

A procedure for testing for Tokyo Type 1 Open-Ended Evolution*

Alastair Channon

School of Computer Science and Mathematics

Keele University

a.d.channon@keele.ac.uk

Abstract

Tokyo type 1 open-ended evolution (OEE) is a category of OEE that includes systems exhibiting the ongoing generation of adaptive novelty and ongoing growth in complexity. It can be considered as a necessary foundation for Tokyo type 2 OEE (ongoing evolution of evolvability) and Tokyo type 3 OEE (ongoing generation of major transitions). This paper brings together five methods of analysis to form a procedure for testing for Tokyo type 1 OEE. The procedure is presented as simply as possible, isolated from the complexities of any particular evolutionary system, and with a clear rationale for each step. Through the development of these steps, we also identify five key challenges in open-ended evolution. The last of these (achieving a higher order of complexity growth within a system exhibiting indefinitely scalable complexity) can be considered a grand challenge for Tokyo type 1 open-ended evolution. Promising approaches to this grand challenge include also achieving one or both of Tokyo types 2 and 3 OEE; this can be seen as one answer to *why* these other types of OEE are important, providing a unified view of open-ended evolution.

Keywords: open-ended evolution, evolutionary activity statistics, unbounded evolutionary dynamics, indefinite scalability, open problems in artificial life, unified views of evolution

1 Introduction

We think of open-ended evolution (OEE) as being defined by its behavioral hallmarks, which give rise to different kinds or categories of open-ended evolution, and about which there is now a good degree of consensus. These behavioral hallmarks are distinct from hypothesized requirements or mechanisms for the generation of open-ended evolution, about which there is currently less of a consensus and there are many different although often complementary ideas. The report from the first workshop on OEE categorized the behavioral hallmarks of OEE into the *York categories of OEE* (Taylor et al., 2016; Packard et al., 2019):

York type 1 OEE: Ongoing generation of adaptive novelty:

York type 1(a) OEE: Ongoing generation of new adaptations

York type 1(b) OEE: Ongoing generation of new kinds of entities

York type 1(c) OEE: Ongoing major transitions in evolution

York type 1(d) OEE: Ongoing evolution of evolvability

York type 2 OEE: Ongoing growth of complexity:

York type 2(a) OEE: Ongoing growth of entity complexity

York type 2(b) OEE: Ongoing growth of interaction complexity

In their paper *Grand Views of Evolution*, de Vladar, Santos and Szathmary succinctly stated the problem that “Biological evolution appears to be open-ended, but models of evolution have so far failed to account for this phenomenon” (de Vladar et al., 2017). They categorized evolutionary open-endedness into three forms: weak (“novel phenotypes can occur indefinitely”), strong (“continual appearance of evolutionary novelties and/or innovations”) and ultimate (“an indefinite increase in complexity”) (de Vladar et al., 2017), which parallel to some degree the York categories of OEE.

*This is the author’s final version. The article has been accepted for publication in Artificial Life.

Following the second and third workshops on OEE, the editorial introduction to the Open-Ended Evolution II Special Issue (Packard et al., 2019) presented the *Tokyo categories of OEE*, in which York types 1a, 1b, 2a and 2b were merged into Tokyo type 1; York types 1d and 1c became Tokyo types 2 and 3; and a new category (type) was added, with the caveat that “perhaps semantic evolution is a special case of Tokyo type 1 OEE”:

Tokyo type 1 OEE: Ongoing generation of interesting new kinds of entities and interactions

Tokyo type 2 OEE: Ongoing generation of evolution of evolvability

Tokyo type 3 OEE: Ongoing generation of major transitions

Tokyo type 4 OEE: Ongoing generation of semantic evolution

OEE can be studied in nature (Bedau et al., 1997, 1998); in systems with ongoing activity from nature (or other systems), such as the global economy, Internet traffic and systems involving user evaluation or interaction, including within cultural evolution (Skusa and Bedau, 2003; Borg et al., 2023) and the evolution of technology (Buchanan et al., 2011; Bedau et al., 2019), all of which examples include ongoing human activity; and in self-contained artificial systems, that is in systems with no ongoing activity from nature or other systems. It is of course possible to study OEE of all Tokyo-types in systems such as nature or the global economy, by taking advantage of nature having already achieved OEE types 1, 2 and 3. However, within the field of Artificial Life (Langton, 1992) we are also interested in developing a sufficient understanding of OEE for the development of self-contained artificial systems that exhibit OEE, and so a more complete understanding of OEE as well as novel and practical applications. Tokyo type 1 OEE is considered (by this author at least) as a necessary foundation for Tokyo types 2 and 3 OEE. It is difficult to conceive of a self-contained system achieving Tokyo type 2 or 3 OEE without first achieving Tokyo type 1 OEE. So the construction *and testing* of self-contained artificial systems to achieve Tokyo type 1 OEE is the most immediate challenge for research into the open-ended evolution of self-contained artificial systems.

