter, when the distribution of a random choice is not stated, it is assumed to be uniform, with equal probability.
   (b) Let $g_i = \{x_{i_1}, x_{i_2}, ..., x_{i_k}\}$ be the genotype encoding of e_i. Compute the mean of g_i, which is equal to the phenotype of e_i:
   $$p_i = \frac{1}{k}\sum_{\ell=1}^{k} x_{i_\ell} \quad ;$$
   this choice of reproduction step simplifies the model while keeping it intuitive and not limiting the claims; it can be generalized in future analyses.
   (c) Compute a new set g' by sampling k real numbers, $g' = \{x'_1, x'_2, ..., x'_k\}$, subject to normal distribution with mean p_i and standard deviation h (and variance h^2) for some fixed global parameter h which specifies and constrains the noise, or error, that may be introduced in the reproduction processes.
   (d) Create an entity e' whose genotype encoding is g', and its phenotype is the mean of the points of g', and add it to c'.
3. After completing n iterations: Mark the current population as being the set c' and discard c.

Bi-parental Reproduction

1. Denote the current population as being the set c and create an empty set c'.
2. Repeat n times to create n children:

    (a) Pick a random pair of two separate individuals, $e_u, e_v \in c,\ u \neq v, 1 \leq u, v \leq n$ out of the $\frac{n(n-1)}{2}$ possible pairs in c. The pair e_u, e_v will be the joint parents of the child created in this iteration; in separate iterations these two individuals may participate in other pairs or again in the very same pair.
    (b) Let the respective genotype encodings of e_u and e_v be
    $g_u = \{x_{u_1}, x_{u_2}, ..., x_{u_k}\}$ and
    $g_v = \{x_{v_1}, x_{v_2}, ..., x_{v_k}\}$, and let p_u and p_v be the respective means of these genotype encodings; again, these mean values are conveniently equal to the respective phenotypes.
    (c) Create two interim ("gamete") encodings g'_u and g'_v by randomly selecting $\frac{k}{2}$ points, under normal distribution with mean p_u and variance h^2, and $\frac{k}{2}$ points under normal distribution with mean p_v and variance h^2, respectively; let g' be the union of these two sets, $g' = g'_u \cup g'_v$. We term this process *merging* of the parents' encodings. Note that this merging involves a sampling step, and does not rely on copying of elements of g_u or g_v into g'.
    (d) Create an entity e' whose genotype encoding is g', and its phenotype is the mean p' of the points of g', and add it to c'.

3. After completing n iterations: Mark the current population as being the set c' and discard c.

As stated earlier, neither MR nor BR employ considerations of Natural Selection and classical fitness metrics like an individual's probability of reproduction and its expected number of offspring over its life time, and BR does not make use of mating types or of mate selection.

3.3 Model realization example

Below we illustrate the above formal model with an example realization in an imaginary species of intelligent and capable animals. One of the traits that sustains this species is their ability to build every year a mud hill of "exactly" a certain height, as traps for some creature that they feed upon. The way they remember the right height is that each mature individual carries with them a bunch of sticks that they got from their parents. The individual builds each mud hill so that its height is close to the mean length of its set of sticks. Old mud hills from the previous year are not available for comparison, as they have been washed away by the rains. When young individuals mature, they leave their parents to look for new territory. Before they leave, the parents (one or two of them, as the case may be) equip the young with a new bunch of sticks, cut as best as they can to the mean length of the parents' bunch. In the case of two parents' each one gives the child half the number of sticks based on that parent's set. The parents cannot give their own sticks to the offspring, as they have to keep their own sticks for several more years. The individuals are solitary most of the time, so each one needs their own set of sticks. Individuals of the species use sets of sticks for this purpose, rather than just one, as the sticks may break or get lost. In addition, the parent's skill and the available material limit the precision of each cut stick; having several sticks, some too short and some too long, is a way for conveying the height of the mud hill that is desired by the individual preparing the new set.

3.4 A numerical example

For illustration, below we go through the core elements of MR and BR using a small-scale example.

