were chosen as a result of investigating how long the genetic algorithm generally takes to converge, in our case the algorithm rarely made progress after approximately 45 generations. An important consideration is that on lower-k landscapes the algorithm normally converged much faster, but to allow for fair comparison across different k-values we decided to keep the process consistent.

To make results more easily interpretable, we aimed to develop a more reliable fitness metric for our evolved walkers than the fitness value after a single run of their strategy. Since any strategy with a Random Walk step will vary greatly in its final fitness, there is a chance for a walker to simply get "lucky" and stumble into a high final fitness. In order to combat the inconsistency this would create, the fitness of an individual is determined by the average final fitness across 25 different runs of the strategy. This allows us to ensure that the final evolved strategy must consistently perform well to get a high fitness score.

Results

Relative Performance of Strategies

To verify that our evolved walker is showing improvement over the baseline walkers, we evolved separate walkers on 500 different landscapes for each k-values 0-14 with a consistent $n = 15$ and compared their results to the baseline strategies. Each strategy (evolved and baseline) was 200 steps long, and the evolved strategies were set to have exactly 40 walk steps in total. This 40 walk constraint was introduced to allow for a more direct comparison between different evolved strategies; the specific value of 40 walks was selected because when the number of walks is unconstrained the population regularly converges to having about 20% walk steps within any reasonable length strategy. The NK landscapes were randomly generated, though the same set of landscapes was used to test each individual type of Walker in order to ensure a fair comparison. The results of this experiment can be seen in Figure 3.

There are a few general trends to note from results in Figure 3. First, the random walker shows no sign of progress throughout its entire lifetime, as expected. This shows that without a productive developmental process, we can expect a final fitness of only slightly more than 0. Second, for the more adaptive strategies, we can see that each is able to solve the $k = 0$ landscape entirely and achieve the maximum fitness of 1, but as the difficulty of the landscape increases the fitness reached by each strategy decreases. This decrease in final fitness is sharp when progressing through lower k values, but eventually levels off at high k-values. This is all as expected - a more difficult landscape is inherently much harder to solve, so we would expect every strategy's performance to degrade as the k-value increases.

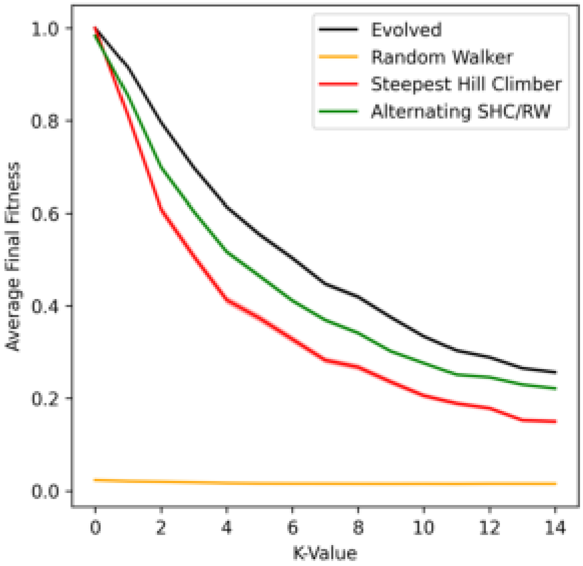


Figure 3: Final fitnesses of strategies across different K values. Average final fitness across 500 different landscapes for k-values from 0 to 14 on a $n = 15$ landscape. Strategy length 200 with 40 walk steps.

Distribution of Look Steps in Evolved Strategies

A more interesting trend appears at k-values greater than zero, where the SHC's strategy begin to perform worse than the Alternating and Evolved Strategies. This is almost certainly the result of local optima being present on these nontrivial landscapes.

When a purely exploitative strategy like SHC encounters a local maxima, it has no recourse. It will exclusively takes actions that greedily improve its fitness, so once it enters a local maxima it won't be able to improve anymore, and will then become 'stuck' for the remainder of its lifetime. Since local optima become more frequent at higher k-values this means that SHC strategy will increasingly get stuck at an earlier step in its lifetime, causing it to lose performance.

A solution to being stuck at a local optima is taking purely random, exploratory steps as the Alternating and Evolved strategies do. Although random walks are by themselves unproductive, a random walk action followed by subsequent exploitative actions gives a strategy the potential to escape that local optima (see Figure 2 for an illustration). This is why the evolved walker and alternating walker offer improvement over SHC - they are able to escape local optima which results in overall better performance. Although this explains why the SHC strategy performs worse than these two other strategies on more difficult landscapes, the reason for the preformance difference between the Evolved and Alternating strategy requires deeper analysis into the structure of these evolved strategies.

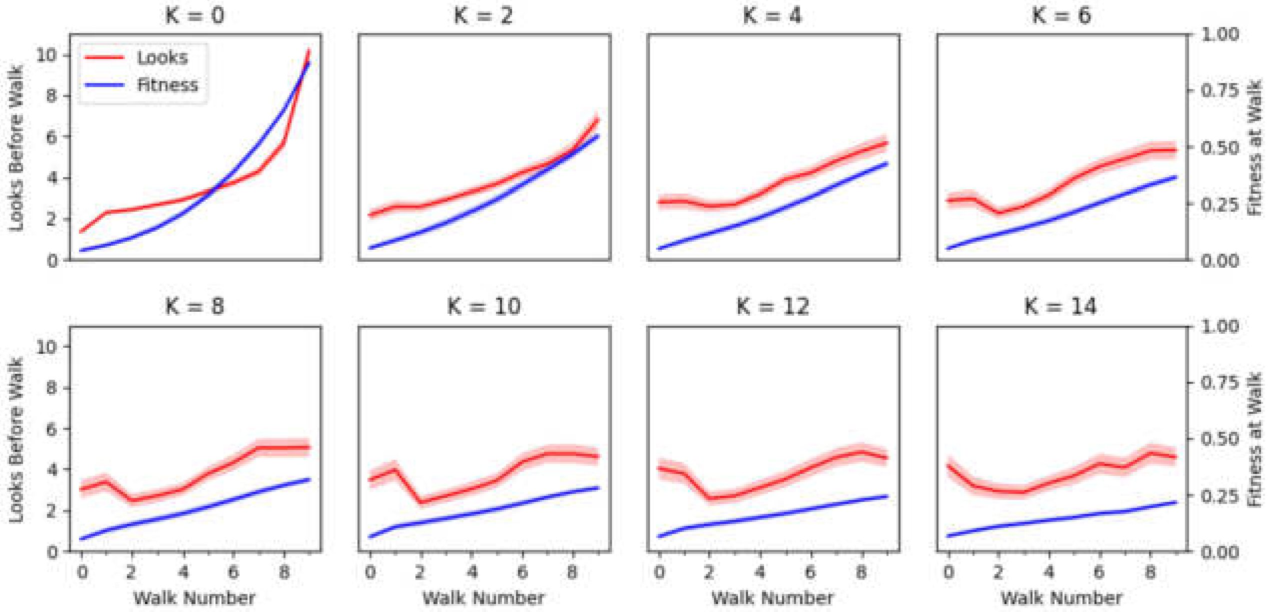


Figure 4: Look before walks and fitness at each walk across K values. Average distribution of Look Steps (in red) and fitness (in blue) of the evolved strategy across 500 landscapes, with error bars (standard deviation). Strategy length at 50 steps with 10 walk steps.

To do this, we will investigate the distribution of look actions in the final evolved strategy in landscapes of certain k-values. We investigated the trends among several sets of parameters and noticed similar structures across most strategy lengths as long as the ratio of walk to total steps remained consistent, so we decided to continue by investigating a shorted length strategy to reduce noise and allow easier data interpretation. These shorter strategies will be 1/4th the length of the previous strategies, with a total length of 50 steps, 10 of which are walks to retain the original 20% walk percentage.

Figure 4 shows the average walk distribution across different landscape difficulties with this more controlled strategy length and reveals several important trends. The first is that across low-difficulty landscapes, we see a preference for early-development exploration (low-look walk steps) and a preference for late-development exploitation (high-look walk steps). This aligns with what is has often been observed in developmental processes (Spreng and Turner, 2021). However, once the landscapes become sufficiently difficult (in this case, $k >= 10$), we see that this trend is significantly weaker, and the level of exploration/exploitation throughout the developmental process doesn't vary nearly as much between early and late development. We believe this is a result of need to escape local optima via use of the exploratory 'Random Walk' steps. At higher k-values we see more frequent local optima, and these exploratory steps become more important at even the late stages of development in order to escape increasing amounts of local optima. The distribution of these purely exploratory random walk steps is discussed in depth in the next section.

Another important trend is how on more difficult landscapes ($k >= 6$) we see that opening two steps of the strategy frequently have a higher preference for exploitation than the following few developmental steps. A likely explanation for this trend is that Random Walk steps aren't necessarily useful if an individual is not currently located at a local optima, and the chance that an individual is at a local optima before taking any exploitative steps is fairly low. This means that we would expect each walker to take a few exploitative steps before preferring exploration on these high-k landscapes, which is the trend we see emerging.

Distribution of Random Walks in Evolved Strategy

Using the same experimental setup as before, with 50-length strategies that each have 10 walks, we now analyze how the 0-look 'Random Walk' steps are distributed (see Figure 5). The most basic case with $k = 0$ shows no exploratory periods beyond the very first step. This is a result of the specific landscape containing a single global optima and no other local optima. This means that any purely exploitative strategy will succeed on this landscape, so there is never a need for the Random Walk steps.

However, once contribution factor dependence is intro-

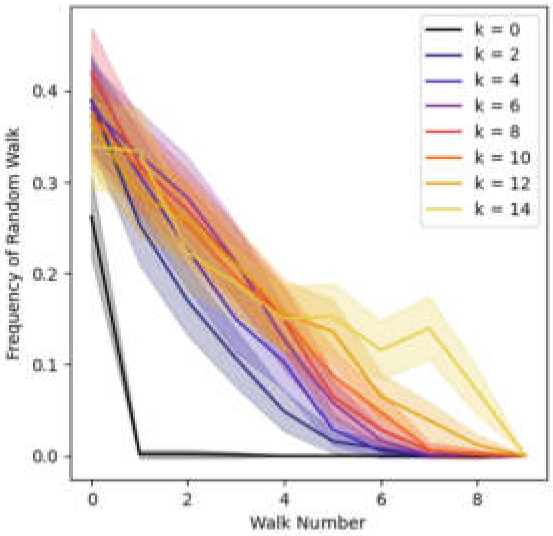


Figure 5: Frequency of Random Walk at Different Walk Steps. Percentage of evolved strategies ($1 = 100\%$) with a Random Walk at each step in their developmental process, with error bars (standard deviation). 500 strategies used at each k-value, and each strategy was evolved on a different landscape to avoid data bias.

duced ($k > 0$), we see the evolutionary process begin selecting for later-in-development exploration steps. These Random Walk steps are present throughout the developmental process for every k-value besides $k = 0$, and as the difficulty continues to increase, we see these exploration actions persist into mid-development, and even into late-development. An important factor to note is that while higher k-valued landscapes do contain more midlife exploration steps, they do not reduce the number of exploration actions in early development. This can be deduced by the fact that for every landscape with k higher than 2 there is a similar chance (±0.05) for a walk step at the first few steps of development. This shows that while we do see exploration actions persisting later in development on harder landscapes, this does not mean that early development exploration actions are unnecessary, rather this shows the contrary– mid-development exploration periods are more effective when paired with exploration earlier in development.

Another factor that reinforces this fact is how in Figure 3, we see a convergence between the final fitness of the evolved and alternating SHC/RW strategies at high k-values. This is representative of the fact that at higher-k landscapes with more local optima, the strategy to alternate between exploration and exploitation becomes more prominent to escape these optima. Eventually, when the k-value gets high enough there are so many local optima that the alternating strategy preforms almost equally as well as our evolved strategy at these extremely high k-values. Although this doesn't necessarily mean that the evolved and alternating strategies are using the same method to achieve their fitness, this does tell us that the importance of escaping local optima becomes increasingly important as the k-value increases.

Discussion

By simulating the evolution of developmental strategies in an abstract model, this paper reveals several insights on the relationship between evolution and development that can apply to various organisms and developmental processes. At the highest level, evolution produces developmental steps that explicitly explore the landscape in a random manner. In fact, the presence of such randomly exploring steps is shown to be necessary for all developmental strategies to survive selection, regardless of their location and frequency within the developmental period. It is interesting that evolution drives organisms to ignore environmental cues at times against the promise of a fitness ascent. However, the superiority of populations that develop with random exploration can be attributed to their divergence across a larger search space in the landscape, avoiding convergence at local optima and redirecting themselves to regions of potentially higher fitness. This may help explain the emergence of developmental phenomena such as child rebellion in humans and other animals (Sachser et al., 2018; Chakradhar, 2018; Spear, 2000), where evolution has selected for populations whose adolescents deviate from developmental trajectories fostered by their parents to achieve greater phenotypic and behavioral diversity. Conversely, in the case that only one global optimum exists, represented by the landscape with $k = 0$ (no interdependent factors), evolution indeed produces developmental strategies with no random exploration.

The location of the exploratory steps offers another insight. Evolution drives development to be time-sensitive, with non-uniform patterns of exploration and exploitation comprising the developmental strategy. The populations of evolved strategies in various landscapes all consistently exhibit different degrees of exploitation (and hence those of exploration) at different points in development, notably with large changes at the early and terminal stages of development. For example, the extensive exploitation towards the end of development can be attributed to its direct impact on the final fitness, which is the only fitness function used to evolve the strategies in this model. Similarly, sensitive periods in biological systems demonstrate time-sensitive development and significantly affect organisms towards the end of their development. As the adult form of many organisms are fixed for the rest of their lifetimes with low plasticity, the terminal stage of development is critical for their evolutionary fitness (Spreng and Turner, 2021; Del Maschio et al., 2018; Brehmer et al., 2014). These are consistent with observations of the uniformly distributed strategy alternating between exploration and exploitation, which was not able to develop higher fitness than the evolved, non-uniform strategies.

Besides the terminal stage, the early and middle stages of development that emerge on different landscapes reveal another perspective. Evolution selects for a more complex developmental strategy marked by multiple transitions between predominantly exploitative and exploratory phases, as the landscape becomes more complex with a greater number of interdependent factors. The early exploitation and middle exploration stages that emerge on high-k landscapes exemplify such phases and their transitions, along with the terminal exploitation stage common to all landscapes. Considering their contrast with the simpler, early exploration stage on low-k landscapes, the multiple transitions can be attributed to a potential mechanism related to the increased number of local optima (and hence the increased risk of suboptimal convergence). The early ascent through exploitation may guide organisms to regions of generally higher fitness, increasing the probability that the following exploration will place the organism near a high-fitness optimum prior to the terminal exploitation. Although it is inconclusive from this work whether this mechanism is truly responsible for the emergence of multiple transitions, it offers a useful direction for future work to verify the presence of such regions or examine completely different hypotheses.

More importantly, it is surprising to note that nature and this highly abstract model share the emergence of a more complex developmental strategy across increasingly complex landscapes. The quantitative development of flatworms exemplifies a simple landscape with few interdependent factors, exhibiting a simple exploitative development increasing or reducing in size based on the nutrition available (Martín-Durán and Egger, 2012). In contrast, the cognitive development of humans has a uniquely high number of interdependent factors due to social influences, evolving multiple distinct stages of development with varying degrees of exploration and exploitation (Thompson, 2021). Along with the aforementioned findings in the value of exploration and the time-sensitivity of development, the emergence of this trend demonstrates the ability of this model to simulate and examine various interactions between evolution and development. Next, we will discuss potential extensions to the model, which would enable future investigations.

Future Work

We see three major directions for future work based on the preliminary results reported on in this paper. First, in our model the only way to escape a local optima is by taking a purely random walk step and getting lucky. A simple extension would be to include a multi-tier look (i.e. change multiple bits in the bitstring) to more intelligently escape optima. This multi-tier look could also incur an additional cost and potentially offer insight into the underlying selection pressures for sensitive periods. This multi-tier look could allow for greater relative sensitivity to an environment (along with greater cost), which in turn could help identify the usefulness of critical periods in a developmental trajectory.

Second, our model currently only evolves the look and walk actions with a fixed starting location. Often, models employing NK fitness landscapes involve the evolution of the starting location. Extending the model to include starting location would allow more thorough exploration of the interaction between evolution and development.

Finally, in addition to the aforementioned extensions of this model, we see an interesting opportunity to conduct experiments in dynamic fitness landscapes where the fitness values shift over time. This could happen within an agent's lifetime, periodically much like seasonal changes, but also could happen across evolutionary time such that a more flexible developmental strategy might be required to sustain a population. Interesting possibilities could include analysis of the relationship between agent lifetime and the scale of the changes in the environment.

Acknowledgements

We would like to thank the reviewers for constructive feedback which greatly improved the final version of this paper. We would also like to thank Robert J Williamson and Mitchel J Daniel for helpful feedback on an earlier draft of this paper. Finally, we would like to thank Eduardo J Izquierdo and Madhavun Candadai for their conceptual contributions to early explorations of this project.

References

Altenberg, L. (1996). NK Fitness Landscapes. *Evolution*, pages 1–11.

Belew, R. K. (1990). Evolution, Learning, and Culture: Computational Metaphors for Adaptive Algorithms. *Complex Systems*, 4:11–49.

Brehmer, Y., Kalpouzos, G., Wenger, E., and Lövdén, M. (2014). Plasticity of brain and cognition in older adults. *Psychological Research*, 78(6):790–802.

Campbell, C. M., Izquierdo, E. J., and Goldstone, R. L. (2020). How much to copy from others? The role of partial copying in social learning. In *CogSci*.

Chakradhar, S. (2018). Animals on the verge: What different species can teach us about human puberty. *Nature Medicine*, 24(2):114–117.

Del Maschio, N., Sulpizio, S., Gallo, F., Fedeli, D., Weekes, B. S., and Abutalebi, J. (2018). Neuroplasticity across the lifespan and aging effects in bilinguals and monolinguals. *Brain and Cognition*, 125:118–126.

Fragata, I., Blanckaert, A., Dias Louro, M. A., Liberles, D. A., and Bank, C. (2019). Evolution in the light of fitness landscape theory. *Trends in Ecology & Evolution*, 34(1):69–82.

Frankenhuis, W. E. and Walasek, N. (2020). Modeling the evolution of sensitive periods. *Developmental Cognitive Neuroscience*, 41:100715.

Geard, N., Wiles, J., Hallinan, J., Tonkes, B., and Skellett, B. (2002). A comparison of neutral landscapes - NK, NKp and NKq. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, volume 1, pages 205–210. IEEE.

Hebbron, T., Bullock, S., and Cliff, D. (2008). NKα: Non-uniform epistatic interactions in an extended NK model. *Artificial Life*, XI.

Hinton, G. and Nowlan, S. (1987). How learning can guide evolution. *Complex Systems*, 1(3):495–502.

Kauffman, S. and Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11–45.

Knudsen, E. I. (2004). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, 16(8):1412–1425.

Lynch, J. A., Brent, A. E., Leaf, D. S., Pultz, M. A., and Desplan, C. (2006). Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia. *Nature*, 439(7077).

Martín-Durán, J. and Egger, B. (2012). Developmental diversity in free-living flatworms. *EvoDevo*, 3(1):7.

Müller, G. B. (2007). Evo–devo: extending the evolutionary synthesis. *Nature Reviews Genetics*, 8(12):943–949.

Panchanathan, K. and Frankenhuis, W. E. (2016). The evolution of sensitive periods in a model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, 283(1823):20152439.

Park, S.-C., Neidhart, J., and Krug, J. (2015). Greedy adaptive walks on a correlated fitness landscape. *Journal of Theoretical Biology*.

Pitzer, E. and Affenzeller, M. (2012). A comprehensive survey on fitness landscape analysis. In *Studies in Computational Intelligence*, volume 378, pages 161–191.

Rhodes, M. L. and Dowling, C. (2018). 'What insights do fitness landscape models provide for theory and practice in public administration?'. *Public Management Review*, 20(7):997–1012.

Roth, S. and Hartenstein, V. (2008). Development of Tribolium castaneum. *Development Genes and Evolution*, 218(3-4):115–118.

Sachser, N., Hennessy, M. B., and Kaiser, S. (2018). The adaptive shaping of social behavioural phenotypes during adolescence. *Biology Letters*, 14(11):20180536.

Smart, J. M. (2019). Evolutionary Development: A Universal Perspective. In *Springer Proceedings in Complexity*, pages 23–92.

Soltoggio, A., Stanley, K. O., and Risi, S. (2018). Born to learn: The inspiration, progress, and future of evolved plastic artificial neural networks. *Neural Networks*, 108:48–67.

Sommer, R. J. (2009). The future of evo-devo: Model systems and evolutionary theory.

Spear, L. (2000). Modeling adolescent development and alcohol use in animals. *Alcohol Research and Health*, 24(2).

Spreng, R. N. and Turner, G. R. (2021). From exploration to exploitation: a shifting mental mode in late life development. *Trends in Cognitive Sciences*, 25(12):1058–1071.

Thompson, M. J. (2021). Piaget's stages of cognitive development and erikson's stages of psychosocial development. In *Child and Adolescent Mental Health: Theory and Practice*.

Todd, G., Candadai, M., and Izquierdo, E. J. (2020). Interaction Between Evolution and Learning in NK Fitness Landscapes. In *The 2020 Conference on Artificial Life*, pages 761–767, Cambridge, MA. MIT Press.

Walasek, N., Frankenhuis, W., and Panchanathan, K. (2021). An evolutionary model of sensitive periods when the reliability of cues varies across ontogeny. *Behavioral Ecology*.

West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences*, 102(suppl_1):6543–6549.

Wilke, C. O. and Martinetz, T. (1999). Adaptive walks on time-dependent fitness landscapes. *The American Physical Society*, 60(2):2154–2159.

Two Theories of Responsiveness

Jonathan Bowen[1,2]

[1]Western University
[2]Rotman Institute of Philosophy
jbowen23@uwo.ca

Abstract

Organisms are responsive--they respond to stimuli. This is a unique mode of causation that we usually only ascribe to organisms. What does it amount to? In this talk, I propose two candidate theories of responsiveness. The first is a functional pathway theory according to which organisms that are responsive are organisms with a certain kind of physiologically realized functional architecture. The second is a vital-integrative theory, according to which responsiveness is a capacity of whole organisms to integrate their activity with the environment in such a manner that their needs are met. I will explain the two views and their underlying rationales. Finally, I will argue that these two theories attribute different kinds of causal structure to the organism, and say divergent things about how their activity is organized. Adjudicating between these views could help to resolve a deeper, older debate between mechanistic and organicist theories of the organism. Therefore, we should find ways to test these theories of responsiveness.

What is Responsiveness?

Organisms are *responsive*–they respond to stimuli. What does this mean?

It cannot *mean* simply that they can be causally impacted by things. Billiard balls can be impacted by other billiard balls and caused to move, but the collision of a moving ball is not a *stimulus* to which the stationary ball *responds*. Nor is it sufficient for the impacting body and the impacted body to be putative responders–two armadillos or pill-bugs could undergo the same communication of motion, and these wouldnot be response phenomena either. So what differentiates response phenomena from this pattern of simple cause-and-effect?

On the other hand, philosophers often distinguish between passive and active movements of organisms (e.g. Dretske 1988) There is a difference between something happening to an organism, and the organism *doing* something. It is common for accounts of the active/passive distinction to identify active behaviour with internally caused movement. Consider: an armadillo goes for a walk. Here, the activity seems to have been initiated spontaneously by the armadillo. In the situation's stipulated simplicity, no externalities are cited–the armadillo forms a desire to go for a walk, and then this state initiates the walking. The chain of causation seems to start in the organism.

This cannot be an account of responsiveness either, for two reasons. The first, more obvious reason is that some responses are instances of reflex action, which on most accounts is not something an organism actively does. But more fundamentally, responses are responses-*to* stimuli. We would not call a behaviour or change in bodily state a response if it truly were initiated without any essential connection to things outside of the organism. A tendency to spontaneously behave in ways that show *no* contingency on externalities would not be responsiveness.

So responsiveness seems to lie somewhere in the crux between these two paradigms of causation–the mechanical cause-and-effect picture, and the internally-caused movement accounts of agency. Like the paradigm of simple cause-and-effect, the response is in some way causally contingent upon the stimulus. But like the paradigm of activity, the response is something the organism does and not something that happens to it. Neither of these concepts are sufficient as an analysis of responsiveness; it needs an analysis of its own. So what essentially is responsiveness?

In this talk I will propose that in psychology there have been two basic underlying theories of responsiveness–a functional pathway theory of responsiveness and a vital-reorganizational theory of responsiveness. I will explain the two views and their underlying rationales. Finally, I will argue that adjudicating between these views could help to resolve a deeper, older debate between mechanistic and organicist theories of the organism. Therefore, we should find ways to test these theories of responsiveness.

The Functional Pathway Theory of Responsiveness

On the functional pathway conception of responsiveness, the stimulus is to be identified with energy applied to parts of an organism called its *receptors*, and the response is to be identified with activity at other parts of the organism called its *effectors*. This input theory of the stimulus and output theory of the response presupposes an input-output conception of the organism.

This scheme for understanding responsiveness has been the dominant causal theory of responsiveness since Descartes' description of the reflex arc, and it will likely be recognizable to the point of seeming truistic today. It is a broadly mechanistic theory, and is deeply embedded in the protocols of experimental psychology. Because of its general acceptance I will explicate it by first briefly describing its development in the history of experimental psychology (I have drawn largely from Boring's 1932 history) and philosophy of mind. Then I will give a formalized statement of it.

From discrete structures to localized functions

The first point is that when we examine our anatomy, we find it comprised of conspicuously distinct discrete structures–we have structures at our extremities which we call *receptors* or sense-organs, which are connected to distinct long cord-like structures called nerves, which are connected to a brain, which is connected to more nerves, which finally terminates in muscles and viscera. It is natural to ask of all these distinct structures what they do and how they work. Nerves all share the same gross basic anatomical features, but Bell and Magendie experimentally demonstrated that they have different functions in the organism in virtue of what they are connected to. Sensory nerves connect sense organs with the brain; motor nerves connect the brain to muscles.

From local function and forward direction of nervous impulse to the reflex arc

Furthermore, there seems to be a *forward direction* in the nervous system–impulses travel in only one direction in each neuron, and the nervous system is just a complex network of these neurons. There may be reafferent connections within the brain, but especially *between* the functional parts of the arc impulses that begin at the receptors travel in a sequence through an arc. So it seems that the whole sequence that culminates in movement follows an *arc*, or a functional pathway.

This distinction of sensory and motor nerve function was used to explain reflex movements. If you remove the brain of a vertebrate but leave the spinal cord, that vertebrate is still capable of reflex activity. In these cases the functional pathway that produces the reflex movement is something like this: sense receptors lead to sensory nerves, which lead to spinal interneurons, which lead to motor nerves, which lead to effector organs. This sequence is called the *reflex arc*. When the brain is not bypassed, another central component is added. The brain, being an elaborate network of more neurons, has its function circumscribed for it by its place in the functional pathway. It needs to produce outputs which culminate in responses on the basis of inputs which come from transduced energy at the receptors. Whatever intellectual, emotive, and volitional capacities there are, they need to be assimilated to a structure whereby impulses enter and exit.

From the reflex arc to the organism as an input-output machine

In philosophy of mind, at the start of the cognitive revolution, there was a movement confoundingly named "functionalism" according to which the physiological structures which gave rise to the reflex arc were seen to be mere *implementation details*. The basic feedthrough conceptual model of the reflex arc was retained, but without essential reference to the physiological structures which gave rise to it. Take Putnam's early arguments against identity theory and formulation of his alternative *machine functionalism* (1967). Putnam explicitly takes the Turing machine as a "model of the organism." Creatures with mental states, on this theory, could be modeled as Turing machines that A) were probabilistic, and B) implemented certain transition tables. The particular physiological facts which gave rise to the reflex arc–the neurons with their forward motion, the central parts with local functions–were abstracted away from the model of the organism, but fundamentally the transition tables are a function from inputs and internal state variables to internal state variables and outputs. The mediating mechanisms may be unspecified, but the fact remains that they are *mediating mechanisms*–mechanisms that produce movements at local functional parts called effectors on the basis of other local functional parts called receptors (or "sensors.")

Formalizing the input-output machine: the functional pathway theory of responsiveness

So there are three conceptual parts to an organism considered as an input-output machine: the input, central processor or controller, and the output. These parts have localized functions and are connected in a sequence via input-output relations, so they can be considered components of a componential mechanism. What must each of these components be?

First, there is the *sensor.* The sensor needs to do two things. First, it needs to *transduce* energy–convert heat, light, or something else that can locally affect a sensory structure of the relevant modality into a format compatible with the next component. Second, it needs to *transmit* this signal to the next part of the functional pathway. Both of these facets are necessary. Disrupt the sensor's transducing function, and it will not produce informative input signals; disconnect the sensor from the sensory nerve or the central component and no amount of transduction will constitute an input. The sensor needs to *play its role in the functional pathway*; it needs to be adjusted appropriately to the central component for it to be a source of input.

Second, there is the *effector*. The effector also needs to do two things. First, it needs to receive output from the central component as input. Second, it needs to transform its input signal into some kind of motion (an "effect"). The obvious example in an animal is a muscle cell that contracts. A robot car's motor would be another example. Again, both of these conditions need to be met for an effector to be an effector; if it does not receive input from the central component (or at least from upstream in the functional pathway!) it is not functioning as an effector, and if its movement is not a function of that input, it is not motor *output*.

Third, we have the *central component*. The central component takes inputs from the sensors, transforms or assimilates them somehow, and produces outputs for the effectors. No further *conceptual* requirements are strictly necessitated by the general input-output conception of the organism, but the adaptive or plastic qualities of responsiveness can be explained in terms relating inputs to outputs–perhaps in computational terms.

Responsiveness as activity produced by a functional pathway

With the functional pathway conception in place, we can now state what responsiveness, stimulus, and response are. On the functional pathway theory of responsiveness, then, a response is an activity produced by effectors, where those effectors receive inputs from a central processor, and that central processor receives inputs from sensors. In this view, a stimulus is an application of energy to the sensor. An organism is responsive when at least some of its activities are responses defined in this sense.

The Vital-Integrative Theory of Responsiveness

The *vital-integrative* theory of responsiveness has it that responsiveness is an attribute of a whole, intact organism whereby it establishes and maintains functional relationships with its environment in a manner that meets its vital needs; a *stimulus* is a breakdown of the organism's integration with its environment consisting of a noticed threat or opportunity that the organism has not adapted to. This is the beginning phase of a coordination; a *response* is the final phase of that coordination which establishes a new state of integration with the environment incorporating the stimulus. I will explain each of these features in turn.

Starting with the whole organism in its conditions of living

The second theory of responsiveness begins to be expressed after the development of the reflex arc scheme with psychology's second school, the Functional Psychologists. It was expressed in part and with varying degrees of explicitness by others as well, including American pragmatists, organismic theorists, and some humanistic psychologists. These schools of thought were broadly united in their methodological commitments, with emphases on the pattern of life of the whole, intact organism (e.g. Holt 1915, Goldstein 1934, Maslow 1943a, 1943b) and the ways in which the organism adapts and integrates into its conditions of living (e.g. Dewey 1896, Angell 1903, Goldstein 1934, Maslow 1943c). Because of their shared methodological commitments, these schools are largely compatible with each other, but collectively fell out of fashion around the end of the Second World War. It is difficult to find an explicit formulation of this theory that synthesizes across these schools, but their insights could be recovered and synthesized to form a second theory of responsiveness, which could be descriptively named a vital-integrative theory of responsiveness. I will describe some of the principles it offers for understanding responsiveness and gesture towards possible ways in which they could be integrated into a resuscitated 21st century theory.

Organisms have needs, and survival requires meeting them

Darwin's theory of natural selection starts with populations of reproducing organisms. It explains the origin of *species*–that is, the emergence of particular *kinds* of organisms over time. Its operands are organisms that survive, vary, and reproduce. Before it can be used to explain the emergence of more complex forms of responsiveness, there needs to be a population of organisms that vary, reproduce, and inherit. What can be said in general about what an organism is, and what it means for an organism to *survive*?

One proposed general feature of organisms that has been relatively uncontroversial at least since Schrodinger's essay (1944) has been the fact that they seem to resist entropy. They persist as complex objects in a way that seems to subvert generalizations of statistical physics, namely, the tendency towards maximum entropy. Organisms don't just resist breakdown, they even repair themselves and *grow*. This is generally accomplished by "drinking orderliness," i.e. taking in nutrients that can be converted into usable energy and assimilated into the organism's own body. These assimilative processes are given the name *metabolism*, and metabolism is a necessary condition of living. When these vital processes stop, the organism *dies*, and while they are operating, the organism is *alive*.

Organisms are not just alive–which they could be while under artificially contrived conditions of life support. They *survive*, which means that they *stay* alive despite obstacles, struggles, and dangers from without, and the consequences of their own metabolic processes from within. And in any realistic environment there are many such dangers. Walter Cannon described the exceedingly narrow parameters under which an organism's vital processes can operate (Cannon 1932). If an organism's internal temperature leaves this narrow range, or if there is a deficiency of some nutrient, or if there is not enough water, it *dies*. This is what a *need* is--a necessary condition for the organism's vital processes to continue, i.e. for the organism to continue surviving. When needs are met, we say that the need is *satisfied*. (Thus defined, the notions of need and satisfaction are not anthropomorphic, but organismic.)

Furthermore, physiological needs like the above are not the only kinds of needs. *There are as many kinds of needs as there are ways to thwart the vital processes in a particular environment*. So in addition to parameters of the internal milieu that may be controlled by internal regulative mechanisms and feeding, drinking, or expulsive behaviours, there are also necessary parameters for the external milieu. (Goldstein 1934). If there are predators which can consume the organism, that organism now additionally has *safety needs* that can be met by, in one way or another, avoiding being eaten by these predators. If the organism's form of life is social and its survival depends on its standing in a community, then it will also have social needs (Maslow 1943b). From the intrinsic vulnerability of the metabolic process and the complexity of the environment, the variety and extent of needs *balloons*. How is the organism to negotiate its standing amidst this seemingly endless set of dangers, and opportunities? This question can be partially addressed from observing their activity. When we do that, we notice that the vital processes themselves incorporate aspects of the environment.

