scientific reports

Check for updates

OPEN

Natural averaging may complement known biological constraints in sexual reproduction's advantages over asexual in conserving species quantitative traits

Assaf Marron[1]✉, Smadar Szekely[1], Irun R. Cohen[2] & David Harel[1]

Commonly recognized effects of sexual reproduction include increased diversity, improved adaptability, enabling of DNA repair, constrained accumulation of deleterious mutations, and species genotype homogenization. Additionally, there are studies that show how sexual reproduction slows down certain evolutionary responses, offering advantages in population cumulative growth and stability over time and other metrics. Here, we contribute an observation of another distinct effect of sexual reproduction, focusing on retaining a species's key traits. In an initial mathematical analysis and simulation, we show that in an environment where copying is prone to error, quantitative polygenic traits that are shared within a parents' generation are transmitted to future generations under sexual reproduction with less deviation than under asexual reproduction. Furthermore, the model shows that this *retention of common traits* (abbr. RoCT), is driven by the very nature of mixing of parental traits, and occurs even before adding effects like trait-specific reproductive advantages, DNA repair, or the raising of reproductive barriers. Since survival of ecosystems depends on the ability of individuals to replace the networked interactions and interdependencies associated with failing, dying, or absent members of the same species, RoCT helps sustain species and ecosystems.

Keywords Sexual reproduction, Asexual reproduction, Clonal reproduction, Bi-parental, Probabilistic sampling

There are multiple theories for the evolution and pervasiveness of sexual reproduction in nature, as contrasted with asexual, clonal reproduction. However, the underlying question is still considered open, as can be attested by recent publications[1–4] (Macpherson[1] specifically lists many additional papers *"who refer to the unresolved paradox of sexual reproduction"*).

The various factors and effects, which in some cases appear as complementing or supporting each other, and in other cases as alternative observations and explanations, include:

1. The contribution of recombination to diversity, which then leads to increased adaptivity to changing conditions, enhanced competitiveness and the ability to co-evolve with pathogens, parasites, or symbionts. Seminal publications studying this issue, each with its finer claims, and environment and time scale considerations, include Fisher's 1930 *The Genetical Theory of Natural Selection* (Chapter VI)[5], van Valen's Red Queen hypothesis[6], Williams's *Sex and Evolution*[7], and Maynard Smith's *The Evolution of Sex*[8].
2. Constraining of the irreversible penetration of deleterious mutations, often termed Muller's ratchet[9]; this occurs by combination of dominant/recessive relations of pairs of alleles, and by the reproductive disadvantage imposed by the deleterious trait at hand, which may be augmented by Kondrashov's Hatchet[10] – interactions between slightly deleterious mutations; and,

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, 76100 Rehovot, Israel. [2]Department of Immunology and Regenerative Biology, Weizmann Institute of Science, 76100 Rehovot, Israel. ✉email: assaf.marron@weizmann.ac.il

3. Preservation of common traits through biological processes including (i) the raising of reproduction barriers between divergent subgroups within a species[11], (ii) DNA repair that occurs as part of meiosis[11–13], (iii) sexual selection which promotes reproduction of individuals with certain preferred traits[11], and (iv) interference with specialization and speciation of variants that find and prefer new niches[11](ch.16).
4. Genetic homogenization. Heng and Gorelick[13] observe that diversity, adaptivity, and constraining of deleterious mutations are generally considered the more prominent factors. They then argue that much greater significance should be assigned to the contribution of sexual reproduction to genotype homogenization, as resulting from mechanisms like DNA repair, and the preservation of structure throughout meiosis. These processes maintain the species identity, i.e., the functions that constitute the delicate interaction networks that sustain the species and its ecosystem. They write: *We do not quibble with Fisher's (1930) notion-which really dates back to Darwin (1859) and Weismann (1891)-that (additive) genetic variance is the engine driving evolutionary change. We only contest that sex helps generate that heritable variance, especially at the genome level*".

This view can be aligned with the observation in[14] that one of the effects of sexual reproduction is *"to maintain a frequency distribution of genes within the species that fits the life style and ecosystem*".

5. Stabilizing population growth rates and reducing fluctuations. Flegr and Ponížil[15] focus on evolution in fluctuating environments, say, when ambient temperature rises and falls. They use mathematical models and simulation to show that sex actually slows down evolution in a way that enables species to survive in such fluctuating environments. This angle aligns with Roughgarden's[16] focus on long term growth rate in fluctuating environments, using the geometric mean (*n*th root of product of *n* numbers) of consecutive growth rates. This computation shows advantage to sexual reproduction when combining multiple increases and reductions in growth rate caused by changing levels of success of adaptation.
6. Recombination as a maintainer of alternatives. Hamilton[17], observed that in an environment where a short-life-cycle parasite evolves to always attack the phenotype with highest frequency, sexual reproduction ensures that an optimal genotype will not immediately dominate; recombination maintains population diversity even without mutations. This enables evasion of the parasite's attacks, akin to password-based defense with *"continual random recreation of 'passwords' by sex and recombination*".

Doebeli[18] shows the advantages of sexual reproduction in reducing fluctuations in population density in host-parasite systems as well as under competition between individuals in the population and between sexuals and asexuals within the population. The focus is on quantitative phenotypic traits, which are manifested as fitness. Here again, it is assumed that mutations may be the source of original variability in the population, but no mutations occur in the time frame in which sexual and asexual reproduction are compared.

An additional window into how published opinions in this area may align with or differ from each other see Williams's review[19] of Maynard Smith's book[8].

Resolving the general paradox or weaving the various explanations into a cohesive theory is still an open research area. Still, such efforts may be aided by additional observations and insights into the effects of sexual reproduction on genotypes and phenotypes, and on species and ecosystems.