One qualitative approach to classifying different types of novelty is to do so “relative to a model” (Rosen, 1978, 1985; Cariani, 1991). Banzhaf et al. (2016) define variation, innovation and emergence respectively as novelty within the model, for example “flipping a bit in GA”; novelty that changes the model, for example “adding a new species in an ecosystem”; and novelty that changes the meta-model (a “model for describing open-ended systems”), for example “when solitary individuals become colonies or societies” or “when a number of species form an organized ecosystem”. The artificial life system Geb has exhibited such examples (Channon, 2006) and the automata chemistry Stringmol has also been shown to exhibit these three types of novelty (Stepney and Hickinbotham, 2020). The approach taken in this paper is instead first quantitative, based on measuring *adaptive novelty*, the *accumulation of adaptive success*, diversity and complexity; qualitative classifications are then determined through the analysis of trends in these underlying measures, for example whether they are bounded or unbounded.

This paper brings together five methods of analysis to form a *procedure for testing for Tokyo type 1 OEE*:

1. Evolutionary activity statistics (Bedau et al., 1997);
2. Component-normalized evolutionary activity statistics (Channon, 2003, 2006);
3. Long-term evolutionary dynamics classification (Bedau et al., 1998);
4. Analysis of Indefinite Scalability in Diversity and Complexity (Ackley and Small, 2014; Channon, 2019);
5. Analysis of the Order of Indefinite Scalability.

This procedure is not intended to discourage other forms of analysis. Authors are encouraged to collect additional statistics from their systems, including diversity of various kinds, complexity of various kinds, functional information and various kinds of evolutionary activity statistics. This is especially important when considering other types of OEE, where the use of additional statistics will be required.

2 Procedure for testing for Tokyo type 1 OEE

The procedure involves proceeding through these five steps sequentially. It is presented here as simply as possible, isolated from the complexities of any particular evolutionary system, and with a clear rationale for each step. It is expected that the procedure will be developed further in the years to come, with the construction and evaluation of evolutionary systems aiding in that development.

Step 1: Compute basic evolutionary activity statistics

At the core of OEE is the ongoing evolution of *adaptive novelty* (York type 1 OEE): “new components flowing into the system and proving their adaptive value through their persistent activity” (Bedau et al., 1998). Components could be, for example, genes, organisms or species. For example, in a system of agents such as Packard’s Bugs (Packard, 1989; Bedau et al., 1997) or Geb (Channon, 2006), components could be entire genotypes (Bedau et al., 1997) or (alternatively) subsequences within genotypes, akin to genes, for example sections of genotype that specify an agent’s neural development (Channon, 2006). Likewise, in a system of self-replicating strings of code such as Tierra (Ray, 1991) or Avida (Adami and Brown, 1994), components could be entire genotypes (Bedau et al., 1997) or (alternatively) subsequences within genotypes, for example sections of code that lie between consecutive no-operation instructions (which are used for pattern-based addressing) in Tierra and Avida.

However, an evolutionary process could continue to generate adaptive novelty but lose what had previously been evolved at the same or a faster rate, cycling or idling with a limited extent of adaptive success. Ongoing progress, an unbounded *accumulation of adaptive success*, is also important. Evolutionary activity statistics provide a measure of exactly that: “a measure of the continual adaptive success of the components in the system” (Bedau et al., 1998), based on *adaptive persistence*.

$$\Delta_i(t) = \begin{cases} 1 & \text{if component } i \text{ exists at } t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$a_i(t) = \begin{cases} \sum_{\tau=0}^t \Delta_i(\tau) & \text{if component } i \text{ exists at } t \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

A component’s *cumulative evolutionary activity* $a_i(t)$ is a measure of the accumulation of its adaptive success (although see step 2 for an improved measure $a_i^N(t)$ that excludes non-adaptive persistence). Specifically, it is the length of time that the component has existed, discounting any periods of absence (equation 2). The sum of component activities (for those components present, in use) is a measure of the system’s accumulation of adaptive success, termed *total cumulative evolutionary activity* ($A_{\text{cum}}(t)$, equation 3). That is all there is to know to compute these basic evolutionary activity statistics! They can be computed for any evolving system with an available record of its components’ existence times, so are widely applicable across artificial and natural systems.

$$A_{\text{cum}}(t) = \sum_i a_i(t) \quad (3)$$

In step 1, these evolutionary activity statistics are computed and a quick check performed to see whether or not total cumulative evolutionary activity is bounded. If it is, there is no potential for a classification of unbounded evolutionary dynamics in step 3 below and so the procedure ends. Of all self-contained systems, only the biosphere (Bedau et al., 1997, 1998), Geb (Channon, 2001, 2006) and Pichler’s computational ecosystem (Pichler and Canamero, 2007, 2008; Pichler, 2009) have demonstrated unbounded total cumulative evolutionary activity (Taylor et al., 2016) (even before normalization using a shadow model). Developing and demonstrating more self-contained systems that exhibit unbounded total cumulative evolutionary activity (and that have good prospects for success at subsequent steps) is a clear priority. Fortunately, the effort required to evaluate a system to this extent is low, given that no random-selection shadow model is required for this first step.