Life processes constitutively involve pivotal outer objects

The reflex arc model itself is *internalist*–it explains responsiveness in terms of processes solely inside the organism. But Holt stresses that behaviour is organized around a *pivotal outer object*. The organism, Holt says, "while a very interesting mechanism in itself, is one whose movements turn on objects outside of itself, much as the orbit of the earth turns upon the sun; and these external, and sometimes very distant, objects are as much constituents of the behaviour process as is the organism which does the turning" (1915). To even characterize what a response is, he argues, you need to ask what aspect of the world that response is an adjustment to.

What this means is that the vital-integrative theory of responsiveness is not purely internalist. But neither is it purely *externalist*. It is rather that the physiological principles are *relational* in nature and involve organizing to or around external things. The external things can thereby become constituents of the vital processes themselves.

This either allows us to extend our conception of the vital processes beyond metabolism, or to extend our conception of metabolism out into the world. As Dewey puts it, we can conceive of the processes of living as "enacted by the environment as truly as by the organism; for they *are* an integration." (1938)

If vital processes *are* an integration with their environments, and activity can be a function of arbitrarily distal, complex, and abstract features of the environment, then perhaps they could approach the task of meeting the organism's ballooning set of needs. Here is an example of Holt applying this logic to a bee:

> The fact is that the specific object on which the bee's activities are focused, and of which they are a function, its ' home,' is a very complex situation, neither hive, locality, coworkers, nor yet flowers and honey, but a situation of which all of these are the related components. In short we cannot do justice to the case of the bee, unless we admit that he is the citizen of a state, and that this phrase, instead of being a somewhat fanciful metaphor or analogy, is the literal description of what the bee demonstrably is and does (1915).

As Holt notes, there is a tendency to read such a claim as that the bee integrates into its home is a mere whimsical description. But on the contrary, if the vital-integrative theory is correct, it is actually a direct statement of a complex relational biological causal process. Perhaps it is a shorthand of an organicist equivalent of a mechanism sketch.

So if the activities of an organism are a kind of integration, what are the stimulus and the response? What does responsiveness amount to?

Life processes are directional

On the vital-integrative theory, responsiveness takes place against the backdrop of the ongoing, total integration an organism has with its environment. Living is described as a process of integration with environmental need-satisfiers and threats. One way that this could work would be akin to early descriptions of homeostatic processes–just as nutrient concentrations, temperature, and the like need to be held constant, so, we might think, an organism might act to keep all of its relationships with important aspects of its environment about the same.

But organisms do not just maintain a static form of integration. Another general feature of organic activity is its *direction*, which Goldstein claims is "the essential characteristic of every vital phenomenon" (1934). This direction manifests itself markedly in at least two ways–first, in the tendency of organisms to *recover* from injuries, which radically alter their state of integration with the environment. When this previously established pattern of living is disrupted by a brain injury, the organism tends to re-establish a *new* total pattern of integration with its environment consistent with meeting as many needs as it can (this process is described in great detail in Goldstein's 1934 book The Organism: A Holistic Approach to Biology Derived from Pathological Data in Man). The second way that direction is manifest in organisms concerns the sheer complexity of the problem of meeting their many needs in their specific environments. This seems to improve over the course of an organism's life; Dewey and Maslow call this process *growth*. Not all needs are equally crucial for survival, and from this fact it is predictable that a hierarchy of "prepotent" needs would emerge–wherein some needs are more basic than others, but once they are adequately met, the organism becomes motivated to satisfy the less-crucial needs. The process of meeting increasingly many needs is the process of integrating more and more with one's environment.

So living is integrating with the environment in a manner that meets the organism's needs. Armed with this conception of the life process, we may be in a good position to interpret Dewey's pregnant but somewhat obscure alternative to the reflex arc in his famous "The Reflex Arc Concept in Psychology" (1896).

Stimulus and response as phases of adaptation

Dewey argues that it was a failure of interpretation to assign psychological notions like perception, cognition, and action to local activities of the reflex arc analysis' component mechanisms. *Sensation* or *perception* are not names of transduction at sensor surfaces; *cognition* is not a name for an associative or integrative process carried out in a central nervous system, and *action* is not a name for movement generated at effectors. In Dewey's memorable phrase, this leaves the organism "a patchwork of disjointed parts, a mechanical conjunction of unallied processes". (Dewey 1896). Instead, perception, cognition, and action are "divisions of labour, functioning factors within the single concrete whole."

As an illustration, consider the act of *chasing*. Is this activity motor? Certainly–a chaser needs to move. But chasing is also sensory. If it were not, in which direction would one chase? Without the perceptual guidance inherent to chasing, one may dart away from one's chasee, rather than towards them. And what if the chase involves some tact, wile, and prediction of the chasee's behaviour? Then the whole activity would be cognitive as well. The "single concrete whole" that has perceptual, cognitive, and motor functionality is what Dewey calls a *coordination*; another term he uses for this unit is "act".

Stimulus and response don't take place at different parts of the organism either. All of the organism is stimulated, and all of the organism responds. They, too, are teleological distinctions, or "parts played with reference to reaching or maintaining an end" (Dewey 1896). What they describe are *stages* in the process of adapting the organism's functional relationships with the world. What are these stages?

Prior to stimulation, the organism has a certain standing in its environment–it is well-fed, or safe, or otherwise is not motivated by any needs related to the stimulus. When the organism is stimulated, some aspect of the situation becomes *problematized* to the organism–that is, there is some aspect of the situation that is pertinent to its ability to meet its needs that it has become aware of. From the moment of stimulation, the organism begins to coordinate with the stimulus. Through a process that Dewey calls "inquiry" in the human case, but ascribes to all responsive entities as well, the organism determines the means of adapting to the stimulus. In the response stage, those means are executed, and the organism thereby establishes an integration with the stimulus such a way that its needs problem is solved and the discomfort dissipates. At the end of the whole sequence of events, the organism has an enlarged total coordination with the environment.

So which theory of responsiveness is better?

The functional pathway theory of responsiveness and the vital-integrative theory of responsiveness describe and explain responsiveness in fundamentally different ways. The functional pathway view seems to follow a *mechanistic* style of explanation, where parts of an organism are decomposed into components with relatively localized functions. The nature of and prospects for mechanistic explanations in the life sciences and cognitive science have been a topic of intense interest in recent history. (Craver 2007; Bechtel & Richardson 2010). The vital-integrative view, in contrast, seems to be committed to a kind of *organicist* and processual biology, which have both seen some recent interest (e.g. Gilbert & Sarkar 2010; Nicholson & Dupre 2018).

One might still wonder: do these theories fundamentally clash, or are they in any way reconcilable? Might these be different ways of describing the same phenomenon, or orthogonal but equally valid kinds of explanation?

One thing is clear. These two theories ascribe causal patterns with radically different profiles to the organism. They also make divergent prescriptions for experimental psychology. The functional pathway picture is concerned with how application of energy at one part of the organism propagates through the system and culminates in movements at another part of the system. This means that we can precisely control the properties of the stimulus. It also means that stimuli can be *applied to* an organism. It does not matter what state the organism is in, or whether the stimulus culminates in a response. As long as the components of the functional pathway are working, an application of energy at the receptors counts as a stimulus. The responses, likewise, are patterns of effector activity that can be described without any essential reference to their stimuli or their manipulanda. The causal pattern described is between activity at sensors and activity at effectors, with intervening activity in the central component. No essential reference is made to the needs of the organism (though it can be an additional fact that the central processes modulate their output in a manner that represents those needs, or happens to conduce to meeting them.) This is not to say that an input-output machine could not meet its needs in virtue of operating as an input-output machine, but that need-satisfaction is incidental to the *causal* structure of responsiveness.

On the vital-integrative picture, on the other hand, it is more difficult to precisely control whether something is a stimulus. Only if something induces the organism to modify its relationships with the need-relevant affordances in its world is that thing a stimulus, and the activities of the organism a response to it. Accordingly, stimuli cannot just be *applied* to an organism. The only way to design a stimulus for an organism is to know something about its needs, its means of attaining those needs, and its disposition to employ them. The causal pattern described is one that is distributed broadly throughout the organism and involves understanding how the organism is integrating itself into its world, and integrating the world into itself in a manner that meets its needs.

How would you synthesize responsiveness?

Which theory of responsiveness we adopt determines what sorts of entities we are trying to make, what their essential organization is, and how we might go about making them. I will say a brief word about what it would mean to create artificial responsiveness according to both theories.

Artificial functional pathways

At the outset, it is clear that the functional pathway system not only follows a familiar style of explanation (i.e. mechanistic), but synthesizing an entity with a fully operational functional pathway also is a familiar process. Any robotic system which has distinct sensors, effectors, and central components that are appropriately wired to each other instantiates this functional architecture. It has a number of advantages. First, the parts can be manufactured and calibrated independently, and assembly would involve connecting these components. Importantly, the outputs of earlier parts of the functional pathway need to be adjusted to the inputs of the next part–so if the central component performs an information processing function, it needs to accept the format of the sensor output; and likewise, the effectors need to accept the format of the output from the central system.

What would be the asymptotic limit of success with such an approach? It would be to implement any possible function between energy at receptors and movement at effectors.

Artificial vital integrators

It is clear at the outset that the vital-integrative theory claims that responsiveness requires a kind of organization that is atypical of current artifacts (although perhaps that could change!) The first thing to do is create an entity with a metabolism that *survives*. Since we are designing these entities, perhaps that gives us some freedom to diverge from the vital processes we are familiar with on earth. But survival at the very least means persisting against the forces of entropy, which means that they, like us, need to be able to take in sources of energy and use that energy. They also need to be able to assimilate this material into their own body.

Once we have an entity with artificial metabolism, we have ipso facto created an entity with *needs*. Those needs already include at least physiological needs and safety needs. To create minimal responsiveness, all we need to do is alter the matrix of vital processes so that its activity is a function of its needs in a manner that conduces to meeting them. That may not require much–perhaps it is a matter of swimming around randomly until the environing nutrient gradient is sufficient (which may work in a relatively limited milieu). Or maybe that requires sensitivity to information in the ambient array that specifies need-satisfying affordances, and coordinated exploratory movements towards them (which may work in a relatively expansive milieu.) We can be as creative and elaborate, or as frugal and ad hoc as the evolutionary process is allowed to be. The ground truth is in whether the organism survives in an environment full of dangers and opportunities.

What would be the asymptotic limit of success for such an approach? This being would meet all of its needs, and thereby survive in its environment.

References

Bechtel, W. and R.C. Richardson, 2010 [1993], Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research, Second Edition. Cambridge, MA: MIT Press/Bradford Books.

Cannon, W. (1932). The wisdom of the body. W W Norton & Co.

Dewey, J. (1896). The reflex arc concept in psychology. Psychological review, 3(4), 357.

Dewey, J. (1938). Art as Experience. New York: Minton, Balch & Company

Dretske, F. (1991). *Explaining behavior: Reasons in a world of causes*. MIT press.

Gilbert, S. F., & Sarkar, S. (2000). Embracing complexity: organicism for the 21st century. Developmental dynamics: an official publication of the American Association of Anatomists, 219(1), 1-9.

Holt, E. B. (1915). The Freudian wish and its place in ethics. H. Holt and Company.

Machamer, P.K., L. Darden, and C.F. Craver, 2000, “Thinking about Mechanisms”, Philosophy of Science, 67:1–25.

Maslow, A. (1943a). The dynamics of personality organization, I and II. Psychological Review (50), 514-539, 541-558.

Maslow, A (1943b). Preface to motivation theory. Psychosomatic Medicine, 5, 85-92.

Maslow, A. (1943c). A theory of human motivation. Psychological Review 50(4), 370.

Nicholson, D. J., & Dupré, J. (2018). Everything flows: towards a processual philosophy of biology (p. 416). Oxford University Press.

Okasha, S. (2018). *Agents and goals in evolution*. Oxford University Press.

Putnam, H. (1967). The nature of mental states. Readings in philosophy of psychology, 1, 223-231.

Schrodinger, E. (1944). *What is life?*. Cambridge University Press, New York.

Tolman, E. C. (1932). *Purposive behavior in animals and men*. Univ of California Press.

Walsh, D. M. (2015). *Organisms, agency, and evolution*. Cambridge University Press.

Self-recognition as Optimisation

Timothy Atkinson and Nihat Engin Toklu

NNAISENSE, Lugano, Switzerland
{timothy, engin}@nnaisense.com

Abstract

We tackle the concept of 'self-recognition' in a simulated setting. We propose an experiment where two simultaneous reinforcement learning environments are controlled by two agents. Although each agent is given the control of its own environment, both agents receive the visual input of the *same* environment. The success threshold depends on self-recognition by definition as the agent must answer: am I seeing a mirror, or am I seeing a camera? We show that this experiment can be posed as an optimisation problem, solvable via evolutionary computation.

Introduction

Self-awareness is defined as "the capacity to become the object of one's own attention" (Morin, 2006). Self-awareness is studied also within the context of computational systems (Lewis et al., 2011). Here, we focus on *self-recognition*, which has been suggested as evidence of self-awareness (Gallup Jr, 1998), or seen as the implication of "some form of self-awareness" (Morin, 2011).

The classic 'mark-test' mirror test (Gallup Jr, 1970), used to detect self-recognition in animals equipped with visual stimuli, broadly proceeds as follows: (*a*) mark the animal with a dye (somewhere not directly visible to the animal) (*b*) observe the animal's frequency of touching the dyed area; (*c*) place a mirror in front of the animal (allowing the animal to see the dyed area) (*d*) observe again the animal's frequency of touching the dyed area. Observing a significant increase in the frequency of touching the dyed area is interpreted as the animal seeing itself in the mirror, observing the presence of the dye and therefore touching it. It is then argued that the animal must have an internalised notion of 'self' as to recognise the being in its visual stimuli as itself. A number of animals have been reported to pass various forms of the mirror test, including orangutans (Suárez and Gallup Jr, 1981), dolphins (Marten and Psarakos, 1994) and magpies (Prior et al., 2008).

Quoting Plotnik et al. (2006), the subject's behaviour is characterised as: (*i*) *social response*; (*ii*) *mirror inspection (looking behind the mirror)*; (*iii*) *repetitive mirror-testing behaviour (where the animal observes the effects of its own actions on the mirror image)*; and (*iv*) *self-directed behaviour (recognition of the mirror image as self)*. We argue that the most interesting parts of this process are (*iii*) and (*iv*). Therefore, we propose a new experimental setting based on reinforcement learning (RL) where expected reward can only be maximised through mirror-testing and ultimately self-recognition in a mirror. Our rewarding mechanism has a clear threshold which can only be surpassed through self-recognition. Experiments demonstrate that existing evolutionary algorithms can readily surpass this threshold. While similar efforts have been made for self-recognising artificial intelligence (Haikonen, 2007; Winfield, 2014; Pipitone and Chella, 2021), these were not framed as a pure optimisation problem.

The source code for our study is available online [1].

Self-recognition in Reinforcement Learning

We propose an experimental setting with two simultaneous RL environments: the 'mirror environment' E_M and the 'camera environment' E_C. Using the same policy (neural network *and* weights), two agents A_M and A_C operate in E_M and E_C, respectively. The distinctive feature is that both A_M and A_C observe E_M (i.e. receive their inputs from E_M). We argue that, by observing E_M, A_M can be thought as 'seeing' itself in the mirror (since its inputs are its own environment and 'body'). Meanwhile, A_C will be seeing the other agent's environment and body. A_M gets rewarded for completing the assigned task in E_M while A_C gets penalised for producing non-zero actions. To reach success, A_M must realise that it is observing itself ('passing the mirror test') and act to complete the task, while A_C must realise that it is observing another agent and stop. Reaching this realisation is non-trivial: initially, an agent has no way of knowing whether or not its observations are coming from the mirror. Therefore, the agent must 'understand' in time if its own actions cause updates on observations. We pose this framework as an optimisation problem of seeking policy parameters which maximise the total reward obtained by A_M and A_C.

[1] https://github.com/nnaisense/self-recog-as-optim.git

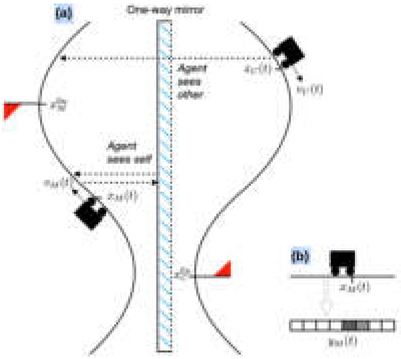


Figure 1: **(a)** The set up of the two environments E_M and E_C, with the agent seeing itself in E_M and the agent seeing the other environment in E_C. **(b)** Conversion of position $x_M(t)$ to visual input $y_M(t)$ passed to both agents.

Our experiments modify the classic mountain car task (Singh and Sutton, 1996). Contained in each environment E_i, $i \in \{M, C\}$, is a car at time t with position $x_i(t)$ and velocity $v_i(t)$. The state of agent A_i in E_i is denoted $s_i(t)$. The system is visualised in Figure 1. In each iteration $t+1$, the agent outputs actions $a_i(t+1) \in [-1, 1]$ which is integrated into the environment state according to,

$$v_i(t+1) = v_i(t) + \alpha a_i(t+1) - \beta \cos\left(3 * x_i(t+1)\right),$$
$$x_i(t+1) = \mathtt{clamp}\left(x_i(t) + v_i(t+1), -1.6, 0.6\right),$$

with $\alpha = 0.001$ and $\beta = 0.0025$. Initial conditions are set as $x_i(0) = -\frac{1}{2}$ and $v_i(0) = 0$ such that their initial dynamics are identical. The critical difference between them lies in the observations. *Both* agents receive a 32-variable visual representation $y_M(t)$ of $x_M(t)$ (see Figure 1b), such that $a_i(t+1) = \mathtt{Policy}(y_M(t), s_i)$. The reward in each time step is, $r_i(t) = -\mathtt{min}\left(\mathtt{abs}(x_i(t) + x_i^{\mathrm{fin}})/D^{\mathrm{fin}}, 1\right)$ where $x_M^{\mathrm{fin}} = \pi/6$ and $x_C^{\mathrm{fin}} = -\frac{1}{2}$ and $D^{\mathrm{fin}} = \frac{x_M^{\mathrm{fin}} - x_C^{\mathrm{fin}}}{2}$. A_M is rewarded for maximising being at the top of the hill in E_M, and A_C for staying at the origin in E_C. The episode terminates at $t = 200$.

There exists a threshold expected episodic reward above which we can definitively claim the presence of self-recognition. Specifically, for an agent which *cannot* discriminate between E_M and E_C such that $x_M(t)$ and $x_C(t)$ follow the same distribution, $r_M(t) + r_C(t) \leq -1$, the total episodic reward across both settings is bounded: $R = \sum_{t \in \{0,1,2,\ldots,200\}} r_M(t) + r_C(t) \leq -200$. Therefore, when we see $R > -200$, the agent has learned some degree of self-recognition.

Experiments and Discussion

To solve the proposed environment, we use Separable Natural Evolution Strategies (SNES) (Schaul et al., 2011) with the ClipUp optimiser (Toklu et al., 2020) with initial radius $r = 4.5$. As in Salimans et al. (2017), we do not adapt σ. The policy is a 64-neuron recurrent network with $\mathtt{ELU}$ activation for the hidden layer (Clevert et al., 2015) and $\mathtt{tanh}$ activation for the output layer. Each hidden activation is passed through layer normalisation (Ba et al., 2016). The agent's initial hidden state $s_i(0)$ is randomised, drawn from $\mathcal{N}(0, I_{64})$ and then passed through $\mathtt{ELU}$ activation and then layer normalisation, allowing the agent to take psuedo-randomised actions for mirror testing.

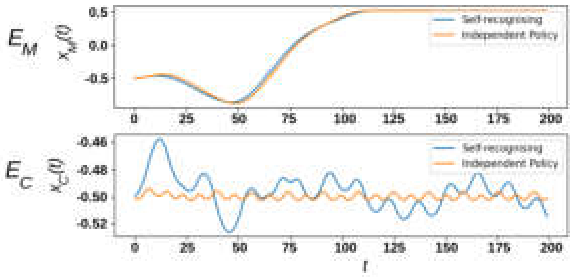


Figure 2: Position trace of learned agents, where 'self recognising' refers to a single policy trained to recognise itself across both environments, and 'independent policy' refers to individually trained policies for both environments. On E_M, both agents climb the hill at a similar rate. On E_C, the self-recognising agent takes much larger movements, suggesting that the agent prioritises E_C for mirror testing. It is worth noting that the self-recognising agent terminates within 0.1 of the goal position in 87% of scenarios, suggesting that performance can be further improved.

The population size is $15,000$ and evolution is run for 4000 generations. Each candidate solution is evaluated for total episodic reward R when controlling each environment E_M or E_C in turn. The experiment is repeated 10 times. For comparison, we run the experiment in a setting with an independent policy for each E_i. In the self-recognition scenario, the median generations for the mean fitness of the population to surpass the threshold of -200 is **967**, demonstrating task solvability with existing techniques. This is substantially higher than the **141** generations needed with independent policies for each E_i meaning that the introduction of 'self-recognition' substantially complicates the overall task. This difference is statistically significant from the non-parameteretric two-tailed Mann-Whitney U-test (Mann and Whitney, 1947) ($p = 10^{-4} \leq 0.05$) and the effect size is large according to the non-parametric Vargha-Delaney A Test (Vargha and Delaney, 2000) ($A = 0.96 \geq 0.71$).

While we are not claiming that the learned simple $64-$neuron RNNs are truly 'self-aware', it is interesting to note that the self-recognition test can readily be solved by existing evolutionary algorithms. From a more practical perspective, the substantial performance degradation obtained through the adaption of an existing environment to a self-recognition setting suggests a general template for describing harder, intrinsically hierarchical, RL benchmarks.

References

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint*. arXiv:1607.06450.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint*. arXiv:1511.07289.

Gallup Jr, G. G. (1970). Chimpanzees: self-recognition. *Science*, 167(3914):86–87. doi: 10.1126/science.167.3914.86.

Gallup Jr, G. G. (1998). Can animals emphatize? Yes. *Scientific American*. Feature article: Animal self-awareness: A debate.

Haikonen, P. O. (2007). Reflections of consciousness: The mirror test. In *AAAI Fall Symposium: AI and Consciousness*, pages 67–71.

Lewis, P. R., Chandra, A., Parsons, S., Robinson, E., Glette, K., Bahsoon, R., Torresen, J., and Yao, X. (2011). A survey of self-awareness and its application in computing systems. In *2011 Fifth IEEE Conference on Self-Adaptive and Self-Organizing Systems Workshops*, pages 102–107. doi: 10.1109/SASOW.2011.25.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, 18(1):50–60.

Marten, K. and Psarakos, S. (1994). Evidence of self-awareness in the bottlenose dolphin (tursiops truncatus). In *Self-awareness in animals and humans: Developmental perspectives*, pages 361–379. Cambridge University Press. doi: 10.1017/CBO9780511565526.026.

Morin, A. (2006). Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consciousness and cognition*, 15(2):358–371. doi: 10.1016/j.concog.2005.09.006.

Morin, A. (2011). Self-recognition, theory-of-mind, and self-awareness: What side are you on? *Laterality*, 16(3):367–383. doi: 10.1080/13576501003702648.

Pipitone, A. and Chella, A. (2021). Robot passes the mirror test by inner speech. *Robotics and Autonomous Systems*, 144:103838. doi: 10.1016/j.robot.2021.103838.

Plotnik, J. M., De Waal, F. B., and Reiss, D. (2006). Self-recognition in an asian elephant. *Proceedings of the National Academy of Sciences*, 103(45):17053–17057. doi: 10.1073/pnas.0608062103.

Prior, H., Schwarz, A., and Güntürkün, O. (2008). Mirror-induced behavior in the magpie (pica pica): evidence of self-recognition. *PLoS biology*, 6(8):e202. doi: 10.1371/journal.pbio.0060202.

Salimans, T., Ho, J., Chen, X., Sidor, S., and Sutskever, I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint*. arXiv:1703.03864.

Schaul, T., Glasmachers, T., and Schmidhuber, J. (2011). High dimensions and heavy tails for natural evolution strategies. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 845–852. doi: 10.1145/2001576.2001692.

Singh, S. P. and Sutton, R. S. (1996). Reinforcement learning with replacing eligibility traces. *Machine learning*, 22(1):123–158. doi: 10.1023/A:1018012322525.

Suárez, S. D. and Gallup Jr, G. G. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of human evolution*, 10(2):175–188. doi: 10.1016/S0047-2484(81)80016-4.

Toklu, N. E., Liskowski, P., and Srivastava, R. K. (2020). Clipup: A simple and powerful optimizer for distribution-based policy evolution. In *International Conference on Parallel Problem Solving from Nature*, pages 515–527. Springer. doi: 10.1007/978-3-030-58115-2$_3$6.

Vargha, A. and Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2):101–132. doi: 10.3102/10769986025002101.

Winfield, A. F. T. (2014). Robots with internal models: A route to self-aware and hence safer robots. In *The Computer After Me: Awareness and Self-Awareness in Autonomic Systems*, page 237–252. Imperial College Press. doi: 10.1142/9781783264186$_0$016.

Gradient Climbing Neural Cellular Automata

Shuto Kuriyama[1], Wataru Noguchi[2], Hiroyuki Iizuka[2,3], Keisuke Suzuki[3] and Masahito Yamamoto[2,3]

[1]Graduate School of Information Science and Technology, Hokkaido University, Japan
[2]Faculty of Information Science and Technology, Hokkaido University, Japan
[3]Center for Human Nature, Artificial Intelligence, and Neuroscience, Hokkaido University, Japan
9ri8ma5296@gmail.com

Abstract

Chemotaxis is a phenomenon whereby organisms like ameba direct their movements responding to their environmental gradients, often called gradient climbing. It is considered to be the origin of self-movement that characterizes life forms. In this work, we have simulated the gradient climbing behaviour on Neural Cellular Automata (NCA) that has recently been proposed as a model to simulate morphogenesis. NCA is a cellular automata model using deep networks for its learnable update rule and it generates a target cell pattern from a single cell through local interactions among cells. Our model, Gradient Climbing Neural Cellular Automata (GCNCA), has an additional feature that enables itself to move a generated pattern by responding to a gradient injected into its cell states.

Introduction

Gradient-climbing is a behaviour where an organism perceives environmental differences such as chemical gradients as well as its own internal chemical reactions, and then moves its body in response to those gradients (Suzuki and Ikegami, 2009; Parent and Devreotes, 1999; Wadhams and Armitage, 2004). The couplings between an environment and internal dynamics of a organism cause self-movement that is one of the key features characterizing life forms and separating them from non-life forms (Suzuki and Ikegami, 2009; Pollack et al., 2004; Varela et al., 1974).

Another fundamental feature unique to living things is the ability to generate their own shapes and maintain them through local interactions among the smallest units composing themselves such as a cell or a chemical substance, i.e., morphogenesis. NCA has recently been proposed as a 2D Cellular Automata (CA) model that simulates morphogenesis applying neural networks to its update rule (Mordvintsev et al., 2020). The 2D NCA model has a unique structure where each cell has a 16-dimensional continuous vector as its cell state, and the first four dimensions (channels) represent an RGBA image. NCA generates a target pattern on the visible four channels starting from a single cell through recurrent local interactions among cells.

The 2D NCA model comprises two main phases: Perception and Update. The former is for each cell to perceive its neighborhood's features as well as its own feature and it is implemented by a 3×3-convolution with three fixed kernels $(Sobel_x, Sobel_y, Identity)$. The latter represents a learnable update rule which consists of 1×1-convolutional neural networks (Dense layers) and outputs an incremental update vector to each cell state. In addition, the cell states are probabilistically updated by adding the update states with $p = 0.5$ (Stochastic Update). Finally, the Alive Masking phase tests whether each cell is alive or dead, and then it allows only alive cells to hold their state vectors' values for the next step by setting all the dead cells' states to zero vectors.

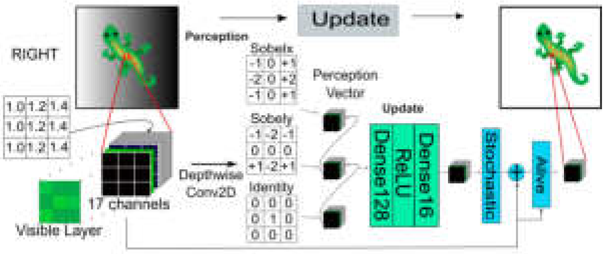


Figure 1: A single step of CA update in GCNCA.

In our research, we simulated the gradient climbing behaviour on NCA. We aimed to build a model that generates a predefined shape of an object on its visible channels and then moves its body in a direction corresponding to a gradient, set on its cell states. The process is carried out by local interactions among internal chemical substances and an environmental gradient.

Gradient Climbing NCA

To simulate self-movement, we modified the Persistent model of NCA by introducing an additional channel to cell states as an environmental gradient that affects the direction of its agent's movement.

We first added the 17th channel to the cell states that represents an environmental gradient. For simplicity, we introduced five types of gradients, zero gradient and gradients increasing in the four directions: right, left, up, and down. A zero gradient has the same value for all grid cells. The

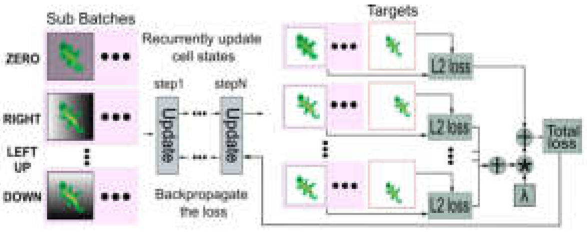


Figure 2: A single training epoch of GCNCA.

rest of the gradients have values increasing towards their direction. The original NCA architecture for CA update—the collection of Perception, Update, Stochastic, and Alive Masking—remains the same in our architecture (See Fig. 1).

For each gradient direction, a set of cell states is sampled as a sub-batch from the Pool holding multiple cell states. A corresponding gradient is set to the 17th channel of each cell states. Then, each sample is updated recurrently for a fixed number of steps while all the values of the 17th channel are constant during an epoch (See Fig. 2).

After the final step of the recurrent updates, the model calculates its loss from these updated cell states. In the original NCA model, MSE (mean squared error) loss was applied between the RGBA channels of a sample and a single predefined target image. Our model, however, has five different target images, each of which corresponds to a sub-batch. Those are an original target image and four additional images shifted by a certain pixels in a direction corresponding to their gradients from the original. The loss function is defined as below.

$$L = L_z + \lambda(L_r + L_l + L_u + L_d), \tag{1}$$

$$L_g = \frac{1}{N_g}\sum_{i=1}^{N_g}\mathrm{MSE}(I_i - T_g),\ g \in \{z, r, l, u, d\}, \tag{2}$$

where L is the total loss, L_z, L_r, L_l, L_u, and L_d are the average losses for sub-batches with a zero, right, left, up, and down gradients respectively. L_g ($g \in \{z, r, l, u, d\}$) is calculated as the MSE between the updated target channels and its corresponding target image; the MSE losses are averaged over N_g samples in the sub-batch. I_i is the RGBA channels of a sample in the sub-batch, T_g is the target image corresponding to the sub-batch. The total loss is backpropagated to the parameters of the convolutional neural networks in the Update phase. A set of cell states in the Pool that were sampled are replaced with the updated samples after they are shifted back by the same pixels as their target images were shifted by but in the opposite direction. Then, model proceeds to the next training epoch.

For the beginning part of its training epochs, only zero gradients were set to all the samples to generate the shape of the object on every sample in the Pool. For the rest of the epochs, all the five types of gradients were used for the training.

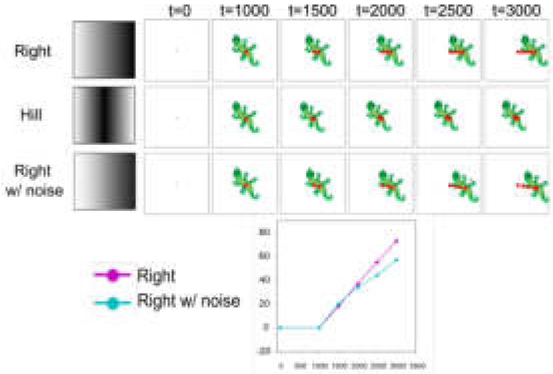


Figure 3: Gradient climbing at step t with a learnt gradient, an unseen gradient, and a perturbated gradient. Each red point is the center coordinates of the object at each step. In the bottom plots, the x-axis is the number of steps and the y-axis is the number of pixels the object moved towards right.

Experiment & Results

We simulated gradient climbing behaviour on GCNCA and demonstrated under three different conditions: the gradients used for the model's training, gradients that it has never encountered, and a perturbated gradient (See Fig. 3).

For the first condition, the object successfully moved in a corresponding direction by responding to the five types of gradients without losing its shape.

Regarding the second point, we applied two unseen gradients, hill gradient and valley gradient. With a hill gradient whose values increase towards the center column from left and right, the object shrank its body towards the middle column and stayed around the center. A valley gradient whose values increase towards left and right from the center column, expanded the shape of the object at the beginning and the object moved to the right after some steps. These results show that the model was trained in a way that the object climbs gradients while maintaining its shape.

For the last, the object processed a gradient with noise that reverses its gradient direction locally at some part of the grid. Nevertheless, the model moved the object by responding to the direction represented by the whole gradient grid.

Conclusion

This work has successfully added the simulation of self-movement to NCA and built a new model GCNCA that can simulate three fundamental features characterizing life forms: self-generation, self-maintenance, and self-movement with self-organization process. Furthermore, the diversity and the robustness of those behaviours were observed by applying unseen or perturbated gradients to our model.

References

Mordvintsev, A., Randazzo, E., Niklasson, E., and Levin, M. (2020). Growing neural cellular automata. *Distill*. https://distill.pub/2020/growing-ca.

Parent, C. A. and Devreotes, P. N. (1999). A cell's sense of direction. *Science*, 284(5415):765–770.

