Studies in History and Philosophy of Biological and Biomedical Sciences 58 (2016) 107–116

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Contingency, convergence and hyper-astronomical numbers in biological evolution

Ard A. Louis

Rudolph Peierls Centre for Theoretical Physics, Univeristy of Oxford, 1 Keble Road, Ox1 3NP, United Kingdom

ARTICLE INFO

Article history:
Available online 8 February 2016

Keywords:
Genotype–phenotype maps
Bias in evolution
Arrival of variation
Extended evolutionary synthesis

ABSTRACT

Counterfactual questions such as “what would happen if you re-run the tape of life?” turn on the nature of the landscape of biological possibilities. Since the number of potential sequences that store genetic information grows exponentially with length, genetic possibility spaces can be so unimaginably vast that commentators frequently reach of hyper-astronomical metaphors that compare their size to that of the universe. Re-run the tape of life and the likelihood of encountering the same sequences in such hyper-astronomically large spaces is infinitesimally small, suggesting that evolutionary outcomes are highly contingent. On the other hand, the wide-spread occurrence of evolutionary convergence implies that similar phenotypes can be found again with relative ease. How can this be? Part of the solution to this conundrum must lie in the manner that genotypes map to phenotypes. By studying simple genotype–phenotype maps, where the counterfactual space of all possible phenotypes can be enumerated, it is shown that strong bias in the arrival of variation may explain why certain phenotypes are (repeatedly) observed in nature, while others never appear. This biased variation provides a non-selective cause for certain types of convergence. It illustrates how the role of randomness and contingency may differ significantly between genetic and phenotype spaces.

1. Introduction

In Charles Darwin’s theory of evolution, random variation provides the raw material for natural selection (Darwin (1989)). Variation proposes and natural selection disposes. But where does the variation come from? The modern evolutionary synthesis (Huxley, 1942) incorporated discrete Mendelian genetics into evolutionary theory, showing that the variation arises from mutations and recombinations that change and rearrange genes, leading to different phenotypes. Early reflections on the exponentially vast numbers of potential gene combinations suggested that variation is plentiful and probably isotropic, so that natural selection is the primary cause of evolutionary change. The size of these hyperspaces also naturally suggests that evolution is contingent because life can only explore a vanishingly small fraction of all genetic possibilities. On the other hand, the increasing evidence for widespread evolutionary convergence (Conway Morris, 2003, 2015; McGhee, 2011) suggests that nature can find the same solutions again and again. If the number of genetic possibilities is so vast, how can history appear to repeat itself?

It is important to remember that random mutations happen at the level of genotypes, while selection happens at the level of phenotypes which describe the different characteristics of organisms. Thus a full understanding of how evolution progresses needs to include a description of how genotypes map onto phenotypes. In this contribution I explore some recent theoretical studies of genotype to phenotype maps that may shed new light on the role of randomness and contingency in evolution. Even though mutations are fundamentally random with respect to outcomes, these studies suggest that there may nevertheless be strong bias in the kind of

E-mail address: ard.louis@physics.ox.ac.uk.

http://dx.doi.org/10.1016/j.shpsc.2015.12.014
1369-8486/

variation that arises. Certain types of variation are much more likely to arise than others, which may influence evolutionary outcomes.

But before proceeding, it may be helpful to reflect on the fundamental reason why the space of possible genotypes, or for that matter almost any combinatorial problem, grows so rapidly.

1.1. The power of exponentiating

Perhaps the earliest known reference to the power of exponential growth comes from an Old Babylonian tablet which dates from 1800–1750 BC. Soubeyran (1984). It describes the doubling of the number of barley corns for thirty consecutive days, ending up with no less than two "thousand," seven "hundred" and thirty-seven talents, half a mina, two and one-third shekels, and four barley-corns, which comes to about 47 metric tons of grain.[1] Such doubling problems were a popular subset of the so-called Silk Road problems (Friberg, 2005). They often took the form of 30 doublings, as in the tablet above, or else 64 doublings. A famous example of the latter comes from the epic poem *Shah-nama* (The Book of Kings) written by the Persian poet Firdausi around the turn of first millennium. It tells of the mythical Indian inventor of chess. Apparently the king was so pleased that he told the sage he could name his reward. The sage then asked to be given one barley corn for the first square of his board, two on the next, four on the next, and so on. While this at first seems like a very modest request, it would have totaled $2^{64}-1 = 18{,}446{,}744{,}073{,}709{,}551{,}615$ grains, weighing about 1,000 times the world's current annual wheat production. The shift from 30 to 64 in these silk-road doubling problems illustrates how quickly exponential growth leads to unearthly large numbers.

In the next section, I will explore how the power of exponential growth may have influenced some of the founding fathers of the neo-Darwinian Modern Synthesis (MS).

