



CONSTRUCTAL EVOLUTION OF GENETIC CODING SYSTEMS

ALBERT T. MAGNELL

Department of Economics, University of Oxford
albert.magnell@new.ox.ac.uk

This paper is a theoretical exploration of how biological evolution may evolve itself according to the Constructal Law. The astonishing diversity of life on Earth reflects how effective evolution is at augmenting survivability within various ecological niches. Yet, the genetic coding system that drives evolution is remarkably conserved among virtually all organisms. This is perhaps indicative that the process of evolution on Earth has so far been powerful enough to meet the planet's ecological challenges. If these challenges become increasingly intense or unstable, then for life to survive, its evolvability must evolve. Such second-order evolution may occur via fundamental changes to the genetic coding system.

Keywords: Constructal Law; Econobiology; Evolution of evolvability; Xenobiology.

1. INTRODUCTION

Virtually all life on Earth uses a remarkably conserved format for encoding genetic information. Genomes consist of strings of DNA nucleotides organized as triplets (codons). Since there are four different types of DNA nucleobases (adenine, guanine, thymine, cytosine), there are $4^3 = 64$ possible types of codons, which map onto the 20 types of canonical amino acids that serve as the building blocks of proteins. From a theoretical perspective, the combinatorics at play are ostensibly arbitrary. Why not have codons that are quadruplets or additional types of nucleobases?

When comparing different coding formats, it becomes apparent that there are functional tradeoffs. A format with high combinatorial diversity can encode genetic information with greater density and potentially more resilience to read/write errors. However, such an elaborate system will pose a more significant energetic burden to the organism. These advantages and disadvantages consequently serve as countervailing forces that delineate a set of quasi-optimal genetic coding formats. Framing this puzzle in terms of tradeoffs extends it beyond the purview of theoretical biology into the realm of economic theory. Since economics is fundamentally about how to optimally deal with scarcity, it is central to the origins of life and evolution. Perhaps surprisingly, though, a detailed framework of economic biology, or econobiology, has yet to be constructed.

In the context of the evolution of life, the genetic coding system is the configuration of coding format and molecular machinery that, together, control evolvability. Evolvability is critical for survival because it gives life greater access to the genetic sequence space encoding fitness adaptations [1,2]. Treating evolution as if it were a problem-solving process used by life to expand into new ecological niches or resist ecological shocks means that we can judge the effectiveness of a genetic coding system in terms of its evolvability. A good system has high evolvability, allowing the life that uses the system to adapt both resiliently and quickly [3]. When considering how a genetic coding system might evolve, a bioevolutionary form of the Constructal Law becomes a guiding principle: For life to persist in time (to thrive), its genetic coding system must evolve in such a way that provides easier evolutionary access to pathways of adaptability [4].

2. CONCEPTUAL MODEL

This model of the potential evolvability of a genetic coding system has two degrees of freedom: the number of nucleobase types (N) and the codon length (L). Although this simplification precludes modeling even mildly alien genetic coding systems, it preserves much-needed tractability. The capacity of evolvability (E) is undergirded by a tradeoff between the benefits (B) and costs (C) of having a more complex genetic coding system. The two main benefits of having a more complex genetic coding format are a higher capacity for diversity (D) and enhanced coding fidelity (F). These come at the cost of a more significant energetic burden (J) and a longer time (T) for the system to operate. We, thus, represent B as a function of D and F , and C as a function of J and T ,

$$E \sim B(D,F) - C(J,T). \quad (1)$$

Moreover, both B and C can be thought of as production functions. D and F are inputs for producing benefits, while J and T are inputs for ‘producing’ costs. Since their inputs are neither perfect substitutes nor complements, both of these production functions can be represented in standard Cobb-Douglas form,

$$E \sim XD^\alpha F^\beta - YJ^\gamma T^\delta. \quad (2)$$

X and Y are the total factors of production for B and C , respectively; however, it may be more helpful to think of them as shadow prices. The four exponents are output elasticities and determine whether B and C exhibit increasing, decreasing, or constant returns to scale in regard to their inputs. D , F , J , and T are all functions of the model’s degrees of freedom, N and L .

3. QUALITATIVE SIMULATIONS

The diminishing returns to F concerning N and L potentially translate into B exhibiting decreasing returns to scale. In contrast, J and T have increasing (negative) returns concerning N and L . Thus, C exhibits increasing (negative) returns to scale. The countervailing forces on B and C mean that a point of optimality and a range of viable diversity exist, as illustrated in Fig. 1, panel a). The exogenous shadow prices of B and C affect their spread, which we may alter when qualitatively simulating hypothetical scenarios of interest. Although this is not formally a dynamic system, the intrinsic survival bias of evolution may render the optimality point an apparent attractor over time. Let us consider how systemic perturbations might steer the evolution of a genetic coding system toward greater evolvability. A sudden and drastic shock, such as a large meteor strike, to an ecosystem can destabilize the dynamic equilibria of niches beyond their resiliency. In order for life to thrive in these new conditions, it needs to be able to test a wider variety of phenotypes. This is manifested through an enhanced capacity for the genetic coding system to explore the genetic sequence space. This essentially raises the shadow value of B and the weight of α . Alternatively, if a positive shock occurs to an ecosystem, then life might more easily tolerate the costs of a more complex genetic coding system, thus decreasing the shadow value of C . If γ and δ also decrease sufficiently, then C could even exhibit negative returns to scale. Our behavior bias toward loss aversion means that we tend to fixate on the upfront negative effects of ecological change; however, a critical feature of life is that some organisms can construct new niches for themselves. From their

perspective, the aftermath of a negative shock to the ecosystem is the beginning of a bull market of evolutionary opportunity.