Pollack, J., Bedau, M. A., Husbands, P., Watson, R. A., and Ikegami, T. (2004). Self-repairing and mobility of a simple cell model.

Suzuki, K. and Ikegami, T. (2009). Shapes and self-movement in protocell systems. *Artificial Life*, 15(1):59–70.

Varela, F. G., Maturana, H. R., and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196.

Wadhams, G. H. and Armitage, J. P. (2004). Making sense of it all: bacterial chemotaxis. *Nature reviews Molecular cell biology*, 5(12):1024–1037.

Towards a Unified Framework for Technological and Biological Evolution

Roger Tucker[1][2]

[1]Independent Researcher, Chepstow, UK
[2]Glean, 4 The Boulevard, Leeds, UK
roger.tucker@gmail.com

Abstract

It has long been observed that human cultural evolution is in some ways analogous to biological evolution, having reproduction with variation and a form of selection, but the fact that both technology and biology are physical brings them much closer than culture in general. Many have observed that they share universal traits that pervade their long-term trends, yet they seem so different. What is at the root of this? This paper considers a number of properties that would seem essential to any evolutionary system which produces real artefacts – construction, search, selection and various aspects of structure and organisation – and explains briefly how each operate in technological and biological evolution. This provides an initial attempt at a basic unified framework which can then be extended. Such a framework would help progress bio-inspired design, and suggest features to study on the way to meet the grand challenge of Open Ended Evolution.

Introduction

Technology has been drawing inspiration from biology for many decades, as evidenced by a number of design movements including bionics, biomechanics, biomimicry, biomimetics and bio-inspired engineering, as well as the field of evolutionary computing. More recently, bio-engineering is incorporating biology rather than just being inspired by it, including the examples of artificial organisms produced from biological cells and fuel cells from microbes mentioned in this conference's call for papers. However, less well investigated in this coming together of technology and biology is a clear understanding of the correspondence between their evolutionary systems.

It has long been observed that human cultural evolution is in some ways analogous to biological evolution, having reproduction with variation and a form of selection. Within the study of cultural evolution, technology is seen as an externalisation of knowledge (Mesoudi *et al.*, 2015), with knowledge, construction techniques and technological artefacts evolving in separate but interrelated ways (Brey, 2008).

But the fact that both technological and biological evolution describe physical artefacts or organisms brings them much closer than culture in general. They both have to overcome problems of transport, energy, processing of materials, creation of machinery etc. Many have observed that they share universal traits that pervade their long-term trends, no matter how different they appear at a lower scale (Vogel, 1999; Zinman, 2000; Solé *et al.*, 2013) and ask what is at the root of these potential universals?

Technological evolution is not just the aggregation of all the clever inventions of humankind over a long period of time. In his book "The Nature of Technology", Brian Arthur explains "the collective of technology builds itself from itself with the agency of human inventors and developers much as a coral reef builds itself from itself from the activities of small organisms." (Arthur, 2009, p. 169). He is not overstating the analogy. Technology is part of an interconnected socio-economic system which progresses ever faster, effecting more and more change in our lives, whether we like it or not. What started out as humankind making a few basic tools has become an emergent system which, like all emergent systems, has a top-down impact on its constituent parts.

There are many intriguing analogies in the two evolutionary systems. Vogel names cultural dissemination, natural selection, the role of isolation, conservative bias, the time course of change (though here you have to speed up the timescale for technology), incremental progress, new uses for old devices, parallel developments and extinction (Vogel, 1999). Zinman adds "diversification, speciation, convergence, stasis, evolutionary drift, satisficing fitness, developmental lock, vestiges, niche competition, punctuated equilibrium, emergence, extinctions, co-evolutionary stable strategies, arms races, ecological interdependence, increasing complexity, self-organisation, unpredictability, path dependence, irreversibility and 'progress'" (Zinman, 2000, p. 5). Clearly the two systems exhibit many similar dynamics. To get to the bottom of these, there have been a number of attempts to map concepts from biological to technological evolution to create a similar evolutionary model for technology, reaching different conclusions on the extent that the genotype-phenotype concepts apply, and to *what* they apply (Brey, 2008).

Conversely, there are also many contrasts between technology and biology. There are the obvious ones - technology has a bias towards flat surfaces, right angles, abrupt corners and the ubiquitous wheel and axle – all rarely found in biology. But many contrasts are much more subtle – dry vs wet; the way round levers are used; preference for stiffness vs strength, to name but a few.

A helpful way of understanding the correspondence between the two is to consider *what* needs to be achieved, and then contrast *how* it is achieved. For instance, technology's moving mechanisms – engines – are based on rotation or expansion; most of biology's are based on sliding or contracting. Loading of materials is usually in compression in technology but in tension in biology (Vogel, 1999). Could there be a set of needs common to the evolutionary systems themselves which together can form the elusive universal framework into which technological and biological evolution can both be fitted?

The most significant contrast, which lies at the root of all the others, is that biology is self-building and self-developing whereas *humans* develop and build technology. Any attempt to properly understand the correspondence of the two evolutionary systems has to embrace and accommodate this difference.

A good starting point is to note that individual technologies are not as much a result of human intention as we might think. There has been a long-running debate over how much technological evolution is driven by the "pull" of the market, and how much it is the "push" of human creativity and imagination (Basalla, 1988). From a number of surprising examples, Basalla successfully shows how even the most outstanding inventions are in fact derived from existing technology. Arthur's later analysis expands on and confirms this. Much technology development is hidden internal improvement, driven by inefficiencies or ineffectiveness in component parts. These improvement opportunities emerge from the evolutionary system itself. Even the development of a new item of technology, be it physical or software, is very much driven by the opening up of a new market opportunity. Humans have lots of good and creative ideas, but technology succeeds only when it matches a market niche, and so it is the formation of these niches, at a time when the technology needed to service them can become available, which drives much new technology development (Arthur, 2009).

From this socio-economic system-level view of technology we can characterise the role of human creativity as one of a search for opportunities and better solutions. We can then characterise biological evolution's reproduction-with-variation also as a search, and compare and contrast the way the two operate (Perkins, 2000). In a similar way, we can characterise developmental biology as construction, the genome as a form of learning from experience and so on. Using concepts important to technology alongside those associated with biology provides a richer way of understanding how the two systems correspond to one another.

These high-level concepts provide the top level of a unified framework into which evolutionary system properties that describe *both* systems can be fitted. It is not straightforward to identify these properties - most of the time a property is better known in one than the other, and it may work in a common way, an equivalent (analogous) way or in a contrasting way.

Table 1 gives an overview of the initial framework explained in this paper, summarising the properties considered and their relationship in the two systems, with colours to highlight whether at that level it is best described as commonality, equivalence or contrasting. It begins with two fundamental properties of any evolutionary system producing physical artefacts – construction and search - which are performed in largely contrasting ways in the two systems. The role of selection is then considered, followed by several organisational properties which are essential to technological evolution and which have a common or equivalent role in biological evolution.

This property list is not at all exhaustive and the explanations given in the sections that follow are necessarily very brief, but should be sufficient to establish the feasibility and utility of building a unified framework in this way.

Property	Relationship: Commonality / Equivalence ⇔ / *Contrasting*	Inspired by:
Construction		
Use of Material	*Uniform vs molecular construction*	Biology
Assembly	*Precise vs adaptive; components-first vs progressive*	Biology
Instability	*Avoidance vs management*	Biology
Interoperability	*Biology interoperable at a much smaller level*	Biology
Search		
Type	*Goal-Directed vs Exploratory Search*	Biology
Model	*Direct vs generative*	Biology
Solution Forms	*Predictability vs adaptability; Single vs multi-level*	Biology
Accumulation of Experience	Theory, methods, components and tools ⇔ Genome	Technology
Selection		
Feature Selection	Designer's judgement + selection ⇔ gene recombination + selection	Technology
Role of Environment	Other technology ⇔ biotic / socio-economics ⇔ abiotic	Biology
Organisation		
Modularity	Self-contained units, weakly coupled to everything else, allow independent development and optimisation	Technology
Hierarchy	Reduction in complexity of design ⇔ simplification of control	Technology
Emergence	Something new emerges at a higher level	Biology
Combination	Filling in gaps in possibility space; novel combination. *Recombining existing genes within an organism vs combining anything*	Technology
Re-use	Modular re-use / copy & modify re-use	Technology

Table 1: Overview of a Basic Unified Framework for Biological and Technological Evolution

Construction

Use of Material

Technologists make their component parts by liquefying a material so it can fill a mould, softening it so it can be shaped, removing unwanted parts or by putting down thin layers (as in 3D printing). The designer then fits these components together to achieve the desired structure. This process may be very sophisticated as in semiconductor fabrication, but the principle remains that technology is made from uniform materials.

Whereas biology self-builds using the fundamental properties of matter to create its structures one molecule at a time. At the lowest level these properties are those of the quantum world, familiar to us through the rules of chemistry; at a molecular level it is a combination of those properties and the physical structure caused by the way a protein folds; at a cellular level it is those properties combined with the properties of soft matter (Newman and Bhat, 2008).

From this one difference stem many others. Although many of biology's fundamental design challenges are shared with technology, their readily available solutions are different to those of technology (Vincent *et al.*, 2006).

Use of resources. This difference in types of solution is clearly demonstrated in the use of resources. Technology is consumptive of resources, some of which are starting to run out. Biology ultimately utilises just sunlight and geothermal energy and makes everything by chemical transformation, with one organism's waste becoming another's resource.

Assembly

Precise vs adaptive. Technology depends on every component being made and assembled with reasonable precision. Biology makes extensive use of exploratory processes, which first generate a very large amount of functional variation, often at random, and then select or stabilize the most useful ones, with the rest disappearing or dying back. Microtubule structures within cells, the vascular system, the nervous system, neurons in the brain – all develop in this way (Gerhart and Kirschner, 2007). Exploratory processes are fundamentally adaptive and allow genes to make changes through simple regulatory control.

Components-first vs progressive. In technology, all the components of a design are made first and then assembled according to the design, which is a direct representation of the item being built. If the design is altered then several, or many, parts need to be changed simultaneously. This adjustment to the design requires a higher level of abstraction of the design space that enables a single change to affect all the necessary parts - an abstraction which usually requires a human intelligence.

Whereas in developmental biology multicellular organisms are built progressively, by starting with a single cell, and differentiating the cells as they divide and reproduce, allowing large changes to be made in the final result by smaller, perhaps single, changes earlier on in the development cycle.

Managing Instability

Engineers choose materials and make designs that give total stability and reliability, with maintenance as infrequent as possible whilst still ensuring adequate performance (notwithstanding the modern trend for built-in obsolescence). In contrast, there is a real sense in which biology lives on the edge of instability. Chemical reactions normally move towards a state of thermodynamic equilibrium, but in a cell everything is kept away from this equilibrium, and it is complex feedback systems which maintain a stable but dynamic equilibrium. When these are perturbed, perhaps through a mutation, it is easily possible to go closer to instability, which is normally destructive but might just occasionally prove beneficial and so generate innovation. The instability of the proteins that make up cells, especially in warm bodied animals, requires them to be continually re-synthesized and replaced, with an average half-life for an adult human of about 80 days (Vogel, 1999).

Nature manages instability at many different levels. At one end of the scale, in the avian compass an entangled pair of electrons somehow survive in the face of decoherence for at least a microsecond (Al-Khalili and McFadden, 2014). At the other end of the scale, the instability of the earth's environment (e.g. earthquakes, forest fires) can force the whole local ecosystem to adapt, or very rarely (e.g. a large meteorite strike), the global ecosystem adapts, removing dominant species and allowing others to develop.

Interoperability

Technology's uniform-material approach bypasses a problem that biology's molecule-by-molecule approach has had to solve, that of the interoperability of its basic building blocks. In fact, the low-level building blocks of life - nucleic acids, proteins and regulatory circuits - are very interoperable. Even though each amino acid in a protein has a different shape, the chemical linkage between them is identical. Nucleic acid strands also use a standardised chemical bond to link their component nucleotides. Regulation circuits use a standardised way to regulate genes based on the principle that regulator proteins interact with specific nucleotide sequences on DNA (Wagner, 2014). Since cell structures are self-built from monomeric components such as amino acids, structures can easily be altered by adjusting just one or more of those molecules.

Mature technology does eventually become standardised, but even then it does not mean that it is easy to fit things together. Parts that fit together have to be designed specifically to do that.

Search

As mentioned in the Introduction, the concept of search is a way to accommodate the fundamental difference that humans develop and build technology, but biology is self-building and self-developing. Perkins introduces three general search strategies that can be observed both in biological evolution and in the development of technology - adaptation by revision, selection and coding (Perkins, 2000). He stresses that whilst these may involve human consciousness, intentionality and imagination, these are not required, and both technology (and human creativity in general) and biological evolution make use of all three strategies. However, the focus here is not on these commonalities, but on what is fundamentally *different* between technology and biology – the role of human intention.

Exploratory vs Goal-Directed Search

What makes nature's and technology's searches distinctive can be best explained in terms of intention. Technology mostly progresses by what can be described as "goal-directed" search – technologists look for a way of achieving a specific outcome within various practical constraints, maybe in small easy steps (as in the modern Agile methodology) or maybe with an ambitious goal which is not easy to realize.

By contrast, nature's reproduction-with-variation and selection, from a population perspective, amounts to a high-level exploratory search yielding improvements that offer better reproductive success. This is usually in small steps, either through improving the effectiveness of an existing capability, or occasionally introducing a completely new one. It is a stochastic search where many different search trajectories are explored in parallel lineages to establish useful traits.

Whereas technology's goal-directed search involves individual designers and only a little time, this high-level exploratory search involves whole populations and a very long time. However, nature is not time constrained, and there are considerable developmental constraints which limit the range of phenotypic variability.

A short-coming of goal-directed development is that its goal is not the *actual* goal of the designer's organisation, which may be profit, security, military supremacy, social good etc. and most of all, survival. This indirectness is the cause of many a tech company's failure, as a brilliantly-engineered product does not equate to market success. Even if the goal were actual market success, as the technological landscape changes ever faster what may have been successful when first conceived can be a poor market fit by the time it is ready for product launch. This is yet another reason for the predominance of agile development, particularly for software products.

Compared to nature's exploratory search, often somewhat disparagingly described as "tinkering" (Jacob, 1977), goal-directed solutions are vastly accelerated by human conceptualisation & capability. But they are also limited by human imagination. The engineer Genrikh Altshuller understood these limitations very well and developed the TRIZ methodology to try to overcome the limitation of normal human thinking (Altshuller, 1999). All the bio-inspiration movements are essentially also a way of doing this.

Of course, technology can also have an exploratory element to its search. Sometimes technologists have a goal but no idea how best to fulfil it. A famous example is Thomas A. Edison who, trying to improve the lifetime of a light bulb filament, tested a vast range of different carbonised materials, extending his search worldwide to test as many grasses and canes as possible – in all testing no fewer than six thousand different species of vegetable growths (Dyer and Martin, 1929, p. 262). Edison's search only varied the filament material within certain limits, but a pure exploratory search would have no preconception of what to vary, or what might bring about a useful improvement. Although much trial and error experimentation does take place in research labs around the world, it is never as open as nature's exploratory search.

Direct vs Generative model

In its most general sense, a "model" is any abstraction of a reality that represents its most important aspects. The purpose of the model in goal-directed search is for the designer to predict the outcome of various designs without actually making them. It can be tangible, like a drawing or calculation, or simply in the mind of the designer. As noted earlier, it is a *direct* representation of the intended result. So any exploratory search in technology can only vary a few parameters of the design, and usually ones that can be varied without affecting too much else.

Whereas the model in a pure exploratory search needs to suitably parameterise the *whole* design. This is achieved in biology by specifying the design as a "recipe", a genomic specification of the components which together assemble the organism. It is a model of how the final result is *generated* rather than what it actually is.

This also extends to the approach animals use to build their "technology" – bird's nests, spider webs, termite mounds etc., where the construction process seems to mostly result from an algorithm that links simple behaviour into a sequence that generates the finished result (Boyd *et al.*, 2013).

Solution Forms

The two search types lead to different forms of solution. Two examples are:

Predictable vs adaptable. Engineers dislike feedback if it can't be carefully controlled, as it can easily lead to instability, but nature thrives on feedback systems. This in turn affects the key quality of adaptability, which in biology is at both organism level and species level, but in technology is modest at best.

Single vs multi-level. Technology focuses at one level (an extension of a single human's capability) but nature can develop solutions that work at all organisational levels simultaneously (Vincent *et al.*, 2006). In effect technology focuses on solving problems for individuals but nature solves problems at the level of the whole population, because that is the level at which it searches.

Accumulation and Application of Experience

Search requires the accumulation of experience to help direct it in potentially fruitful ways. Technology has theory (science and maths), methods, components and tools. All these encapsulate experience in a way that makes it possible to construct new technology with relatively little personal experience.

Biology accumulates experience primarily in the genome. It represents all that has been learned throughout the history of that lineage, expressed in terms of how to build it from a single cell of that organism. The most hydrodynamic body shapes in fish, muscle structures, brain structures, how to create a camera eye – all this has been derived from experience and encapsulated in the DNA of the genome, which is biology's alternative to components, tools and methods.

Selection

Feature Selection

Selection in nature is powerful because it can work at many levels. Yes, it can detect if a specific change is an improvement or not, but this is the least interesting aspect of it. In a constantly changing environment it can select for processes and regulatory architectures that best allow adaptation and maybe even evolvability itself (Payne and Wagner, 2019).

In technology the inefficiency of pure selection is bypassed by the judgement of designers, who use a combination of imagination and trial and error to decide what will work best before turning anything into a product. But ultimately it is the market that decides what succeeds and what doesn't, and this is remarkably hard for humans to predict. Over time it is usage that determines which architectures, components, methods take hold and it is this selection process that ultimately drives technological evolution.

For all this to happen, selection must work at a feature/gene level and not just at a whole system (product/organism or technology/population) level. Selection acts on individuals, so those individuals must exhibit different combinations of features so over time and in populations useful features can succeed and detrimental ones disappear. One important way biology achieves this is through gene recombination in sexual reproduction, which allows selection to operate at the level of individual alleles. In asexual organisms Horizontal Gene Transfer within or between species can sometimes achieve this. It may not be in the organism's interest to share its genes with competing species, but it is in the gene's interest. HGT has been known for a while to be particularly important in prokaryotes, effectively allowing them to share their discoveries with one another, for instance antibiotic resistance, making the phylogenetic tree more like a network. Whether or not it plays a major role in eukaryotes remains controversial.

The Role of the Environment

In this context, the environment can be seen as everything which determines the solution space of potential improvements (to fitness or market success) for selection to act upon.

For biology, this solution space is determined both by the inanimate environment and the impact of other life, often referred to as the "abiotic" and "biotic" environments. These are in effect nature's specification for the genome-based exploratory search for improvements. The feedback within both environments ensures they are always changing, in addition to the natural instability of the earth and its weather systems. Varying environments accelerate complex adaptation (Pál and Papp, 2017) allow newly acquired genes to exploit new niches (Gogarten *et al.*, 2002), can become highly selective and thereby accelerate microevolution (Newman, 2006), help develop a modular design (Parter *et al.*, 2008) and generally allow a continual exploration of phenotype space (Taylor, 2018).

In technology, the environment is other technology, possibly competing, embedded in a complex mix of social, cultural and economic factors. In addition to these there are the less market-driven and more artificial factors of military requirements and governmental regulations (Basalla, 1988). Innovations and new technologies can directly impact all these factors, creating potentially rapid change. They also feed back into existing technology and replace older technology, which further increases the speed of that change.

Technology Research

The 20th Century has seen the rise of technology research worldwide, undertaken by both industry and universities, with ambitious goals such as machine vision, automatic speech recognition (ASR) and artificial intelligence, paving the way for even more ambitious goals such as self-driving cars and humanoid robots. Many decades of research passed by with few products in the market, and a different form of selection has been needed to drive pre-market development. As an example, we will look at how this developed in the case of ASR.

In the 1980s deployed systems were few and far between and only worked in niche applications with barely adequate performance. Progress towards the goal of unconstrained speech recognition was much slower than anyone had anticipated. The problem was that although researchers met up at conferences and published their results, these results were not easily comparable because each system was tested on proprietary data. The only way to find out whose techniques worked the best was to implement and test each one.

In 1988 the first of a sequence of speech databases, TIMIT, was created and made available to the whole community. The Linguistic Data Consortium (LDC) was formed to administer and create a revenue stream to fund the databases. Having databases that everyone used meant that results could be compared directly, revealing the most useful techniques. However, there was also a tendency for systems to become tuned to the exact data, so over time this informal comparison developed into annual competitions where previously unseen data was provided strictly only for the final evaluation. Each participant reported their methods and results in a special session at the annual speech technology conference (see (Reddy *et al.*, 2021) for a recent one).

The community effectively created an artificial selection system so that genuine advancements would stand out against less effective ones. However, this had an unintentional side-effect. In his keynote speech at Eurospeech in 1995, Hervé Bourlard spoke on "Towards *increasing* speech recognition error rates" (Bourlard, 1995). He was referring to the problem that completely new approaches inevitably lead to an increase in error rate, whereas the focus of the ASR community was now on those approaches that reduce word error rate. Boulard was one of the few researchers at that time using Artificial Neural Networks (ANNs), when almost everyone else exclusively used Hidden Markov Models. Now, many years later, the ANNs of deep learning not only pervade ASR, but the whole field of AI. Selection must not be so strong that it stifles important novelty.

Structural and Organisational Development

For technology, structure and organisation is the same in the model (design) and the product itself, since they have a 1-1 correspondence. However, the organisation evident in biology's product (phenotype) is not usually very evident in its model (genotype), with a few notable exceptions like the ancient Hox genes. This lack of 1-1 correspondence has made it very hard to discover exactly how biological organisation has occurred. However, enough is known for us to see how the properties discussed in this section, which are foundational for technology, are not only also evident in biology, but perform similar roles.

Modularity

Modules are reasonably self-contained units, weakly coupled to everything else, which allows them to develop and be optimised independently. In technology this means that the designer does not need to understand the internal workings of the module, in biology it means that regulatory control is simplified. More will be said on this in the Re-use section below.

Modularisation has a lot of organisational advantage. In biology, it may be the result of direct selection for stability, robustness or evolvability (Kashtan *et al.*, 2009) or it may be a dynamic side-effect of evolution, a result of the duplication of genes or subsystems. However it has come about, modularity appears as essential to biology as it is to technology (Gilbert, 2000).

Hierarchy

Hierarchy is a natural consequence of components being grouped together to form larger units. The ability to use or control a component without knowledge of its internal workings reduces the complexity of both goal-directed design and exploratory search. Designers only need to think (and need skill) at the level they are working, and can treat all the modules and components they work with as "black boxes". For example, an architect works at a different level to a builder, who works at a different level to the manufacturer of the building materials.

Nature also exhibits hierarchy. Organisms are constructed of units that are self-contained and yet part of a larger unit. This ranges from organelles to cells, tissues, organs and organ systems. The hierarchy also extends beyond the organism through populations, communities and ecosystems to the whole biosphere.

Emergence

Emergence is a special case of hierarchy which forms from individual artefacts/organisms joining together to form a larger unit which has properties the parts did not have on their own. In technology, emergence happens when someone spots the joining-together potential to build something that already exists better, or maybe a known concept which has had no way of being built until then. Emergence has taken place when this new functionality, as it is incrementally extended, has enough impact for attention to switch away from the joined parts to the functionality they enable. A good example is how transistors were used to formed logic gates, which formed adders, CPU and memory, which formed microprocessors, which formed computers, which formed networks, and are now embedded in so many different devices that we have the “internet of things”.

Emergence in nature is a remarkable phenomenon that has been recognised as a three-stage process along the lines of formation, maintenance, and transformation (Szathmáry, 2015). Other aspects of the three stages are fitness (Okasha, 2005) and causality (Deacon, 2003), both of which are important because the emergent entity must exert a top-down causality and also have a fitness independent of the fitness of its parts. Szathmáry lists no less than 7 emergences, including eukaryotes, multicellularity and animal societies.

Combination

The central thesis of Brian Arthur's book “The Nature of Technology” is that novel technologies arise by combination of existing technologies (Arthur, 2009).

It's useful to distinguish between standard combination that produces something similar to before, filling in gaps in the possibility space, and novel combination that produces innovation. Standard combination allows straightforward adaptation to changing environments/requirements. In technology it is the designer using the tools and methods of his trade to meet a particular need, in biology, it is male and female gene recombination driven by sexual selection. Sexual reproduction constitutes a surprisingly efficient trade-off between exploiting alleles that were fit on average in the past and sampling alleles in new combinations. (Watson and Szathmáry, 2016).

But there is also novel combination which results in innovation. The combinatorial engine of biology is at the genome level, creating new reactions, proteins and complex regulation networks, all facilitated by the interoperability of these low-level building blocks of life described earlier. Gene regulation networks play a particularly important part in this - most of the many and varied anatomical and physiological traits that have evolved in the last 500 million years are mainly the result of changes in regulation networks, according to the theory of Facilitated Variation (Gerhart and Kirschner, 2007).

In technology almost anything can be combined, and it is this which leads Arthur to his thesis that technology is a result of “combinatorial evolution”. In biology, combination is restricted to the components that are already represented in the genome, with useful mutations only occasionally creating new ones. The difference in biology is that all these components are very adaptable, so as the regulatory networks change the way they are combined, new possibilities open up quite easily.

Re-use

In technology, re-use ideally involves modularising and componentisation. As stated above, it is advantageous for easy usage that functionality is encapsulated, so internal workings do not need to be understood, and a component can be used in different applications. In software engineering, where re-use can just involve copy and paste with modification, it is still considered good practice to create a module with a more generalised and encapsulated functionality than to keep on copying and slightly modifying code, though both approaches are common.

In biology, both modular and copy & modify kinds of re-use are important. In the genotype, copy and modification is seen in the duplication of genes, and even whole genomes have been duplicated (Crow and Wagner, 2006). A most remarkable modularity example is the phenomenon of weak regulatory linkage in the animal phenotype, where core conserved processes are controlled with a very simple input signal. These processes are so well encapsulated that although the signal seems superficially to control the response, it invariably turns out that the responding core process can produce its output by itself but inhibits itself from doing so (Gerhart and Kirschner, 2007). This allows the regulatory network to easily try new combinations of core components in new amounts and states, at new times and places in the animal.

A completely different form of module re-use in unicellular organisms like bacteria is when HGT involves the introduction of complex, multigene pathways (Gogarten *et al.*, 2002). For instance, a study on the phylogeny of the flagellum in bacteria suggests that there was a transfer of the entire flagellar gene complexes between proteobacterial lineages after their separation from other major bacteria groups (Liu and Ochman, 2007).

The prevalence of hierarchy means that there are multiple levels of re-use in both tech and biology. The functional unit of one level may make a good component of the level above; Watson observed that in evolutionary systems, selection at one level of organisation can operate like unsupervised learning at a higher level of organisation (Watson and Szathmáry, 2016). This is because anything with a useful function which operates robustly and reliably (both of which will be selected for) is a good candidate for a component of a bigger system, no matter what it actually does. Each one of the component levels in computing mentioned above – transistors, logic gates, arithmetic units and memory, instructions, routines, programs - started out as a specific solution to a particular problem, but in time were adapted to become a more general component of a larger system.

Re-use for multicellularity. A very significant example of this is found in the transition to multicellularity. Newman and Bhat have identified 8 or 10 "Dynamical Patterning Modules" which together lay the foundation of the complex morphology observed in multicellular organisms: adhesion, alternative cell states, phase separation, tissue multi-layering, topological change, interior cavities, tissue elongation, tissue solidification and elasticity, pattern formation, segmentation and periodic patterning. (Newman and Bhat, 2008, 2009). The molecules of the DPMs "mostly evolved in single-celled organisms prior to the evolution of the metazoa, and only took on their DPM-associated roles with the change of spatial scale that was a consequence of multicellularity". That is, they evolved because of the capabilities they gave a unicellular organism, but those capabilities then allowed the transition to multicellularity.

More on the transition to multicellularity has come to light in a more recent study of the genome of 21 species of single-celled choanoflagellates, the closest living relative to animals (Richter *et al.*, 2018). They share some very important genes with animals, including genes essential for early development and genes that help the immune system detect pathogens. The study found that around 372 gene families previously thought to be animal-specific, including Notch, Delta, and homologs of the animal Toll-like receptor genes, instead evolved prior to the animal-choanoflagellate divergence. They conclude "it appears that the single-celled ancestor of animals was already well-equipped for multicellular life".

It seems as if over many millions of years single celled organisms were steadily accumulating base capabilities needed for multicellular life, even though at that stage they may have been used for something different. Then, when enough components were in place to make multicellular life viable, the transition could take place. This long wait also enabled the accumulation of a rich set of other genes which have helped give multicellular life the diversity we observe today.

Discussion

The approach to a unified framework in this paper has been to identify a number of essential properties in both evolutionary systems and to consider how they are accomplished in each. Very often nature and technology contrast with one another, but it is significant that there is always nuance in that contrast, either biology has notable exceptions that work technology's way, or technology on occasion looks a little more like biology. These exceptions are important for a unified framework - expanding on these exceptions would open up the possibility of making technology more biology-like, or even biology more technology-like, through bio-engineering.

The framework, even at this very basic initial stage, could also be useful for Open Ended Evolution simulations (Packard *et al.*, 2019), which was described in 2017 as "the last grand challenge you've never heard of" (Stanley *et al.*, 2017). Are the organisational features described here essential for open-endedness?, If so, it could be helpful to study them individually as stepping stones on the way to a system which can produce them all. A simulation is usually a mixture of extrinsic and intrinsic processes (Taylor, 2018), so it should be possible to have some of these features extrinsic, effectively creating a hybrid between nature's and technology's approach in the simulation. As Channon observes "it is clearly necessary to skip over or engineer in at least some complex features that arose through major transitions in our universe" (Channon, 2019).

To build a full unified framework from this beginning will require the incorporation of many more of the key concepts from both biological and technological evolution. To take some biological examples suggested by a reviewer; *mutational robustness* could be incorporated as a necessary quality of Exploratory Search, *heritability* would help expand out Accumulation and Application of Experience with interesting equivalents in technology, but *plasticity* requires an addition to the framework. The principle is always to consider what is being achieved and then how that is achieved in the other system. What is more explicit in one may help reveal what is much less obvious in the other. As in the above examples, some concepts will be able to fit into one of the main properties introduced here, others will require completely new ones. Some, like Accumulation and Application of Experience, will need expanding with sub-properties – it is a very significant topic particularly for technological evolution. Indeed, Universal Darwinism holds that an inferential system is key to facilitating complex order in many other domains besides biology and culture (Campbell and Price, 2019).

A possible criticism of the framework would be that the emphasis is on the physical, whereas modern technology is dominated by software and algorithms (in a rather analogous way to how nature is dominated by brainpower). The intention has been for most of what has been discussed to apply equally well to both, except of course the subsection on Use of Material, but it maybe that at some point in the future this will need to be addressed explicitly.

What of all those analogies listed in the Introduction? Most of these have still not been addressed, and so are we any further forward answering the key question of what is at the root of these universals? This paper proposes that at the root of them are a number of evolutionary system properties which are common to both systems, but which may operate in different ways. Each analogy needs to be carefully considered as to whether it either reveals a new property which should be fitted into the framework, or is simply a by-product of properties already in the framework. Thus this basic framework can be extended to consider and hopefully encompass these and the many other observed analogous aspects of technological and biological evolution.

Acknowledgements

Thanks to Wolfgang Banzhaf and Andrew Halestrap for their critique and helpful comments on the initial version of this paper. Thanks also to all the reviewers for taking the time to give thoughtful critical feedback and useful suggestions.

References

Al-Khalili, J. and McFadden, J. (2014) *Life on the Edge - the coming of Age of Quantum Biology*. Random House.

Altshuller, G.S. (1999) 'The innovation algorithm: TRIZ, systematic innovation and technical creativity', *Technical Innovation Center, Inc* [Preprint].

Arthur, W.B. (2009) *The Nature ofTechnology: What It Is and How It Evolves*. Penguin.

Basalla, G. (1988) *The evolution of technology*, *The Evolution of Technology*. Cambridge University Press.

Bourlard, H. (1995) 'Towards Increasing Speech Recognition Error Rates', in *4th European Conference on Speech Communication and Technology EUROSPEECH '95*, pp. 883–894.

Boyd, R., Richerson, P. and Henrich, J. (2013) 'The cultural evolution of technology: facts and theories', *Cultural Evolution: Society, Technology, Language, and Religion*, pp. 119–142.

Brey, P. (2008) 'Technological Design as an Evolutionary Process', in *Philosophy and Design: From Engineering to Architecture. Springer.*

Campbell, J.O. and Price, M.E. (2019) 'Universal darwinism and the origins of order', *Springer Proceedings in Complexity*, pp. 261–290.

Channon, A. (2019) 'Maximum Individual Complexity is Indefinitely Scalable in Geb', *Artificial Life*, 25(2), pp. 134–144.

Crow, K.D. and Wagner, G.P. (2006) 'What is the role of genome duplication in the evolution of complexity and diversity?', *Molecular Biology and Evolution*, 23(5), pp. 887–892.

Deacon, T.W. (2003) 'The Hierarchic Logic of Emergence: Untangling the Interdependence of Evolution and Self-Organization', in Weber, B. and Depew, D. (eds) *Evolution and learning: The Baldwin effect reconsidered*. MIT Press.

Dyer, F.L. and Martin, T.C. (1929) *Edison: His Life and Inventions*. Harper and Brothers Publishing.

Gerhart, J. and Kirschner, M. (2007) 'The theory of facilitated variation.', *Proceedings of the National Academy of Sciences of the United States of America*, 104 Suppl, pp. 8582–8589.

Gilbert, S.F. (2000) 'Modularity: The Prerequisite for Evolution through Development', in *Developmental Biology*. 6th edn. Sinauer Associates.

Gogarten, J.P., Doolittle, W.F. and Lawrence, J.G. (2002) 'Prokaryotic evolution in light of gene transfer', *Molecular Biology and Evolution*, 19(12), pp. 2226–2238.