In this paper we add such an observation regarding the retention, in a child generation, of traits that are common in the parent generation; the effect is abbreviated hereafter as *Retention of Common Traits* (RoCT). Using generalized statistical analysis and simulation, we show that in a population where any kind of copying is imperfect, say, due to mutations, sexual reproduction that includes merging of two encodings (i.e., concrete representations) of inherited traits, conserves in the offspring generation traits that are shared within the parents' generation better than does asexual reproduction. Furthermore, we show that this effect is a general statistical property, driven by the very mixing of parental traits, even before adding the contribution of DNA repair, effects on reproductive rates, and the raising of reproductive barriers,

The model shows this for one transition between generations, when then extends to conservation of traits across multiple generations. Given that survival of organisms, species, and ecosystems depends on interaction networks, and that all organisms eventually die, higher fidelity RoCT helps sustain species by enabling offspring to step in and carry out interaction roles performed by their ancestors.

Certain aspects of the topic were discussed as far back as a century ago. For example, in 1918, Fisher[20] established the correlation between the phenotypic expression of quantitative traits (also termed *characters*) in offspring and that of the parents. In 1886, Galton[21] reported observations that for quantitative traits, offspring traits tended towards the mean of the population, eliminating extremes that may have been manifested by their parents, a phenomenon termed "regression towards the mean". In his experiment measuring human stature he also observed population-wide processes which he summarized as *"two opposite sets of actions, one concentrative and the other dispersive*", and concluded that *"they necessarily neutralise one another, and fall into a state of stable equilibrium*"[21](p.256).

Our model focuses on the effects of noisy copying and merging of encodings, before the application of repair mechanism and before considering the longer term effects of the change in trait on reproductive success or on the possibility of raising of reproductive barriers among sub-groups.

RoCT may be correlated with homogenization of offspring genotype, but these are distinct concepts. Indeed, the fact that sexual reproduction results in homogenization of the genome in the offspring generation 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 all are confined within a geographical region (see, e.g., the abstract of[22]); it is also often

implicit from discussions of gene flow in structured populations (see, e.g., ch.16, p.441 in[11]), or in pondering the puzzle of sympatric speciation: why panmictic populations sometimes give rise to diversity and speciation even within a given location[22,23]. However, mere homogenization within the offspring generation alone does not immediately imply retention of parental traits and may even dilute such traits.

As in the research by Doebeli[18], our analysis and simulation is constrained to quantitative traits, which by and large are polygenic. Quantitative traits constitute the majority of the identifiable traits in organisms[24](p.239), and our knowledge about how various genes affect each such phenotypic trait is very limited. Indeed, the very existence of quantitative phenotypic traits whose manifestation can take into account summing or averaging the traits of the two parents is an important factor that enables the retention of what is common or shared in the entire generation of these parents. This choice connects the present work more broadly to quantitative genetics[11] $(ch.14,16,17,19,26)$,[20,25]. While *blending inheritance* theory, where parents traits are mixed to form the offspring features, is long considered obsolete when discussing individual genes and alleles[26](p.52), quantitative genetics studies the many quantitative phenotypic traits which may manifest averaging effects. Such traits include lifespan, weight, strength, speed, dimensions of various organs, and many others. Quantitative traits 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 polygenic aggregation 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[27,28], and cannot yet be reduced to a small set of rules; the model makes a modest attempt to reflect this complexity.

The present work is part of a larger research project[29,30] exploring 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*. It is manifested in the genes, physiology and environment of the individuals, defines the species's identity, and enables its transmittal to future generations. Conservation of common traits thus help conserve the species. we use the term *an encoding* to refer to an object or structure that encodes, or represents, another object or concept. The terms "encoding" and "species interaction code" are distinguished from the concept of "a code" used by Barbieri and others to refer to an entire mapping between two domains[31].

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[32]; 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 use, in the context of the mathematical model, of terms that are commonly applied to living organisms, is for convenience, and alludes to the model's applicability to biology.

The paper is structured as follows: In Section 2 we introduce our model for genotype, phenotype and sexual and asexual transmittal of quantitative characters. In Section 3 we prove, using basic probability and the central limit theorem in the above model, the inevitability of conservation of quantitative parental traits as a general effect of sexual reproduction. In Section 4 we illustrate the differences between the RoCT effect under sexual and asexual reproduction with examples drawn from simulation runs that reproduce diverse objects including real numbers, images of printed text, and images of flowers. The latter simulations also illustrate how conservation of many distinct traits, like individual Red-Green-Blue (RGB) values in thousands of pixels, can affect conservation of emergent traits, like recognizing that an image is of a flower. In Section 5 we discuss the positioning and significance of the contribution relative to ideas and methodological approaches in published works, limitations of our model, and related future research directions.

The model

Below we describe the model and specific choices used in our analysis and simulation. Possible extensions are discussed in Section !5.

Population and trait representation

We begin with a population of entities, also referred to as individuals. The size of the population is kept fixed at a value n. We label the entities in a cohort as $\{e_1, e_2, ..., e_n\}$. The population reproduces in methods discussed in detail below; most basically, in every reproduction step, a set of n entities, referred to as the cohort of the next/offspring/child generation, appears instantly and synchronously.

The analysis does not include other considerations of time.

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.

All entities are associated with exactly one quantitative phenotypic trait P, which is manifested here as a real number. Different entities may have the same or different values for P. Let p denote any value of P, and, for example, p_i is the value of P in the entity e_i.