Step 2: Compute component-normalized evolutionary activity statistics

Following success at step 1, step 2 involves first the implementation of a “shadow” model (population and system) that is identical to the real evolutionary system (running in parallel) except that whenever selection operates in the real system, random selection should be employed in the shadow; and, second, implementation of the shadow-resetting method and component-normalized evolutionary activity statistics as described below.

Bedau et al. (1997) defined the first neutral shadow models, of Evita (a system of self-replicating strings of code, akin to Tierra (Ray, 1991) and Avida (Adami and Brown, 1994)) and Packard's Bugs model (a system of agents that exist in a spatial grid):

"We also define a neutral analogue of Evita, which differs from Evita only in that there is no chance that a genotype's presence or concentration in the population is due to its adaptive significance. Nominal "programs" exist at grid locations, reproduce and die. The neutral model has two parameters: the number of mutations in the population per timestep (possibly a vector), and the number of "programs" that reproduce per timestep (possibly a vector). When the neutral model is due to have a reproduction event, the self-reproducing "program" is chosen at random from the population (with equal probability). When a "program" reproduces, its oldest neighboring "program" dies and the new child occupies the newly emptied grid location. Each "program" has a nominal "genotype" which it's children inherit. Whenever a mutation strikes a "program" it is assigned a new "genotype". . . . None of the dynamic of a genotype in the neutral analogue is due to that genotype's adaptive significance for the genotypes have no adaptive significance whatsoever. By recording mutation rates and reproduction rates from an actual Evita run, the neutral analogue can then be run with these vectors as parameters." (Bedau et al., 1997, p. 128)

"As we did with Evita, we also create a neutral analogue of the Bugs model, which differs from the Bugs only in that a genotype's persistence is no reflection of its adaptive significance. Nominal "agents" are born, live, reproduce, and die at rates determined exactly by the values of those variables measured in a particular run of the normal Bugs model. . . . The distinctive feature of the neutral analogue is that birth, reproduction and death events happen to "agents" chosen at random from among those present in the population. Each "agent" has a nominal "genotype" which it inherited from its parent unless it suffered a mutation at birth (mutation rate is another model parameter). . . . As with Evita's neutral analogue, none of the dynamic of a given genotypes in this neutral analogue of the Bugs model is due to that genotype's adaptive significance for it has no adaptive significance." (Bedau et al., 1997, p. 129)

Bedau et al. (1998) made the link between a shadow and its real (normal) run's birth, reproduction and death events more explicit when defining a neutral shadow of Holland's Echo model (Holland, 1975):

"The timing and number of birth and death events in the neutral shadow are directly copied from those in the normal Echo run, as is the neutral shadow's mutation rate. When some creature is born in the normal Echo run, it is time for a birth event in the shadow model, so a shadow parent chosen at random (with equal probability) from the shadow population reproduces. The new shadow child inherits its parent's genotype unless a mutation gives the child a new, unique genotype. When some creature dies in the normal Echo run, it is time for a death event in the shadow model and a "creature" is chosen at random (with equal probability) from the shadow population and killed. Thus, all selection in the neutral shadow is random." (Bedau et al., 1998, p. 232)

These examples set the standard for defining neutral shadow models, which others then followed. For example, Channon (2006) followed this approach to define the neutral shadow model of Geb. When some agent is born in the normal Geb run, it is time for a birth event in the shadow model, so shadow parents chosen at random (each with equal probability, avoiding the same individual being chosen for the second parent) from the shadow population reproduce using the same crossover and mutation procedures and rates as in the normal model. When some agent dies in the normal Geb run, it is time for a death event in the shadow model and an "agent" is chosen at random (with equal probability) from the shadow population and killed. Thus, all selection in the neutral shadow is random.