Jacob, F. (1977) 'Evolution and Tinkering', *Science*, 196, pp. 1161–1166.

Kashtan, N., Parter, M., Dekel, E., Mayo, A.E. and Alon, U. (2009) 'Extinctions in heterogeneous environments and the evolution of modularity', *Evolution*, 63(8), pp. 1964–1975.

Liu, R. and Ochman, H. (2007) 'Stepwise formation of the bacterial flagellar system', *PNAS*, 104(17), pp. 7116–7121.

Mesoudi, A. *et al.* (2015) 'The Cultural Evolution of Technology and Science', *Cultural Evolution*, (November), pp. 193–216.

Newman, S.A. (2006) 'The Developmental Genetic Toolkit and the Molecular Homology—Analogy Paradox', *Biological Theory*, 1(1), pp. 12–16.

Newman, S.A. and Bhat, R. (2008) 'Dynamical patterning modules: Physico-genetic determinants of morphological development and evolution', *Physical Biology*, 5(1).

Newman, S.A. and Bhat, R. (2009) 'Dynamical patterning modules: A "pattern language" for development and evolution of multicellular form', *International Journal of Developmental Biology*, 53(5–6), pp. 693–705.

Okasha, S. (2005) 'Multilevel selection and the major transitions in evolution', *Philosophy of Science*, 72(5), pp. 1013–1025.

Packard, N. *et al.* (2019) 'An Overview of Open-Ended Evolution: Editorial Introduction to the Open-Ended Evolution II Special Issue', *Artificial Life*, 25(2), pp. 93–103.

Pál, C. and Papp, B. (2017) 'Evolution of complex adaptations in molecular systems', *Nat Ecol Evol.*, 1, pp. 1084–1092.

Parter, M., Kashtan, N. and Alon, U. (2008) 'Facilitated variation: How evolution learns from past environments to generalize to new environments', *PLoS Computational Biology*, 4(11).

Payne, J.L. and Wagner, A. (2019) 'The causes of evolvability and their evolution', *Nature Reviews Genetics*, 20(1), pp. 24–38.

Perkins, D. (2000) 'The evolution of adaptive form', in Zinman, J. (ed.) *Technological Innovation as an Evolutionary Process*. Cambridge University Press, pp. 159–173.

Reddy, C.K.A. *et al.* (2021) 'INTERSPEECH 2021 Deep Noise Suppression Challenge', in *Interspeech 2021*, pp. 2796–2800.

Richter, D.J., Fozouni, P., Eisen, M.B. and King, N. (2018) 'Gene family innovation, conservation and loss on the animal stem lineage', *eLife*, 7, pp. 1–43.

Solé, R. V *et al.* (2013) 'The Evolutionary Ecology of Technological Innovations', *SFI WORKING PAPER: 2012-12-022*, 000(00), pp. 1–13.

Stanley, K.O., Lehman, J. and Soros, L. (2017) *Open-endedness: The last grand challenge you've never heard of.* Available at: https://www.oreilly.com/ideas/open-endedness-the-last-grand-challenge-youve-never-heard-of (Accessed: 4 March 2021).

Szathmáry, E. (2015) 'Toward major evolutionary transitions theory 2.0', *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), pp. 10104–10111.

Taylor, T. (2018) 'Routes to Open-Endedness in Evolutionary Systems', in *Third Workshop on Open-Ended Evolution (OEE3), Tokyo, Japan*.

Vincent, J.F.V., Bogatyreva, O.A., Bogatyrev, N.R., Bowyer, A. and Pahl, A.-K. (2006) 'Biomimetics: its practice and theory', *Journal of The Royal Society Interface*, 3(9), pp. 471–482.

Vogel, S. (1999) *Cats' Paws and Catapults*. Penguin Books.

Wagner, A. (2014) *Arrival of the Fittest*. Oneworld.

Watson, R.A. and Szathmáry, E. (2016) 'How Can Evolution Learn?', *Trends in Ecology and Evolution*, 31(2), pp. 147–157.

Zinman, J. (2000) *Technological Innovation as an Evolutionary Process*. Cambridge University Press.

Hereditary Stratigraphy:
Genome Annotations to Enable Phylogenetic Inference over Distributed Populations

Matthew Andres Moreno, Emily Dolson, and Charles Ofria

Michigan State University, East Lansing, MI 48103
mmore500@msu.edu

Abstract

Phylogenies provide direct accounts of the evolutionary trajectories behind evolved artifacts in genetic algorithm and artificial life systems. Phylogenetic analyses can also enable insight into evolutionary and ecological dynamics such as selection pressure and frequency-dependent selection. Traditionally, digital evolution systems have recorded data for phylogenetic analyses through perfect tracking where each birth event is recorded in a centralized data structure. This approach, however, does not easily scale to distributed computing environments where evolutionary individuals may migrate between a large number of disjoint processing elements. To provide for phylogenetic analyses in these environments, we propose an approach to enable phylogenies to be inferred via heritable genetic annotations rather than directly tracked. We introduce a "hereditary stratigraphy" algorithm that enables efficient, accurate phylogenetic reconstruction with tunable, explicit trade-offs between annotation memory footprint and reconstruction accuracy. In particular, we demonstrate an approach that enables estimation of the most recent common ancestor (MRCA) between two individuals with fixed relative accuracy irrespective of lineage depth while only requiring logarithmic annotation space complexity with respect to lineage depth. This approach can estimate, for example, MRCA generation of two genomes within 10% relative error with 95% confidence up to a depth of a trillion generations with genome annotations smaller than a kilobyte. We also simulate inference over known lineages, recovering up to 85.70% of the information contained in the original tree using 64-bit annotations.

Introduction

In traditional serially-processed digital evolution experiments, phylogenetic trees can be tracked perfectly as they progress (Bohm et al., 2017; Wang et al., 2018; Lalejini et al., 2019) rather than reconstructed afterward, as must be done in most biological studies of evolution. Such direct phylogenetic tracking enables experimental possibilities unique to digital evolution, such as perfect reconstruction of the sequence of phylogenetic states that led to a particular evolutionary outcome (Lenski et al., 2003; Dolson et al., 2020). In a shared-memory context, it is not difficult to maintain a complete phylogeny by ensuring that offspring retain a permanent reference to their parent (or vice versa). As simulations progress, however, memory usage would balloon if all simulated organisms were stored permanently. Garbage collecting extinct lineages and saving older history to disk greatly ameliorates this issue (Bohm et al., 2017; Dolson et al., 2019).

If sufficient memory or disk space can be afforded to log all reproduction events, recording a perfect phylogeny in a distributed context is also not especially difficult. Processes could maintain records of each reproduction event, storing the parent organism (and its associated process) with all generated offspring (and their destination processes). As long organisms are uniquely identified globally, these "dangling ends" could be joined in postprocessing to weave a continuous global phylogeny. Of course, for the huge population sizes made possible by distributed systems, such stitching may become a demanding task in and of itself. Additionally, even small amounts of lost or corrupted data could fundamentaly degrade tracking by disjoining large tree subsections.

However, if memory and disk space are limited, distributed phylogeny tracking becomes a more burdonsome challenge. A naive approach might employ a server model to maintain a central store of phylogenetic data. Processes would dispatch notifications of birth and death events to the server, which would curate (and gabage collect) phylogenetic history much the same as current serial phylogenetic tracking implementations. Unfortunately, this server model approach would present scalability challenges: burden on the server process would worsen in direct proportion to processor count. This approach would also be similarly brittle to any lost or corrupted data.

A more scalable approach might record birth and death events only on the process(es) where they unfold. However, lineages that went extinct locally could not be safely garbage collected until the extinction of their offspring's lineages on other processors could be confirmed. Garbage collection would thus require extinction notifications to wind back across processes each lineage had traversed. Again, this approach would also be brittle to loss or corruption of data.

In a distributed context — especially, a distributed, best-effort context — phylogenetic reconstruction (as opposed to tracking) could prove simpler to implement, more efficient at runtime, and more robust to data loss while providing sufficient information to address experimental questions of interest. However, phylogenetic reconstruction from genomes with a traditional model of divergence under grandual accumulation of random mutations poses its own difficulties, including

- accounting for heterogeneity in evolutionary rates (i.e., the rate at which mutations accumulate due to divergent mutation rates or selection pressures) between lineages (Lack and Van Den Bussche, 2010),
- performing sequence alignment (Casci, 2008),

- mutational saturation (Hagstrom et al., 2004),
- appropriately selecting and applying complex reconstruction algorithms (Kapli et al., 2020), and
- computational intensity (Sarkar et al., 2010).

The computational flexibility of digital artificial life experiments provides a unique opportunity to ovecome these challenges: designing heritable genome annotations specifically to ensure simple, efficient, and effective phylogenetic reconstruction. For maximum applicability of such a solution, these annotations should be phenotypically neutral heritable instrumentation (Stanley and Miikkulainen, 2002) that can be applied to any digital genome.

In this paper, we present "hereditary stratigraphy," a novel heritable genome annotation system to facilitate post-hoc phylogenetic inference on asexual populations. This system allows explicit control over trade-offs between space complexity and accuracy of phylogenetic inference. Instead of modeling genome components diverging through a neutral mutational process, we keep a record of historical checkpoints that allow comparison of two lineages to identify the range of time in which they diverged. Careful management of these checkpoints allows for a variety of trade-off options, including:

- linear space complexity and fixed-magnitude inference error,
- constant space complexity and inference error linearly proportional to phylogenetic depth, and
- logarithmic space complexity and inference error linearly proportional to time elapsed since MRCA (which we suspect will be the most broadly useful trade-off).

In Methods we motivate and explain the hereditary stratigraphy approach. Then, in Results and Discussion we simulate post-hoc inference on known phylogenies to assess the quality of phylogenetic reconstruction enabled by the hereditary stratigraphy method.

Methods

This section will introduce intuition for the strategy of our hereditary stratigraph approach, define the vocabulary we developed to describe aspects of this approach, overview configurable aspects of the approach, present mathematical exposition of the properties of space complexity and inference quality under particular configurations, and then recap digital experiments that demonstrate this approach in an applied setting.

Hereditary Strata and the Hereditary Stratigraphic Column

Our algorithm, particularly the vocabulary we developed to describe it, draws loose inspiration from the concept of geological stratigraphy, inference of natural history through analysis of successive layers of geological material (Steno, 1916). As an introductory intuition, suppose a body of rock being built up through regular, discrete events depositing of geological material. Note that in such a scenario we could easily infer the age of the body of rock by counting up the number of layers present. Next, imagine making a copy of the rock body in its partially-formed state and then moving it far away. As time runs forward on these two rock bodies, independent layering processes will cause consistent disparity in the layers forming on each forwards from their point of separation.

To deduce the historical relationship of these rock bodies, we could simply align and compare their layers. Layers from their base up through the first disparity would correspond to shared ancestry; further disparate layers would correspond to diverged ancestry. Figure 1 depicts the process of comparing columns for phylogenetic inference.

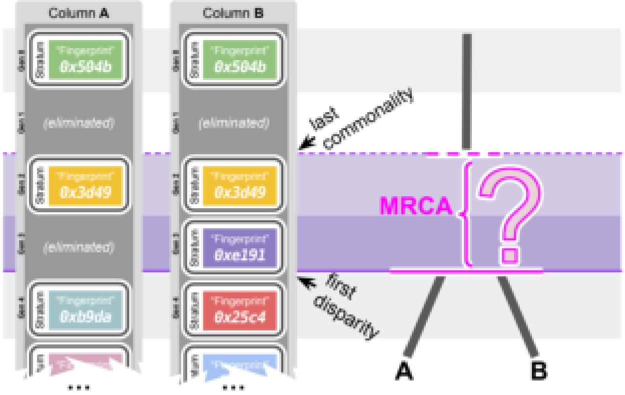


Figure 1: Inferring the generation of the most-recent common ancestor (MRCA) of two hereditary stratigraphic columns "A" and "B". Columns are aligned at corresponding generations. Then the first generation with disparate "fingerprints" is determined. This provides a hard upper bound on the generation of the MRCA: these strata *must* have been deposited along separate lines of descent. Searching backward for the first commonality preceding that disparity provides a soft lower bound on the generation of the MRCA: these strata evidence common ancestry but *might* collide by chance. Some strata mmay have been eliminated from the columns, as shown, in order to save space at the cost of increasing uncertainty of MRCA generation estimates.

Shifting now from intuition to implementation, a fixed-length randomly-generated binary tag provides a suitable "fingerprint" mechanism mirroring our metaphorical "rock layers." We call this "fingerprint" tag a *differentia*. The width of this tag controls the probability of spurious collisions between independently generated instances. At 64 bits wide the tag effectively functions as a UID: collisions between randomly generated tags are so unlikely ($p < 5.42 \times 10^{-20}$) they can essentially be ignored. At the other end of the spectrum, collision probability would be $1/256$ for a single byte and $1/2$ for a single bit. In the case of narrow differentia, in order to set a lower bound for the MRCA generation, you would have to backtrack common strata from the last commonality until the probability of that many successive spurious collisions was enough to satisfy your desired confidence level (e.g., 95% confidence). Even then, there would be a possibility of the the true MRCA falling before the estimated lower bound. Note, however, that no matter the width of the differentia the generation of the first discrepancy provides a hard upper bound on the generation of the MRCA.

In accordance with our geological analogy, we refer to the packet of data accumulated each generation as a *stratum*. This packet contains the differentia and, although not employed in this work, could hold other arbitrary user-defined data (i.e., simulation timestamp, phenotype characteristics, etc.). Again in accordance with the geological analogy, we refer to the chronological stack of strata that accumulate over successive generations as a *hereditary* *stratigraphic column.*

Stratum Retention Policy

As currently stated, strata in each column will accumulate proportionally to the length of evolutionary history simulated. In an evolutionary run with thousands or millions of generations, this approach would soon become intractable — particularly when columns are serialized and transmitted between distributed computing elements. To solve this problem, we can trade off precision for compactness by strategically deleting strata from columns as time progresses. Figure 2 overviews how stratum deposit and stratum elimination might progress over two generations under the hereditary stratigraphic column scheme.

Different patterns of deletion will lead to different trade-offs, both in terms of the scaling relationship of column size to generations elapsed and in terms of the arrangement of inference precision over evolutionary history (i.e., focusing precision on more recent evolutionary history versus spreading it evenly over the entire history).

We refer to the rule set used to selectively eliminate strata over time as the "stratum retention policy." We explore several different retention policy designs here, and implement our software to allows for free, modular interchange of retention policies.

Our software allows specification of a policy as either a "predicate" or a "generator." The predicate method requires a function that takes the generation of a stratum and the current number of strata deposited and returns whether that stratum should be retained at that point in time. The generator method requires a function that takes the current number of strata deposited and yields the set of generations that should be deleted at that point in time. Although the predicate form of a policy is useful for analyzing and proving properties of policies, the generator form is generally more efficient in practice. We provide equivalent predicate and generator implementations for each stratum retention policy discussed here.

Strata elimination causes a stratum's position within the column data structure to no longer correspond to the generation it was deposited. Therefore, it may seem necessary to store the generation of deposit within the stratum data structure. However, for all deterministic retention policies a perfect mapping exists backwards from column index to recover generation deposited without storing it. We provide this formula for each stratum retention policy surveyed here. Finally, for each policy we provide a formula to calculate the exact number of strata retained under any parameterization after n generations.

The next subsections introduce several stratum retention policies, explain the intuition behind their implementation, and elaborate their space complexity and resolution properties. For each policy, patterns of stratum retention are illustrated in Figure 3. The formulas for number of strata retained after n generations, the formulas to calculate stratum deposit generation from column index, and the retention predicate specifications of each policy are available in Supplementary Listing 5 (Moreno et al., 2022). The generator specification of each policy is available in Supplementary Listing 1 (Moreno et al., 2022). For tapered depth-proportional resolution and recency-proportional resolution, the accuracy of MRCA estimation can also be explored via an interactive in-browser web applet at `https://hopth.ru/bi`.

Fixed Resolution Stratum Retention Policy

The fixed resolution retention policy imposes a fixed absolute upper bound r on the spacing between retained strata. The strategy is simple: permanently retain a stratum every rth generation. (For arbitrary reasons of implementation convenience, we also require each stratum to be retained during at least the generation it is deposited). See top panel of Figure 3.

This retention policy suffers from linear growth in a column's memory footprint with respect to number of generations elapsed: every rth generation generation a new stratum is permanently retained. For this reason, it is likely not useful in practice except potentially in scenarios where the number of generations is small and fixed in advance. We include it here largely for illustrative purposes as a gentle introduction to retention policies.

Depth-Proportional Resolution Stratum Retention Policy

The depth-proportional resolution policy ensures spacing between retained strata will be less than or equal to a proportion $1/r$ of total number of strata deposited n. Achieving this limit on uncertainty requires retaining sufficient strata so that no more than n/r generations elapsed between any two strata. This policy accumulates retained strata at a fixed interval until twice as many as r are at hand. Then, every other retained stratum is purged and the cycle repeats with a new twice-as-wide interval between retained strata. See second from top panel of Figure 3.

When comparing stratigraphic columns from different generations, the resolution guarantee holds in terms of the number of generations experienced by the older of the two columns. Because this retention policy is deterministic, for two columns with the same policy, every stratum that is held by the older column is also guaranteed to be present in the younger column (unless it hasn't yet been deposited on the younger column). Therefore, the strata that would enable the desired resolution when comparing two columns of the same age are guaranteed to be available, even when one colummn has elapsed more generations.

Because the number of strata retained under this policy is bounded as $2r+1$, space complexity scales as $O(1)$ with respect to the number of strata deposited. It follows that the MRCA generation estimate uncertainty scales as $O(n)$ with respect to the number of strata deposited.

Tapered Depth-Proportional Resolution Stratum Retention Policy

This policy refines the depth-proportional resolution policy to provide a more stable column memory footprint over time. The naive depth-proportional resolution policy builds up strata until twice as many are present as needed then purges half of them all at once. The tapered depth-proportional resolution policy functions identically to the depth-proportional policy except that it removes unnecessary strata gradually from back to front as new strata are deposited, instead of eliminating them simultaneously. See third from top panel of Figure 3.

The column footprint stability of this variation makes it easier to parameterize our experiments to ensure comparable end-state column footprints for fair comparison between retention policies, in addition to making this policy likely better suited to most use cases.

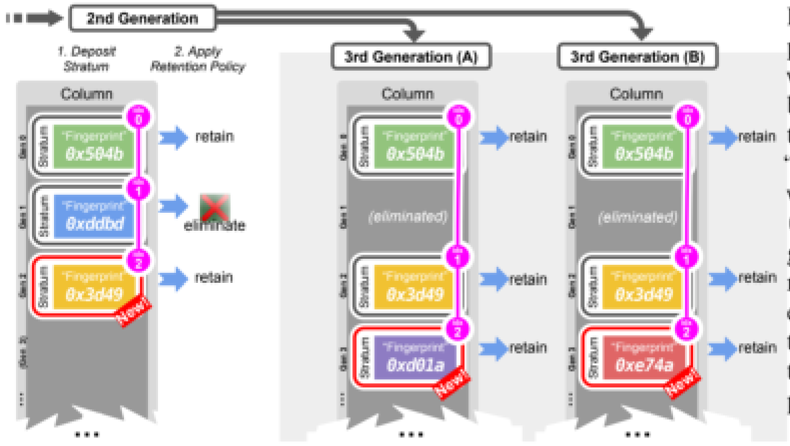


Figure 2: Cartoon illustration of stratum deposit process. This process marks the elapse of a generation when a hereditary stratigraphic column is inherited by an offspring. First, a new stratum is appended to the end of the column with a randomly-generated "fingerprint." This "fingerprint" distinguishes strata that were generated along disparate lines of descent (e.g., `0xd01a` for 3rd Generation A and `0xe74a` for 3rd generation B). Then, the column's configured stratum retention policy is applied to "prune" the column by eliminating strata from specific generations. Although this cartoon depicts an empty space for eliminated strata, the underlying data structure behind a column (i.e., the pink overlay) can condense to reduce space complexity.

Num Gens Elapsed	Guaranteed MRCA-Recency-Proportional Resolution			
	1	4	10	100
1.0×10^3	18	26	41	80
1.0×10^6	32	50	85	184
1.0×10^9	51	79	134	293
1.0×10^{12}	64	102	177	396

Table 1: Number strata retained after one thousand, one million, one billion, and one trillion generations under the recency-proportional resolution stratum retention policy. Four different policy parameterizations are shown, the first where MRCA generation can be determined between two extant columns with a guaranteed relative error of 100%, the second 25%, the third 10%, and the fourth 1%. A column's memory footprint will be a constant factor of these retained counts based on the fingerprint differentia width chosen. For example, if single byte differentia were used, the column's memory footprint in bits would be $8\times$ the number of strata retained.

By design, this policy has the same space complexity and MRCA estimation uncertainty scaling relationships with number generations elapsed as the naive depth-proporitonal resolution policy.

MRCA-Recency-Proportional Resolution Stratum Retention Policy

The MRCA-recency-proportional resolution policy ensures distance between the retained strata surrounding any generation point will be less than or equal to a user-specified proportion $1/r$ of the number of generations elapsed since that generation.

This policy can be constructed recursively. So, to begin, let's consider setting up just the *first* generation g of the stratum after the root ancestor we will retain when n generations have elapsed. A simple geometric analysis reveals that providing the guaranteed resolution for the worst-case generation within the window between generation 0 and generation g (i.e., generation $g-1$) requires

$$g \leq \lfloor n/(r+1) \rfloor.$$

We now have an upper bound for the generation of the first stratum generation we must retain. However, we must guarantee that strata at these generations are actually available for us to retain (i.e., haven't been purged out of the column at a previous time point). We will do this by picking the generation that is the highest power of 2 less than or equal to our bound. If we repeat this procedure as we recurse, we are guaranteed that this generation's stratum will have been preserved across all previous timepoints.

Why does this work? Consider a sequence where all elements are spaced out by strictly nonincreasing powers of 2. Consider the first element of the list. All multiples this first element will be included in the list. So, when we ratchet up g to $2g$ as n increases, we are guaranteed that $2g$ has been retained. This principle generalizes recursively down the list. This is a similar principle to the approach of strictly-doubling interval sizes used in the Depth-Proportional Resolution stratum retention policies described above.

This step of truncating to the nearest less than or equal to power of 2 affects our recursive step size is at most halved. So, because step size is a constant fraction of remaining generations n (at worst $\frac{n}{2(r+1)}$), the number of steps made (and number of strata retained) scales as $O(\log(n))$ with respect to the number of strata deposited. Table 1 provides exact figures for the number of strata retained under different parameterizations of the recency-proportional retention policy between one thousand and one trillion generations.

As for MRCA generation estimate uncertainty, in the worst case it scales as $O(n)$ with respect to the greater number of strata deposited. However, with respect to estimating the generation of the MRCA for lineages diverged any fixed number of generations ago, uncertainty scales as $O(1)$.

How does space complexity scale with respect to the policy's specified resolution r? Through extrapolation from OEIS sequences A063787 and A056791 via guess and check (OEIS, 2021b,a), we posited the exact number of strata retained after n generations as

$$\mathrm{HammingWeight}(n) + \sum_{1}^{r} \lfloor \log_2(\lfloor n/r \rfloor) \rfloor + 1.$$

This expression has been unit tested extensively to ensure perfect reliability. Approximating and applying logarithmic properties, this policy's space complexity can be calculated within a constant factor as

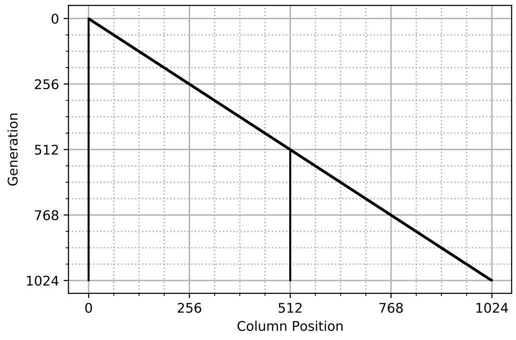
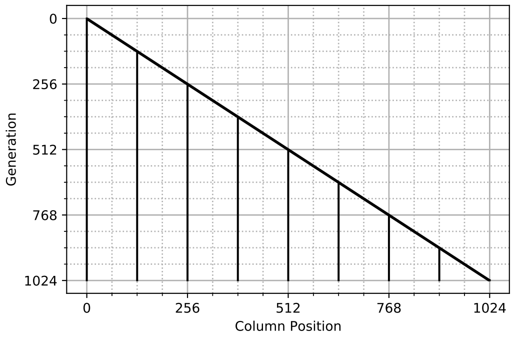
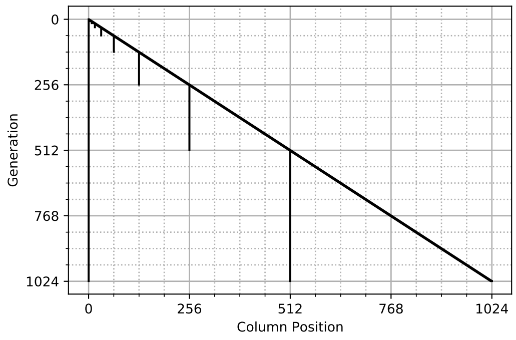
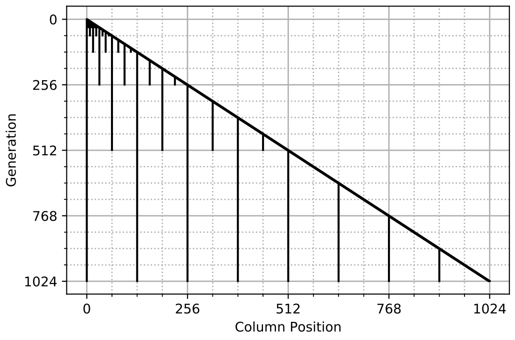
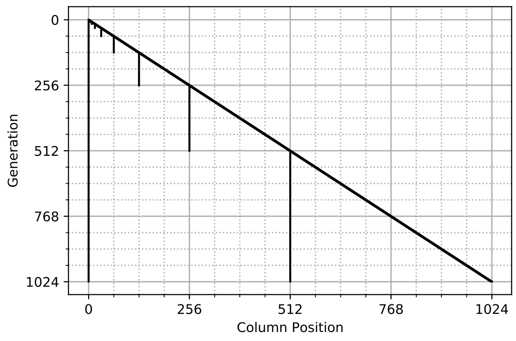
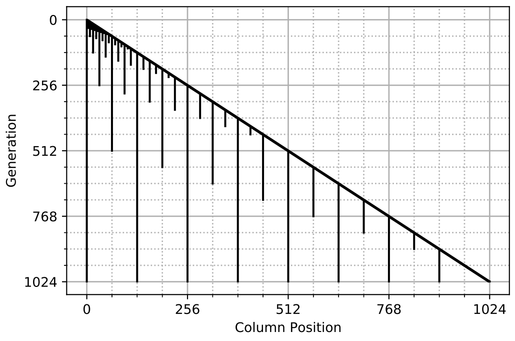
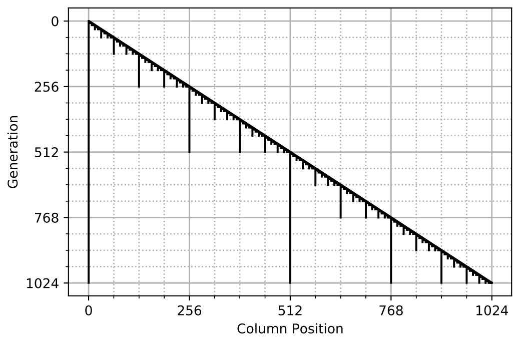
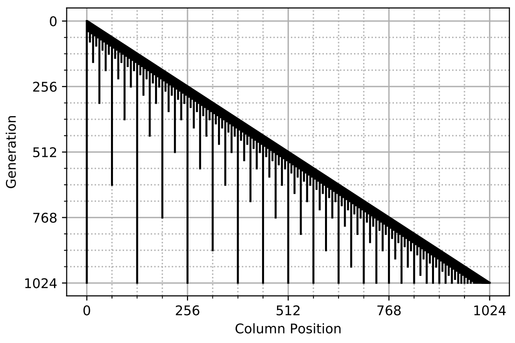
Policy	Lower-Resolution Parameterization	Higher-Resolution Parameterization	Properties
Fixed Resolution			**Space Complexity** $O(n)$ **MRCA Uncertainty** $O(1)$ where n is gens elapsed.
Depth-Proportional Resolution			**Space Complexity** $O(1)$ **MRCA Uncertainty** $O(n)$ where n is gens elapsed.
Tapered Depth-Proportional Resolution			**Space Complexity** $O(1)$ **MRCA Uncertainty** $O(n)$ where n is gens elapsed.
Recency-Proportional Resolution			**Space Complexity** $O(\log(n))$ **MRCA Uncertainty** $O(m)$ where m is gens since MRCA and n is total gens elapsed.

Figure 3: Comparison of stratum retention policies. Policy visualizations show retained strata in black. Time progresses along the y-axis from top to bottom. New strata are introduced along the diagonal and then "drip" downward as a vertical line until eliminated. The set of retained strata present within a column at a particular generation g can be read as intersections of retained vertical lines with a horizontal line with intercept g. Policy visualizations are provided for two parameterizations for each policy: the first where the maximum uncertainty of MRCA generation estimates would be 512 generations and the second where the maximum uncertainty of MRCA generation estimates would be 128 generations.

$$\log(n)+\log\left(\frac{n^r}{r!}\right).$$

To analyze the relationship between space complexity and resolution r, we will examine the ratio of space complexities induced when scaling resolution r up by a constant factor $f>1$. Evaluating this ratio as $r \to \infty$, we find that space complexity scales directly proportional to f,

$$\lim_{r\to\infty} \frac{\log(n)+\log\left(\frac{n^{fr}}{(fr)!}\right)}{\log(n)+\log\left(\frac{n^r}{r!}\right)} = f.$$

Evaluating this ratio as $n \to \infty$, we find that this scaling relationship is never worse than directly proportional for any r,

$$\begin{aligned}\lim_{n\to\infty} \frac{\log(n)+\log\left(\frac{n^{fr}}{(fr)!}\right)}{\log(n)+\log\left(\frac{n^r}{r!}\right)} &= \frac{fr+1}{r+1}\\ &= f\frac{r+1/f}{r+1}\\ &\leq f.\end{aligned}$$

Computational Experiments

In order to assess the practical performance of the hereditary stratigraph approach in an applied setting, we simulated the process of stratigraph propagation over known "ground truth" phylogenies extracted from pre-existing digital evolution simulations (Hernandez et al., 2022). These simulations propagated populations of between 100 and 165 bitstrings between 500 and 5,000 synchronous generations under the NK fitness landscape model (Kauffman and Weinberger, 1989). In order to ensure coverage of a variety of phylogenetic conditions, we sampled a variety of selection schemes that impose profoundly different ecological regimens (Dolson and Ofria, 2018),

- EcoEA Selection (Goings et al., 2012),
- Lexicase Selection (Helmuth et al., 2014),
- Random Selection, and
- Sharing Selection (Goldberg et al., 1987).

Supplementary Table 7 provides full details on the conditions each ground truth phylogeny was drawn from. The phylogenies themselves are available with our supplementary material (Moreno et al., 2022).

For each ground truth phylogeny, we tested combinations of three configuration parameters:

- target end-state memory footprints for extant columns (64, 512, and 4096 bits),
- differentia width (1, 8, and 64 bits), and
- stratum retention policy (tapered depth-proportional resolution and recency-proportional resolution).

Stratum retention policies were parameterized so that the maximum number of strata possible were present at the end of the experiment without exceeding the target memory footprint. If the target memory footprint is exceeded by the sparsest possible parameterization of a retention policy, then that sparsest possible parameterization was used. Supplementary Tables tables 2 to 6 provide the calculated paramaterizations and memory footprints of extant columns (Moreno et al., 2022).

In order to assess the viability of phylogenetic inference using hereditary stratigraphic columns from extant organisms, we used the end-state stratigraphs to reconstruct an estimate of the actual ground truth phylogenetic histories. The first step to reconstructing a phylogenetic tree for the history of an extant population at the end of an experiment is to construct a distance matrix by calculating all pairwise phylogenetic distances between extant columns. We defined phylogenetic distance between two extant columns as the sum of each extant organism's generational distance back to the generation of their MRCA, estimated as the mean of the upper and lower 95% confidence bounds. Supplementary Figure 6 provides a cartoon summary of the process of calculating phylogenetic distance between two extant columns (Moreno et al., 2022).

We then used the unweighted pair group method with arithmetic mean (UPGMA) reconstruction tool provided by the BioPython package to generate estimate phylogenetic trees (Cock et al., 2009; Sokal, 1958). After generating the reconstructed tree topology, we performed a second pass to adjust branch lengths so that each internal tree node sat at the mean of its estimated 95% confidence generation bounds.

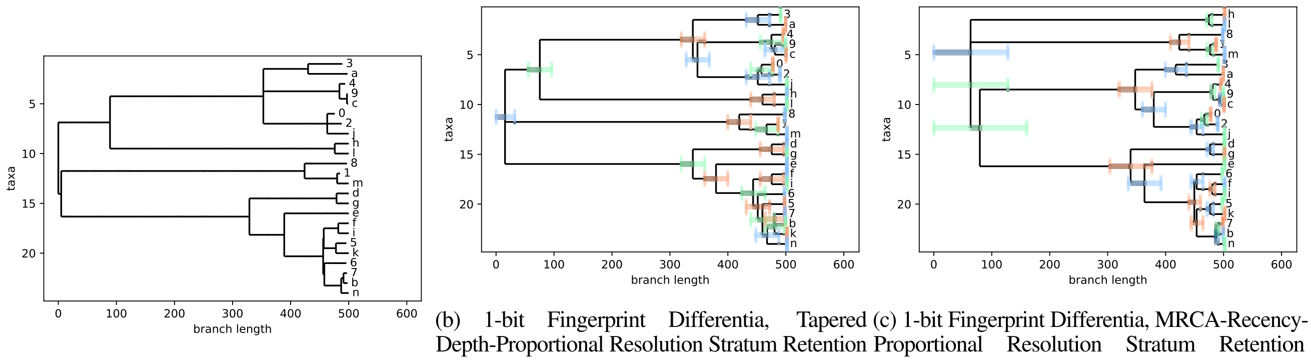
Software and Data

As part of this work, we published the `hstrat` Python library with a stable public-facing API intended to enable incorporation in other projects with extensive documentation and unit testing on GitHub at `https://github.com/mmore500/hstrat` and on PyPI.In the near future, we intend to complete and publish a corresponding C++ library.