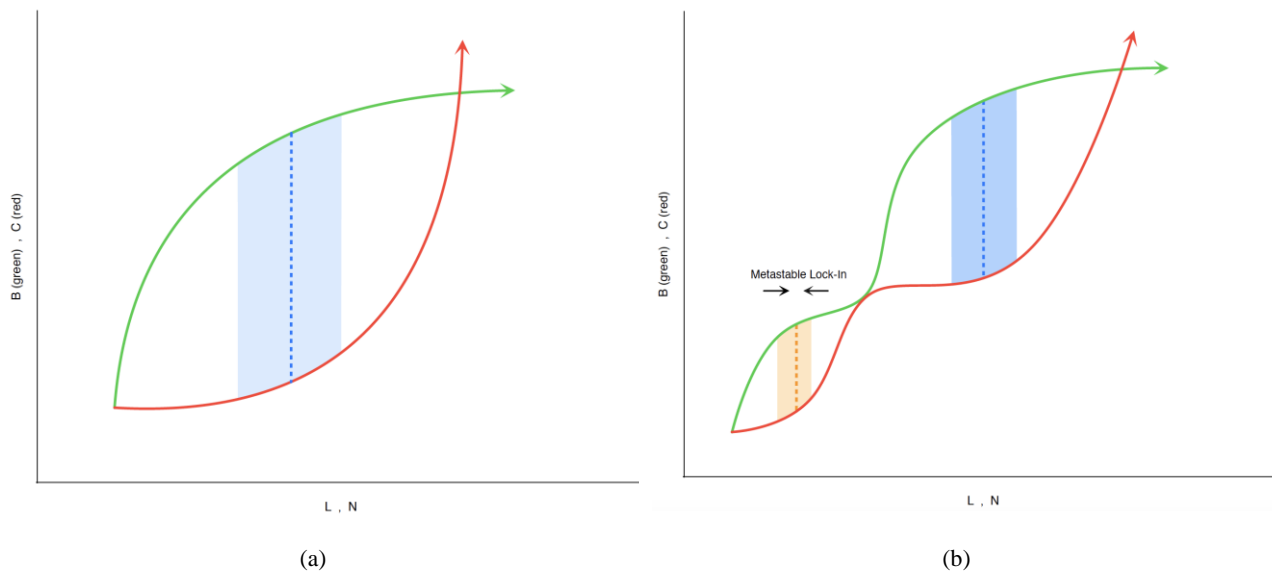


Fig. 1 – (a) Countervailing returns to scale of B and C . The dashed blue line corresponds to the combination of L and N that yields optimal net benefits (the green curve minus the red curve) to evolvability. The lighter blue region represents a zone of viable diversity. (b) Pinched node prevents the genetic coding system from evolving out of the orange zone of locally viable diversity to the blue zone of absolutely viable diversity.

An important upshot of all of this is the interesting situation shown in Fig. 1, panel b), in which the generally decreasing returns to scale for B and the generally increasing returns to scale for C vacillate. If this happens, then two optimalities arise, one local and the other absolute. However, due to the node between them, the local optimality is locked in a state of metastability [5]. This lock-in may be escaped through a combination of stochastic ecological shocks that simultaneously smooth the dimples in the returns to scale of B and C . Given that there are not significantly different genetic coding systems on Earth, such an escape route may be extraordinarily rare via acts of nature. Bioengineering, in contrast, may feasibly open up such a path.

4. DISCUSSION AND CONCLUSIONS

As a proof-of-concept that economic theory and methods can yield new insights in theoretical biology, this paper should evoke a net surplus of questions directed not only at the origins and future of life on Earth but also toward other aspects of biological theory that implicitly involve economization, such as bioenergetics, intracellular trafficking, and intercellular communication. The formal cause of life is one of the most profound scientific problems because what we learn may alter our understanding of how we *Homo sapiens* see ourselves about our tree of life.

Indeed, we may find it to be just one small tree within an expansive forest. Our growing sophistication in synthetic biology has put us in the particular situation of being able to jailbreak our genetic coding system effectively [6]. With concerns on the rise that looming environmental disasters will haringer the planet's next

mass extinction period, our newfound ‘unnatural’ powers may be what allow Earthly nature to flourish well beyond any expiration date.

REFERENCES

1. Maynard Smith J., Natural selection and the concept of a protein space, *Nature*, **225**, 5232, pp. 563–564 (1970).
2. Kauffman S., Levin S., Towards a general theory of adaptive walks on rugged landscapes, *Journal of Theoretical Biology*, **128**, 1, pp. 11–45 (1987).
3. Dawkins R., *The evolution of evolvability* in *Artificial Life*, Langton, C., Ed., Addison-Wesley Publishing Company: Reading, Massachusetts, United States, 1989, pp. 201–220.
4. Bejan A., Zane J.P., *Design in Nature: How the Constructal Law governs evolution in biology, physics, technology, and social organizations*, Doubleday, New York, United States, 2012.
5. Arthur W.B., Competing technologies, increasing returns, and lock-in by historical events, *The Economic Journal*, **99**, 394, pp. 116–131 (1989).
6. de la Torre D., Chin J.W., Reprogramming the genetic code, *Nature Reviews Genetics*, **22**, 3, pp. 169–184 (2021).