1.2. Hyper-astronomical numbers in the modern synthesis

Although Gregor Mendel and Charles Darwin were contemporaries, Darwin remained unaware of the far-reaching implications of Mendel's experiments in genetics for his theory of evolution by natural selection.[2] When Mendelian genetics was rediscovered in the late nineteenth and early twentieth centuries, it appeared to raise a significant objection to biometric formulations of Darwinian evolution: How can changes in discrete genes lead to the small continuous changes in phenotypic variation that Darwinian natural selection was said to act on?[3] This apparent conflict was solved between 1918 and the early 1930s by a triumvirate of great mathematical biologists—R. A. Fisher, J. B. S. Haldane, and Sewall Wright—who showed that if the traits of an organism are affected by many genetic loci, then, by the laws of statistics, many separate discrete changes will translate into effectively continuous variation, as required.

This early work on population genetics helped give birth to the fully fledged MS, also known as the neo-Darwinian synthesis. The term was coined by Julian Huxley with his book *Evolution, The Modern Synthesis* (1942) and other major figures in the movement included the aforementioned trio of Fisher, Haldane, and Wright, as well as the paleontologist George Gaylord Simpson, the ecologist E. B. Ford, the geneticist Theodosius Dobzhansky, and the evolutionary generalist Ernst Mayr who did much to cement the way the history of the MS is recounted.[4]

An important early set of arguments that fed into the MS can be derived from the geometry of discrete genetic spaces. Once you have many genes, it is natural to ask how many ways you can arrange them. Such thought experiments quickly lead to exponential growth and comparisons to the size of our universe, as can be seen in the following piece from Sewall Wright's hugely influential 1932 paper:

> "*Estimates of the total number of genes in the cells of higher organisms range from 1000 up…. With 10 allelomorphs in each of 1000 loci, the number of possible combinations is 10^{1000} which is a very large number. It has been estimated that the total number of electrons and protons in the whole visible universe is much less than 10^{100}.*" (Wright, 1932, pp. 356)

Stuart Kauffman (1995, pp. 167) has described such numbers as *hyper-astronomical* because they are beyond even the kinds of gigantic numbers that are used in astronomy. Of course such large numbers also easily obtain in all kinds of other combinatorial problems in physical sciences and engineering.[5]

Reflections on the hyper-astronomical size of these spaces may have influenced the further development of the MS. Here is Wright again in the same article: "*The population is thus confined to an infinitesimal portion of the field of possible gene combinations*" (Wright, 1932, pp. 356). These spaces are so (exponentially) vast that even over 3.8 billion years, life won't explore much more than a tiny fraction of all possible genetic combinations. Thus it naturally follows nature can only explore an unimaginably small fraction of all theoretically possible genomes.

What is not known, of course, is what fraction of that theoretical space of all genotypes can generate viable phenotypes. Given the fact that the majority of mutations appear to be deleterious, perhaps it is only an extremely small fraction. The rest of the space may be biologically sterile, "*… the howling wildernesses of the maladaptive, the 99.9% recurring of biological space where things don't work, the Empty Quarters of biological non-existence.*" (Conway Morris (2003, pp. 309))

In Fig. 1 three schematics of the way the viable genotypes could be distributed are depicted. In each case only a small fraction of genotypes is viable. This immediately raises the question: how did life ever find the first viable genotypes? Given that we have evidence of life just a few hundred million years after the end of the violent "late heavy bombardment" of meteorites that characterized the birth of our planet, it would seem that life can find this viable region fairly quickly. With regard to Fig. 1, such an argument might favor the middle panel with a larger initial target, or the right panel with many small targets spread across the space.

Wright also pointed out that any genotype would have an exponentially large number of neighbors. Since he is also known for his advocacy of the role of genetic drift, random non-adaptive changes in genomes (Crow (2010)), he likely assumed that even if only a very small fraction of genetic possibilities are viable, the total space of potentially fruitful gene combinations remains unimaginably vast. Thus what we observe today in nature is only a small

---

[1] If one assumes that one barley corn, an ancient but tiny measure of weight, is about 0.05 g, then the total on day 30 is about 47 tons of barley. Interestingly, the scribe got his sums wrong, due perhaps to the complexity of Mesopotamian number systems (Soubeyran, 1984).

[2] As usual the history is more complex. There is evidence that Darwin had at least some indirect interaction with Mendel's work (see e.g. Sclater, 2006). Be that as it may, for all practical purposes, Mendel was forgotten.

[3] "*Natural selection can act only by the preservation and accumulation of infinitesimally small inherited modifications*" (Darwin, 1859, pp. 95).

[4] See Amundson (2007) for an opinionated counterpoint to Mayr's telling of the history of the MS.

[5] Another subfield of biology where the qualifier hyper-astronomical is frequently used is the number of possible connections in the brain (e.g. Edelman, 2001, pp. 38).