Supporting software materials can be found on GitHub at `https://github.com/mmore500/hereditary-stratigraph-concept` Supporting computational notebooks are available for in-browser use via BinderHub at `https://hopth.ru/bk` (Ragan-Kelley and Willing, 2018). Our work benefited from many pieces of open source scientific software (Sukumaran and Holder, 2010; Virtanen et al., 2020; Hunter, 2007; Virtanen et al., 2020; Waskom, 2021; Bostock et al., 2011; Meurer et al., 2017; Smith, 2020b; Paradis et al., 2004; Ushey et al., 2022; Wickham et al., 2022). The ground truth phylogenies used in this work as well as supplementary figures, tables, and text are available via the Open Science Framework at `https://osf.io/4sm72/` (Foster and Deardorff, 2017; Moreno et al., 2022). Phylogenetic data associated with this project is stored in the Alife Community Data Standards format (Lalejini et al., 2019).

Results and Discussion

In this section, we analyze the quality of reconstructions of known phylogenetic trees using hereditary stratigraphy. Figure 4 compares an example reconstruction from columns using tapered depth-proportional stratum retention, an example reconstruction using recency-proportional stratum retention, and the underlying ground truth phylogeny. Interactive in-browser visualizations comparing all reconstructed phylogenies to their corresponding ground truth are available at `https://hopth.ru/bi`.



(a) Ground truth phylogeny.

(b) 1-bit Fingerprint Differentia, Tapered Depth-Proportional Resolution Stratum Retention Predicate, 64 bit target column footprint.

(c) 1-bit Fingerprint Differentia, MRCA-Recency-Proportional Resolution Stratum Retention Predicate, 64 bit target column footprint.

Figure 4: Example phylogeny reconstructions of ground-truth lexicase selection phylogeny from inference on extant hereditary stratigraphic columns. Shaded error bars on reconstructions indicate 95% confidence intervals for the true generation of tree nodes. Arbitrary color is added to enhance distinguishability.

Reconstruction Accuracy

Measuring tree similarity is a challenging problem, with many conflicting approaches that all provide different information (Smith, 2020a). Ideally, we would use a metric of reconstruction accuracy that 1) is commonly used so that there exists sufficient context to understand what constitutes a good value, 2) behaves consistently across different types of trees, and 3) behaves reasonably for the types of trees common in artificial life data. Unfortunately, these objectives are somewhat in conflict. The primary source of this problem is multifurcations, nodes from which more than two lineages branch at once. In reconstructed phylogenies in biology, multifurcations are generally assumed to be the result of insufficient information. It is thought that the real phylogeny had multiple bifurcations that occurred so close together that the reconstruction algorithm is unable to separate them. In artificial life phylogenies, however, we have the opposite problem. When we perfectly track a phylogeny, it is common for us to know that a multifurcation did in fact occur. However, it is challenging for our reconstructions to properly identify multifurcations, because it requires perfectly lining up multiple divergence times. Many of the most popular tree distance metrics interpret the difference between a multifurcation and a set of bifurcations as a dramatic change in topology. For some use cases, this change in topology may indeed be meaningful, although research on the extent of this problem is limited. Nevertheless, we suspect that for the majority of use cases, the tiny branch lengths between the internal nodes will make this source of error relatively minor.

To overcome this obstacle, we have measured our reconstruction accuracy using multiple metrics. We will primarily focus on Mutual Clustering Information (as implemented in the R TreeDist package) (Smith, 2020a), which is a direct measure of the quantity of information in the ground truth phylogeny that was successfully captured in the reconstruction. It is relatively unaffected by the failure to perfectly reproduce multifurcations. For the purposes of easy comparison to the literature, we also measured the Clustering Information Distance (Smith, 2020a).

Across ground truth phylogenies, we were able to reconstruct the phylogenetic topology with between 47.75% and 85.70% of the information contained in the original tree using a 64-bit column memory footprint, between 47.75% and 80.36% using a 512-bit column memory footprint, and between 51.13% and 83.53% using a 4096-bit column memory footprint. While the Clustering Information Distance reached its maximum possible score (1.0) for the heavily-multifurcated EcoEA phylogeny, it agreed with the Mutual Clustering Information score for less multifurcated phylogenies, such as fitness sharing. Using the Recency Proportional Resolution retention policy and a 4096-bit column memory footprint, we were able to reconstruct a fitness sharing phylogeny with a Clustering Information Distance of only 0.2923471 from the ground truth. For context, that result is comparable to the distance between phylogenies reconstructed from two closely-related proteins in H3N2 flu (0.25) (Jones et al., 2021). To build further intution, we strongly encourage readers to refer to our interactive web reconstruction. Figure 5 summarizes error reconstructing the fitness sharing selection phylogeny in terms of the mutual clustering information metric (Smith, 2022). The phylogenies reconstructed from the EcoEA condition performed comparably, with lexicase and random selection faring somewhat worse (Moreno et al., 2022). In the case of random selection, we suspect that this reduced

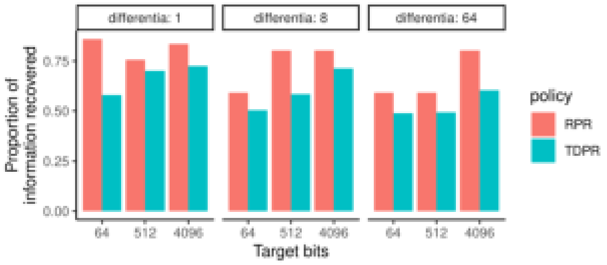


Figure 5: Proportion of information present in the ground-truth ftness sharing phylogeny that was captureed by our reconstruction, across various retention policies. High is better (1 is perfect). RPR is recency-proportional resolution policy and TDPR is tapered depth-proportional resolution policy.

performance is the result of having many nodes that originated very close together at the end of the experiment. As expected, we did observe overall more accurate reconstructions from columns that were allowed to occupy larger memory footprints.

Differentia Size

Among the surveyed ground truth phylogenies and target column footprints, we consistently found that smaller differentia were able to yield more or as accurate phylogenetic reconstructions. The stronger performance of narrow differentia was particularly apparent in low-memory-footprint scenarios where overall phylogenetic inference power was weaker. Overall, single-bit differentia outperformed 64-bit differentia under 20 condtions, and were indistinguishable under 7 conditions, and were worse under 3 conditions. Full results are available in Supplementary Section . Although narrower differentia have less distinguishing power on their own, their smaller size allows more to be packed into the memory footprint to cover more generations, which seems to help reconstruction power. We must note that narrower differentia can pack more thoroughly into the footprint caps we imposed on column size, so their extant columns tended to have slightly more overall bits. However, this was a small enough imbalance (in most cases $< 10\%$) that we believe it is unlikely to fully account for the stronger performance of narrow-differentia configurations.

Retention Policy

Across the surveyed ground truth phylogenies and target column memory footprints, we found that the recency-proportional resolution stratum retention policy generally yielded better phylogenetic reconstructions. Phylogenetic reconstruction quality was better in 28 conditions, equivalent in 14 conditions, and worse in 3 conditions. Again, this effect was most apparent in the small-stratum-count scenarios where overal inference power was weaker. Full results are available in Supplementary Section . The stronger performance of recency-proportional resolution is likely due to the denser retention of recent strata under the recency-proportional metric, which help to resolve the more numerous (and therefore typically more tightly spaced) phylogenetic events in the near past (Zhaxybayeva and Gogarten, 2004). Recency-proportional resolution tended to be able to fit fewer strata within the prescribed memory footprints (except in cases where it could not fit within the footprint) so its stronger performance cannot be attributed to more retained bits in the end-state extant columns.

Conclusion

To our knowledge, this work provides a novel design for digital genome components that enable phylogenetic inference on asexual populations. This provides a viable alternative to perfect phylogenetic tracking, which is complex and possibly cumbersome in distributed computing scenarios, especially with fallible nodes. Our approach enables flexible, explicit trade-offs between space complexity and inference accuracy. Hereditary stratigraphic columns are efficient: our approach can estimate, for example, the MRCA generation of two genomes within 10% error with 95% confidence up to a depth of a trillion generations with genome annotations smaller than a kilobyte. However, they are also powerful: we were able to achieve tree reconstructions recovering up to 85.70% of the information contained in the original tree with only a 64-bit memory footprint.

This and other methodology to enable decentralized observation and analysis of evolving systems will be essential for artificial life experiments that use distributed and best-effort computing approaches. Such systems will be crucial to enabling advances in the field of artificial life, particularly with respect to the question of open-ended evolution (Ackley and Cannon, 2011; Moreno et al., 2021b,a) Mork work is called for to further enable experimental analyses in distributed, best-effort systems while preserving those systems' efficiency and scalability. As parallel and distributed computing becomes increasingly ubiquitous and begins to more widely pervade artificial life systems, hereditary stratigraphy should serve as a useful technique in this toolbox.

Important work extending and analyzing hereditary stratigraphy remains to be done. Analyses should be performed to expound MRCA resolution guarantees of stratum retention policies when using narrow (i.e., single-bit) differentia. Constant-size-complexity stratum retention policies that preferentially retain a denser sampling of more-recent strata should be developed and analyzed. Extensions to sexual populations should be explored, including the possibility of annotating and tracking individual genome components instead of whole-genome individuals. An alternate approach might be to define a preferential inheritance rule so that at each generation slot within a column, a single differentia sweeps over an entire interbreeding population. Optimization of tree reconstruction from extant hereditary stratigraphs remains an open question, too, particularly with regard to properly handling multifurcations. It would be particularly valuable to develop methodology to annotate inner nodes of trees reconstructed from hereditary stratigraphs with confidence levels.

The problem of designing genomes to maximize phylogenetic reconstructability raises unique questions about phylogenetic estimation. Such a backward problem — optimizing genomes to make analyses trivial as opposed to the usual process of optimizing analyses to genomes — puts questions about the genetic information analyses operate on in a new light. In particular, it would be interesting to derive upper bounds on phylogenetic inference accuracy given genome size and generations elapsed.

Acknowledgment

This research was supported in part by NSF grants DEB-1655715 and DBI-0939454 as well as by Michigan State University through the computational resources provided by the Institute for Cyber-Enabled Research. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1424871. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Thank you to Santiago Rodriguez Papa for manuscript comments.

References

Ackley, D. H. and Cannon, D. C. (2011). Pursue robust indefinite scalability. In *13th Workshop on Hot Topics in Operating Systems (HotOS XIII)*.

Bohm, C., Hintze, A., et al. (2017). Mabe (modular agent based evolver): A framework for digital evolution research. In *ECAL 2017, the Fourteenth European Conference on Artificial Life*, pages 76–83. MIT Press.

Bostock, M., Ogievetsky, V., and Heer, J. (2011). D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309.

Casci, T. (2008). Lining up is hard to do. *Nature Reviews Genetics*, 9(8):573–573.

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423.

Dolson, E., Lalejini, A., Jorgensen, S., and Ofria, C. (2020). Interpreting the tape of life: Ancestry-based analyses provide insights and intuition about evolutionary dynamics. *Artificial Life*, 26(1):58–79.

Dolson, E. and Ofria, C. (2018). Ecological theory provides insights about evolutionary computation. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 105–106.

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.

Foster, E. D. and Deardorff, A. (2017). Open science framework (osf). *Journal of the Medical Library Association: JMLA*, 105(2):203.

Goings, S., Goldsby, H., Cheng, B. H., and Ofria, C. (2012). An ecology-based evolutionary algorithm to evolve solutions to complex problems. In *ALIFE 2012: The Thirteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 171–177. MIT Press.

Goldberg, D. E., Richardson, J., et al. (1987). Genetic algorithms with sharing for multimodal function optimization. In *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, volume 4149. Hillsdale, NJ: Lawrence Erlbaum.

Hagstrom, G. I., Hang, D. H., Ofria, C., and Torng, E. (2004). Using avida to test the effects of natural selection on phylogenetic reconstruction methods. *Artificial life*, 10(2):157–166.

Helmuth, T., Spector, L., and Matheson, J. (2014). Solving uncompromising problems with lexicase selection. *IEEE Transactions on Evolutionary Computation*, 19(5):630–643.

Hernandez, J. G., Lalejini, A., and Dolson, E. (2022). What Can Phylogenetic Metrics Tell us About Useful Diversity in Evolutionary Algorithms? In Banzhaf, W., Trujillo, L., Winkler, S., and Worzel, B., editors, *Genetic Programming Theory and Practice XVIII*, pages 63–82. Springer Singapore, Singapore.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Jones, J. E., Le Sage, V., Padovani, G. H., Calderon, M., Wright, E. S., and Lakdawala, S. S. (2021). Parallel evolution between genomic segments of seasonal human influenza viruses reveals rna-rna relationships. *Elife*, 10:e66525.

Kapli, P., Yang, Z., and Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7):428–444.

Kauffman, S. A. and Weinberger, E. D. (1989). The nk model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of theoretical biology*, 141(2):211–245.

Lack, J. B. and Van Den Bussche, R. A. (2010). Identifying the confounding factors in resolving phylogenetic relationships in vespertilionidae. *Journal of Mammalogy*, 91(6):1435–1448.

Lalejini, A., Dolson, E., Bohm, C., Ferguson, A. J., Parsons, D. P., Rainford, P. F., Richmond, P., and Ofria, C. (2019). Data standards for artificial life software. In *ALIFE 2019: The 2019 Conference on Artificial Life*, pages 507–514. MIT Press.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423(6936):139–144.

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J. K., Singh, S., et al. (2017). Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103.

Moreno, M. A., Dolson, E., and Ofria, C. (2022). Hereditary stratigraph concept supplement. Available at https://doi.org/10.17605/osf.io/4sm72.

Moreno, M. A., Papa, S. R., and Ofria, C. (2021a). Case study of novelty, complexity, and adaptation in a multicellular system. In *OEE4: The Fourth Workshop on Open-Ended Evolution*.

Moreno, M. A., Papa, S. R., and Ofria, C. (2021b). Conduit: a c++ library for best-effort high performance computing. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1795–1800.

OEIS (2021a). Sequence a056791. The on-line encyclopedia of integer sequences. Available at https://oeis.org/A056791.

OEIS (2021b). Sequence a063787. The on-line encyclopedia of integer sequences. Available at https://oeis.org/A063787.

Paradis, E., Claude, J., and Strimmer, K. (2004). Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290.

Ragan-Kelley, B. and Willing, C. (2018). Binder 2.0-reproducible, interactive, sharable environments for science at scale. In *Proceedings of the 17th Python in Science Conference (F. Akici, D. Lippa, D. Niederhut, and M. Pacer, eds.)*, pages 113–120.

Sarkar, S., Majumder, T., Kalyanaraman, A., and Pande, P. P. (2010). Hardware accelerators for biocomputing: A survey. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 3789–3792. Ieee.

Smith, M. R. (2020a). Information theoretic generalized robinson-foulds metrics for comparing phylogenetic trees. *Bioinformatics*, 36(20):5007–5013.

Smith, M. R. (2020b). ms609/treedistdata: v1.0.0.

Smith, M. R. (2022). Robust analysis of phylogenetic tree space. *Systematic Biology*.

Sokal, R. R. (1958). A statistical method for evaluating systematic relationships. *Univ. Kansas, Sci. Bull.*, 38:1409–1438.

Stanley, K. O. and Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127.

Steno, N. (1916). *The prodromus of Nicolaus Steno's dissertation concerning a solid body enclosed by process of nature within a solid*, volume 11. University of Michigan Press.

Sukumaran, J. and Holder, M. T. (2010). Dendropy: a python library for phylogenetic computing. *Bioinformatics*, 26(12):1569–1571.

Ushey, K., Allaire, J., and Tang, Y. (2022). *reticulate: Interface to 'Python'*. https://rstudio.github.io/reticulate/, https://github.com/rstudio/reticulate.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wang, R., Clune, J., and Stanley, K. O. (2018). Vine: an open source interactive data visualization tool for neuroevolution. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1562–1564.

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

Wickham, H., François, R., Henry, L., and Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

Zhaxybayeva, O. and Gogarten, J. P. (2004). Cladogenesis, coalescence and the evolution of the three domains of life. *TRENDS in Genetics*, 20(4):182–187.

Growing Isotropic Neural Cellular Automata

Alexander Mordvintsev,[1] Ettore Randazzo,[1] and Craig Fouts[2]

[1] Google Research
{moralex, etr}@google.com

[2] The Ohio State University
fouts.57@osu.edu

Abstract

Modeling the ability of multicellular organisms to build and maintain their bodies through local interactions between individual cells (morphogenesis) is a long-standing challenge of developmental biology. Recently, the Neural Cellular Automata (NCA) model was proposed as a way to find local system rules that produce a desired global behaviour, such as growing and persisting a predefined target pattern, by repeatedly applying the same rule over a grid starting from a single cell. In this work, we argue that the original Growing NCA model has an important limitation: anisotropy of the learned update rule. This implies the presence of an external factor that orients the cells in a particular direction. In other words, "physical" rules of the underlying system are not invariant to rotation, thus prohibiting the existence of differently oriented instances of the target pattern on the same grid. We propose a modified Isotropic NCA (IsoNCA) model that does not have this limitation. We demonstrate that such cell systems can be trained to grow accurate asymmetrical patterns through either of two methods: **(1)** by breaking symmetries using structured seeds or **(2)** by introducing a rotation-reflection invariant training objective and relying on symmetry-breaking caused by asynchronous cell updates.

Introduction

Every multicellular organism begins its life as a single cell. Descendants of this egg cell reliably form complex structures of an organism through a process of division and differentiation, also known as morphogenesis (Turing, 1990). In many cases, this process doesn't require any external control or orchestration and is described as self-organizing; cells communicate with their neighbors to make collective decisions about the overall body layout and composition. Understanding this process is an active area of research (Pezzulo and Levin, 2016) with a number of models proposed to explain the development procedure of various tissues (Malheiros et al., 2020) and organisms.

There is a wide spectrum of existing models that either attempt to reproduce biological processes or simplify them in order to accomplish engineering tasks or simulate artificial life. One widespread approach is the use of Artificial Gene Regulatory Networks (GRN) (Cussat-Blanc et al., 2019; de Jong, 2002), which have, for instance, been used to model cell growth (Schramm et al., 2012; Jacobsen et al., 1998). Discovering effective models with these networks often relies on manual engineering or genetic algorithms. Although these models have demonstrated impressive results, we think that genetic algorithms are limited in the model complexity and fitness accuracy that is possible to achieve within a plausible time and compute budget. Differentiable programming has shown itself as a powerful and versatile approach to solving complex engineering problems, including morphogenesis modeling.

One recent approach to modelling morphogenesis is based on Neural Cellular Automata (NCA) (Mordvintsev et al., 2020). Here, the authors represent a growing organism with a uniform grid of raster cells where the state of each cell is characterized by a set of scalar values. Cells repeatedly update their states using a rule defined by a small neural network that takes as input information collected from each cell's neighbors at the current moment in time. Backpropagation through time is used to learn the local update rule that satisfies the global objective of growing a predefined target pattern, allowing for the discovery of much more complex rules than what has been possible with existing strategies. While this is an attractive direction to pursue for modeling complex systems, we observe that this model is still not *fully* self-organising.

Anisotropy of Neural CA

Figure 1 demonstrates the weakness of the original Growing NCA model which challenges the claims of fully self-organizing pattern growth achieved by this model. This model can only grow and persist patterns in a specific orientation that is determined by properties of the space itself rather than the intrinsic states of the cells occupying said space. The NCA anisotropy stems from the axis-aligned Sobel filters that are used to model cell perception. In the last experiment, [1] authors show that altering properties of the grid (Sobel filter directions) leads to rotations of the resultant pattern, but don't address the main concern that pattern

[1] `https://distill.pub/2020/growing-ca/#experiment-4`

orientation should be defined by the configuration of cells occupying the space and not a property of the space itself. In the follow up work on NCA texture synthesis (Niklasson et al., 2021b), the same group of authors experiment with varying the filter directions across space, reinforcing the idea of external control on each cell's perception.

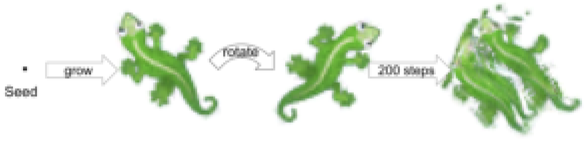


Figure 1: Anistropy of the Growing NCA model. Cells rely on externally-provided global cell alignment and are unable to sustain a pattern if cell states are re-sampled in a rotated coordinate frame. In contrast, real living creatures can usually tolerate rotation without exploding.

In this work we argue that the original Growing NCA model is not *fully* self-organizing due to a limitation in the model's architecture. The learned update rule is anisotropic, which implies the presence of an external factor that orients all cells in a particular direction. In other words, "physical" rules of the underlying system are not invariant to rotation, prohibiting the formation of differently oriented instances of the target pattern on the same grid. This would be akin to an animal only able to grow, or even exist, when facing north. We aim to relax this limitation with following contributions:

- Propose a simplification to the original NCA update rule to make it isotropic, so the perception of each cell is invariant to rotation or reflection of the grid.
- Show that this invariance enables us to perform rotations, reflections, and other augmentations on structured seeds that predictably influence the model's behavior.
- Design a rotation-reflection invariant training objective that steers the system towards reliable growth of asymmetric, anisotropic patterns through symmetry-breaking rather then external guidance.
- Demonstrate the robustness of the learned NCA rule to out-of-training grid structures.

Isotropic Neural CA model

The Isotropic NCA (IsoNCA) model described here can be seen as a more restricted version of the Growing NCA model, where the Sobel X and Y perception filters are replaced with a single Laplacian filter. This section covers key features of the model design.

Grid Cells exist on a regular Cartesian grid; the state of each cell is represented by a vector

$$\mathbf{s} = [s^0 = R, s^1 = G, s^2 = B, s^3 = A, ..., s^{C-1}]$$

where C is the number of channels and the first four channels represent a visible RGBA image. Initially, the whole grid is set to zeros, except the seed cell in which $A = 1$. Cells iteratively update their states using only the information collected from their 3x3 Moore neighbourhood.

Stochastic updates Cell updates happen stochastically; at every NCA step each cell is updated with probability p_{upd} (we use value 0.5 in our experiments). This stochasticity is meant to eliminate dependence on a global shared clock that synchronizes the updates between cells. Previous work on NCA discusses the impact of this strategy on NCA robustness (Niklasson et al., 2021a). In IsoNCA models, asynchronicity plays a critical role in the symmetry breaking process (see the Results section). This asynchronicity can be seen as a strategy for generating noise, which has been documented to help biological systems construct complex functions in simple ways (Samoilov et al., 2006).

"Alive" and "empty" cells The alpha channel ($s^3 = A$) plays a special role in determining whether a cell is currently "alive" or "empty"; each cell is alive if $A > 0.1$ or if it has at least one alive cell in its 3x3 neighbourhood. The state of empty cells is explicitly set to zeros after each CA step.

Perception Each cell collects information about the state of its neighborhood using a per-channel discrete 3x3 Laplacian filter. This filter computes the difference between the state of the cell and the average state of its neighbours. A cell's perception vector is the concatenation of its own current state and per-channel Laplacians of its neighbourhood:

$$\mathbf{p} = concat(\mathbf{s}, K_{lap} * \mathbf{s})$$

where $\mathbf{s}$ denotes the cell's state and K_{lap} is given by

$$K_{lap} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Update rule Cells stochastically update their states using a learned rule that is represented by a two-layer neural network:

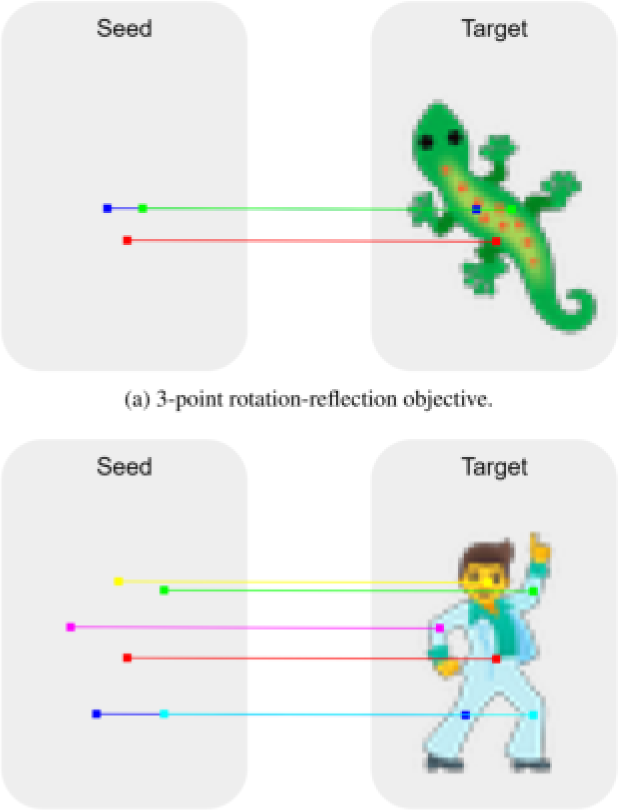
$$\mathbf{s}_{t+1} = \mathbf{s}_t + relu(\mathbf{p}_t W_0 + b_0) W_1$$

where parameters W_0, b_0, and W_1 have shapes $(32, 192)$, (192) and $(192, 16)$ respectively, which gives a total of 9408 learned parameters.

Training IsoNCA

The original Growing NCA model was trained to learn an update rule starting from a single seed cell and based on a target pattern fixed in a specific orientation. Under this training regime, our isotropic restrictions prevent the model from breaking spatial symmetries, demanding we either relax the target objective or further specify the initial conditions. Here, we propose and present implementations of both of these strategies.

Structured seed strategy



(a) 3-point rotation-reflection objective.

(b) 6-point structural mutation objective.

Figure 2: Mapping between structured seed points and their respective structural features in the target pattern. Points are placed manually to correspond with these features and distinct point colors are generated by converting equidistant HSV hues into RGB encodings.

One way to train our isotropic model is by introducing a more comprehensive seed structure (e.g. by using three or more distinct points) in order to break symmetries of the system from the initial state. This is inspired by synthetic reaction-diffusion networks used to govern complex pattern growth (Scalise and Schulman, 2014). A *structured seed* is defined by its number of points, the initial channel encodings of those points, and their positions relative to one another. Given that an isometry relating any pair of congruent triangles is unique (Coxeter, 1963), we use 3 non-collinear points to define the seed's orientation, with points distributed uniformly on a circular edge of predefined radius. To establish directional responsibility, points are distinguished by their RGB encodings. An example of how one such seed maps to a target pattern is shown in Figure 2a. Training our model to grow in alignment with this seed then proceeds identically to the training of the original Growing NCA's single-seed scenario.

Alternatively, structured seeds can be manually engineered to map key features of the target pattern to specific points of the seed. So long as there are three or more non-collinear points, this configuration enables the model to break symmetries similarly to the triangular seed. Figure 2b shows one such mapping between the appendages of a dancer pattern and their respective point assignments in the corresponding seed. These points are reconfigurable and consequently used to grow predictable out-of-training structural mutations of the target pattern. Changing the configuration of a structured seed is performed by modifying the positions and channel encodings of its composite points. For example, if a structured seed is comprised of points A and B, then replacing point A with point B involves adjusting the RGB encoding of point A to match that of point B.

Single-seed strategy

Similar to the Growing NCA work, we also train NCA models that grow and persist a predefined pattern on a plane starting from a single seed cell. We use pixel-wise differences to match the pattern produced by the trained model to the target, but modify the loss function to make it rotation-reflection invariant. Due to the rotation symmetry of the cell perception we are unable (and do not want) to enforce a particular pattern orientation and chirality during training. Instead we select an individual rotation and reflection of the target that minimizes the pixel-wise loss value for each NCA-generated sample in the training batch. We expect that this would cause the model to amplify noise coming from stochastic cell updates to break the rotational symmetry and make collective decision on the grown pattern orientation.

A naive implementation of the previously described procedure would require matching against densely sampled rotated and reflected instances of the training pattern, which is computationally inefficient. Instead, we use a polar coordinate transformation and FFT to efficiently compute the discrepancy across different target rotations. Figure 3 shows the architecture of the proposed loss function; here, we match patterns in the polar coordinate system, where rotations become horizontal translations. For each possible rotation angle θ, we compute the sum of squared pixel-wise differences between the NCA-generated pattern S and the target T across all radius values r and target channels c. Target channels include RGBA color representations and extra

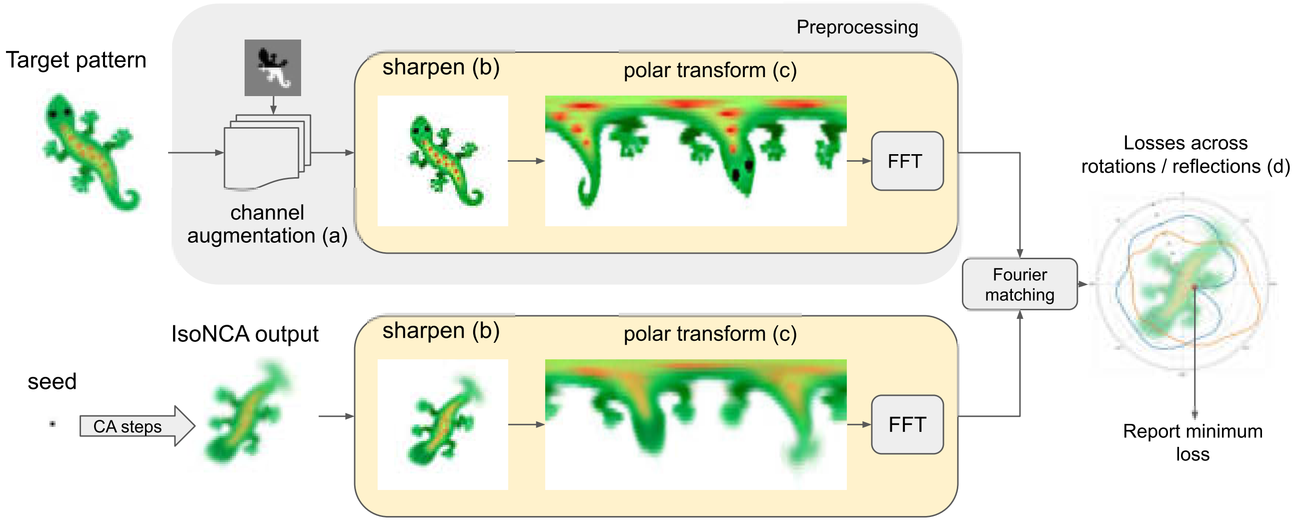


Figure 3: Rotation / reflection-invariant IsoNCA training pipeline. The target pattern is augmented with extra channels **(a)** to break symmetries that may interfere with training. Both the target and NCA-grown patterns are sharpened **(b)** to steer optimization into preserving fine details. A polar transformation **(c)** is applied to turn the unknown rotation between two images into a horizontal shift. Fourier-domain image matching enables efficient computation of pixel-matching losses across all orientations and reflections **(d)**. The blue plot shows losses with respect to the original pattern, and the yellow with respect to its reflected version. The minimal loss value is selected for backpropagation.

auxiliary channels that help the optimization break symmetries and escape sub-optimal local minima (see Figure 4). The rotation-invariant loss can be expressed as follows:

$$L(S,T) = \min_{\theta} \sum_{r,c} L_{r,c,\theta}$$

$$L_{r,c,\theta} = \sum_{\theta'} (S_{r,c,\theta'} - T_{r,c,\theta'-\theta})^2 =$$

$$\sum_{\theta'} S^2_{r,c,\theta'} + \underbrace{\sum_{\theta'} T^2_{r,c,\theta'-\theta}}_{\text{doesn't depend on } \theta} \underbrace{-2\sum_{\theta'} S_{r,c,\theta'} T_{r,c,\theta'-\theta}}_{\text{1D convolution}}$$

The 1D convolution term in this expression can be efficiently computed for all values of θ using the convolution theorem:

$$s * t = \mathcal{F}^{-1}(\mathcal{F}(S) \cdot \mathcal{F}(T)),$$

where $\mathcal{F}$ denotes the Fourier transform and "$\cdot$" indicates elementwise multiplication. To make the pattern-matching loss reflection-invariant, we also compute the loss for a reflected version of the target pattern, and select the minimum between the two:

$$L_{\text{inv}} = \min(L(S,T), L(S, \mathit{reflect}(T)))$$

Auxiliary channels Having a rotation-invariant loss causes some patterns to exhibit strong local minima. For instance, models trained on the lizard pattern are be unable to reliably break one symmetry and often create a mixture of head-tails along with an assortment of limbs because they fail to differentiate up from down (see Figure 7). This issue can be rectified using an auxiliary loss by adding a new "target channel" to the rotation-reflection invariant loss in which the image is split in half and the upper and lower parts of the image have target values of -0.5 and $+0.5$ respectively (see

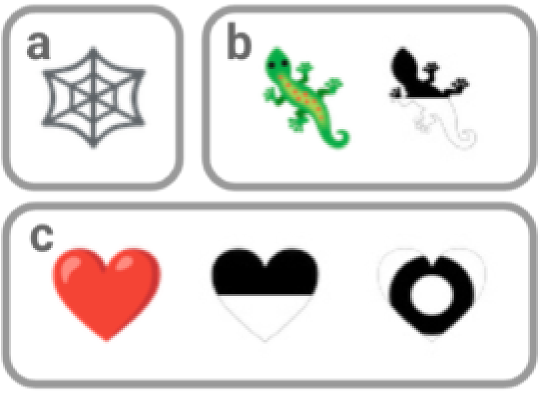


Figure 4: Target patterns and auxiliary targets used for the ration-reflection invariant training. **(a)** the spiderweb does not use auxiliary targets. **(b)** the lizard uses the binary auxiliary target. **(c)** the heart uses both binary and radial encoding auxiliary targets.