Having first implemented a shadow model, it is now possible to implement the shadow-resetting method. This resets the shadow system's state (population) and evolutionary activity history to those of the real (running) system immediately after each snapshot (when an entry is made in the component existence record). This provides the data required for the computation of component-normalized evolutionary activity statistics (Channon, 2003, 2006). Specifically, it enables comparison of inter-snapshot changes in activity in the real run with changes we would expect from neutral (random) selection.

$$\Delta_i^R(t) = \begin{cases} 1 & \text{if component } i \text{ exists in the real run at } t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$\Delta_i^S(t) = \begin{cases} 1 & \text{if component } i \text{ exists in the shadow at } t \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$\Delta_i^N(t) = \Delta_i^R(t) - \Delta_i^S(t) \quad (6)$$

$$a_i^N(t) = \begin{cases} \sum_{\tau=0}^t \Delta_i^N(\tau) & \text{if component } i \text{ exists in the real run at } t \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

The shadow is used to normalize (exclude non-adaptive) evolutionary activity at the component level (“component activity normalization”, equations 4-6), giving a measure of each component’s *adaptive* evolutionary activity $a_i^N(t)$ (equation 7) and so also component-normalized (adaptive) measures of total, mean and median cumulative evolutionary activity (equations 8-10).

$$A_{\text{cum}}^N(t) = \sum_{\substack{i: \text{component } i \text{ exists} \\ \text{in the real run at } t}} a_i^N(t) \quad (8)$$

$$\bar{A}_{\text{cum}}^N(t) = \frac{A_{\text{cum}}^N(t)}{D^R(t)} \quad (9)$$

$$\tilde{A}_{\text{cum}}^N(t) = \underset{\substack{i: \text{component } i \text{ exists} \\ \text{in the real run at } t}}{\text{Median}} a_i^N(t) \quad (10)$$

Ongoing adaptive novelty is determined through *new activity* $A_{\text{new}}^N(t)$ (equation 12): the sum of *newly adaptively-significant* components’ activities, divided by component diversity (the number of components present, in use in the real run). A component is considered adaptively significant if its activity is above a threshold, taken to be the absolute value of the most negative component-normalized evolutionary activity so as to screen out non-adaptive activity (Channon, 2006).

$$D^R(t) = \#\{i : a_i(t) > 0\} \quad (11)$$

$$A_{\text{new}}^N(t) = \frac{1}{D^R(t)} \sum_{i: \text{component } i \text{ 'new'}} a_i^N(t) \quad (12)$$

For the purpose of this test for Tokyo type 1 OEE, an entity that either is or contains a new adaptively-significant component (a new component whose component-normalized evolutionary activity exceeds the threshold) is considered to be *an interesting new kind of entity*. This carries forward the basis of York type 1 OEE (adaptive novelty) into Tokyo type 1 OEE. It does not capture ways of being *interesting* or of a *new kind* that confer no adaptive advantage to any entity. The *new kind* is the class of entities containing the new component, and the *interesting new kinds* are those for which the corresponding new component is adaptively significant. At the system level, we consider both *ongoing adaptive novelty* (ongoing positive new evolutionary activity) and *accumulation of adaptive success* (unbounded total and median cumulative evolutionary activity) to be *interesting*.

Stout and Spector (2005) conducted a series of experiments carefully designed in an attempt to “break” the classification scheme (step 3 below) by achieving a classification of unbounded evolutionary dynamics in “intuitively unlikelike” systems. Their attempts included populations undergoing genetic drift; evolution with static fitness functions; and experiments in which the fitness function was changed at regular intervals, with first simple and then more complex (“factored”) fitness functions. The classification scheme proved robust against

these attempts, with none achieving a classification of unbounded evolutionary dynamics. Stout and Spector concluded that component activity normalization was “of particular importance to the scheme’s robustness” as it “assures that the neutral shadow remains a faithful shadow of the real population, canceling out the potential for spurious results arising from the (random) divergence of the real and shadow populations”. Bedau et al.’s reasoning that “the mere fact that a family appears in the fossil record is good evidence that its persistence reflects its adaptive significance” (Bedau et al., 1998) (as “[s]ignificantly maladaptive taxonomic families would likely go extinct before leaving a trace in the fossil record” (Bedau et al., 1998)) is generally accepted. However, this is not true for all systems or for all choices of component class. Stout and Spector’s conclusions support the case that, in general (where appearance in the component record cannot be taken as conclusive evidence of adaptive significance), component activity normalization should be employed, in order to avoid spurious classifications of unbounded evolutionary dynamics.

Dolson et al. (2019) implemented a generic “persistence filter”, intended as an alternative to implementing system-specific shadow models, limiting analysis to “those components whose descendants persist for a substantial number of generations”. They noted that such a mechanism had previously been used “with $t=1$ ” in Channon (2006) to “screen out . . . isolated occurrences: when a component occurs in the current snapshot but not the previous one”. However, the very purpose of component activity normalization is to exclude non-adaptive persistence from measures of adaptive persistence, from the lowest measure (each component’s *adaptive* evolutionary activity $a_i^N(t)$) to whole-system measures ($A_{cum}^N(t)$, $\bar{A}_{cum}^N(t)$, $\tilde{A}_{cum}^N(t)$ and $A_{new}^N(t)$) derived from it. So persistence alone is not an adequate substitute. Nevertheless, Dolson et al.’s idea presents an interesting direction toward the possible invention of a future filter that excludes non-adaptive persistence, which would be an adequate substitute for the use of shadow models and component activity normalization.