1. **Parameter Setting:** Assume that in the parents' cohort c , P is normally distributed with mean $p_c = 2.0$, with standard deviation $\sigma_c = 0.2$ The number of points in a genotype is $k = 4$; the points in a genotype set are shown here in a sorted order for easier reading. The standard deviation of the normal distribution of the reproduction noise is $h = 0.15$
2. Let individual e_1 have a genotype encoding of $g_1 = \{1.70, 1.80, 1.90, 2.10\}$ whose mean and phenotype is $p_1 = 1.875$;
3. **MR of e_1:** Compute the mean of the genotype, yielding again $p_1 = 1.875$. Draw a sample of $k = 4$ points around this value as a mean with the above standard deviation $h = 0.15$, yielding say, $g' = \{1.61, 1.71.1.91, 2.01\}$, whose mean (and phenotype of child $e'_{1,\text{MR}}$) is p’=1.81;
4. Let individual e_2 in the parent cohort c have genotype encoding of $g_2 = \{1.95, 2.15, 2.25, 2.35\}$. whose mean and phenotype is $p_2 = 2.175$.
5. **BR of e_1 and e_2:** We sample $\frac{k}{2} = 2$ points around the mean p_1, and 2 points around the mean p_2, both with the noise standard deviation h, yielding, say, $g'_1 = \{1.8, 2.1\}$, and $g'_2 = \{2.02, 2.18\}$, respectively. The union of these two samples yields $g' = \{1.8, 2.02, 2.1, 2.18\}$,

whose mean (and phenotype of the joint offspring $e'_{1,2,\mathrm{BR}}$) is $p' = 2.025$.

4 Comparing the Two Reproduction Methods

In comparing MR and BR in a particular transition from a parent cohort c to a child cohort c', we are initially interested in a basic metric d defined as the distance between the phenotype properties of a random individual in c', denoted p'_i, and the common property p_c of the parents' cohort:

$$d_i = |p'_i - p_c| \ .$$

For a given parent cohort c, let d_{MR} and d_{BR} be the expected values $E[d_i]$ under the MR and BR processes, respectively, when considering all possible $c \to c'$ transitions that could emanate from c in one reproduction step, and random choices of e'_i individuals within c'.

For a given parent cohort c, let σ^2_{MR} and σ^2_{BR} be the expected values $E[\sigma^2_{c'}]$, the variance of c' under the MR and BR processes, respectively, when considering all possible $c \to c'$ transitions that could emanate from c in one reproduction step under the respective reproduction process.

Proposition 1. Let c be a cohort of n entities in the above model; assume that n is large and that the distribution of the p_i values of individuals in c approximates a random sample from a normal distribution with mean p_c and some variance σ^2_c; Then, $d_{\mathrm{BR}} < d_{\mathrm{MR}}$.

Proof.

A. Compute d_{MR}.

Let c'_{MR} be a random child cohort of c, selected from all possible MR transitions with c as a parent.

Let $e'_{i_{\mathrm{MR}}}$ be a random individual in c'_{MR}.

Let $g'_{i_{\mathrm{MR}}}$ be the genotype encoding of $e'_{i_{\mathrm{MR}}}$ and let $p'_{i_{\mathrm{MR}}}$ be its phenotype.

Let e_i be the element of c that served as the single parent of $e'_{i_{\mathrm{MR}}}$ in this MR transition.

Let p_i be the phenotype of e_i (it is also the mean of the genotype encoding g_i of e_i).

The genotype encodings $g'_{i_{\mathrm{MR}}}$ of all possible direct MR children of e_i, can be considered as samples of size k from a normal distribution with mean p_i and variance h^2, where h is the noise parameter defined above.

According to the central limit theorem (Blitzstein and Hwang, 2015, p.435), the means of these samples, namely $p'_{i_{\mathrm{MR}}}$, are distributed normally with mean p_i and standard deviation $\frac{h}{\sqrt{k}}$ (and variance $\frac{h^2}{k}$).