The genotype encoding of the quantitative phenotype trait p_i in each individual e_i is modeled as a set g_i of k real values which are also referred to as the points in g_i. The phenotype expression of the trait P in e_i is the mean of the points in g_i, i.e.,

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

These points represent the loci, or QTL, contributing to the manifestation of the phenotype value p_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. In the analysis, the integer k is assumed to be large, say, $k \geq 100$; we also constrain k to be an even number. We note that in general, encodings do not have to be compact or minimalistic, as they may involve redundancy to overcome decoding errors, or as the physical medium may dictate, for example, when capturing a single handwritten digit in an image with thousands of pixels, or as happens in nature, when a large number of DNA loci contribute to a phenotypic trait that is measured by humans as one number.

Reproduction

In this model, we use the terms sexual and asexual reproduction (abbreviated XR and AXR, respectively) as follows:

Sexual reproduction (XR) involves two parent entities of a single mating type (as appears, e.g., in fungi[33]); each parent contributes equal amounts of genetic material to the offspring; and, the copying and transmittal of genetic material is subject to error, thus modeling some of the effects of mutation.

Asexual reproduction (AXR) involves one parent and transmittal of the full genetic material, subject to the same kind of error as in XR. We do not use the term clonal here in order to emphasize that offspring are rarely, if ever, perfect clones of their parents.

Below we describe one reproduction step of an entire generation, termed *a generational transition*, in XR and AXR.

Asexual reproduction (AXR) generational transition

1. The input is a set c of the current population; 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 pre-existing genotype encoding of e_i. Compute the mean of g_i, which is equal to the phenotype of e_i:

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

   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. This step aims to model the creation, in a parent organism, of a single cell that gives rise to the child organism. See "Discussion" for discussion of other possible choices in modeling this process.

   (c) 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: discard the parent set c; rename the child set c' to c; discard c'.*Sexual reproduction (XR) generational transition*

1. As in AXR, the input is a set c of the current population; 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.
Create two interim 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; this represents the creation of gametes in our model; again, see Section 5 for discussion of other possible modeling choices.

   (c) Let g' be the union of these two sets, $g' = g'_u \cup g'_v$. 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. As in AXR, after completing n iterations: discard the parent set c; rename the child set c' to c; discard c'.

In the analysis below, in each cohort *c* we are interested in the mean p_c of the p_i values within the cohort: $p_c = \frac{1}{n}\sum_{i=1}^{n} p_i$, which represents a common trait, or capability, that is manifested with some diversity in the living individuals.

To emphasize our claim that certain mathematical effects of XR are intrinsic to the merging process, this model assumes also that (i) the rate of reproduction is constant and uniform in the population, and there are no selective pressures, competition and fitness considerations; (ii) copying errors are confined to trait values and not structural ones, and thus they do not prevent subsequent merging; and (iii) the contributions of the distinct points in the genotype g_i to the phenotype p_i are independent of each other. See Section 5 for discussion of modeling choices related to aspects of fitness and natural selection, DNA repair, epistasis, and more.

Model realization example

Demonstrating the model in a particular setting of an organism's quantitative trait and its QTL genetic representation is beyond the scope of this paper. Still, for further clarification, below we illustrate the above 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.

A numerical example

For illustration of the various calculations, below we go through the core elements of AXR and XR using a small-scale example.

1. **Parameter setting:** Assume that in the parents' cohort *c* , *P* is normally distributed in the parent generation with mean $p_c = 2.0$, and 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. **AXR 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 $e'_{1,\mathrm{AXR}}$) 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. **XR 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 $e'_{1,2,\mathrm{XR}}$) is $p' = 2.025$.

Comparing the two reproduction methods

In comparing AXR and XR in a particular generational 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 trait of a random individual in c', denoted p'_i, and the common trait p_c of the parents' cohort:

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

For a given parent cohort *c*, let d_{AXR} and d_{XR} be the expected values $E[d_i]$ under the AXR and XR processes, respectively, when considering all possible $c \to c'$ generational 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_{AXR} and σ^2_{XR} be the expected values $E[\sigma^2_{c'}]$, the variance of *P* within c' under the AXR and XR processes, respectively, when considering all possible $c \to c'$ generational 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, $d_{\mathrm{XR}} < d_{\mathrm{AXR}}$.

Proof A. Compute d_{AXR}.

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

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

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

Let e_i be the element of c that served as the single parent of $e'_{i_{\mathrm{AXR}}}$ in this AXR generational 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 $g'_{i_{\mathrm{AXR}}}$ of all possible direct AXRchildren 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[32](p.435), the means of these samples, namely $p'_{i_{\mathrm{AXR}}}$, 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_c^2, and (ii) based on the selected p_i and the noise parameter, randomly selecting the sample g' of size k and computing its mean $p'_{i_{\mathrm{AXR}}}$.

We can now examine the distribution of the random variable $p'_{i_{\mathrm{AXR}}}$ as drawn from the original distribution of c. According to the law of total variance (a.k.a. Eve's law[32](p.401)), this is a normal distribution with mean p_c and variance that is the sum of the variances, hence

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

In general, the average absolute deviation of a random sample from the mean of a normal distribution with standard deviation σ is $\sigma \cdot \sqrt{\frac{2}{\pi}}$ [34]. Therefore, the expected value of the absolute distance of $p'_{i_{\mathrm{AXR}}}$ from p_c is

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

B. Compute d_{XR}.

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

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

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

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

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

The genotype encoding $g'_{j_{\mathrm{XR}}}$ 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 $p'_{j_{\mathrm{XR}}}$ 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:

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

Let us now analyze the random selection of $p'_{j_{\mathrm{XR}}}$ 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 that is the sum $\sigma_c^2 + \frac{h^2}{\frac{k}{2}}$.