Step 3: Long-term evolutionary dynamics classification

After determining long-term trends in component-normalized evolutionary activity statistics, including *new activity* and total, mean and median *cumulative evolutionary activity*, the system’s long-term evolutionary dynamics can be classified. The hallmark of unbounded evolutionary dynamics is ongoing positive new evolutionary activity $A_{new}^N(t)$ in combination with unbounded total ($A_{cum}^N(t)$) and median ($\bar{A}_{cum}^N(t)$) cumulative evolutionary activity (Bedau et al., 1998; Channon, 2006).

A classification of unbounded evolutionary dynamics, using component-normalized evolutionary activity statistics, provides a test for York type 1a OEE (without ruling out other types). Only one self-contained artificial system, Geb (Channon, 2003, 2006), has demonstrated this (Taylor et al., 2016). Developing more self-contained systems that do is a further clear priority; any such system would be a significant new contribution to the field.

Further, the one self-contained artificial system we have at this level (Geb) has poor behavioral transparency, such that the direct observation of artifacts and behaviors beyond the early stages of evolution soon becomes increasingly less practical. A very significant advance would be made by the development of a self-contained artificial system that demonstrates unbounded evolutionary dynamics (using component-normalized evolutionary activity statistics) *and* in which long (evolutionary) sequences of evolved artifacts or behaviors and the evolution of more complex artifacts and behaviors can be *readily seen*, evidenced by phenotypes rather than just by metrics. While it can be informative to analyze evolutionary systems in terms of metrics (as in this test for OEE), it is wise to also observe evolved artifacts and behaviors to verify that such analysis corresponds to what has evolved (and indeed to study these for their own sake). This could lead to either increased confidence in or the revelation of a flaw in the method of analysis (here in this test for Tokyo type 1 OEE). This highlights the need to develop future systems such that behavioral descriptions are as easy to generate as possible, for example by constructing systems such that behaviors will be transparent to human observers.

Step 4: Analysis of Indefinite Scalability in Diversity and Complexity

If an evolutionary system exhibits unbounded evolutionary dynamics in step 3, it would be natural to want to know whether or not the system also exhibits ongoing growth in maximum individual (or group or system) complexity, i.e. York type 2 OEE. This carries forward the basis of York type 2 OEE (growth of complexity) into

Tokyo type 1 OEE, into which York type 2 was merged along with York types 1a and 1b. So, in all, we consider *ongoing adaptive novelty*, *accumulation of adaptive success* and *growth of complexity* to be *interesting*.

The diversity of components (the number of different components) in an individual (or group or species) is one measure of its complexity. This is particularly appropriate when a component is analogous to a gene. Schadt et al. (2011) demonstrated that organism complexity correlates significantly with gene number in the absence of plant genomes, and Chen et al. (2014) reached the same finding.

In Bedau, Snyder and Packard's classification of long-term evolutionary dynamics (Bedau et al., 1998), the class of systems with unbounded evolutionary dynamics is divided into three subclasses: (a) those with unbounded diversity of adaptive components but bounded adaptive success (cumulative evolutionary activity) per component; (b) those with bounded diversity but unbounded adaptive success per component; and (c) those with unbounded diversity and unbounded adaptive success per component. Where complexity is measured as diversity of components, ongoing growth in system complexity is equivalent to unbounded diversity, so these subclasses relate directly to ongoing growth in system complexity, i.e. York type 2 OEE, and subclass c is implied by ongoing growth in any of individual, group or system complexity when observed together with unbounded adaptive success per component.

While adaptive success per component can be truly unbounded (over unbounded time), the diversity of adaptive components is necessarily bounded: in artificial systems by unavoidable physical limits such as computer memory, and in nature again by physical limits such as number of atoms. A claim of unbounded diversity in the biosphere is really a claim that diversity is not practically bounded, or that it has not reached the upper bound yet. A more precise notion than "unbounded" diversity is needed. Ackley's concept of *indefinite scalability* (Ackley and Small, 2014) provides this. The key criteria for indefinite scalability is that should an upper bound be reached, increasing the values of physical limitations (such as available matter, population size or memory) should enable an unbounded sequence of greater upper bounds to be achieved (after sufficient increases in the limitations and evolutionary time); in the case of diversity this means an unbounded sequence of greater upper bounds on diversity.