We are facing now a compound, nested distribution, that involves (i) the random selection of an e_i, which is close enough for our purposes to drawing a random value p_i from the original normal distribution of c with mean p_c and variance σ^2_c, and (ii) based on the selected p_i and the noise parameter, randomly selecting the sample g' of size k and computing its mean $p'_{i_{\mathrm{MR}}}$.

We can now examine the distribution of the random variable $p'_{i_{\mathrm{MR}}}$ as drawn from the original distribution of c. According to the law of total variance (a.k.a. Eve's law) (Blitzstein and Hwang, 2015, Sec.9.5), this is a normal distribution with mean p_c and variance that is the sum of the variances, hence

$$\sigma^2_{c'_{\mathrm{MR}}} = \sigma^2_c + \frac{h^2}{k}$$

In general, the average absolute deviation of a random sample from the mean of a normal distribution with standard deviation σ is $\sigma \cdot \sqrt{\frac{2}{\pi}}$ (Geary, 1935). Therefore, the expected value of the absolute distance of $p'_{i_{\mathrm{MR}}}$ from p_c is

$$d_{\mathrm{MR}} = E[|p'_{i_{\mathrm{MR}}} - p_c|] = \left(\sqrt{\sigma^2_c + \frac{h^2}{k}}\right) \cdot \sqrt{\frac{2}{\pi}} \quad .$$

B. Compute d_{BR}.

We now examine the expected effects of all possible BR transitions with c as a parent cohort.

Let c'_{BR} be a random child cohort of c within all possible BR transitions.

Let $e'_{j_{\mathrm{BR}}}$ be a random individual in c'_{BR}.

Let $g'_{j_{\mathrm{BR}}}$ be the genotype encoding of $e'_{j_{\mathrm{BR}}}$ and and let $p'_{j_{\mathrm{BR}}}$ be its phenotype.

Let e_u and e_v be the two elements of c that served as the parents of $e'_{j_{\mathrm{BR}}}$ in this BR transition, and let p_u and p_v be the means of their genotype encodings, respectively.

The genotype encoding $g'_{j_{\mathrm{BR}}}$ was created as a union of two random samples g'_u and g'_v, each of size $\frac{k}{2}$, around the means p_u and p_v, respectively. Let p'_u and p'_v be the means of g'_u and g'_v, respectively.

The phenotype $p'_{j_{\mathrm{BR}}}$ is the mean of the union of g'_u and g'_v, and is thus the average of the means of these two equal size samples:

$$p'_{j_{\mathrm{BR}}} = \frac{p'_u + p'_v}{2}$$

Let us now analyze the random selection of $p'_{j_{\mathrm{BR}}}$ compounding the randomness inherent in the way g'_u and g'_v were sampled, and the nested randomness of p_u and p_v as selected from the original distribution of c.

The distribution from which both p_u and p_v were sampled is normal with mean p_c and variance σ_c^2.

The sets g'_u and g'_v were drawn from distributions with means p_u and p_v respectively, both with variance h^2. According to the central limit theorem again, the distributions of p'_u and p'_v are normal with mean p_u and p_v respectively, and standard deviation $\frac{h}{\sqrt{\frac{k}{2}}}$ (and variance $\frac{h^2}{(\frac{k}{2})}$).

Compounding each of these two distributions separately with the underlying distribution of c we get that p'_u and p'_v are distributed normally with mean p_c and variance $\sigma_c^2 + \frac{h^2}{\frac{k}{2}}$

Clearly the mean of the random variable $p'_{j_{\mathrm{BR}}}$ is p_c since it is *half* of the sum $E[p'_u]$ and $E[p'_v]$ (each of which is equal to p_c), and its variance is a quarter of the variance of their sum, as it is the square of the scaling by half of the variable and the mean (Blitzstein and Hwang, 2015, p.159):

$$\sigma_{\mathrm{BR}}^2 = \frac{1}{4} \cdot 2(\sigma_c^2 + \frac{h^2}{\frac{k}{2}}) = \frac{\sigma_c^2}{2} + \frac{h^2}{k}.$$