Figure 4b). We call this additional channel the "binary auxiliary channel." Including it enables models to learn to break symmetries and results in a much smoother loss. We have not observed any need to add an additional perpendicular auxiliary target to break the left-right symmetry.

We have also observed that IsoNCA struggles to generate uniform patterns, such as the heart emoji, suffering from stability issues and being unable to form the proper shape. We hypothesized that the shape of the pattern does not facilitate smooth training and decided to enhance the target image in a similar way to how we break symmetries using another auxiliary loss. Inspired by positional encodings used in the transformer architecture (Vaswani et al., 2017), we add target channels generated by aliased concentric circles with the following rule: given a point with distance r from the center and mode n, the positional encoding of the point is given by

$$\begin{cases} sign(sin(r*n*\pi))*0.5, & \text{if n is even} \\ sign(cos(r*n*\pi))*0.5, & \text{otherwise} \end{cases}$$

Thus, the image can have arbitrarily many auxiliary channels with different modes to generate radial encodings. For the heart emoji, one radial encoding with mode 4 is sufficient (Figure 4c).

In this paper, we showcase three example patterns: the spiderweb, where no auxiliary loss is needed; the lizard, where the binary auxiliary channel is needed; and the heart, where both the binary and radial encoding auxiliary channels are needed (see Figure 4).

Results

Structured seed experiments

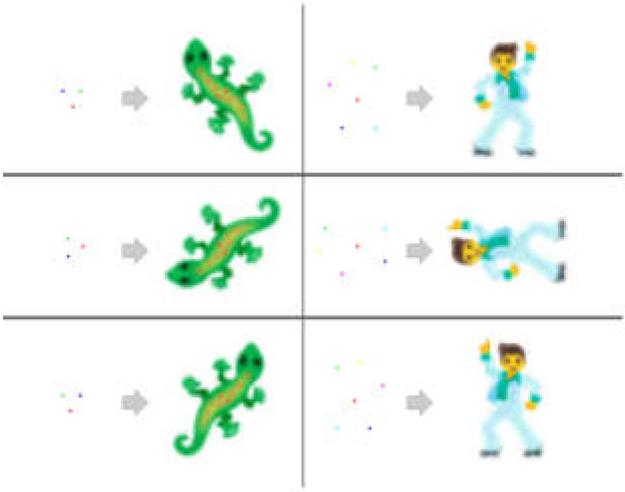


Figure 5: Pairs of initial seed configurations and their resulting unfoldings at step 5000 for several out-of-training isometries.

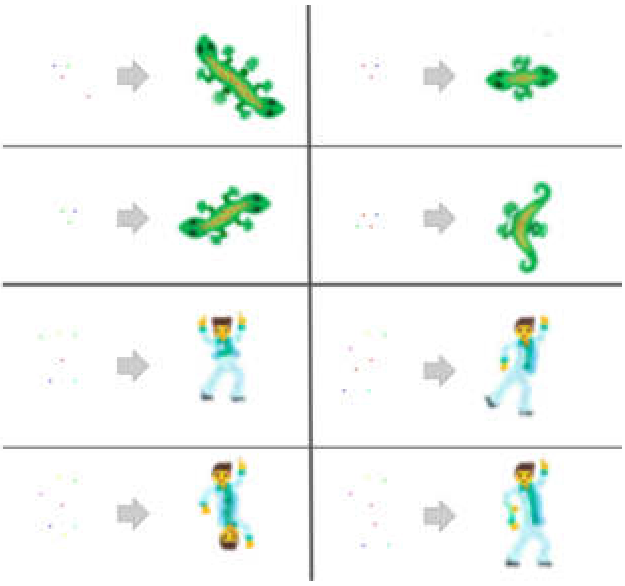


Figure 6: Pairs of initial seed configurations and their resulting unfoldings at step 5000 for several out-of-training mutations.

In the original Growing NCA work, it is shown that the model can learn to grow at predefined angles, but it fails to extrapolate to unseen orientations after training is complete. Our model not only successfully addresses this limitation, but expands upon how orientation can be implicitly encoded using structured seeds.

Distributing growth responsibility between the points of a structured seed enables us to manipulate the model's behavior by performing plane isometries on the seed. In performing such an isometry, we see the transformation propagate throughout the model's growth of the resultant pattern without loss of stability. As shown in Figure 5, IsoNCA exhibits stable rotations and reflections irrespective of the fixed orientation of the target pattern.

Apart from plane isometries, we observe that the isotropic properties of our model emerge within sub-regions of the growing pattern. Reconfiguring the structured seed after training thus causes our model to manifest structural mutations corresponding to the modified seed configuration. As shown in Figure 6, our model grows a variety of out-of-training patterns exhibiting seed-directed mutations while maintaining the cohesion of the resultant shape. The stability of such irregular modifications is observed to be less reliable than that of plane isometries, however, and further seed manipulation often yields improved results. For example, by adjusting the supplanted seed point's relative position to more closely resemble that of its original counterpart, we often find that mutations stabilize with more consistently desirable behavior.

Single seed experiments

Target augmentation As discussed in the ”auxiliary channels” section, we augment some target patterns with extra channels to facilitate optimization convergence to a desirable solution. The lizard pattern is a particularly interesting case; without augmentation, the optimization fails to break the symmetry between the head and the tail, which leads to formation of the ghostly symmetrical pattern seen in Figure 7, left. What's more, the heart pattern fails to converge to a definite shape at all without both radial and top-bottom contrast augmentations.

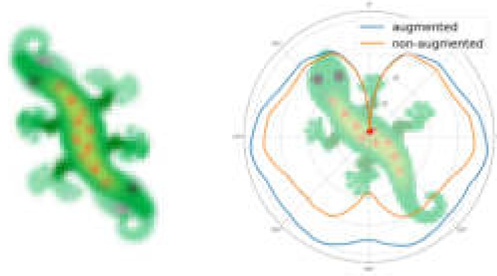


Figure 7: **Left:** the result of training IsoNCA to grow a non-augmented lizard pattern. Optimization struggles to break the head-tail symmetry. **Right:** matching losses computed between the target pattern and it's rotated instances. The non-augmented pattern has a strong local minima corresponding to a 180° rotation. Augmenting the target with a non-symmetrical auxiliary channel flattens the spurious minima and facilitates convergence to the correct solution.

Stochastic symmetry breaking Growing non-rotationally-symmetrical patterns starting from a single seed is fundamentally different from the structured seed scenario. For example, on a regular square grid, both a cell's perception field and starting condition have symmetries that need to be broken during pattern development. We use stochastic asynchronous cell updates, which cells may rely on as a source of randomness. During training, cells successfully develop a protocol for making a collective decision about the final pattern layout that appears to rely on this randomness. To validate this assumption, we ran the IsoNCA rule, trained in the stochastic update regime, in a fully synchronous setting ($p_{\text{upd}} = 1$). We expected that our model wouldn't be able to break symmetries between grid directions, and would produce some 90° rotation- and reflection-symmetrical patterns. To our surprise, we instead observed that after approximately 100-150 iterations, asymmetries develop, and eventually our model produces an incomplete, unstable, but definitely not symmetrical version of the target pattern (see Figure 8, top). Puzzled by this behaviour, we discovered that our model was able to exploit the non-associativity of floating point number accumulation in the convolution with the Laplacian filter to break the symmetry. We then implemented an associative version of the Laplacian filter by performing the convolution using fixed-point number representation. This time, as expected, the model produced a symmetrical pattern that vanished after oscillating for about 700 steps (see Figure 8, bottom).

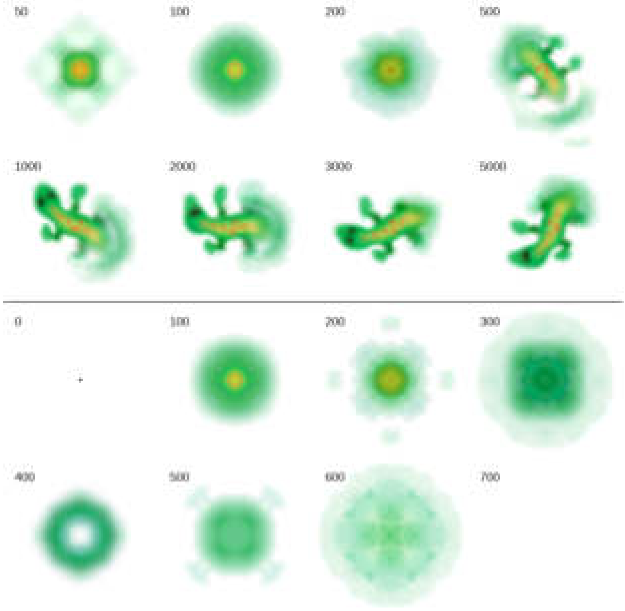


Figure 8: **Top:** IsoNCA manages to break the symmetry even in the case of deterministic synchronous cell updates by exploiting the non-associativity of floating-point number addition in the Laplacian filter convolution. **Bottom:** the evolution of IsoNCA in the case of fully synchronous cell updates and perfectly symmetric perception. As expected, the model can't break symmetries to develop features of the lizard. The pattern produced by this particular checkpoint deterministically vanishes after about 700 steps. Other checkpoints may produce stable or exploding behaviours in this out-of-training regime.

Figure 9 shows the unfolding of three different trained IsoNCA rules, starting from the original seed and running for up to 5000 steps. Here, we can observe how some patterns are more stable than others. For instance, in this training run, the lizard pattern appears to rotate over long time periods. This can sometimes happen as the rotation-reflection invariant loss used is consequently invariant to rotations over time. We observe these models to be stable for any number of steps and choose to visualize 5000 steps only because no further changes manifest beyond this point (besides rotations in the lizard case). Note in Figure 10 how, for single seed experiments, every run of the same model results in a different pattern rotation and reflection.

Figure 11 shows how the lizard model learns to form a

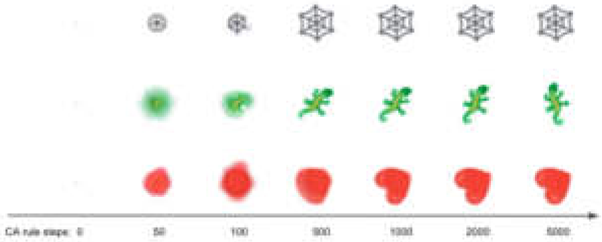


Figure 9: Unfolding of three trained Isotropic NCA rules with single seeds.

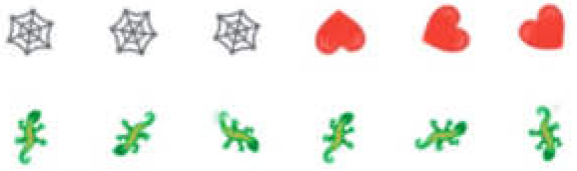


Figure 10: Step 5000 for different runs of three single seed models.

stable pattern during training where the x-axis denotes the number of training steps (up to 5000) and the y-axis indicates the checkpoint of the model at the given training step. The model appears to first learn to generate a green, unstable blob before learning to break symmetries and, eventually, to form lizard-like features, albeit with some instability; over time, the pattern becomes more and more stable. Note how on training step 10000 the model is somewhat stable but first creates two heads and then one of them becomes a tail. Training the model for longer tends to speed up convergence to the desired form and may suppress this behaviour.

Non-regular Grids

Previous work discusses the robustness of NCA models to out-of-training grid structures (Niklasson et al., 2021b). For example, it was found that models trained on a square grid can be executed on hexagonal grids if 3x3 convolutional perception filters are replaced with their hexagonal counterparts. Recently, it was also demonstrated that neural models relying on a diffusion operation for communication are effectively discretization agnostic (Sharp et al., 2022). In this work, we demonstrate that IsoNCA can be effectively transferred from a regular square grid to a non-regular one. To build a non-regular grid, we sample a set of points on a plane using a fast Poisson-disk sampling algorithm (Bridson, 2007). Cells are then constructed by computing a Voronoi diagram of the sampled points. The Laplacian operator is defined as the difference between a cell's own state and the weighted average of the state vectors of its adjacent

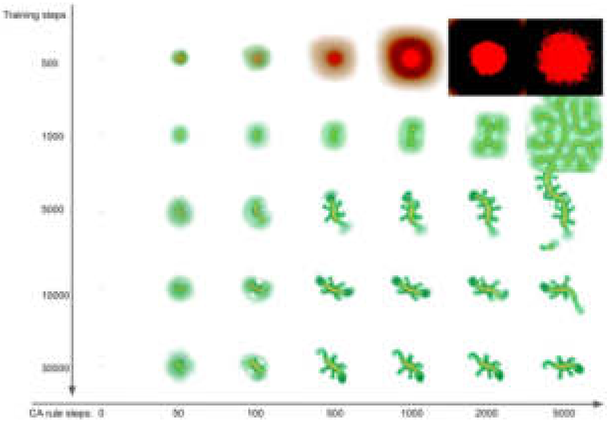


Figure 11: Evolution of the lizard IsoNCA rule during training (y-axis) versus the unfolding of each NCA rule checkpoint (x-axis).

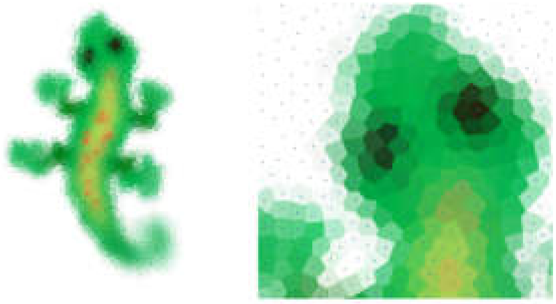


Figure 12: Example of running the lizard IsoNCA on a non-regular grid constructed using a Voronoi diagram on Poisson-disk point samples. **Left**: full pattern. **right**: close-up with voronoi cell centers shown. Running the unmodified IsoNCA rule on a non-regular grid is more prone to artifacts and instabilities. For example, in this case, the tail curl direction doesn't match that of the target pattern.

cells:

$$w_{i,j} = l_{i,j} / \sum_j l_{i,j}$$

where $l_{i,j}$ is the length of the Voronoi diagram edge shared by cells i and j (equal to zero if there is no such edge). Figure 12 shows an example of the lizard pattern grown on a non-regular grid substrate by a square-grid-trained model using the modified Laplacian operator.

Conclusion and future work

In this work we demonstrated the capability of fully isotropic NCA models in reliably growing complex asymmetric patterns even when the perception field of each cell is fully symmetric. We believe this creates an important practical lower bound on the requirements of cell communication capabilities for morphogenesis simulations. Moreover, we think that coupling IsoNCA with physically grounded models of cell division and migration may elicit exciting possibilities for the accurate reproduction of body growth and regeneration phenomena and even enable new bio-engineering applications in the future.

References

Bridson, R. (2007). Fast poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH 2007 sketches on - SIGGRAPH '07*, New York, New York, USA. ACM Press.

Coxeter, H. S. M. (1963). *Introduction to Geometry*. Wiley New York.

Cussat-Blanc, S., Harrington, K., and Banzhaf, W. (2019). Artificial Gene Regulatory Networks—A Review. *Artificial Life*, 24(4):296–328.

de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103. PMID: 11911796.

Jacobsen, T. L., Brennan, K., Arias, A. M., and Muskavitch, M. A. (1998). Cis-interactions between delta and notch modulate neurogenic signalling in drosophila. *Development*, 125(22):4531–4540.

Malheiros, M. d. G., Fensterseifer, H., and Walter, M. (2020). The leopard never changes its spots: realistic pigmentation pattern formation by coupling tissue growth with reaction-diffusion. *ACM Trans. Graph.*, 39(4).

Mordvintsev, A., Randazzo, E., Niklasson, E., and Levin, M. (2020). Growing neural cellular automata. *Distill*. https://distill.pub/2020/growing-ca.

Niklasson, E., Mordvintsev, A., and Randazzo, E. (2021a). Asynchronicity in neural cellular automata. volume ALIFE 2021: The 2021 Conference on Artificial Life of *ALIFE 2021: The 2021 Conference on Artificial Life*. 116.

Niklasson, E., Mordvintsev, A., Randazzo, E., and Levin, M. (2021b). Self-organising textures. *Distill*. https://distill.pub/selforg/2021/textures.

Pezzulo, G. and Levin, M. (2016). Top-down models in biology: explanation and control of complex living systems above the molecular level. *J. R. Soc. Interface*, 13(124).

Samoilov, M. S., Price, G., and Arkin, A. P. (2006). From fluctuations to phenotypes: the physiology of noise. *Sci. STKE*, 2006(366):re17.

Scalise, D. and Schulman, R. (2014). Designing modular reaction-diffusion programs for complex pattern formation. *TECHNOLOGY*, 02(01):55–66.

Schramm, L., Jin, Y., and Sendhoff, B. (2012). Evolution and analysis of genetic networks for stable cellular growth and regeneration. *Artif. Life*, 18(4):425–444.

Sharp, N., Attaiki, S., Crane, K., and Ovsjanikov, M. (2022). DiffusionNet: Discretization agnostic learning on surfaces. *ACM Trans. Graph.*, 41(3):1–16.

Turing, A. M. (1990). The chemical basis of morphogenesis. 1953. *Bull. Math. Biol.*, 52(1-2):153–97; discussion 119–52.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need.

Finding Chemical Organisations in Matter-Conserving AChems

Jonathan Young[1] and Simon Colton[1,2]
[1]EECS, Queen Mary University of London, London, UK
[2]SensiLab, Faculty of IT, Monash University, Melbourne, Australia
j.a.young@qmul.ac.uk

Abstract

Chemical Organisation Theory (COT) provides a way of understanding the evolution of collectively self-maintaining sets of molecular species. Adding conservation of matter to an Artificial Chemistry (AChem) can increase evolutionary activity, so it may be useful to understand the evolution of organisations under these conditions. We show how in a reaction network generated by a matter-conserving chemistry, every edge within an organisation must be a part of a cycle in the organisation's bipartite representation. A consequence of this fact is used to alter an existing algorithm to more efficiently discover the complete set of organisations. The altered algorithm is shown to be faster than the original when tested on reaction networks generated by the Spiky-RBN AChem. Overall, this paper contributes useful tools for analysing chemical evolution in matter-conserving chemistries.

Introduction

Artificial Chemistries (AChems) are commonly abstract computational models that are employed to better understand generic principles of life and its origin which are not contingent upon terrestrial conditions and history (Dittrich et al., 2001; Langton, 1989). Evolving self-maintaining or self-replicating networks of molecular species may have preceded the development of life (Kauffman, 1986; Hordijk et al., 2010). Chemical Organisation Theory (COT) can be used to identify subnetworks of a reaction network that do not create any molecular species outside the subnetwork and that are able to maintain their existence over time (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008). These subnetworks are called *organisations*. The concept of an organisation was originally introduced as part of a minimal theory of biological organisation (Fontana and Buss, 1994). Hence, COT is an appropriate means of analysing the evolution of chemical organisations in models of prebiotic evolution (Matsumaru et al., 2006b). Algorithms for finding the complete set of organisations in a reaction network can be expensive (Centler et al., 2008), but there are more efficient algorithms for reaction networks that are generated under specific conditions (Milreu et al., 2010; Speroni di Fenizio, 2015). For example, the set of organisations generated by an AChem forms a lattice if every molecular species has a nonzero probability of disappearing from the reaction vessel (Speroni di Fenizio, 2015). An algorithm that takes advantage of this property can find the complete set of organisations much faster than a brute force approach (Speroni di Fenizio, 2015).

In some AChems, the total amount of matter is conserved over time. In this paper, *matter* refers to an indivisible unit which everything within the system is made from. We describe matter as being *conserved* in a chemistry if the total number of these units within the system cannot change over time and every reaction consumes the same amount of units that it produces. AChems have been shown to behave differently when matter-conservation is added. For example, conserving the total matter in the Stringmol (Hickinbotham et al., 2010) AChem can increase evolutionary activity (Hickinbotham and Stepney, 2015). The Spiky-RBN (Krastev et al., 2016) AChem is also found to behave differently in matter-conserving reaction vessels (Krastev et al., 2017). Milreu et al. (2010) hypothesise that more efficient algorithms for finding organisations may exist for matter-conserving chemistries. But to the best of our knowledge, no such algorithm has been proposed. Such an algorithm would make it easier to characterise the evolution of organisations under such conditions. This could help facilitate research, examples of which may include comparing the evolutionary behaviours of matter-conserving AChems, or comparing the behaviour of an AChem with and without matter-conservation.

Here we prove that in a reaction network generated by a matter-conserving chemistry (artificial or natural), every edge within an organisation must be a part of a cycle in the organisation's bipartite representation. A consequence of this fact is used to alter the constructive algorithm proposed by Centler et al. (2008) such that the complete set of organisations can be discovered more efficiently. We demonstrate that this altered algorithm is approximately nine times faster than the unaltered version when tested on reaction networks generated by the Spiky-RBN (Krastev et al., 2016) AChem.

In this paper, we discuss COT and introduce its core definitions, then provide all relevant theorems before providing the altered version of the constructive algorithm, and then present experimental results and discuss conclusions.

Chemical Organisation Theory

Minimal biological systems have been characterized as chemical systems that are capable of self-maintenance and reproduction. Fontana and Buss (1994) introduced the concept of a biological organisation as an operationally closed and self-maintaining chemical system. This work was further developed into COT, the goal of which is to identify subnetworks of a reaction network that do not create any molecular species outside the subnetwork and that are able to maintain themselves over time (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008). COT can be applied to any system in which entities can interact and form new entities (Dittrich and Speroni di Fenizio, 2007). In particular, It has been used to study models of the martian atmosphere (Centler and Dittrich, 2007), HIV infection (Matsumaru et al., 2006a), ecological systems (Veloz, 2020) and political systems (Dittrich and Winter, 2008). It has also been used to study the evolution of chemical organisations in an AChem model of prebiotic evolution (Matsumaru et al., 2006b) since self-maintenance may be a prerequisite for minimal biological systems (Fontana and Buss, 1994). COT has been shown to overlap with theories relating to autocatalytic sets of species. For example, a reflexively autocatalytic set that is generated by a food set and cannot generate species outside of the set is also an organisation (Hordijk et al., 2018). Autocatalytic sets that are organisations also contain catalytic loops or cycles (Contreras et al., 2011).

We will formally define what a chemical organisation is after some definitions are introduced for reaction networks generally. A reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ is defined as a set of reactions $\mathcal{R}$ that occur among molecular species $\mathcal{M}$. Each reaction within $\mathcal{R}$ is an ordered pair consisting of a set of *reactants* and a set of *products*. Therefore $\mathcal{R} \subseteq \mathcal{P}_M(\mathcal{M}) \times \mathcal{P}_M(\mathcal{M})$, where $\mathcal{P}_M(\mathcal{M})$ are all multisets that can be created from elements in $\mathcal{M}$ and $X \times Y$ consists of all ordered pairs (x, y) that exist with $x \in X$ and $y \in Y$. Reactions are denoted $(A \to B) \in \mathcal{R}$ hereafter, where $A \in \mathcal{P}_M(\mathcal{M})$ is the multiset of reactant species and $B \in \mathcal{P}_M(\mathcal{M})$ is the multiset of product species. They are multisets because there can be multiple instances of the same element in a multiset, just like there can be multiple instances of the same species consumed or produced in a reaction. We also define a $|\mathcal{M}| \times |\mathcal{R}|$ stoichiometric matrix $\mathcal{S} = (s_{i,j})$ where $|s_{i,j}|$ is the number of instances of species $i \in \mathcal{M}$ that are produced or consumed in reaction $j \in \mathcal{R}$. If $s_{i,j} > 0$, then species i is produced in the reaction, but if $s_{i,j} < 0$, it is consumed (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008). Given a subset of species $\mathcal{M}' \subseteq \mathcal{M}$, we say the reaction network *implied* by $\mathcal{M}'$ is $R(\mathcal{M}') = \{(A \to B) \in \mathcal{R} \mid A \in \mathcal{P}_M(\mathcal{M}')\}$.

A chemical organisation is a set of species $\mathcal{O} \subseteq \mathcal{M}$ that is *closed* and *self-maintaining* as defined below. A set of species is considered closed if there are no reactions using species within the set that can react to form any species outside of the set (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008).

Definition 1. *$\mathcal{O} \subseteq \mathcal{M}$ is **closed** if for all reactions $(A \to B) \in R(\mathcal{O})$ then $B \in \mathcal{P}_M(\mathcal{O})$.*

A set of species is self-maintaining because all species within the set can sufficiently be produced such that none of them decays (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008).

Definition 2. *A set of species $\mathcal{O} \subseteq \mathcal{M}$ is **self-maintaining** if there exists a vector of reaction rates (called a flux vector) $r = (r_1, \dots, r_n) \in \mathbb{R}^n_{\geq 0}$ such that the following conditions hold: (1) For every reaction $j \in R(\mathcal{O})$, its corresponding flux is $r_j > 0$. (2) For every reaction $j \in \mathcal{R} \setminus R(\mathcal{O})$, its corresponding flux is $r_j = 0$. (3) For every species $i \in \mathcal{O}$, the concentration change $(\mathcal{S}r)_i$ is nonnegative $(\mathcal{S}r)_i \geq 0$.*

We define $\Pi(\mathcal{R}')$ with $\mathcal{R}' \subseteq \mathcal{R}$ as the set containing every flux vector $r \in \Pi(\mathcal{R}')$ that exists with a positive flux $r_j > 0$ for every reaction $j \in \mathcal{R}'$ and zero flux $r_j = 0$ for every reaction $j \in \mathcal{R} \setminus \mathcal{R}'$. Note that one organisation can be a subset of another, so organisations can have a hierarchical structure (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008). Such a hierarchical structure is considered a requirement for evolvability in autocatalytic sets and organisations, since it shows that they can combine and grow over time (Hordijk et al., 2012; Speroni di Fenizio, 2015; Hordijk et al., 2018). The following additional definitions are useful when discussing algorithms designed to find organisations within a reaction network (Dittrich and Speroni di Fenizio, 2007; Centler et al., 2008).

Definition 3. *A species $i \in \mathcal{O}$ with $\mathcal{O} \subseteq \mathcal{M}$ is **consumed within** $R(\mathcal{O})$ if there exists a reaction $j \in R(\mathcal{O})$ with $s_{i,j} < 0$ and is **produced within** $R(\mathcal{O})$ if there exists a reaction $j \in R(\mathcal{O})$ with $s_{i,j} > 0$.*

Definition 4. *A set of species $\mathcal{O} \subseteq \mathcal{M}$ is called **semi-self-maintaining** if all species $s \in \mathcal{O}$ that are consumed within implied network $R(\mathcal{O})$ are also produced within it.*

Note that a self-maintaining set of species is also semi-self-maintaining. A set of species that is closed and semi-self-maintaining is a *semi-organisation*.

Definition 5. *A set of species $\mathcal{O} \subseteq \mathcal{M}$ is **reactive connected** if every species is a product or reactant of at least one reaction, and if for any two species $o_i \in \mathcal{O}$ and $o_j \in \mathcal{O}$ there exists a sequence of n species $(o_1, \dots, o_n)$ with $o_k \in \mathcal{O}$ for $k = 1, \dots, n$ such that $o_i = o_1$, $o_j = o_n$ and o_k is directly connected to o_{k+1} for $k = 1, \dots, n-1$. Two species $o_l \in \mathcal{O}$ and $o_m \in \mathcal{O}$ are directly connected if there exists a reaction $(A \to B) \in R(\mathcal{O})$ such that $o_l \in A \cup B$ and $o_m \in A \cup B$.*

Several algorithms have been proposed for finding organisations within any given reaction network (Centler et al., 2008; Milreu et al., 2010; Speroni di Fenizio, 2015). This static approach to finding organisations is complemented by a dynamic approach which can relate a reaction vessel's state at any point in time to an organisation (Dittrich and Speroni di Fenizio, 2007; Matsumaru et al., 2006b). Although it is possible to precisely determine the evolution of organisations through time, predictions about the evolution of organisations can still be made by attaining the complete set of organisations within a reaction network and determining their hierarchical structure (Speroni di Fenizio, 2015).

Organisations in Matter-conserving Chemistries

Milreu et al. (2010) define a reaction network as being *mass-consistent* if there exists a mass vector $\mathbf{m}$ containing a mass $m_i > 0$ for each species i in the network such that $\mathbf{m}\mathcal{S} = 0$. In such a reaction network, there can be no *inflow* or *outflow* reactions. An inflow reaction $(\emptyset \to B)$ produces products B without consuming any species, and an outflow reaction $(A \to \emptyset)$ consumes reactants A without producing any species (Dittrich and Speroni di Fenizio, 2007). Here, the term *matter-conserving* is used to describe a mass-consistent reaction network or chemistry that generates such a network.

In an organisation within a matter-conserving reaction network (MCRN), reactions cannot exist that will eventually deplete any species. A reaction network or subnetwork can be transformed into a bipartite graph to demonstrate that the network can only be an organisation if every edge within the directed graph is a part of a cycle. Let $\langle \mathcal{M}, \mathcal{R} \rangle$ be a MCRN and let $\mathcal{O} \subseteq \mathcal{M}$ be a closed set of species with implied reaction network $R(\mathcal{O}) \subseteq \mathcal{R}$.

Definition 6. *The* ***bipartite representation*** $G(\mathcal{O}) = (V, E)$ *of* $\mathcal{O}$ *is a directed graph with set of nodes* V *and set of edges* E *that are constructed using the following steps:*

1. *Add a node* v_i *to* V *for each species* $i \in \mathcal{O}$.

2. *Add a node* v_j *to* V *for each reaction* $j \in R(\mathcal{O})$.

3. *Add a directed edge to* E *starting at* v_i *and ending at* v_j *for each combination of species and reaction* $(i, j) \in \mathcal{O} \times R(\mathcal{O})$ *that exists with stoichiometric coefficient* $s_{i,j} < 0$.

4. *Add a directed edge to* E *starting at* v_j *and ending at* v_i *for each combination of species and reaction* $(i, j) \in \mathcal{O} \times R(\mathcal{O})$ *that exists with stoichiometric coefficient* $s_{i,j} > 0$.

Let $G(\mathcal{O}) = (V, E)$ be the bipartite representation of $\mathcal{O}$. Note that all $(v_1, v_2) \in E$ either start from a reaction node $v_1 = v_{A\to B}$ and end at a species node $v_2 = v_i$, or start at a species node $v_1 = v_i$ and end at a reaction node $v_2 = v_{A\to B}$. Hereafter a directed edge is notated as an ordered pair of nodes (x, y) with the first element of the pair x being the start of the edge and the second element y being the end of the edge. To demonstrate that every edge $(v_1, v_2) \in E$ must be a part of a cycle for $\mathcal{O}$ to be an organisation, we must first provide several definitions.

Definition 7. *A sequence of nodes* $(v_1, \ldots, v_n)$ *with* $v_j \in V$ *for* $j = 1, \ldots, n$ *is a* ***path*** *if all nodes in the sequence are distinct and there exists an edge* $(v_j, v_{j+1}) \in E$ *for* $j = 1, \ldots, n-1$. *Note that each node* $v_j \in V$ *could be a reaction node* $v_{A\to B}$ *or a species node* v_i.

Definition 8. *A sequence of nodes* $(v_1, \ldots, v_n)$ *with* $v_j \in V$ *for* $j = 1, \ldots, n$ *is a* ***cycle*** *if nodes* $v_1, \ldots, v_{n-1}$ *are distinct,* $v_1 = v_n$ *and there exists an edge* $(v_j, v_{j+1}) \in E$ *for* $j = 1, \ldots, n-1$. *Note that all cycles will have a minimum length of three because a reaction node or species node cannot be connected to itself.*

Definition 9. *A reaction* $(A \to B) \in R(\mathcal{O})$ ***contains*** *an edge* $(v_1, v_2) \in E$ *if the edge ends or starts at reaction node* $v_{A\to B}$. *Formally,* $(A \to B)$ *contains* (v_1, v_2) *if* $v_1 = v_{A\to B}$ *and therefore* $v_2 = v_i$ *with* $i \in B$ *or if* $v_2 = v_{A\to B}$ *and therefore* $v_1 = v_i$ *with* $i \in A$.

Definition 10. *A sequence* $(x_1, \ldots, x_n)$ ***contains*** *an ordered pair* (y, z) *if there exists an element of the sequence* x_j *such that* $y = x_j$ *and* $z = x_{j+1}$.

Definition 11. *A node* $y \in V$ *is* ***reachable*** *from node* $x \in V$ *if* $y = x$ *or there exists a path* $(v_1, \ldots, v_n)$ *with* $v_1 = x$ *and* $v_2 = y$.

Definition 12. *A node* $y \in V$ *is* ***conversely reachable*** *from node* $x \in V$ *if* $y = x$ *or there exists a path* $(v_1, \ldots, v_n)$ *with* $v_1 = y$ *and* $v_2 = x$.

Let $V^{\leftarrow}(v)$ be the set of all nodes that are reachable from $v \in V$ and let $V^{\rightarrow}(v)$ be the set of all nodes that are conversely reachable from $v \in V$. To demonstrate that every edge $(v_1, v_2) \in E$ must be a part of a cycle, we must identify why the presence of an edge that is not part of a cycle prevents $\mathcal{O}$ from being an organisation. A formal definition of such an edge is provided.

Definition 13. *An edge* $(v_1, v_2) \in E$ *is an* ***acyclic edge*** *if there is no cycle within* $G(\mathcal{O})$ *that contains* (v_1, v_2).

Lemma 1. *Given any acyclic edge* $(v_1, v_2) \in E$, *the set of nodes conversely reachable from* v_1 *do not intersect with the set of nodes reachable from* v_2, *otherwise a path would exist from* v_2 *to* v_1 *which would mean that the edge is part of a cycle. Formally,* $V^{\rightarrow}(v_1) \cap V^{\leftarrow}(v_2) = \emptyset$.

We can partition the nodes of a bipartite representation $G(\mathcal{O}) = (V, E)$ containing an acyclic edge into two sets of nodes (V_s, V_t) such that there is no edge in E starting from a node in V_t that ends at a node in V_s. This will be used to demonstrate that the set of species represented by nodes within V_s cannot collectively maintain themselves.