A practical (and the most literal) interpretation of indefinite scalability is that the sequence of greater upper bounds (on increasing the values of physical limitations) continues to an unknown length, i.e. that no end to it has been found. It is therefore best to qualify any empirical claims by quantifying the extent to which indefinite scalability has been established. Claims about systems can be expressed and evaluated in terms such as a metric (for example a measure of adaptive success per component) increasing apparently without bound *up to* a certain system time (or number of generations, etc.); or a metric (for example diversity) increasing *up to* certain value(s) of system parameter(s) being reached, where it was necessary to increase these to establish increases in scale (for example of diversity) over successive runs.

So, in step 4 the aim is to demonstrate a sequence of greater upper bounds on diversity (on increasing the values of physical limitations) that increases without any known bound, qualifying the extent (for example number of generations or values of physical limitations) to which this has been established. This step (step 4) is likely to be developed further in future years, as to date only one system (Geb) has been evaluated at this level (Channon, 2019). Developing more evolutionary systems that exhibit indefinitely scalable complexity is a further clear priority; any such system would be a very significant new contribution to the field.

Step 5: Analysis of the Order of Indefinite Scalability

While Geb exhibited indefinite scalability in diversity and complexity, the analysis in (Channon, 2019) also revealed that complexity scaled only logarithmically with (the lower of) maximum population size and maximum number of neurons per individual. It was noted that the evolution of artifacts and behaviors of much greater complexity, for example comparable to those in nature, within feasible timescales, will almost certainly require a higher order of complexity scaling.

To appreciate the extent to which a higher order of complexity scaling is required, consider increasing Geb's limiting parameters (*worldlength*, *maxneurons*) from (80, 160), their largest values for the runs reported in (Channon, 2019), to for example (80 million, 160 million or more). For example, increase maximum population size (*worldlength* squared) from 6,400 (80^2) to 6.4 quadrillion ($80,000,000^2$) and allow each agent up to 160 million (or more) neurons; it would also be necessary to allow sufficient evolutionary time to establish

maximum individual complexity. Although fanciful, being far beyond what is practicable, we can nevertheless consider this as a thought experiment. If the logarithmic scaling identified holds to such an extent, this would increase maximum individual complexity (in Geb, the number of genome subsequences involved in an agent's neural development) to approximately 53 ($10.167525 + 1.866825 \log_2(8000000)$), that is to a maximum of just 53 active 'genes' per individual. By contrast, *C. elegans*, mice and humans each have approximately 20,000 protein-coding genes (Davis et al., 2022; Frankish et al., 2023). Gould (1996) noted that lower-complexity organisms have not been displaced on Earth, and described the distribution of organism complexity as having spread out from, and always including, a lower bound (that of the simplest living things) and becoming increasingly skewed in favor of simpler organisms. A population size of 6.4×10^{15} is still far below that of prokaryotes on Earth, estimated to be $4 - 6 \times 10^{30}$ (Whitman et al., 1998), so this is not necessarily at variance with the complexity scaling of life on Earth. However, from a practical perspective, needing to increase population sizes into the quadrillions in order to achieve even such a low level of complexity would be a daunting, impractical prospect.

Achieving a higher order of complexity growth, within a system exhibiting indefinite scalability in complexity, can be considered a grand challenge for Tokyo type 1 open-ended evolution.

3 Conclusions

The above procedure for testing for Tokyo type 1 OEE brings together five methods of analysis and has a clear rationale for each step. Further, these steps have enabled us to identify five key challenges in open-ended evolution, of increasing significance:

1. Developing and demonstrating more self-contained systems (beyond the biosphere, Geb and Pichler's computational ecosystem) that exhibit unbounded total cumulative evolutionary activity (and that have good prospects for success at subsequent steps; section 2, step 1);
2. Developing more self-contained systems (beyond Geb) that achieve a classification of unbounded evolutionary dynamics, using component-normalized evolutionary activity statistics (i.e. York type 1a OEE; section 2, step 3);
3. Developing the first self-contained artificial system to demonstrate unbounded evolutionary dynamics (using component-normalized evolutionary activity statistics) *and* in which long (evolutionary) sequences of evolved artifacts or behaviors and the evolution of more complex artifacts and behaviors can be *readily seen*, evidenced by phenotypes rather than just by metrics (section 2, step 3);
4. Developing more evolutionary systems (beyond Geb) that exhibit indefinitely scalable complexity (section 2, step 4);
5. Achieving a higher order of complexity growth (than demonstrated in Geb) within a system exhibiting indefinitely scalable complexity (section 2, step 5).

The last of these can be considered a grand challenge for Tokyo type 1 open-ended evolution. Promising approaches to this grand challenge include also achieving one or both of Tokyo type 2 OEE (ongoing evolution of evolvability) and Tokyo type 3 OEE (ongoing generation of major transitions). Indeed, this can be seen as one answer to *why* these other types of OEE are important, providing a unified view of open-ended evolution.