As before,

$$d_{\mathrm{BR}} = E[|p'_{j_{\mathrm{BR}}} - p_c|] = \left(\sqrt{\frac{\sigma_c^2}{2} + \frac{h^2}{k}}\right) \cdot \sqrt{\frac{2}{\pi}} \quad .$$

C. Comparing d_{BR} and d_{MR}.

Since the first term in the expression for d_{BR} is $\frac{\sigma_c^2}{2}$ and for d_{MR} it is σ_c^2, we get that

$$d_{\mathrm{BR}} < d_{\mathrm{MR}} \ . \qquad \square$$

.

Note that an intermediate result in the above calculations is that $\sigma_{\mathrm{BR}}^2 < \sigma_{\mathrm{MR}}^2$. This aligns well with the established fact, as was discussed earlier, that in nature, sexual reproduction yields a lower variance *within* the new generation than does mono-parental reproduction. In the present context, such smaller values of $\sigma_{c'}$ help individuals in c' step in and replace *each other* when needed, independently of the magnitude of their differences from their parents and from earlier generations.

5 Visual Illustration by Simulation

Below we present the results of a few simulation run examples. These simulations are meant to serve as an accessible visual illustration for the general effects predicted by the above mathematical analysis; they were not subjected to elaborate quantitative analysis. In all examples we show here just one run; multiple runs using the same parameters and differing only by the pseudo-random numbers used at various steps in the reproduction process, yielded similar results.

5.1 Reproducing a single real number via a set of points

The first set of simulations follows directly the model described in Section 3 demonstrating the reproduction of a real number through an intermediate encoding by a cohort of points. The program source is available in (Marron et al., 2024) Figure 1 depicts the results. The orange color graphs refer to mono-parental reproduction (MR), and the blue to bi-parental reproduction (BR). The simulation was run with the following parameters.

- Cohort size: $n = 32$
- Encoding set size: $k = 30$
- Number of generations: 1000
- The Initial population, i.e., the cohort of generation 1, was created by random sampling of points from a normal distribution with mean 2.0 and standard deviation of $sd_c = 0.2$.
- Initial phenotype is thus: ~ 2.00.
- Reproduction error was produced by random sampling from a normal distribution where the mean is the parent's phenotype, and the standard deviation is $h = 0.15$.

We observe the following: (a) The average of the phenotypes of the entire cohort is closer to the original phenotype of approximately 2.00 under BR than under MR. (b) The standard deviation within each cohort generation is smaller in BR

(a) (b)

(c) (d)

Figure 1: Simulating points reproduction. See explanation in text.

than under MR; i.e., the similarity of members of a given generation to each other is higher under BR than under MR, and thus they are better positioned to replace each other within the ecosystem. (c) On average, the phenotypes of individuals of each generation, are closer to the average of the immediately preceding parent generation in BR than in MR. (d) On average, under BR, the phenotypes of individuals of each generation, are closer to the average of the original ancestral generation than under MR.

5.2 Reproducing images of printed text

The second set of simulations is inspired by the way the survival of ancient manuscripts over hundreds and thousands of years was enabled by reproduction, i.e., copying. Both "mono-parental" and "bi-parental" techniques were employed by scribes: most often copying from a single source, but sometimes collating multiple sources (Skeat, 1999, pp.585;617). (Note: we distinguish here between collation of several sources in the process of copying text, from the collation of sources by history researchers reconstructing an original document.) In this example we applied the reproduction techniques described above for the reproduction of images of text containing one word. We introduced a rudimentary measure of the preservation of properties as the success rate (across the cohort) of image recognition software (gpt-4-turbo model by OpenAI) in recognizing the text. This metric also hints at importance of traits in an organism's interactions, which in this case can be seen as the sending and receiving of messages. The details of the process are as follows. Consider first the following two functions:

Function F1. Noisy encoding of an image. Given an input image, create a copy of it; the copy may be of smaller, same or larger resolution; This step can be seen as a representation of the action of a person wishing to preserve a precious image or text by hand-copying it. Then, add noise to the copy by sampling random values from a normal distribution with mean 0 and standard deviation h, and adding them respectively to each of the three RGB values of all pixels of the copy; round the result to integer and clip the values at 0 and 255. This step may represent the imperfection involved with physical copying; image resizing also adds noise, but since the process is deterministic, it is identical in in all copies, and hence is less significant for our purposes.