Clearly the mean of the random variable $p'_{j_{\mathrm{XR}}}$ 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[32](p.159):

$$\sigma^2_{\mathrm{XR}} = \frac{1}{4} \cdot 2(\sigma_c^2 + \frac{h^2}{\frac{k}{2}}) = \frac{\sigma_c^2}{2} + \frac{h^2}{k} \ .$$

As before, the average absolute deviation is given by

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

C. Comparing d_{XR} and d_{AXR}.

The expressions for d_{AXR} and d_{XR} are the same except for the first term in d_{XR} being $\frac{\sigma_c^2}{2}$ where in d_{AXR} it is σ_c^2. Hence,

$$d_{\mathrm{XR}} < d_{\mathrm{AXR}} \ .$$

. □

Note that an intermediate result in the above calculations is that $\sigma^2_{\mathrm{XR}} < \sigma^2_{\mathrm{AXR}}$. 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.

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. For each simulation case 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. The program source for the simulations is available in[35].

For simplicity, we equate the essence of each individual with the trait at hand. For example, we say that the runs show reproduction of a population of real numbers, or a population of images, as opposed to reproduction of a population of individuals whose phenotype trait is expressed as a real number or as an image, respectively.

Reproducing a population of real numbers

The first set of simulations follows directly the model described in "The model", demonstrating the reproduction of a population of real numbers through intermediate encodings as sets of points. As detailed in Fig. 1, these simulations show that under XR, the traits of the child generation, both as individuals and as a cohort, are closer to the traits of their direct parents' and of their earlier ancestors, than under AXR.

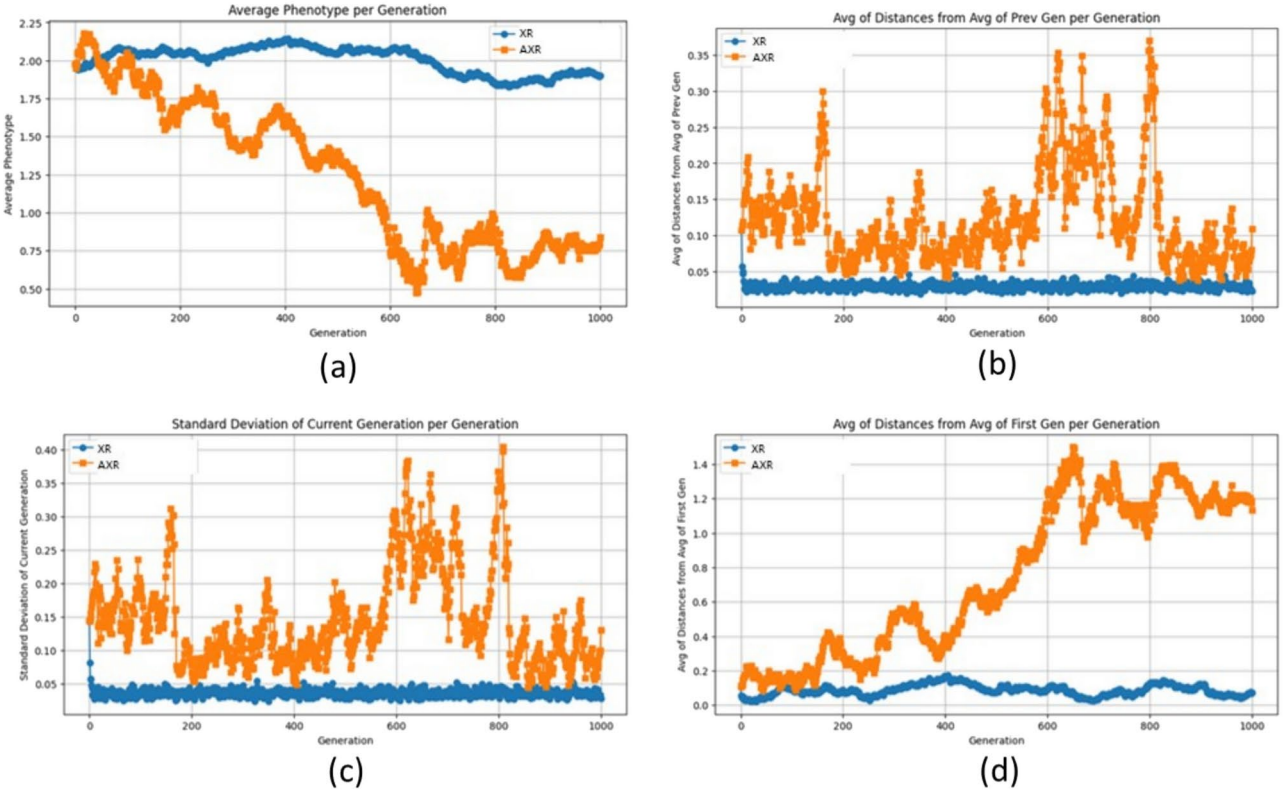

Fig. 1. Reproduction of a population of real numbers. Orange lines with square marks: Sexual reproduction (XR); Blue lines with circle marks: Asexual reproduction (AXR). Simulation parameters: Cohort size=32; Encoding set size=30; Number of generations = 1000; Initial population phenotype trait mean ≈ 2.0; Initial population standard deviation=0.2; Reproduction error standard deviation=0.15. Results: In XR vs. AXR, (**a**) The average of the phenotypes of the entire cohort is closer to the original common phenotype; (**b**) on average, phenotypes of individuals of each generation, are closer to the average of the immediately preceding generation; (**c**) the standard deviation within each generation cohort is smaller; (d) On average, the phenotypes of individuals of each generation, are closer to the mean of the original ancestral generation.

Reproducing populations of 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/asexual" and "bi-parental/sexual" techniques were employed by scribes: most often copying from a single source, but sometimes collating multiple sources[36](p.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 traits 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 the importance of traits in an organism's interactions; in the present case, the interaction 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 copying 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 all copies, and hence is less significant for our purposes.