References

- Ackley, D. H. and Small, T. R. (2014). Indefinitely scalable computing = artificial life engineering. In Sayama, H., Rieffel, J., Risi, S., Doursat, R., and Lipson, H., editors, *Artificial Life XIV*, pages 606–613, Cambridge, MA. MIT Press.
- Adami, C. and Brown, C. T. (1994). Evolutionary learning in the 2D artificial life system 'Avida'. In *Artificial Life IV*, pages 377–381, Cambridge, MA. MIT Press.

- Banzhaf, W., Baumgaertner, B., Beslon, G., Doursat, R., Foster, J. A., McMullin, B., de Melo, V. V., Miconi, T., Spector, L., Stepney, S., and White, R. (2016). Defining and simulating open-ended novelty: requirements, guidelines, and challenges. *Theory in Biosciences*, 135(3):131–161.
- Bedau, M. A., Gigliotti, N., Janssen, T., Kosik, A., Nambiar, A., and Packard, N. (2019). Open-ended technological innovation. *Artificial Life*, 25(1):33–49.
- Bedau, M. A., Snyder, E., Brown, C. T., and Packard, N. H. (1997). A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In *Proceedings of the Fourth European Conference on Artificial Life*, pages 125–134. MIT Press.
- Bedau, M. A., Snyder, E., and Packard, N. H. (1998). A classification of long-term evolutionary dynamics. In Adami, C., Belew, R. K., Kitano, H., and Taylor, C. E., editors, *Artificial Life VI*, pages 228–237, Cambridge, MA. MIT Press.
- Borg, J. M., Buskell, A., Kapitany, R., Powers, S. T., Reindl, E., and Tennie, C. (2023). Evolved open-endedness in cultural evolution: A new dimension in open-ended evolution research. *Artificial Life*, pages 1–22.
- Buchanan, A., Packard, N. H., and Bedau, M. A. (2011). Measuring the evolution of the drivers of technological innovation in the patent record. *Artificial Life*, 17(2):109–122.
- Cariani, P. (1991). Emergence and artificial life. In Langton, C. G., Taylor, C., Farmer, J. D., and Rasmussen, S., editors, *Artificial Life II*, pages 775–797, Redwood City, CA. Addison-Wesley.
- Channon, A. (2001). Passing the ALife test: Activity statistics classify evolution in Geb as unbounded. In Kelemen, J. and Sosík, P., editors, *Advances in Artificial Life: Proceedings of the Sixth European Conference on Artificial Life*, pages 417–426, Berlin, Heidelberg. Springer-Verlag.
- Channon, A. (2003). Improving and still passing the ALife test: Component-normalised activity statistics classify evolution in Geb as unbounded. In Standish, R. K., Bedau, M. A., and Abbass, H. A., editors, *Artificial Life VIII*, pages 173–181, Cambridge, MA. MIT Press.
- Channon, A. (2006). Unbounded evolutionary dynamics in a system of agents that actively process and transform their environment. *Genetic Programming and Evolvable Machines*, 7(3):253–281.
- Channon, A. (2019). Maximum individual complexity is indefinitely scalable in Geb. *Artificial Life*, 25(2):134–144.
- Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A., and Urrutia, A. O. (2014). Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Molecular Biology and Evolution*, 31(6):1402–1413.
- Davis, P., Zarowiecki, M., Arnaboldi, V., Becerra, A., Cain, S., Chan, J., Chen, W. J., Cho, J., da Veiga Beltrame, E., Diamantakis, S., Gao, S., Grigoriadis, D., Grove, C. A., Harris, T. W., Kishore, R., Le, T., Lee, R. Y. N., Luypaert, M., Müller, H.-M., Nakamura, C., Nuin, P., Paulini, M., Quinton-Tulloch, M., Raciti, D., Rodgers, F. H., Russell, M., Schindelman, G., Singh, A., Stickland, T., Van Auken, K., Wang, Q., Williams, G., Wright, A. J., Yook, K., Berriman, M., Howe, K. L., Schedl, T., Stein, L., and Sternberg, P. W. (2022). Wormbase in 2022—data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics*, 220(4):iyac003.
- de Vladar, H. P., Santos, M., and Szathmáry, E. (2017). Grand views of evolution. *Trends in Ecology & Evolution*, 32(5):324–334.
- Dolson, E. L., Vostinar, A. E., Wiser, M. J., and Ofria, C. (2019). The MODES toolbox: Measurements of open-ended dynamics in evolving systems. *Artificial Life*, 25(1):50–73.