Function F2. Noisy computation of a "phenotype". Given n copies of an image, as may be created using F1 above, compute the average of all respective pixel values; add noise to the result in the same manner as in F1, and resize the resulting image to the original resolution. This step may represent the collation process done by a person creating a fresh copy based on several sources, where each may have mistakes and defects.

The parameters used in the simulation are: seed image is an image of the word "TEACHING", typed in the font Calibri Light at resolution 100x100 pixels; copies generated by F1 are 400x400 pixels; noise standard deviation is $h = 20$; cohort size is $n = 8$; and, the number of copies in the "image genotype" encoding is $k = 8$. The simulation steps are as follows:

1. Start with a seed image.
2. [Creating base cohort.] Repeat the following n times:
   (a) [Compute "child genotype".] Create k copies, applying F1 above k times to the seed image.
   (b) [Compute "child phenotype".] Create one child image by applying F2 to the above image genotype.
3. Copy the above base cohort to create generation 1 of the BR process and generation 1 of the MR process.
4. [BR process.] Repeat the following for 90 generations
   (a) Repeat n times:
      i. Randomly select two individuals from the current generation.
      ii. Apply F1 to each of these two individuals $\frac{k}{2} = 4$ times.
      iii. Apply F2 to the $k = 8$ copies, yielding a new child individual.
   (b) Every 10th generation run image recognition on each of the cohort's $n = 8$ images, using OpenAI API with the model gpt-4-turbo and the prompt: "Your role is to identify the word in the image. Please provide a 1-word answer."
5. [MR process.] Repeat the following for 90 generations
   (a) [Create next generation.] Repeat $n = 8$ times:
      i. Randomly select one individual from the current generation.
      ii. Apply F1 to this individual $k = 8$ times.
      iii. Apply F2 to the $k = 8$ copies, yielding a new child individual.
   (b) Every 10th generation run the same image recognition task on the $n = 8$ cohort's individuals as under BR.

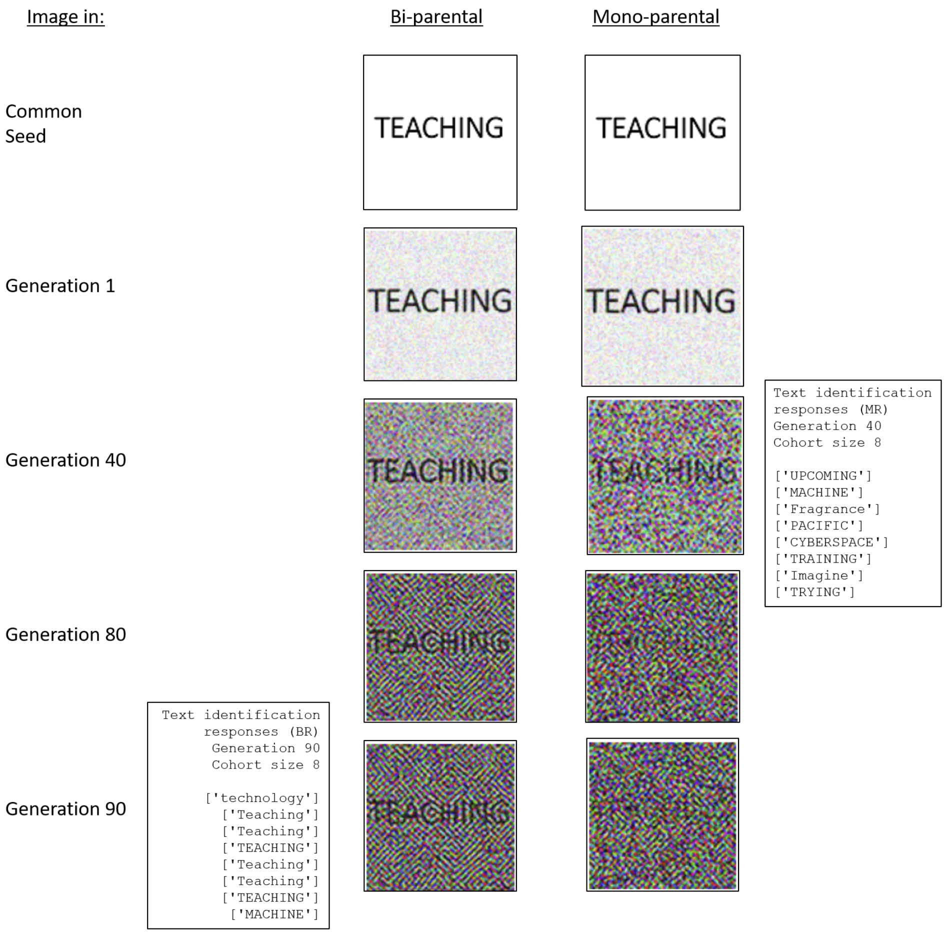
Figure 2 depicts a sample of the resulting images, with one image from each generation.