Function F2. Noisy computation of a phenotype. Given k 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, and this final copying may be imperfect as well.

The parameters used in the simulation are listed in the image caption. The simulation steps are as follows:

1. Start with a seed image.
2. [Creating base cohort.] Repeat the following $n = 8$ 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 XR process and generation 1 of the AXR process.
4. [XR process.] Repeat the following for 90 generations
   (a) [Create next generation.] Repeat $n = 8$ 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. [AXR 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 XR.

Figure 2 depicts a sample of the resulting process and corresponding image content assessment. Clearly, fidelity of image properties is preserved better by XR than by AXR, both according to image-recognition software and to casual human observers.

Additionally, as shown in Figure 3, in child generations, under XR, human observers could recognize conservation of both ancestral and newly emergent properties, like foreground-background contrast and lines and circles that did not exist in the original image; under AXR, we could not readily find any such effects.

Reproducing populations of 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 reproducing images of a flower under XR and under AXR. The simulation uses the same process as in the case of text image, with the following parameters: The starting image was one labeled in its repository as an anemone flower. The image recognition prompt is: "Your role is to identify the object in the image. Please provide a 1-word answer.". The simulation showed that

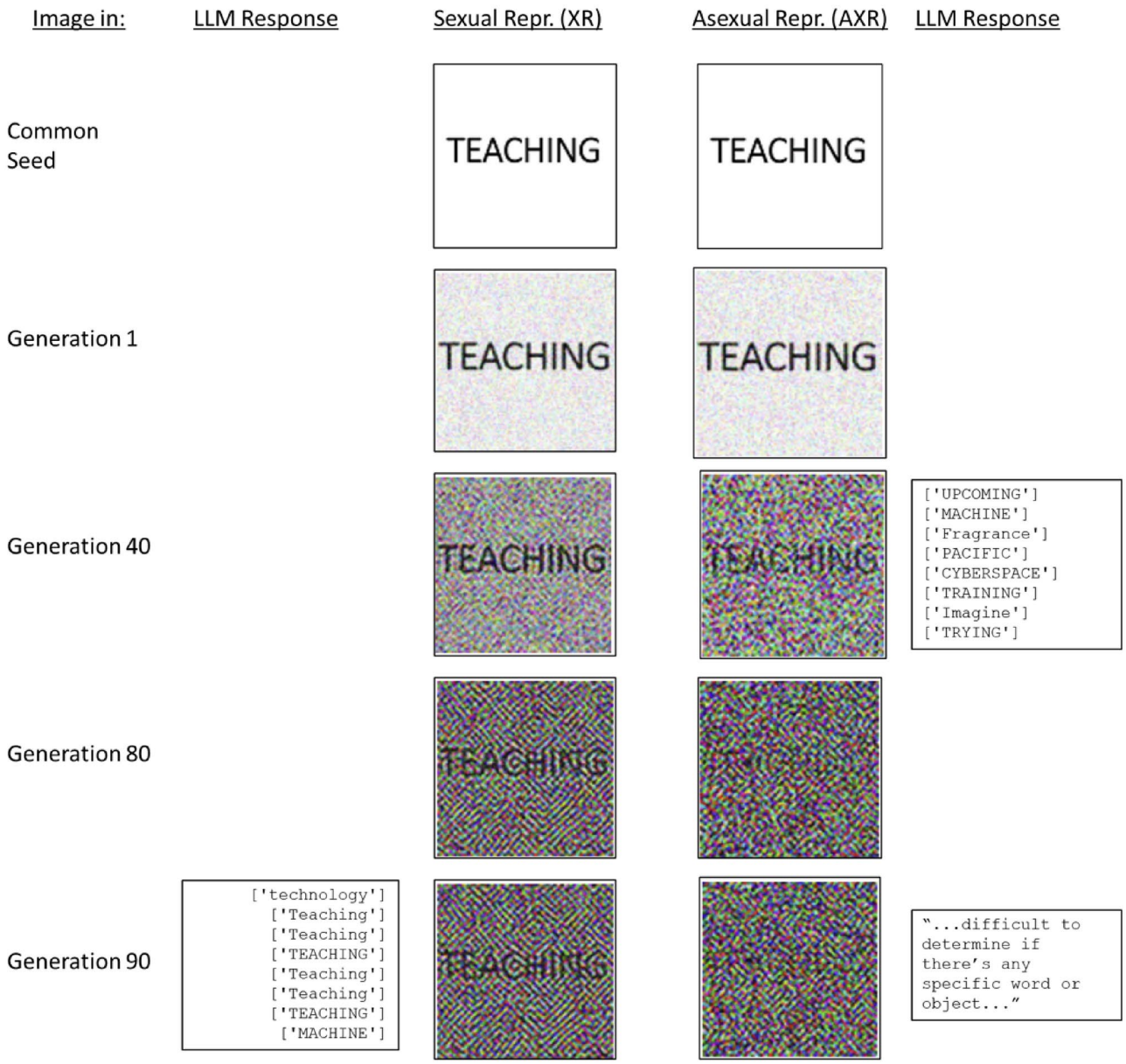

Fig. 2. Simulating reproduction of an image of printed text using sexual and asexual reproduction, alongside corresponding image recognition results using ChatGPT. Showing one image from each generation, for generations 0, 1, 40, 80 and 90. Results: (i) Under XR, the LLM identified the word "TEACHING" throughout more than 70 generations, where under AXR the identification of the word failed completely on or before generation 40. (ii) Under XR, in the 90th generation, the LLM identified that the image is of text, and had some success in identifying the original word, while under AXR, the LLM did not identify any text. Simulation parameters: Text = the word "TEACHING"; Font = Calibri Light; Color = black and white only; Resolution of original image = 100 × 100 pixels; Cohort size = 8; Number of copies in each image "genotype" encoding = 8; Resolution of image copies that constitute the encoding = 400 × 400; Noise standard deviation = 20 (within pixel RGB values of 0-255); ChatGPT Model = gpt-4-turbo.

recognition quality was maintained for more generations under XR than under AXR. We ignore the fact that the LLM recognized the flower as a poppy rather than as an anemone.