Definition 14. *An* ***acyclic cut*** *partitions* V *into two subsets* (V_s, V_t) ***based on*** *an acyclic edge* $(e_1, e_2) \in E$ *such that*

V_t is all nodes reachable from node e_2 and V_s is all other existing nodes. Formally, $V_t = V^{\leftarrow}(e_2)$ and $V_s = V \setminus V_t$.

Let (V_s, V_t) be an acyclic cut based on acyclic edge $(e_1, e_2) \in E$ and let E_c be the set of all edges $(v_1, v_2) \in E$ with $v_1 \in V_s$ and $v_2 \in V_t$. Note that $(e_1, e_2) \in E_c$ because $e_1 \in V_s$ and $e_2 \in V_t$.

Lemma 2. *All edges in E_c are also acyclic edges.*

Proof. All edges $(v_1, v_2) \in E_c$ are acyclic edges because $v_1 \in V_s$ and $v_2 \in V_t$. If (v_1, v_2) is part of a cycle then v_1 would be reachable from $v_2 \in V_t$ and therefore $v_1 \in V_t$ because V_t contains all nodes reachable from v_2. □

Lemma 3. *There is no edge in E starting from a node in V_t that ends at a node in V_s.*

Proof. There cannot exist an edge $(v_1, v_2) \in E$ with $v_1 \in V_t$ and $v_2 \in V_s$ because if $v_1 \in V_t$ then $v_2 \in V_t$ since v_2 is reachable from v_1 which is reachable from e_2 and $V_t = V^{\leftarrow}(e_2)$. □

Lemma 4. *A reaction $j \in R(\mathcal{O})$ cannot contain both an edge in E_c that starts at v_j and an edge in E_c that ends at v_j.*

Proof. For all edges $(v_1, v_2) \in E_c$, $v_1 \in V_s$ and $v_2 \in V_t$. So a reaction $j \in R(\mathcal{O})$ cannot contain an edge $(v_1, v_2) \in E_c$ with $v_1 = v_j$ and an edge $(w_1, w_2) \in E_c$ with $w_2 = v_j$ because this would mean $v_j \in V_t$ and $v_j \in V_s$ which is impossible because $V_t \cap V_s = \emptyset$. □

Let $V(X)$ be the set of species nodes for all species $i \in X$ with $X \subseteq \mathcal{O}$ and let $\mathcal{O}(Y)$ be the set of species for all nodes $v_i \in Y$ with $Y \subseteq V$ and $i \in \mathcal{O}$. Formally, $V(X) = \{v_i \mid i \in X\}$ and $\mathcal{O}(Y) = \{i \in \mathcal{O} \mid v_i \in Y\}$. Let $m(Y, j)$ be the net change in mass for reaction $j \in R(\mathcal{O})$ and all species $i \in \mathcal{O}(Y)$ with $Y \subseteq V$ and let $m(Y, r)$ be the net change in mass for flux vector r and all species $i \in \mathcal{O}(Y)$ with $Y \subseteq V$. Formally, $m(Y, j) = \sum_{i \in \mathcal{O}(Y)} s_{i,j} m_i$ and $m(Y, r) = \sum_{i \in \mathcal{O}(Y)} (\mathcal{S}r)_i m_i$. Also, let $U(X)$ be the set of distinct elements within multiset X. We can now show that there is a net loss of matter for all species $i \in \mathcal{O}(V_s)$ for all reactions involving these species.

Theorem 5. *For any reaction $j \in R(\mathcal{O})$ containing an edge in E_c, $m(V_s, j) < 0$ and $m(V_t, j) > 0$.*

Proof. According to lemma 4, a reaction $j \in R(\mathcal{O})$ can contain edges in E_c that end at v_j if there are no edges in E_c that start at v_j. Conversely, j can contain edges in E_c that start at v_j if there are no edges in E_c that end at v_j. So the theorem must be proven in both scenarios. There are no other scenarios in which j can contain an edge in E_c.

Consider the first scenario in which a reaction $(A \to B) \in R(\mathcal{O})$ contains at least one edge $(v_i, v_{A \to B}) \in E_c$ with $i \in A$. For every species $i \in A$ with $(v_i, v_{A \to B}) \in E_c$, node $v_i \in V_s$ because edges in E_c are acyclic (see lemma 2) and $s_{i,A \to B} < 0$ (see definition 6). For every species $i \in B$, $s_{i,A \to B} > 0$ (see definition 6) and node $v_i \in V_t$ because v_i is reachable from node $v_{A \to B} \in V_t$. Hence, $m(V_s, A \to B) < 0$ and therefore $m(V_t, A \to B) > 0$ because all species produced by the reaction are in $\mathcal{O}(V_t)$, at least one species consumed by the reaction is in $\mathcal{O}(V_s)$ and in a MCRN $\mathbf{m}\mathcal{S} = 0$.

Consider the second scenario in which a reaction $(A \to B) \in R(\mathcal{O})$ contains at least one edge $(v_{A \to B}, v_i) \in E_c$ with $i \in B$. For every species $i \in B$ with $(v_{A \to B}, v_i) \in E_c$, node $v_i \in V_t$ because edges in E_c are acyclic (see lemma 2) and $s_{i,A \to B} > 0$ (see definition 6). For every species $i \in A$, $s_{i,A \to B} < 0$ (see definition 6) and node $v_i \in V_s$ because v_i is conversely reachable from node $v_{A \to B} \in V_s$. Hence, $m(V_t, A \to B) > 0$ and therefore $m(V_s, A \to B) < 0$ because all species consumed by the reaction are in $\mathcal{O}(V_s)$, at least one species produced by the reaction is in $\mathcal{O}(V_t)$ and in a MCRN $\mathbf{m}\mathcal{S} = 0$. □

Theorem 6. *In $R(\mathcal{O})$ there is a net loss of matter for all species $i \in \mathcal{O}(V_s)$. Formally, $m(V_s, r) < 0$ for any flux vector $r \in \Pi(R(\mathcal{O}))$.*

Proof. We have already proven in theorem 5 that $m(V_s, A \to B) < 0$ and $m(V_t, A \to B) > 0$ for any $(A \to B) \in R(\mathcal{O})$ containing an edge in E_c. This means that $U(A) \cap \mathcal{O}(V_s) \neq \emptyset$ because there must be at least one species $i \in \mathcal{O}(V_s)$ with $s_{i,A \to B} < 0$ which makes i a reactant and $U(B) \cap \mathcal{O}(V_t) \neq \emptyset$ because there must be at least one species $i \in \mathcal{O}(V_t)$ with $s_{i,A \to B} > 0$ which makes i a product. Hence, $\mathcal{O}(V_s) \neq \emptyset$ and $\mathcal{O}(V_t) \neq \emptyset$ because E_c contains at least (e_1, e_2) which acyclic cut (V_s, V_t) is based on, and there exists at least one reaction containing an edge in E_c since all edges start or end at a reaction node.

If a reaction $j \in R(\mathcal{O})$ does not contain an edge in E_c then j must either exclusively contain edges $(v_1, v_2) \in E$ with $\{v_1, v_2\} \subset V_s$ or exclusively contain edges $(v_1, v_2) \in E$ with $\{v_1, v_2\} \subset V_t$ since there cannot exist an edge $(v_1, v_2) \in E$ with $v_1 \in V_t$ and $v_2 \in V_s$ (see lemma 3). The equality $\sum_{i \in \mathcal{O}(V_s)} s_{i,j} r_j m_i = 0$ holds for any flux $r_j > 0$ and for all reactions $j \in R(\mathcal{O})$ that exclusively contain edges $(v_1, v_2) \in E$ with $\{v_1, v_2\} \subset V_s$. This equality also holds for any flux $r_j > 0$ and for all reactions $j \in R(\mathcal{O})$ that exclusively contain edges $(v_1, v_2) \in E$ with $\{v_1, v_2\} \subset V_t$ because $\mathcal{O}(V_t) \cap \mathcal{O}(V_s) = \emptyset$. The inequality $\sum_{i \in \mathcal{O}(V_s)} s_{i,j} r_j m_i < 0$ with $r_j > 0$ holds for all reactions $j \in R(\mathcal{O})$ that contain at least one edge in E_c (of which there is at least one). Hence, $m(V_s, r) < 0$ for any flux vector $r \in \Pi(R(\mathcal{O}))$. □

Lastly, we prove that an organisation cannot contain an acyclic edge.

Theorem 7. *An organisation within a MCRN cannot contain an edge within its bipartite representation that is not part of a cycle.*

Proof. If a vector is taken from $\mathcal{S}$ for a species $i \in \mathcal{O}$ and multiplied by some positive scalar $x > 0$, then the sign of the concentration change will not be altered. Formally, $\text{sgn}(x(\mathcal{S}r)_i) = \text{sgn}((\mathcal{S}r)_i)$ where $\text{sgn}(y) = 1$ if $y > 0$, $\text{sgn}(y) = 0$ if $y = 0$ and $\text{sgn}(y) = -1$ if $y < 0$. Hence, if $(\mathcal{S}r)_i \geq 0$ then $x(\mathcal{S}r)_i \geq 0$ and if $(\mathcal{S}r)_i < 0$ then $x(\mathcal{S}r)_i < 0$. If $\mathcal{O}$ is a closed set of species with a bipartite representation $G(\mathcal{O}) = (V, E)$ containing an acyclic edge $(e_1, e_2) \in E$, then there exists an acyclic cut (V_s, V_t) based on (e_1, e_2). According to theorem 6, there is a net loss of matter $m(V_s, r) < 0$ for any flux vector $r \in \Pi(R(\mathcal{O}))$. If $m(V_s, r) < 0$ for any such flux vector then there always exists at least one species $i \in \mathcal{O}$ with $(\mathcal{S}r)_i m_i < 0$ and therefore $(\mathcal{S}r)_i < 0$, because $\text{sgn}((\mathcal{S}r)_i m_i) = \text{sgn}((\mathcal{S}r)_i)$. The existence of a negative concentration change $(\mathcal{S}r)_i < 0$ for a species i within closed set $\mathcal{O}$ means that $\mathcal{O}$ cannot be self-maintaining. Hence, any closed set $\mathcal{O}$ cannot be an organisation if its bipartite representation contains an edge that is not part of a cycle. □

A decomposition theorem of organisations (Veloz et al., 2011; Veloz and Razeto-Barry, 2017) provides a way of partitioning a reaction network or subnetwork into modules to determine if it is an organisation. Matter-conservation has implications for this theorem. The *potential fragile-circuit* of a set of species $\mathcal{O}$ is defined as all species within $\mathcal{O}$ apart from *non-reactive*, *overproduced* or *catalyst* species (Veloz et al., 2011; Veloz and Razeto-Barry, 2017). A species is non-reactive if it is neither a reactant nor a product of any reaction in $R(\mathcal{O})$. A species $i \in \mathcal{O}$ is overproduced by flux vector $r \in \Pi(R(\mathcal{O}))$ if $(\mathcal{S}r)_i > 0$. A species $i \in \mathcal{O}$ is a catalyst if $s_{i,j} = 0$ for every reaction $j \in R(\mathcal{O})$ and $i \in A$ for at least one reaction $(A \to B) \in R(\mathcal{O})$. The potential fragile-circuit of $\mathcal{O}$ is $\mathcal{O} - (N \cup C \cup F)$ where $N \subseteq \mathcal{O}$, $C \subseteq \mathcal{O}$ and $F \subseteq \mathcal{O}$ is the set of non-reactive, catalyst and overproduced species, respectively. An organisation within a MCRN has a net production of zero as it cannot contain an overproduced species.

Lemma 8. *If $\mathcal{O}$ is an organisation within a MCRN then $(\mathcal{S}r)_i = 0$ for every species $i \in \mathcal{O}$ and any flux vector $r \in \Pi(R(\mathcal{O}))$.*

Proof. The equality $\sum_{i \in \mathcal{O}} (\mathcal{S}r)_i m_i = 0$ holds for any flux $r \in \Pi(R(\mathcal{O}))$ because a MCRN is mass-consistent. This means that if $(\mathcal{S}r)_i m_i > 0$ for some species $i \in \mathcal{O}$, then there must be at least one other species $k \in \mathcal{O}$ with $(\mathcal{S}r)_k m_k < 0$. Since $sgn((\mathcal{S}r)_l m_l) = sgn((\mathcal{S}r)_l)$ for any species $l \in \mathcal{O}$, if $(\mathcal{S}r)_i > 0$ for some species $i \in \mathcal{O}$ there must be at least one other species $k \in \mathcal{O}$ with $(\mathcal{S}r)_k < 0$ which means $\mathcal{O}$ cannot be an organisation because it is not self-maintaining. Hence, if $\mathcal{O}$ is an organisation then $(\mathcal{S}r)_i = 0$ for every species $i \in \mathcal{O}$. □

One consequence of lemma 8 is that aside from non-reactive species and catalysts, all other species in $\mathcal{O}$ constitute the potential fragile-circuit of $\mathcal{O}$. If $\mathcal{O}$ is an organisation, every species within the potential fragile-circuit must be both consumed and produced within $R(\mathcal{O})$ (see lemma 4 in Veloz et al., 2011). Note that in a MCRN, an acyclic edge can exist within $\mathcal{O}$'s bipartite representation which prevents $\mathcal{O}$ from being self-maintaining even if every species within $\mathcal{O}$'s potential fragile-circuit is both consumed and produced within $R(\mathcal{O})$. For example, in the following network

$$\mathcal{O} = \{a, b, c, d\} \tag{1}$$

$$R(\mathcal{O}) = \{a \to b, b \to a, b \to c, c \to d, d \to c\} \tag{2}$$

every species within $\mathcal{O}$ is both consumed and produced within $R(\mathcal{O})$, but reaction $b \to c$ induces an acyclic edge. A potential fragile-circuit can be decomposed into many independent smaller fragile-circuits which are linked by catalysts or overproduced species (Veloz et al., 2011; Veloz and Razeto-Barry, 2017). Another consequence of lemma 8 is that in a MCRN, these smaller circuits can only be linked by catalysts if the larger potential fragile-circuit is within an organisation.

Algorithm Optimisation

We propose an altered version of an algorithm given in Centler et al. (2008) for computing reactive connected organisations of a given reaction network. The original algorithm is known as the *constructive approach*, and uses the reaction network's structure to find chemical organisations more efficiently than simply testing all 2^n combinations of n species. We propose a version of the algorithm for MCRNs that more efficiently computes the organisations by taking advantage of the fact that an organisation within a MCRN cannot contain an acyclic edge in its bipartite representation. The algorithm consists of four main functions which are shown as pseudo-code in Figs. 1-4. The algorithm first finds all semi-organisations in the reaction network and then determines which ones are organisations.

We will first redefine the concepts of semi-self-maintenance and semi-organisation for MCRNs. The bipartite representation of $\mathcal{O} \subseteq \mathcal{M}$ contains an edge that is not part of a cycle if $\mathcal{O}$ contains a species that is only consumed or only produced within $R(\mathcal{O})$ (see definitions 3 and 6). According to theorem 7, $\mathcal{O}$ cannot be an organisation if its bipartite representation contains an edge that is not part of a cycle. Hence, semi-self-maintenance can be redefined as follows:

Definition 15. *Within a MCRN $\langle \mathcal{M}, \mathcal{R} \rangle$, a set of species $\mathcal{O} \subseteq \mathcal{M}$ is called* ***semi-self-maintaining*** *if all species $s \in \mathcal{O}$ that are consumed within implied network $R(\mathcal{O})$ are also produced within it and if all species $s \in \mathcal{O}$ that are produced within implied network $R(\mathcal{O})$ are also consumed within it.*

A *semi-organisation* in a MCRN is closed and semi-self-maintaining according to definition 15. Since this definition of semi-self-maintenance is more restrictive, the al-

```
Function computeOrgs:
    Input: MCRN $\langle \mathcal{M}, \mathcal{R} \rangle$
    Output: set of all organisations in result
    result ← $\emptyset$;
    SOsToCheck ← {closure({})};
    while SOsToCheck $\neq \emptyset$ do
        current ← getSmallestSO(SOsToCheck);
        SOsToCheck ← SOsToCheck ∪
         CSOsDirectlyAbove(current);
        SOsToCheck ← SOsToCheck \ {current};
        result ← result ∪ {current};
    end
End Function
```

Figure 1: Returns all organisations within a reaction network

```
Function CSOsDirectlyAbove:
    Input: semi-organisation so, MCRN $\langle \mathcal{M}, \mathcal{R} \rangle$
    Output: set of all semi-organisations directly above
             so in result
    result ← $\emptyset$;
    usable ← $\emptyset$;
    if so = {} then
        usable ← $\cup_{s \in \mathcal{M}} \{\{s\}\}$;
    else
        foreach u ∈ $\mathcal{R}$ with reactants(u) $\not\subseteq$ so do
            if $\exists\, s \in$ so with $s \in$ reactants(u) ∪
             products(u) then
                usable ← usable ∪
                  {reactants(u)\so};
            end
        end
    end
    foreach set ∈ usable do
        result ← result ∪
         SOsDirectlyAboveContaining(so,
         set);
    end
End Function
```

Figure 2: Returns all semi-organisations directly above the input one

gorithm can ignore sets of species that only satisfy the original definition of semi-self-maintenance. Note that semi-self-maintenance according to definition 15 does not imply self-maintenance. For example, each species is both produced and consumed within the reaction network shown in equations 1 and 2, so the network is semi-self-maintaining, but reaction $b \to c$ induces an acyclic edge, so the network cannot be self-maintaining according to theorem 7.

Function computeOrgs (Fig. 1) represents the main loop of the algorithm. It is identical to the computeSemiOrganisations function (Centler et al., 2008) except it operates on a MCRN. The function finds all semi-organisations within the reaction network. The set SOsToCheck is initialised with the smallest semi-organisation which is the closure of the empty set (closure({})). In a MCRN there are no inflow reactions so closure({}) is the empty set. closure(*set*) computes the smallest closed set containing the set of input species set. This is done by adding all species to set that can be produced from all reactions $(A \to B) \in \mathcal{R}$ with $A \subseteq$ set and then repeating this process until no novel species can be added. The function shown in Fig. 3 also makes use of the closure function.

The SOsDirectlyAbove function (Centler et al., 2008) is called by the computeSemiOrganisations function and returns at least all semi-organisations *directly above* input semi-organisation so. A semi-organisation is directly above so if it contains so and does not contain any other semi-organisations that contain so. However, a combinatorial explosion can lead to a very large number of organisations. For example, a reaction network containing n species but no reactions contains 2^n organisations because every combination of species is closed and self-maintaining. Hence, it can be computationally costly to compute all organisations. Also, enumerating sets of non-interacting species is unlikely to be of interest to researchers investigating the chemical origins of life. The ConnectedSOsDirectlyAbove function (Centler et al., 2008) is the same as the original SOsDirectlyAbove function, except it only returns reactive connected semi-organisations, thereby ensuring the algorithm only returns organisations containing interacting species. The CSOsDirectlyAbove function (Fig. 2) is identical to the original ConnectedSOsDirectlyAbove function (Centler et al., 2008) except it operates on a MCRN and is called by computeOrgs. We choose to use CSOsDirectlyAbove instead of SOsDirectlyAbove because of its practical advantages. However, it could be swapped with a function identical to SOsDirectlyAbove that operates on a MCRN.

Function computeOrgs takes the smallest semi-organisation current from SOsToCheck, adds the semi-organisations returned by CSOsDirectlyAbove(*current*) to SOsToCheck, removes current from SOsToCheck and then repeats this entire process until SOsToCheck is empty. In practice, SOsToCheck would be some kind of hash structure that ensures the same semi-organisation cannot be added to SOsToCheck more than once.

Function CSOsDirectlyAbove(*so*) iterates through sets of species that are directly connected to species in so and for each set finds all organisations directly above so that contain the set. To do this, CSOsDirectlyAbove calls SOsDirectlyAboveContaining(*so,species*)

```
Function SOsDirectlyAboveContaining:
    Input: semi-organisation so, species set species to
           be contained in new semi-orgs., MCRN
           $\langle \mathcal{M}, \mathcal{R} \rangle$
    Output: set of all semi-organisations directly above
            so that contain species in result
    result ← ∅;
    closure ← closure(so ∪ species);
    if closure is semi-self-maintaining then
        result ← {closure};
    else
        cSets ←
          producerAndConsumerSets(closure);
        foreach set ∈ cSets do
            result ← result ∪
              SOsDirectlyAboveContaining(so,
              species ∪ set);
        end
    end
End Function
```

Figure 3: Takes an input semi-organisation and returns all semi-organisations directly above it containing additionally specified species

to find a semi organisation containing both so and species set species. The SOsDirectlyAboveContaining function (Fig. 3) is almost identical to the original SOsDirectlyAboveContaining function (Centler et al., 2008) except it functions slightly differently because it operates on a MCRN. First, the closure of the union of so and species is computed and stored in closure which is then tested to see if it is semi-self-maintaining. If it is semi-self-maintaining then a semi-organisation has been found and the function returns.

If closure is not semi-self-maintaining then function producerAndConsumerSets (Fig. 4) is used to retrieve all possible species combinations that produce or consume the species in closure that are only consumed or only produced in closure's implied reaction network respectively. producerAndConsumerSets is similar to the original producerSets function (Centler et al., 2008). The difference is that producerSets does not necessarily operate on a MCRN and therefore only returns all possible species combinations that produce the species in closure that are only consumed in closure's implied reaction network. For each species combination returned by producerAndConsumerSets, SOsDirectlyAboveContaining(so,species) is called recursively but with the species combination added to species.

Each semi-organisation returned by computeOrgs that is self-maintaining is an organisation. Checking if each semi-organisation is an organization is a linear programming problem in which a flux vector r must be found that satisfies the conditions of self-maintenance (Centler et al., 2008). Finding an optimal flux vector is not necessary since meeting the condition of self-maintenance only requires $(\mathcal{S}r)_i$ to be non-negative (Centler et al., 2008). After all organisations have been identified, the algorithm completes.

```
Function producerAndConsumerSets:
    Input: semi-organisation so and matter-conserving
           reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$
    Output: set of all species combinations in result
            that can produce (consume) all species that
            are only consumed (produced) in so
    result ← ∅;
    foreach $m \in$ so where $m$ is consumed and not
      produced within $R$(so) do
        cSets$_m$ ← ∅;
        foreach $u \in \mathcal{R}$ with $s_{m,u} > 0$ do
            cSets$_m$ ← cSets$_m$ ∪ {reactants(u) \so};
        end
    end
    foreach $m \in$ so where $m$ is produced and not
      consumed within $R$(so) do
        cSets$_m$ ← ∅;
        foreach $u \in \mathcal{R}$ with $s_{m,u} < 0$ do
            cSets$_m$ ← cSets$_m$ ∪ {reactants(u) \so};
        end
    end
    repeat
        current ← ∅;
        foreach $m \in$ so for which ∃cSets$_m$ do
            select a set cSetsM from cSets$_m$;
            current ← current ∪ cSetsM;
        end
        result ← result ∪ {current};
    until all possible set combinations have been
      considered;
End Function
```

Figure 4: Returns all combinations of species facilitating reactions consuming (producing) all species that are only produced (consumed) by the input semi-organisation

Experimental Results

We conducted a short experiment to verify that the altered algorithm presented in this paper works and is faster than the original version of the algorithm. We randomly generate 100 Spiky-RBN (Krastev et al., 2016) chemistries, each with 12 instances of 12 unique atomic particle species. We will briefly summarise the Spiky-RBN AChem. Each atomic particle species is a Random Boolean Network (RBN) with 12 nodes and 2 inputs per node. A bonding site, or 'interaction list', is a grouping of RBN nodes that can interlink with the nodes of another interaction list. Each of these in-

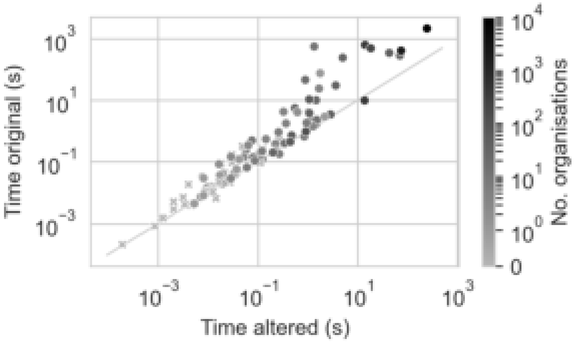


Figure 5: Execution time of both algorithms and number of organisations found per reaction network

teraction lists has a ‘spike’ which is determined by the dynamics of the RBN. We attempt 10,000 reactions for each chemistry in an aspatial, matter-conserving reactor (Krastev et al., 2017). A single reaction attempt consists of randomly picking two atomic particles, attempting to randomly choose an unlinked interaction list on both particles, and interlinking them if their spikes sum to zero. We infer that a new link cannot form between two particles that are already linked. Link formation changes the underlying RBN and can therefore change the spikes. So, directly after link formation every linked pair of interaction lists in the new composite particle is checked to see if the spikes still sum to zero. Any link that does not meet this criterion is removed by reversing the interlinking process. A reaction network is generated for all reactions that take place. A reaction is ignored if the multiset of products is equal to the multiset of reactants. Two particles or composite particles are considered the same species if their underlying RBNs are identical. We find the complete set of reactive connected organisations within each chemistry’s reaction network using Centler et al.’s original algorithm for finding reactive connected organisations (Centler et al., 2008) and the altered version of this algorithm proposed in this paper. We will refer to Centler et al.’s algorithm as the *original* algorithm, and our altered version as the *altered* algorithm hereafter. For each reaction network, both algorithms found the exact same set of organisations.

Fig. 5 shows the execution time of the original algorithm (vertical axis) and the altered algorithm (horizontal axis) for every reaction network, each of which is represented by a single data point. A line shows where the execution times of the algorithms would be equal. The shade of each data point represents how many organisations containing at least two species were found (crosses indicate zero found). Logarithmic scales were used except for organisation counts between zero and one which were made linear (due to zero counts). The plot clearly shows that in the majority of cases the execution time of the altered algorithm is much faster than the execution time of the original algorithm. This is not true for 16 out of the 100 reaction networks. In these cases the cost of the altered algorithm’s iteration over combinations of consuming sets may outweigh the cost of the original algorithm’s iteration over semi-organisation candidates ignored by the altered algorithm. The altered algorithm was 8.92 times faster than the original on average. This average was obtained by dividing the execution time of the original algorithm by the execution time of the altered one for each reaction network and calculating the mean of these values. The number of organisations found appears to increase with execution time. This implies that there are more organisations to be found in reaction networks which are more difficult for the algorithm to process. The mean number of organisations is 143.38, although no organisations were found in 31 out of 100 reaction networks. The difference in execution time between algorithms appears to increase with the number of organisations found, which implies that the benefit of using the altered algorithm increases as the reaction network becomes more difficult to process.

Conclusions and Future Work

We have shown that it is possible to compute the complete set of organisations in a reaction network more efficiently if we know the network has been generated by a matter-conserving chemistry. Adding conservation-of-matter to an AChem has been shown to produce interesting behaviour (Hickinbotham and Stepney, 2015; Krastev et al., 2017) and the theorems and algorithm we provide could facilitate further research into these behaviours. For example, researchers may want to understand if adding conservation-of-matter to an AChem changes the diversity and evolvability of organisations. Such research could contribute knowledge about what factors may prohibit or permit the evolution of self-maintaining sets of molecular species on the way to the emergence of minimal artificial lifeforms. We are currently using the altered algorithm to compare the diversity and evolvability of organisations between different variants of the Spiky-RBN AChem (Krastev et al., 2016). The tools provided here can also be applied to reaction networks for natural systems and may therefore be useful in other fields of research. For example, COT has been used to study atmospheric photochemistry models of Mars (Centler and Dittrich, 2007) and models of sugar metabolism in E. coli (Centler et al., 2007). Models of natural reaction networks may exist which are matter-conserving and could therefore be analysed using our more efficient algorithm. In future research, theorem 7 could be further exploited to find the complete set of organisations even faster.

Acknowledgements

We thank the anonymous reviewers for helpful comments. Jonathan is funded by the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology.

References

Centler, F. and Dittrich, P. (2007). Chemical organizations in atmospheric photochemistries—a new method to analyze chemical reaction networks. *Planetary and Space Science*, 55(4):413–428.

Centler, F., Kaleta, C., Speroni di Fenizio, P., and Dittrich, P. (2008). Computing chemical organizations in biological networks. *Bioinformatics*, 24(14):1611–1618.

Centler, F., Speroni di Fenizio, P., Matsumaru, N., and Dittrich, P. (2007). Chemical organizations in the central sugar metabolism of escherichia coli. In *Mathematical Modeling of Biological Systems, Volume I*, pages 105–119. Springer.

Contreras, D., Pereira, U., Hernandez, V. C., Reynaert, B., and Letelier, J.-C. (2011). A loop conjecture for metabolic closure. In *ECAL*, pages 176–183.

Dittrich, P. and Speroni di Fenizio, P. (2007). Chemical organisation theory. *Bulletin of mathematical biology*, 69(4):1199–1231.

Dittrich, P. and Winter, L. (2008). Chemical organizations in a toy model of the political system. *Advances in Complex Systems*, 11(04):609–627.

Dittrich, P., Ziegler, J. C., and Banzhaf, W. (2001). Artificial chemistries—a review. *Artificial Life*, 7(3):225–275.

Fontana, W. and Buss, L. W. (1994). 'The arrival of the fittest': Toward a theory of biological organization. *Bulletin of Mathematical Biology*, 56(1):1–64.

Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., and Young, P. (2010). Specification of the stringmol chemical programming language version 0.2. Technical report, Technical Report YCS-2010-458, Dept Computer Science, University of York.

Hickinbotham, S. and Stepney, S. (2015). Conservation of matter increases evolutionary activity. In *Artificial Life Conference Proceedings 13*, pages 98–105. MIT Press.

Hordijk, W., Hein, J., and Steel, M. (2010). Autocatalytic sets and the origin of life. *Entropy*, 12(7):1733–1742.

Hordijk, W., Steel, M., and Dittrich, P. (2018). Autocatalytic sets and chemical organizations: modeling self-sustaining reaction networks at the origin of life. *New Journal of Physics*, 20(1):015011.

Hordijk, W., Steel, M., and Kauffman, S. (2012). The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta biotheoretica*, 60(4):379–392.

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *Journal of theoretical biology*, 119(1):1–24.

Krastev, M., Sebald, A., and Stepney, S. (2016). Emergent bonding properties in the spiky RBN achem. In *Proceedings of the Artificial Life Conference*, pages 600–607.

Krastev, M. S., Sebald, A. A.-M., and Stepney, S. (2017). Functional grouping analysis of varying reactor types in the spiky-RBN achem. In *European Conference on Artificial Life*, pages 247–254.

Langton, C. (1989). *Artificial Life: Proceedings Of An Interdisciplinary Workshop On The Synthesis And Simulation Of Living Systems*. Santa Fe Institute Series. Avalon Publishing.

Matsumaru, N., Centler, F., Speroni di Fenizio, P., and Dittrich, P. (2006a). Chemical organization theory applied to virus dynamics. *IT-Information Technology*, 48(3):154–160.

Matsumaru, N., Speroni di Fenizio, P., Centler, F., and Dittrich, P. (2006b). On the evolution of chemical organizations. In *Proc. of the 7th German Workshop of Artificial Life*, pages 135–146.

Milreu, P. V., Acuña, V., Birmelé, E., Crescenzi, P., Marchetti-Spaccamela, A., Sagot, M.-F., Stougie, L., and Lacroix, V. (2010). Enumerating chemical organisations in consistent metabolic networks: Complexity and algorithms. In *International Workshop on Algorithms in Bioinformatics*, pages 226–237. Springer.

Speroni di Fenizio, P. (2015). The lattice of chemical organisations. In *Artificial Life Conference Proceedings 13*, pages 242–248. MIT Press.

Veloz, T. (2020). The complexity–stability debate, chemical organization theory, and the identification of non-classical structures in ecology. *Foundations of Science*, 25(1):259–273.

Veloz, T. and Razeto-Barry, P. (2017). Reaction networks as a language for systemic modeling: On the study of structural changes. *Systems*, 5(2):30.

Veloz, T., Reynaert, B., Rojas-Camaggi, D., and Dittrich, P. (2011). A decomposition theorem in chemical organizations. In *ECAL*, pages 820–827.

Evolutionary stability of host-endosymbiont mutualism is reduced by multi-infection

Shakeal Hodge[1], Zhen Ren[1], Anya Vostinar[1] and Emily Dolson[2,3,4]

[1]Department of Computer Science, Carleton College, Northfield, MN, 55057
[2]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824
[3]Ecology, Evolution and Behavior, Michigan State University, East Lansing, MI, 48824
[4]BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI, 48824
dolsonem@msu.edu

Introduction

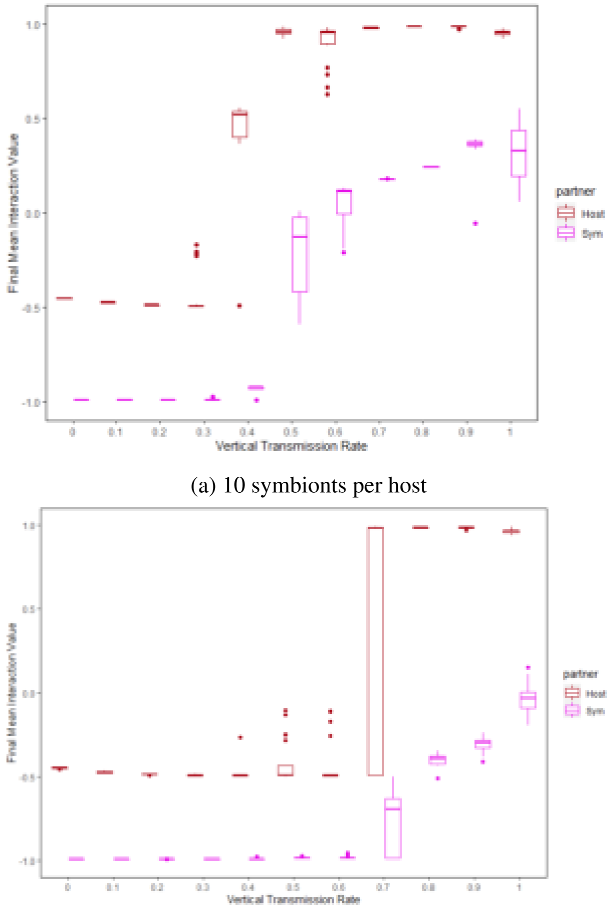
The human gut microbiome is a complex network of bacteria, phage, and other microorganisms, all interacting with each other in different ways. Many of these interactions occur between a host organism (such as a bacterium) and an endosymbiont (such as a phage) that lives inside it. These host-endosymbiont interactions range from mutualistic to parasitic. In a mutualistic interaction, the host and endosymbiont cooperate with each other and both benefit. In contrast, in a parasitic interaction the endosymbiont steals resources from the host and the host expends resources attempting to defend itself, so both suffer. Over the course of many generations, the behavior of hosts and endosymbionts can evolve.