Under BR the LLM identified the word "TEACHING" throughout more than 70 generations, while under MR it failed the identification already on or before generation 40. In the 90th generation, under BR, the LLM did not identify the correct word but still identified a word (e.g., "TECHNOLOGY" instead of "TEACHING"), while under MR it did not identify that the image contained text and responded that "The image appears to be a stereogram or a pattern designed to create a visual illusion making it difficult to determine if there's any specific word or object within it. If there's a particular way I should analyze this please let me know!".)

We have also looked casually for common traits across the various cohorts and noticed that traits like contrast between the letters and their background, or the horizontal line of the letter T were preserved better under BR than under MR (see Figure 3). Under BR, we have also noticed conspicuous and sustained emergent traits, like a circle shape inside the letter C or a third vertical line in the letter H, where under MR we could not readily find any such emergent properties.

5.3 Reproducing images of flowers

In this illustration example we were inspired by object recognition in nature, as is done, for example, by insects species that pollinate only certain plant species. Figure 4 depicts the simulation of BR and MR of an image of a flower. The simulation uses the same process as in the case of text image, with the following parameters: The starting image is of an anemone flower, downloaded from https://it.pinterest.com/pin/red-anemones--797207571571931709/, and used under fair use license; the image was then cropped and resized to 100 x 100 pixels; noise standard variation is $h = 10$; cohort size is $n = 8$; Encoding set size is $k = 8$; Number of generations is 80; image recognition prompt is: "Your role is to



```
Text identification
responses (MR)
Generation 40
Cohort size 8

['UPCOMING']
['MACHINE']
['Fragrance']
['PACIFIC']
['CYBERSPACE']
['TRAINING']
['Imagine']
['TRYING']
```

```
Text identification
     responses (BR)
      Generation 90
        Cohort size 8

   ['technology']
     ['Teaching']
     ['Teaching']
     ['TEACHING']
     ['Teaching']
     ['Teaching']
     ['TEACHING']
      ['MACHINE']
```

Figure 2: Simulating reproduction of an image of printed text. See explanation in the body of the article.

Figure 3: The 8 individuals of generation 40 under BR and MR. See discussion in the body of the article about retention of common traits, both ancestral and emergent.

identify the object in the image. Please provide a 1-word answer.";

One can see that, under BR, more of the images were recognized like the first one ("Poppy") for more generations than under MR. We also observe that certain core properties, like being a flower were better preserved under BR.