Note: The images shown here are illustrative examples. While in all runs the image traits were sustained for longer under XR than under AXR, some intermediate generations showed a local advantage for AXR. 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 traits in the next generation.

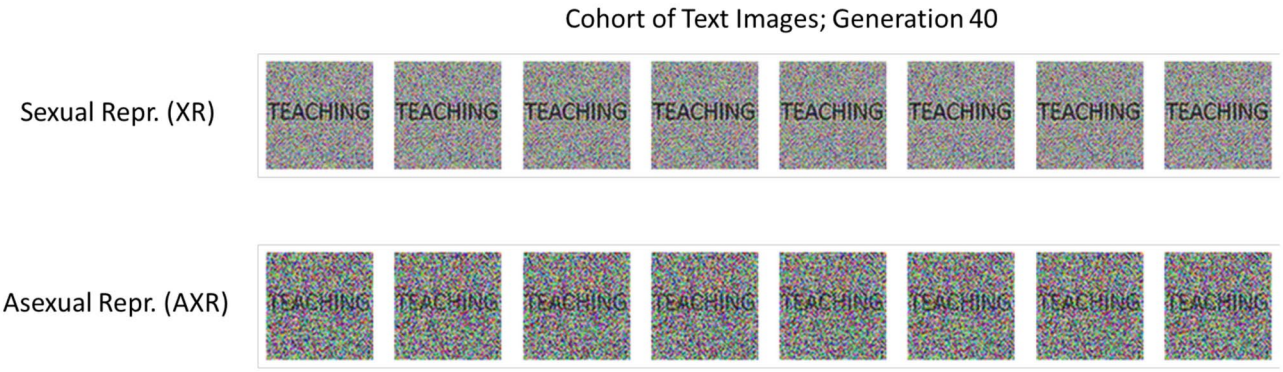

Fig. 3. Ancestral and emergent properties across the 8 individuals of generation 40 from the process depicted in Fig. 2, under sexual and asexual reproduction. Human observation and assessment: (i) Under XR, ancestral traits like the contrast between the letters and their background, or the horizontal line of the letter T, appear better preserved than under AXR. (ii) Under XR, conspicuous and sustained emergent traits appear; for example, a light circle shape inside the letter C, or a third dark vertical line in the letter H. Under AXR we could not readily find any such emergent traits.

Discussion

Our observation that sexual reproduction may contribute to conserving polygenic quantitative traits that are common to a cohort of individuals thus joins existing empirical observations, models, and theories regarding the effects of sexual reproduction, as compared with asexual reproduction.

The analysis in the theorem and proof above differs from other cited works mainly in that it provides a computational model that draws attention to what happens in a single generational transition, before adding consideration of competition, fitness and natural selection, and before the environment changes perceptibly. It also isolates a preservation mechanism that is separate, and in addition to, existing structure conserving mechanisms in meiosis. Another difference from many cited works is the focus in traits that are polygenic and quantitative.

With regard to reproductive success, and success in general, we note that many published studies focus on whether a particular mutation offers a competitive advantage or is deleterious. By contrast, here, we are focused on the fact that existing species' traits are likely to be important in sustaining the species in its current environment.

This one contribution does not aim to answer what are considered open questions in the evolution of sex. Furthermore, quantitative assessment of diversity, homogeneity, preservation of traits, and, of course, survival, depends on what exactly is measured in a particular model, and multiple opposing effects may coexist as suggested by Galton[21]. In particular, diversity and RoCT may co-exist in orthogonal dimensions, or may be nested within one another. For example, a particular common trait may be conserved across the population, while its combination with other traits may vary. Or, the small variations in a retained trait may be aggregated in ways that refine the trait (and its combination with other traits) into diverse clusters.

While some of our assumptions aim to simplify the context in order to enable detailed analysis, we use modeling principles similar to those in published works.

One can draw support for the claims that XR yields RoCT under our model, from an analogy to the mathematical concept of random walks. Consider the combination of two one-dimensional random walks, with the same step size and equal probabilities for stepping in either direction. When two such random walks, starting at different origins, are combined by stepping in parallel and averaging the location, the resulting mean position is the average of the two origins. More importantly, the position variance is half the (equal) variance of the two original random walks. This is also aligned with the known fact that sexual reproduction reduces the variance within the offspring generation.

The significance of recognizing the retention of common species traits as one of the effects of sexual reproduction is twofold: First, it flows from the above discussion and references that it might give us additional angles for understanding of the evolution of sex. Second, since species' traits are manifested by interactions (within organisms, between organisms, or between organisms and the environment), RoCT aligns with the idea expressed in[29,30], that networked ecosystems are sustained by their ongoing interactions and that sexual reproduction helps new generations continue the interactions of earlier ones. More generally, one may infer that sustaining fragile ongoing networked interactions may be no less important than continuous competitive adaptation and innovation, with implications to human well being, social behavior, and interaction with the environment.