- Frankish, A., Carbonell-Sala, S., Diekhans, M., Jungreis, I., Loveland, J., Mudge, J., Sisu, C., Wright, J., Arnan, C., Barnes, I., Banerjee, A., Bennett, R., Berry, A., Bignell, A., Boix, C., Calvet, F., Cerdán-Vélez, D., Cunningham, F., Davidson, C., Donaldson, S., Dursun, C., Fatima, R., Giorgetti, S., Giron, C., Gonzalez, J., Hardy, M., Harrison, P., Hourlier, T., Hollis, Z., Hunt, T., James, B., Jiang, Y., Johnson, R., Kay, M., Lagarde, J., Martin, F., Gómez, L., Nair, S., Ni, P., Pozo, F., Ramalingam, V., Ruffier, M., Schmitt, B., Schreiber, J., Steed, E., Suner, M.-M., Sumathipala, D., Sycheva, I., Uszczyńska-Ratajczak, B., Wass, E., Yang, Y., Yates, A., Zafrulla, Z., Choudhary, J., Gerstein, M., Guigo, R., Hubbard, T. J. P., Kellis, M., Kundaje, A., Paten, B., Tress, M., and Flicek, P. (2023). GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Research*, 51(D1):D942–D949.
- Gould, S. J. (1996). *Full House: The Spread of Excellence from Plato to Darwin*. Harmony Books, New York. Published in the UK under the title *Life's Grandeur*.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. MIT Press, 2nd edition edition.
- Langton, C. G. (1992). Artificial life. In Nadel, L. and Stein, D. L., editors, *1991 Lectures in Complex Systems*, pages 189–241. Addison-Wesley.
- Packard, N., Bedau, M. A., Channon, A., Ikegami, T., Rasmussen, S., Stanley, K. O., and Taylor, T. (2019). An overview of open-ended evolution: Editorial introduction to the open-ended evolution ii special issue. *Artificial Life*, 25(2):93–103.
- Packard, N. H. (1989). Intrinsic adaptation in a simple model for evolution. In *Artificial Life I*, pages 141–155. Addison-Wesley.
- Pichler, P.-P. (2009). *Natural selection, adaptive evolution and diversity in computational ecosystems*. PhD thesis, University of Hertfordshire.
- Pichler, P.-P. and Canamero, L. (2007). An evolving ecosystems approach to generating complex agent behaviour. In *Proceedings of the First IEEE Symposium on Artificial Life*, pages 303–310. IEEE.
- Pichler, P.-P. and Canamero, L. (2008). Evolving morphological and behavioral diversity without predefined behavior primitives. In Bullock, S., Noble, J., Watson, R., and Bedau, M. A., editors, *Artificial Life XI*, pages 404–411, Cambridge, MA. MIT Press.
- Ray, T. S. (1991). An approach to the synthesis of life. In *Artificial Life II*, pages 371–408. Addison-Wesley.
- Rosen, R. (1978). *Fundamentals of measurement and representation of natural systems*. North-Holland.
- Rosen, R. (1985). *Anticipatory Systems*. Pergamon Press.
- Schad, E., Tompa, P., and Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biology*, 12(12):R120.
- Skusa, A. and Bedau, M. A. (2003). Towards a comparison of evolutionary creativity in biological and cultural evolution. In Standish, R. K., Bedau, M. A., and Abbass, H. A., editors, *Artificial Life VIII*, pages 233–242, Cambridge, MA. MIT Press.
- Stepney, S. and Hickinbotham, S. (2020). Innovation, variation, and emergence in an automata chemistry. In Bongard, J., Lovato, J., Soros, L., and Hebert-Dufrésne, L., editors, *ALIFE 2020: The 2020 Conference on Artificial Life*, pages 753–7601, Cambridge, MA. MIT Press.
- Stout, A. and Spector, L. (2005). Validation of evolutionary activity metrics for long-term evolutionary dynamics. In Beyer, H.-G., O'Reilly, U.-M., Arnold, D. V., Banzhaf, W., Blum, C., Bonabeau, E. W., Cantu-Paz, E., Dasgupta, D., Deb, K., Foster, J. A., de Jong, E. D., Lipson, H., Llorca, X., Mancoridis, S., Pelikan, M., Raidl, G. R., Soule, T., Tyrrell, A. M., Watson, J.-P., and Zitzler, E., editors, *Proceedings of the 2005 Genetic and Evolutionary Computation Conference (GECCO 2005)*, pages 137–142, New York. ACM.

Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickinbotham, S., Ikegami, T., McMullin, B., Packard, N., Rasmussen, S., Virgo, N., Agmon, E., Clark, E., McGregor, S., Ofria, C., Ropella, G., Spector, L., Stanley, K. O., Stanton, A., Timperley, C., Vostinar, A., and Wiser, M. (2016). Open-ended evolution: Perspectives from the OEE workshop in York. *Artificial Life*, 22(3):408–423.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, 95(12):6578–6583.