Note: The images shown here are illustrative examples. While in all runs the image properties were sustained for longer under BR than un-

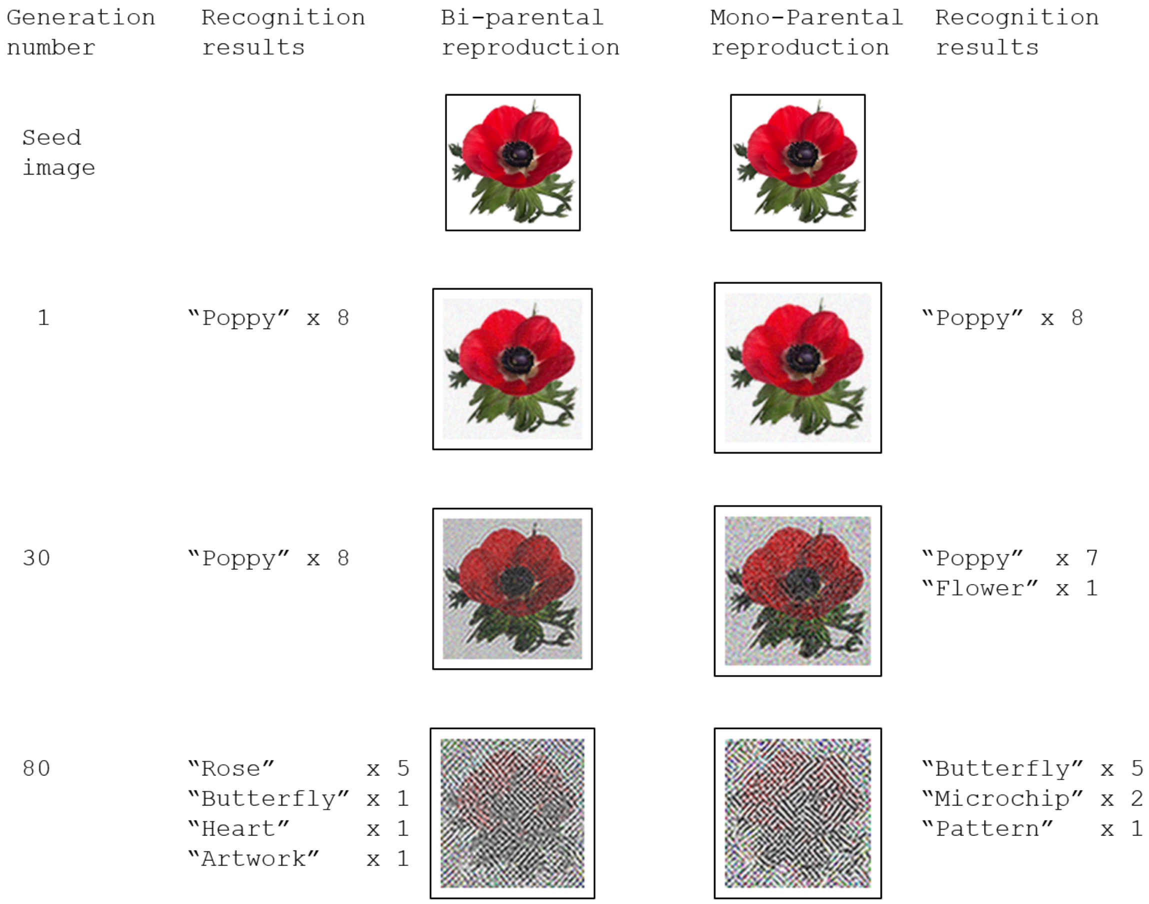


Figure 4: Simulating reproduction of an image of a flower. See explanation in text.

der MR, some intermediate generations showed a local advantage for MR. Such results are to be expected, first, due to the random nature of the inserted noise, and perhaps more so, due to the possibility that in a particular cohort, an individual may reproduce several times, dominating the properties in the next generation.

6 Conclusion and Next Steps

We have shown that mathematically, bi-parental, or sexual, reproduction can contribute to conservation in the offspring of traits that are common and important in sustainment of the parent generation. Confirming the observation in nature, and investigating it in more elaborate models, including multiple properties, multiple generations, and more detailed reflection of biological aspects, remains as future work.