Model limitations and future research

While highlighting our main message, the above initial model is indeed initial, and can be extended in future research in several directions, implied by choices of model definitions described above. Some key directions include:

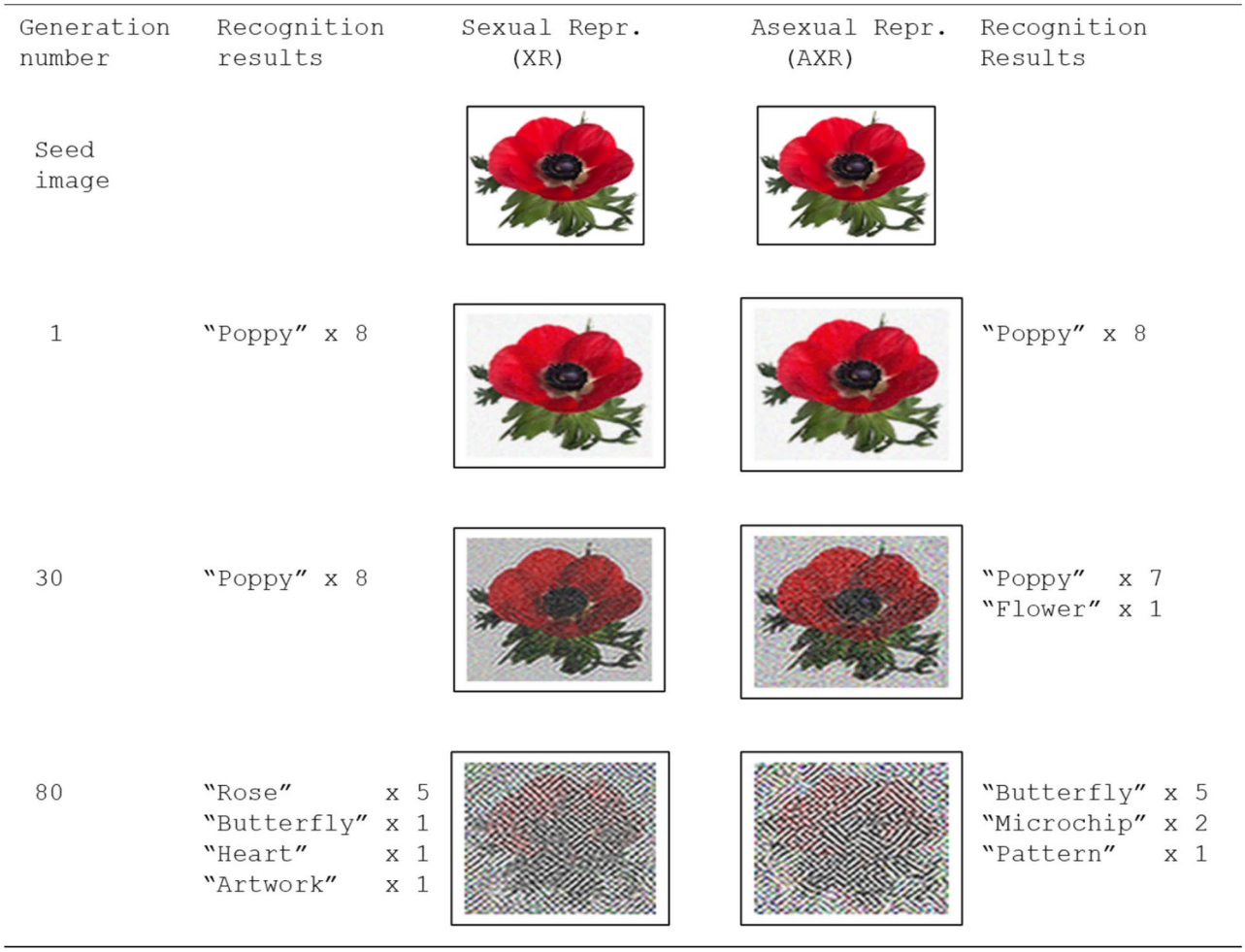

Fig. 4. Simulating reproduction of an image of a flower. Showing one sample image from each generation, for generations 0, 1, 30, and 80. Results: (i) Under XR, images were all assigned the same label as the first one, "Poppy", for more generations than under AXR. (ii) Under XR, certain core traits, like being identified as a flower (the "Rose" attempt), were better preserved in generation 80 than under AXR. Starting image was downloaded from (https://it.pinterest.com/pin/red-anemones--797207571571931709/), under fair use license; the image was cropped and resized to 100x100 pixels. The simulation uses the same overall process as in the case of images of text. Simulation parameters: Noise standard deviation = 10 (within pixel RGB values of 0-255); Cohort size = 8; Encoding set size = 8; Number of generations = 80.

1. Adding richer composition methods of loci/QTL contributions to the quantitative traits; similarly, add considerations of Mendelian traits, or ones where very few genes are involved.
2. Modeling with multiple species and environments, multiple traits per species, activating the conserved traits in the respective mutual interaction networks, and demonstrating how the ongoing interactions enabled by these traits contribute to sustainment of both the species and the networks. In such a model it is also possible to add the effects of competition and variation in reproduction rates.
3. For greater precision or composition with other aspects of sexual reproduction, one may add considerations of mating types, DNA repair, etc.