Acknowledgements

We thank Netzer Moriya and Hagai Cohen for valuable discussions. This research was funded in part by an NSFC-ISF grant to DH, issued jointly by the National Natural Science Foundation of China (NSFC) and the Israel Science Foundation (ISF grant 3698/21). Additional support was provided by a research grant to DH from Louis J. Lavigne and Nancy Rothman, the Carter Chapman Shreve Family Foundation, Dr. and Mrs. Donald Rivin, and the Estate of Smigel Trust.

REFERENCES

Barton, N. H., Briggs, D., Eisen, J., Goldstein, D., and Patel, N. (2007). *Evolution*. Cold Spring Harbor Laboratory Press.

Bayer, Y., Harel, D., Marron, A., and Szekely, S. (2023). Toward automated modeling of abstract concepts and natural phenomena: Autoencoding straight lines. In *MODELSWARD*, pages 275–282.

Bernstein, H., Byers, G. S., and Michod, R. E. (1981). Evolution of sexual reproduction: importance of dna repair, complementation, and variation. *The American Naturalist*, 117(4):537–549.

Billiard, S., López-Villavicencio, M., Hood, M., and Giraud, T. (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of evolutionary biology*, 25(6):1020–1038.

Blitzstein, J. and Hwang, J. (2015). Texts in statistical science.

Cohen, I. R. and Marron, A. (2020). The evolution of universal adaptations of life is driven by universal properties of matter: energy, entropy, and interaction. *F1000Research*, 9.

Cohen, I. R. and Marron, A. (2023). Evolution is driven by natural autoencoding: reframing species, interaction codes, cooperation and sexual reproduction. *Proceedings of the Royal Society B*, 290(1994):20222409.

Cui, Y., Casella, G., and Wu, R. (2004). Mapping quantitative trait loci interactions from the maternal and offspring genomes. *Genetics*, 167(2):1017–1026.

Fisher, R. (1918). The correlation among relatives on the supposition of mendelian in heritance. *Trans. Royal Soc. of Edinburgh*, 52:399–433.

Fitzpatrick, B., Fordyce, J., and Gavrilets, S. (2008). What, if anything, is sympatric speciation? *Journal of evolutionary biology*, 21(6):1452–1459.

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, 27(3/4):310–332.

Heitman, J. (2015). Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biology Reviews*, 29(3-4):108–117.

Kang, M. S. (2020). *Quantitative genetics, genomics and plant breeding*. CABI.

Keeling, P. J. and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, 9(8):605–618.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, 336(6198):435–440.

Koonin, E. V. (2010). Taming of the shrewd: novel eukaryotic genes from rna viruses. *BMC biology*, 8:1–4.

Livnat, A. and Papadimitriou, C. H. (2016). Sex as an algorithm: the theory of evolution under the lens of computation. *Commun. ACM*, 59(11):84–93.

MacPherson, B., Scott, R., and Gras, R. (2021). Sex and recombination purge the genome of deleterious alleles: an individual based modeling approach. *Ecological Complexity*, 45:100910.

MacPherson, B., Scott, R., and Gras, R. (2023). Using individual-based modelling to investigate a pluralistic explanation for the prevalence of sexual reproduction in animal species. *Ecological Modelling*, 475:110191.

Marron, A., Szekely, S., Cohen, I. R., and Harel, D. (2024). Source code for simulation runs comparing biparental and monoparental reproduction. https://colab.research.google.com/drive/1dbbTXI7IjPercLNJcHFHo5IRLbUW63Wb?usp=sharing . Accessed 5/2023.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist*, 66(703):118–138.

Niu, H., Ge, Q., Shang, H., and Yuan, Y. (2022). Inheritance, qtls, and candidate genes of lint percentage in upland cotton. *Frontiers in Genetics*, 13:855574.

Skeat, T. (1999). The codex sinaiticus the codex vaticanus and constantine. *The journal of theological studies*, 50(2):583–625.

Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, (1):1–30.