Conclusion

Reproduction inevitably produces change, causing offspring to differ from their parents to various degrees. Scientific discussions of sexual and asexual reproduction mechanisms in nature explore their operational differences, and their range of effects on the species, most prominently on adaptability and sustainment of the species over extended periods of time. The various arguments for the pervasiveness of sexual reproduction are yet to be integrated into a single cohesive theory that includes also interplay and relative priority among the different observed or modeled effects. We have contributed to these discussions an initial mathematical model, including a theorem and a proof, that compares aspects of natural sexual and asexual reproduction, and shows that merging and averaging in sexual reproduction can contribute to longer and more faithful retention, in offspring generations, of quantitative traits shared by the parent generation. The effect can be seen even in one generation, and complements other features of sexual reproduction that are known to conserve species's

traits. While the model can still be extended in various ways, it can already add to our understanding of the preservation of complex interactions networks in natural ecosystems that depend on constant turnover of the participating individuals. This, in turn, can carry implications to understanding sustainment of interactions among humans and between humans and the environment.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 14 November 2024; Accepted: 16 April 2025
Published online: 25 April 2025

References

1. MacPherson, B., Scott, R. & Gras, R. Using individual-based modelling to investigate a pluralistic explanation for the prevalence of sexual reproduction in animal species. *Ecol. Model.* **475**, 110191 (2023).
2. MacPherson, B., Scott, R. & Gras, R. Sex and recombination purge the genome of deleterious alleles: an individual based modeling approach. *Ecol. Model.* **45**, 100910 (2021).
3. Livnat, A. & Papadimitriou, C. H. Sex as an algorithm: the theory of evolution under the lens of computation. *Commun. ACM* **59**, 84–93 (2016).
4. Billiard, S., López-Villavicencio, M., Hood, M. & Giraud, T. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *J. Evolut. Biol.* **25**, 1020–1038 (2012).
5. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford University Press, 1930).
6. Van Valen, L. A new evolutionary law. *Evolutionary Theory* 1–30 (1973).
7. Williams, G. C. *Sex and Evolution* (Princeton University Press, 1975).
8. Smith, J. M. & Maynard-Smith, J. *The Evolution of Sex*. Vol. 4 (Cambridge University Press, 1978).
9. Muller, H. J. Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
10. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
11. Barton, N. H., Briggs, D., Eisen, J., Goldstein, D. & Patel, N. *Evolution* (Cold Spring Harbor Laboratory Press, 2007).
12. Bernstein, H., Byers, G. S. & Michod, R. E. Evolution of sexual reproduction: Importance of dna repair, complementation, and variation. *Am. Nat.* **117**, 537–549 (1981).
13. Gorelick, R. & Heng, H. H. Sex reduces genetic variation: A multidisciplinary review. *Evolution* **65**, 1088–1098 (2011).
14. Cohen, I. R. Updating Darwin: Information and entropy drive the evolution of life. *F1000Research* **5** (2016).
15. Flegr, J. & Ponížil, P. On the importance of being stable: evolutionarily frozen species can win in fluctuating environments. *Biol. J. Linn. Soc.* **125**, 210–220 (2018).
16. Roughgarden, J. The evolution of sex. *Am. Nat.* **138**, 934–953 (1991).
17. Hamilton, W. D. Sex versus non-sex versus parasite. *Oikos* 282–290 (1980).
18. Doebeli, M. Quantitative genetics and population dynamics. *Evolution* **50**, 532–546 (1996).
19. Williams, G. C. *Mysteries of Sex and Recombination* (1978).
20. Fisher, R. The correlation among relatives on the supposition of mendelian in heritance. *rans. R. Soc. Edinb* **52**, 399–433 (1918).
21. Galton, F. Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. Great Britain Ireland* **15**, 246–263 (1886).
22. Fitzpatrick, B., Fordyce, J. & Gavrilets, S. What, if anything, is sympatric speciation? *J. Evolut. Biol.* **21**, 1452–1459 (2008).
23. Mayr, E. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Harvard University Press, 1982).
24. Neal, D. *Introduction to Population Biology* (Cambridge University Press, 2018).
25. Kang, M. S. *Quantitative Genetics, Genomics and Plant Breeding* (CABI, 2020).
26. Dobzhansky, T. *Genetics and the Origin of Species* (Cambridge University Press, 1951).
27. Cui, Y., Casella, G. & Wu, R. Mapping quantitative trait loci interactions from the maternal and offspring genomes. *Genetics* **167**, 1017–1026 (2004).
28. Niu, H., Ge, Q., Shang, H. & Yuan, Y. Inheritance, qtls, and candidate genes of lint percentage in upland cotton. *Front. Genet.* **13**, 855574 (2022).
29. Cohen, I. R. & Marron, A. Evolution is driven by natural autoencoding: reframing species, interaction codes, cooperation and sexual reproduction. *Proc. R. Soc. B* **290**, 20222409 (2023).
30. Cohen, I. R. & Marron, A. The evolution of universal adaptations of life is driven by universal properties of matter: Energy, entropy, and interaction. *F1000Research* **9** (2020).
31. Barbieri, M. *Code Biology: A New Science of Life* (Springer, 2015).
32. Blitzstein, J. & Hwang, J. *Texts in Statistical Science* (2015).
33. Heitman, J. Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biol. Rev.* **29**, 108–117 (2015).
34. Geary, R. C. The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika* **27**, 310–332 (1935).
35. Marron, A., Szekely, S., Cohen, I. R. & Harel, D. *Source Code for Simulation Runs Comparing Biparental and Monoparental Reproduction* (2024). https://colab.research.google.com/drive/1dbbTXI7IjPercLNJcHFHo5IRLbUW63Wb?usp=sharing.
36. Skeat, T. The codex sinaiticus the codex vaticanus and constantine. *J. Theol. Stud.* **50**, 583–625 (1999).

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.

Author contributions

A.M. initiated, designed and led this work, carried out the mathematical analysis, developed a preliminary version of the simulation software, analyzed the results, and wrote the manuscript; S.S. contributed to the design and analysis of the simulation experiments, and developed the software; I.R.C. conceived ideas underlying this research direction; I.R.C. and A.M. developed the encoding-based research focus; D.H. contributed to the underlying research and the drafting of the manuscript; the research was conducted and funded within the group

of D.H.; all authors discussed the results and commented on the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.