A central evolutionary question in this context is whether hosts and endosymbionts will coevolve towards mutualism or antagonism. To date, evolutionary theory concerning this topic has focused on pairwise interactions (*i.e.* between a single host and endosymbiont). However, most hosts are home to vast numbers of endosymbionts, which co-evolve with each other just as they co-evolve with the host. The evolutionary consequences of this multi-level co-evolution are not well understood (but see (Nelson and May, 2017) for a preliminary mathematical modeling framework). Here we conduct initial inquiries into how existing host-symbiont co-evolution theory must change to accommodate the presence of multiple symbionts.

A factor that is known to be important in determining the course of pairwise host-symbiont co-evolution is vertical transmission rate: the probability of the endosymbiont's offspring ending up in the host's offspring as a result of the host's reproduction process. Prior research has shown that higher vertical transmission rates promote the evolutionary stability of mutualism (Vostinar and Ofria, 2019). However, it is unclear whether this relationship will persist when multiple endosymbionts are allowed to inhabit the same host.

Methods

We addressed this question using Symbulation, an agent based model of co-evolution between host organisms and symbionts, as a simple environment in which to observe the



(a) 10 symbionts per host

(b) 20 symbionts per host

Figure 1: **Evolved interaction values with 10 symbionts per host (top) and 20 symbionts per host (bottom).** Boxplots show distribution of evolved interaction values for hosts and symbionts across vertical transmission rates.

evolution of interactions between these organisms (Vostinar and Ofria, 2019). In Symbulation, the behavior of each individual is characterized by an "interaction value" between -1 (antagonistic) and 1 (mutualistic). This value determines the relationship between each organism and its partners, if it has any. At every time step, the host receives resources. Some of these resources may be passed on to its symbiont(s), based on the host's interaction value. Similarly, if the symbiont receives resources from the host, its interaction value determines whether it returns any resources back to its host. A highly antagonistic host will invest its resources into its own defense and not cooperate with the symbiont, while a highly cooperative host will donate some of its resources to its symbiont partner(s). When an individual accumulates enough resources, it reproduces. On reproduction, the offspring's interaction value is mutated by a value drawn from a normal distribution; thus, interactions evolve over time.

We conducted a series of experiments (n=20 per condition) in which we varied 1) the vertical transmission rate, and 2) the number of symbionts that were allowed to inhabit each host. In each experiment, we observed the final evolved interaction values for hosts and symbionts.

Results

Every increase in the number of symbionts per host led to a further increase in the level of vertical transmission required for mutualism to be an evolutionarily stable strategy. Figure 2 shows results for 10 symbionts per host and 20 symbionts per host, but the effect was consistent across the full range from 1 to 100 symbionts per host. After that point the effect largely saturated, with mutualism persisting at low levels only when the vertical transmission rate was 100%.

We hypothesize that this effect is driven by a dynamic in which parasites can be thought of as cheaters not only with respect to the host, but also with respect to the other symbionts. A mutualistic host is essentially a public good to its internal symbiont population. By parasitizing the host, a symbiont achieves short-term gain for itself at the long-term cost of dis-incentivizing host mutualism. If a host population is invaded by parasites, it will be incentivized to become antagonistic, even if this hurts any remaining mutualistic endosymbionts. These preliminary results highlight the potential for rich interactions among endosymbionts to play a critical role in the evolution of host-symbiont interactions.

References

Nelson, P. G. and May, G. (2017). Coevolution between Mutualists and Parasites in Symbiotic Communities May Lead to the Evolution of Lower Virulence. *The American Naturalist*, 190(6):803–817. Publisher: The University of Chicago Press.

Vostinar, A. E. and Ofria, C. (2019). Spatial Structure Can Decrease Symbiotic Cooperation. *Artificial Life*, 24(4):229–249.

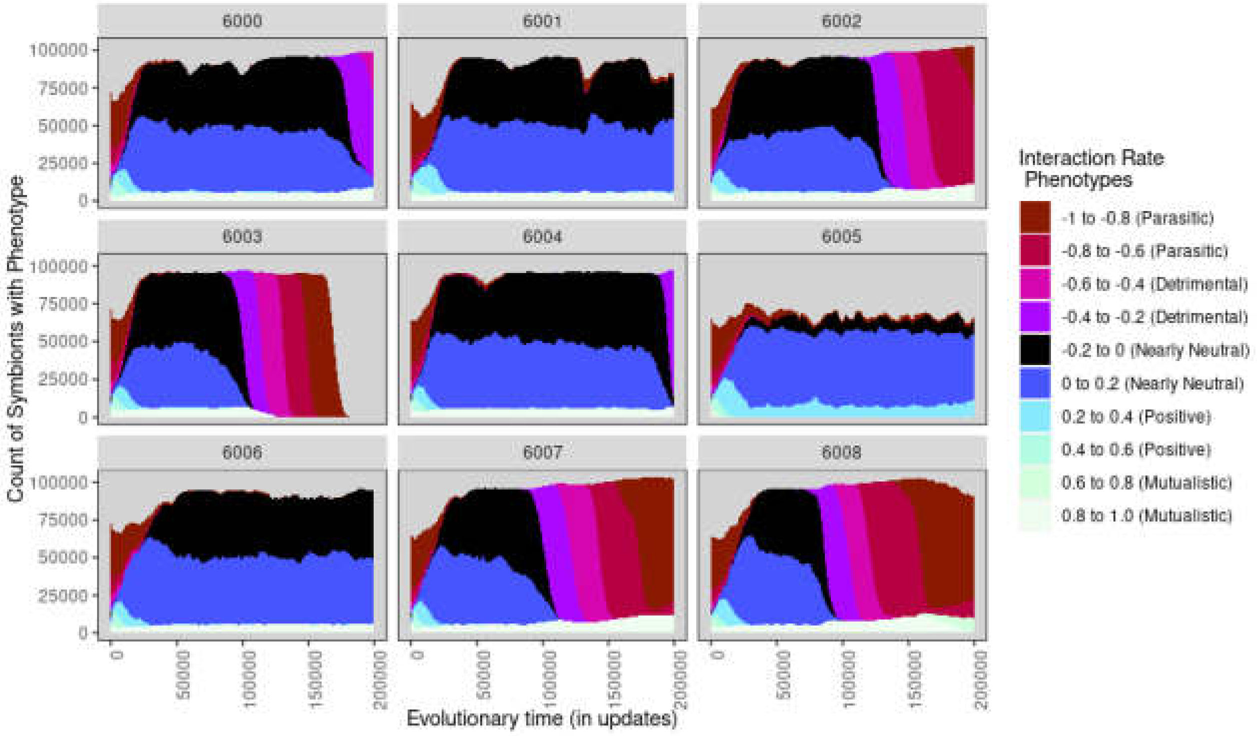


Figure 2: **Population dynamics of representative runs at vertical transmission rate of 60%.**

Analogical comparison of circuits generating a multiply realizable walking behavior

Kira Breithaupt[1] and Abe Leite[2]

[1]Indiana University, Bloomington, IN 47405 USA
[2]Stony Brook University, Stony Brook, NY 11794 USA
[1]kbreitha@iu.edu [2]abrahamjleite@gmail.com

Abstract

An understanding of analogy and the multiple realizability of concepts, ideas, and experience is necessary to understand cognition and the generation of behavior even at the most abstract levels. One of the most fundamental questions one can ask about a pair of neural circuits is whether they are doing the same thing or different things. Our work addresses this question by applying a model of sequential narrative analogy, Net-MATCH, to neural circuits evolved to perform a simple locomotion task. Along the way, we develop a measure of the "experience" of a neural circuit performing a behavior we call its functional trace. We find (i) that Net-MATCH reports strong analogies between some, but not all, neural circuits that perform the walking behavior, (ii) that it finds stronger analogies between circuits of the same class (as reported in previous work on this problem space) than circuits of different classes, and (iii) that it reveals strong analogies between circuits of the previously-reported BS-switch and SW-switch classes, even though these classes are of different circuit sizes. We conclude that Net-MATCH is a powerful tool for understanding the multiple realizability of behavior.

Introduction

Our work aims to address the question of how different individuals can qualitatively have the same experience, even if it is realized differently on both the conceptual level and the neural level. Among humans, by far the most common way in which we share an experience with others is through communication. If a crowd of people is told a story, no two people will make sense of it in quite the same way. The sound impulses hitting each person's ear will be different, the way these impulses are transduced into neural patterns will be different, and the neural pathways that translate these fleeting auditory patterns into words and then concepts are radically different. Still, in spite of all these differences, each person's experience will be linked with the others' in its deep and underlying structure.

While the multiple realization of ideas is commonly facilitated by communication, it can be understood even in individual, isolated organisms or agents, provided that those individuals are representing the same information or generating the same behavior as each other. In fact, the isolated case is an ideal proving ground for understanding multiple realizability and analogy on an abstract level: it is easily controllable, and model agents need not possess sophisticated communication machinery. Living organisms represent an extraordinary variety of stimuli, and generate an extraordinary variety of behavior; as such, one organism might realize the same experience or behavior as a vastly different organism. One important question in theoretical neuroscience is whether and how one can detect that two neural systems are having analogous experiences given only their patterns of activity.

Approach

In this work, we study how a single behavior can be generated in multiple ways. There are many ways that neural circuits can solve a simple walking task, but we would still consider each of these circuits to generate the same behavior. We draw on the idea of analogy to make sense of this phenomenon, proposing that when two systems realize the same concept, there ought to be a strong analogy between the two systems' realizations of the concept; when they realize different concepts, there ought to be weaker analogies. To operationalize this intuition, we use a computational model of analogy between sequential narratives to qualitatively and quantitatively investigate analogies between the activities of different neural circuits that solve the same problem. To obtain a sequential narrative from neural activity, we introduce a measure, functional traces, inspired by differential Hebbian learning. Our findings validate this model of analogy and reveal its tradeoffs with other ways of understanding the space of pattern generators for the legged walking task.

Our contributions in this work are threefold: (1) we propose an existing model of analogical comparison between sequential narratives as a framework to characterize ways that small neural circuits generate behavior; (2) we introduce a mechanism, the functional trace, to transform a pattern of neural activity into a sequential narrative; and (3) we make use of analogical comparison to reveal structural similarities between classes of 3- and 4-neuron circuits that generate the legged walking behavior.

Prior work

The single-legged walking behavior

In this work we use 3-, 4-, and 5-neuron CTRNNs that generate the single-legged walking behavior to model neural behavior generation more generally. The single-legged walking behavior (Beer and Gallagher, 1992) is an abstraction of earlier computational models of insect locomotion, representing the propulsion of a suspended body by a single leg. While it may seem to trivialize the problem, this simplification in fact makes it much more revealing as a substrate to investigate behavior generation more broadly.

While the walking behavior is discussed extensively in a number of well-known references, we will briefly describe the task and a typical controller architecture for it. A single-legged walker has a body attached to a frictionless one-dimensional suspended track. Attached to its body is a single straight leg. The leg's tip can be pressed to the ground (foot down state) or allowed to hang freely (foot up state); this implies that the leg's length is variable. The leg has opposing muscles that can either exert a torque to swing the leg backwards, so that the foot moves from a position in front of the body to a position behind the body, or to swing it forwards. When the foot is pressed to the ground, backwards or forwards torques to the leg will propel the organism's body forwards or backwards, respectively. For technical details of this system and its simulation, refer to Beer et al. (1999).

CTRNNs The legged walker is typically paired with a dynamical neural network such as a CTRNN (Beer and Gallagher, 1992). A CTRNN, or continuous-time recurrent neural network, is perhaps the simplest generalization of a neural network as a continuous dynamical system. It consists of a set of interconnected neurons, each of which has an unbounded real-valued internal state that is expressed through its sigmoid-transformed output (ranging from 0 to 1). Each neuron has a bias value as well as incoming connections from the other neurons linearly weighted by a weight vector; these together with any external input to the neuron determine the neuron's target state, which it approaches according to the neuron's unique time constant. Thus, an n-neuron CTRNN has $n^2 + 2n$ parameters: its weight matrix, bias vector, and time constant array. While CTRNNs are highly simple, easy to program, and tractable to analyze, they are also powerful and capable of serving as universal dynamics approximators (Funahashi and Nakamura, 1993).

When using a CTRNN as a central pattern generator for the single-legged walking behavior, one typically uses three motor neurons. One (**FT** or "foot down") controls whether the foot is pressed to the ground or not. When FT's output is above 0.5, the foot is considered "down" and leg movements will propel the body; when it is below 0.5, the foot is "up" and the body is immobile (losing all momentum, in fact). The others (**BS** or "back-swing" and **FS** or "fore-swing") control, in opposition, the torque applied to the leg. When the sigmoid output of BS is 1 and FS is 0, the leg will swing back maximally, and when the output of FS is 1 and BS is 0, it will swing forwards maximally.

A neural system can generate the walking behavior by alternating between having neurons FT and BS on and FS off (to propel the body forward) and having FT and BS off and FS on (to return the leg to the forwards position). The former state is called the **stance**, and the latter the **swing**. In this work, we also refer to the prior as the BS phase and the latter as the FS phase. This behavior is shown in Fig. 2.

Dynamics classes Previous analysis of the legged walking behavior has split the most adaptive CTRNN pattern generators into categories based on their dynamic modules (Beer et al., 1999). In this study, they found significant heterogeneity in solutions, such that the average (in parameter space) of their solutions was not itself a solution capable of generating the behavior. Suspecting that there might be a multimodal structure to the systems' parameter space, the researchers found two primary classes of systems comprised of 3 neurons – **FS-switch** and **BS-switch** circuits. The distinction between the two is made on the basis of which neuron triggers each transition between stages: because all of the neural units in the system are motor neurons, one of them must act as a timer and trigger the transition at the correct time.

A 4-neuron pattern generator, on the other hand, has the benefit of a physically inert interneuron (**INT1**). As such, using a motor neuron for timing is no longer necessary, and all of the most adaptive 4-neuron circuits rely on INT1 for the timing function. INT1 changes states to trigger the transition from the BS phase to the FS phase, and switches back to trigger the reverse transition. Accordingly, the researchers found a different primary distinction for 4-neuron pattern generators, namely, whether the interneuron turns on during the stance (BS) phase or during the swing (FS) phase. These classes are given the names "stance switch" (**ST-switch**) and "swing switch" (**SW-switch**) accordingly. Notably, averaging the parameters of each class generated functional walking circuits.

The dynamical modules framework allows one to study circuits more deeply by investigating how subsets of the neurons trigger other subsets to transition using a dynamical systems theory technique known as quasistatic approximation. However, this technique sadly becomes more difficult to apply as the number of neurons that make their transitions simultaneously increases. Additionally, and perhaps even more problematically, the number of possible patterns of interaction between neurons grows exponentially with the number of neurons, utterly preventing the sort of neat dichotomy that can be achieved in the 3- and 4-neuron cases.

Analogical comparison

Analogy is central to the study of multiple realizability; namely, analogy is the mechanism by which one can claim that two physical entities instantiate the same phenomenon. Analogy has long been a focus of research in the cognitive sciences (Gentner et al., 2001), and it has been proposed as the "fuel and fire" of thinking (Hofstadter and Sander, 2013). While different models of analogy-making incorporate a variety of analogy's sub-processes – including retrieval, mapping, abstraction, representation, and encoding – virtually all aim to capture the fundamental mapping process that defines analogy (Gentner and Forbus, 2011).

Computational models often side-step another component of analogy-making – perception. SME (Falkenhainer et al., 1989), a seminal model of analogy-making, does so by using predicate logic representations, whereas some Bayesian approaches rely on a grammar for rules relevant to a particular problem domain. On the other end of the spectrum, deep neural network approaches to analogy-making often handle perception internally, at the expense of losing interpretability, often failing to delineate perception from analogy-making, and requiring significant domain-specific training. While SME and deep learning approaches make different trade-offs, the field is still far from generating and analyzing analogies at a human-like level (Mitchell, 2021).

In this paper, we make use of a recently-presented model of the analogical comparison of sequential narratives (Breithaupt and Leite, 2021). This model, which we term Network Model of Aligned Transitions over Core Histories or "Net-MATCH", is designed to assess the strengths of analogies, where analogues can be grounded in multiple different domains and modalities. The tool mitigates the possibility of incompatibility between domains by representing potential analogues with **sequential narratives**: that is, sequences of directed, weighted networks. (There is no need for a sequential narrative to be a temporal sequence, although it is one for the purposes of this paper.) Representing an entity of interest as a sequential narrative requires domain-specific knowledge, but many domains yield natural network constructions and stage segmentations.

Net-MATCH dismisses the need for experiences to resemble each other at the surface level (identity of components of behaviors or events) and instead puts emphasis on the deeper level of persistent structure throughout the narrative derived from the experience (relationships between the components over time). We see this as a key feature that enables us to seek common patterns across the activity of neural circuits of different sizes. In the context of this paper, analogical comparison describes the process of extracting the "experience" of a behaving neural circuit by seeking structural patterns similar to those found in the dynamical modules analysis, and understanding how these patterns shift over time. Two sequential narratives are analogous if their underlying structures shift with similar trajectories over time.

While Net-MATCH exclusively focuses on the mapping stage of analogy, it expects less of its perceptual system by operating on the level of activation patterns rather than the higher level of semantic information. Unlike deep-learning models of analogy, which start with raw sensory inputs and require training, Net-MATCH relies on the intrinsic structure of its input, which has generally undergone some form of neural pre-processing. This activation-based approach to analogical comparison makes Net-MATCH well-suited for comparing behaving neural systems.

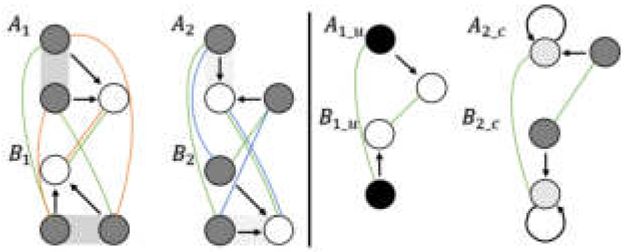


Figure 1: Figure 1 depicts sequential narratives (A and B, left), and the narrative when collapsed by A_1 and B_1's community structure (right). Both networks in A are isomorphic to the corresponding networks in B, but only the isomorphism shown in green is common between the stages. When the narratives are collapsed by community structure prior to alignment, the found isomorphism continues to work across both stages. This indicates a strong analogy between narrative A and narrative B, and the transformation that takes A_1 to A_2 is the same one that takes B_1 to B_2.

Relative to other existing models of analogy-making, Net-MATCH is particularly suitable for the domain of neural circuits due to (i) its applicability to temporal structures, (ii) its invariance to the size of the system it is applied to, and (iii) its domain-flexible network-based approach.

Computation The Net-MATCH algorithm consists of three steps: (i) find the underlying structures of the networks in each sequence; (ii) find the best alignment of the underlying structural elements at each stage of the sequence; and (iii) evaluate how well the alignments from each stage translate to the other stages in the sequence.

To unpack this process further and motivate its steps, we provide a minimal example (Fig. 1). The colored sets of lines going between A_1 and B_1 indicate two equally valid alignments of the networks at stage 1 of the narrative. However, only the alignment shown in green agrees with an alignment between the networks at stage 2 of the narrative.

(i) To account for the possibility of there being several strong ways of aligning the nodes between the networks at each of the stages, the tool first identifies communities of nodes that behave similarly to each other (shaded together in A_1 and B_1) on the basis of similar incoming and outgoing connections, and collapses them together to form a single node (Fig. 1, center right). Doing so both allows the tool to be more robust to ambiguities in the alignments and allows it to handle networks of very different sizes. Similarity in connections is assessed by summing the cosine similarity of the nodes' incoming weight vectors with that of their outgoing weight vectors (with the exception that near-zero vectors have a similarity of 1 with each other and 0 with all other vectors), and then communities are detected by applying Louvain community detection (Blondel et al., 2008) to the resulting similarity matrix. Communities are collapsed into single nodes by summing incoming and outgoing weights.

(ii) Next, having grouped the communities of nodes together, the tool aligns the corresponding networks at each stage of the narrative. (A_{1_u} and B_{1_u} show the strong alignment between the networks of stage 1.) The alignment is performed to minimize the Euclidean distance between the weight matrices of the two networks. In the event that one weight matrix is larger than the other, the least similar nodes are discarded. In our current implementation, this is done exhaustively using a backtracking algorithm with factorial time complexity; however, heuristic graph alignment algorithms could be used at this stage if there are many nodes.

(iii) To evaluate how well the underlying structure from each stage translates to each other stage, the tool finally applies the communities and alignments derived from the networks at one stage (A_{1_u} and B_{1_u}) to the corresponding networks of another stage (A_{2_c} and B_{2_c}) and scores the alignment on the other stage using Euclidean distance. The example demonstrates how the operation going from A_2 to A_{2_c} and B_2 to B_{2_c} effectively applies both possible alignments (green and orange) from the first stage of each narrative to the second. Repeating this cross-evaluation across the length of the narrative and summing the resulting alignment scores yields a final distance from a perfect shared trajectory.

Generalized analogy strength and significance testing To reduce the effect of a narrative's particular weight values and normalize analogies across a range of networks, we have introduced two statistical measures on top of Net-MATCH. These enable the experimenter to understand how strong a given analogy is relative to versions of the same mapping where any common trajectories over the course of the narrative have been disrupted. Concretely, this describes the process of swapping out the identities of nodes at each stage of the narrative so that the alignment at one stage no longer carries over to the rest of the sequence.

We define the measure of **analogy strength** as being the number of standard deviations the distance score of the original analogy is below the mean of the scores of the disrupted versions of the same pairing. To represent the probability of detecting an analogy of this strength by chance, we assign a p-value to analogies to denote the proportion of the time that the distance score of the original analogy is greater (the analogy is poorer) than those of random analogies.

These are each somewhat computationally intensive measures in that they rely on Monte Carlo sampling to represent the space of possible analogy scores. We performed them with 1000 repetitions, which yielded good numerical stability while still allowing our paper's analyses to complete within a few hours of computing time.

Methods

Functional network traces

Legged walking is a cyclic behavior for which each cycle of behavior (one step) is comprised of two distinct behavioral states (the stance and the swing) and many of the most interesting network dynamics take place during the transitions between these states. Taken together with the fact that Net-MATCH operates on sequences of directed weighted networks intended to reflect temporality, our first task becomes to split the behavior's timescale into multiple discrete periods. To do so, we break the behavior of the walkers into four stages – Back Swing (BS), Transition between Back Swing and Front Swing (T1), Front Swing (FS), and Transition between FS and BS (T2).

These stages are split based on the following criteria: (i) when $BS - FS > 0.5$ and $FT > 0.5$, we are in the BS (stance) stage; (ii) when $FS - BS > 0.5$ and $FT < 0.5$, we are in the FS (swing) stage; (iii) in any other time, we are in a transition stage. Additionally, in order to ensure that the entirety of any essential transitions falls within the transition phase, the transition phase is expanded until the total rate of change of all neurons reaches a local minimum (or a threshold value set to ensure that the FS and BS stages are never consumed by T1 and T2).

To obtain a meaningful sequence of functional networks out of each circuit's behavior, we derive the **functional trace** of the circuit during each stage using a technique inspired by differential Hebbian learning (Zappacosta et al., 2018). In order to approximate which neurons cause others to fire, we ask which neurons are active when other neurons change their state. By integrating this measure over the period of the stage, we get a trace of directed correlations between different neurons' activations and rates of change. The computation of a functional trace is illustrated in Fig. 2:

$$F_{i \to j} = \int_{t_i}^{t_f} O_i \frac{dO_j}{dt} dt \qquad (1)$$

We want to emphasize two points regarding this novel functional trace method. First, the functional trace does not rely on any knowledge of the inner workings of the neural circuit it analyzes beyond each neuron's instantaneous activation. Second, the functional trace is a correlative and not causative measure; it reflects patterns in how neurons change configuration over time, but not necessarily ways that neurons cause one another to change configuration.

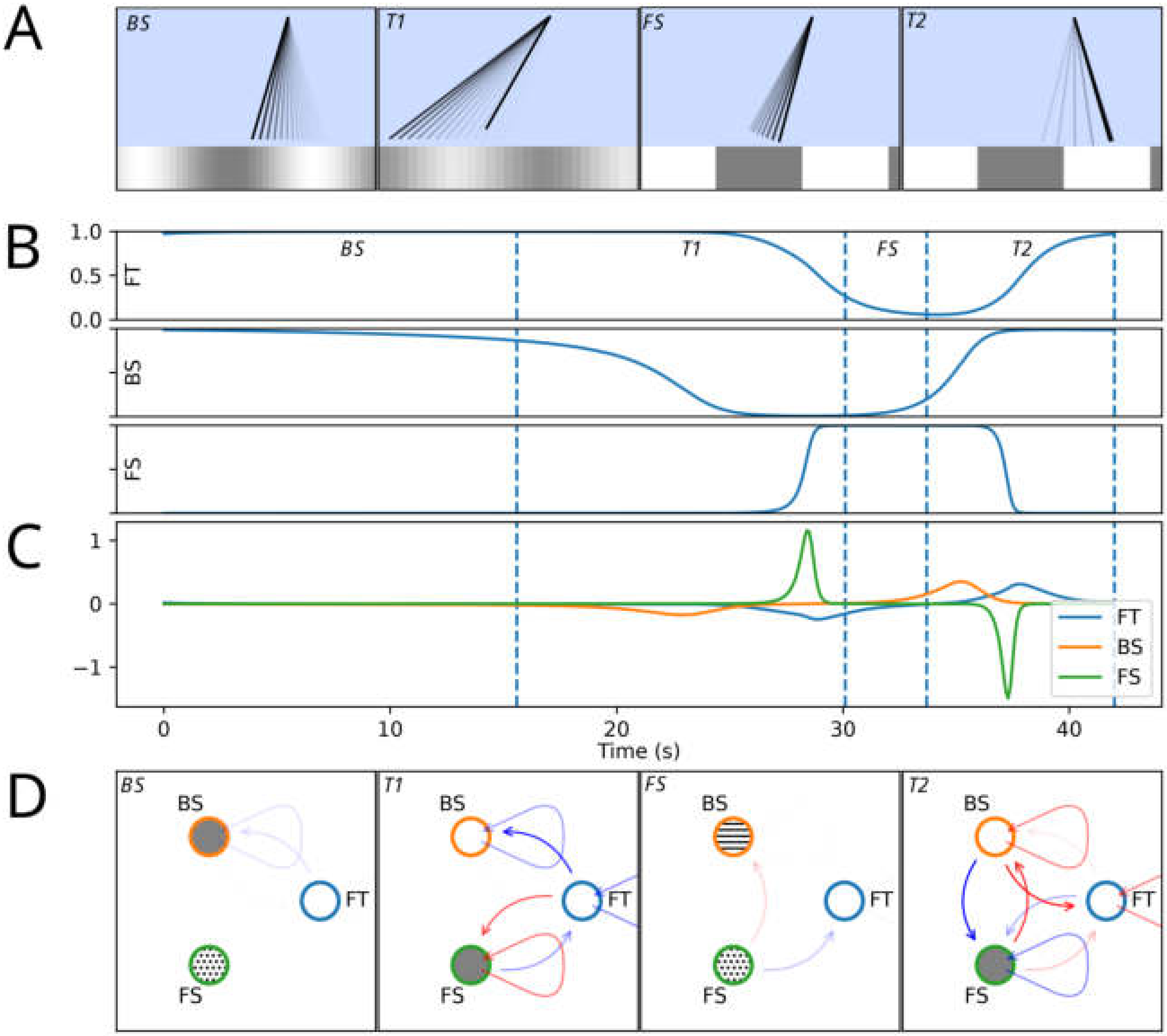


Figure 2: Evolved BS-switch circuit (seed 48) behavior (A), neural activities (B), neural activity rate of change (C), and functional traces during each stage (D). Columns represent stages of behavior (*BS*, *T1*, *FS*, and *T2*, italic type). Neuron identities are FT, BS, and FS (non-italic).

In functional traces, outline color represents neuron identity, lines represent directed functional trace weight between neurons. Red lines represent positive weights and blue lines represent negative weights. Weight magnitude is represented by opacity. Fill pattern represents neuron community, based on incoming and outgoing weights in the functional trace.

The functional trace from neuron i to neuron j (D) is obtained by multiplying neuron i's activation (B) and neuron j's transition rate (C) at each time during a stage, and integrating the result over the length of the stage.

Experiments and results

In order to analyze how well Net-MATCH and the functional trace method apply to the legged walking behavior, we evolved 150 CPG CTRNN circuits for single-legged walking, with 50 circuits each of size 3 (where the neurons are linked with the motor actuators FT, BS, and FS), 4 (with one interneuron), and 5 (with two interneurons). We used a hybrid evolutionary algorithm with a population size of 250 and an elitist fraction of 10 individuals, where the non-selected individuals were subject to mutation noise and recombination with the elitist fraction, with both increasing as the non-selected individual's rank in the population decreases. Our run indices refer to these original 150 seeded runs, with indices 0-49 referring to 3-neuron circuits, 50-99 referring to 4-neuron circuits, and 100-149 referring to 5-neuron circuits. Following Beer et al. (1999), we selected the best 20 circuits of each size for further analysis. Every one of the best 60 circuits quickly converges on a repeated cyclic pattern of activation that corresponds with the walking behavior. We split the behavior into sequential narratives as discussed above, aligned each narrative to the second full BS-T1-FS-T2 cycle, and took functional traces.

Similarity within circuits

Our first experiment validates our intuition that Net-MATCH applies to the small-circuit generation of behavior.

In this experiment, we evaluate the strength of the analogy between a neural circuit's functional trace during a single cycle of the walking behavior with its functional trace during the following cycle. (As this is a cyclic behavior and the neural circuit's activation from one cycle to the next is practically identical, this provides a ceiling for the analogy strength of a system of a given complexity.) Next, we evaluate the strength of the analogy between its functional trace during a single cycle with its functional trace during the following cycle *offset* by either one, two, or three stages. (That is, the BS phase would line up with the T1, FS, or T2 phase.) We evaluated the self-analogy strength at each offset for all sixty top circuits, thus obtaining 4 sets of 60 analogy strength values. Our expectation in designing this experiment was that the properly aligned self-analogy would provide a benchmark for "strong analogy", and that the offset analogies would be weaker, though at this point we had little reason to guess whether or not they would be significant.

In line with our expectations, we found that aligned within-circuit analogies were the strongest, with a mean strength of 3.38. (That is, the analogy distance was 3.38 s.d.s below the mean analogy distance when the persistent structure of each narrative sequence was disrupted.) All 60 circuits' properly aligned self-analogies were significant to a $p < 0.05$ threshold.

This was followed by offset-two analogies (lining BS up with FS), which had a mean strength of 0.58 and were 55% significant (33/60 circuits). As seen in Fig. 3, there is significant variability and the appearance of bimodality among these analogies, with a cluster of analogies centered around 2.5 and a cluster centered at 0. The difference in means between offset 0 and offset 2 is significant to 95% confidence.

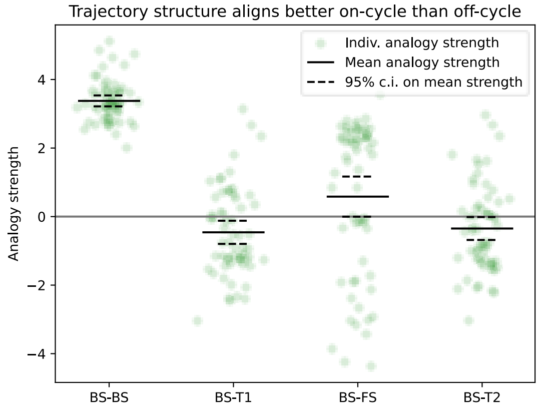


Figure 3: The strength of on-cycle and off-cycle analogies within individual circuits. On-cycle analogies consistently display high analogy scores. Individual circuits' analogy strengths shown as green transparent points. Mean analogy strength for each alignment shown as black solid line; 95% confidence interval for each mean shown with dashed lines.

Next were offset-one and offset-three analogies, lining BS up with T1 and T2, which had mean strengths of -0.46 and -0.35 respectively. Only 4 circuits had significant offset-one analogies, and 8 circuits had significant offset-three analogies. (Given our threshold of 0.05, this is not above chance levels.) The difference in means between offset 2 and offsets 1 and 3 is significant to 95% confidence; the difference in means between offsets 1 and 3 is not significant.

Cross-circuit analogies

Our second experiment shows that Net-MATCH can detect meaningful similarities in the neural activity of different circuits engaged in the same behavior, but that it does not invariably do so.

To execute this experiment, we lined up each circuit's functional trace during a single cycle of walking behavior with each other circuit's, for a total of 1830 unique analogies: 1770 between pairs of circuits and 60 self-analogies.

These analogies had a relatively high mean strength of 1.79 but a great degree of heterogeneity, with standard deviation 1.40, minimum -3.40, and maximum 5.32. Out of the 1170 unique analogies between pairs of circuits, 1120 were significant to a $p < 0.05$ threshold. Certain circuits, particularly those with the five-neuron architecture, were particularly distinctive with strong patterns of analogy and disanalogy. In our next experiment, we explore how this finding reflects the dynamical properties of the circuits in question.

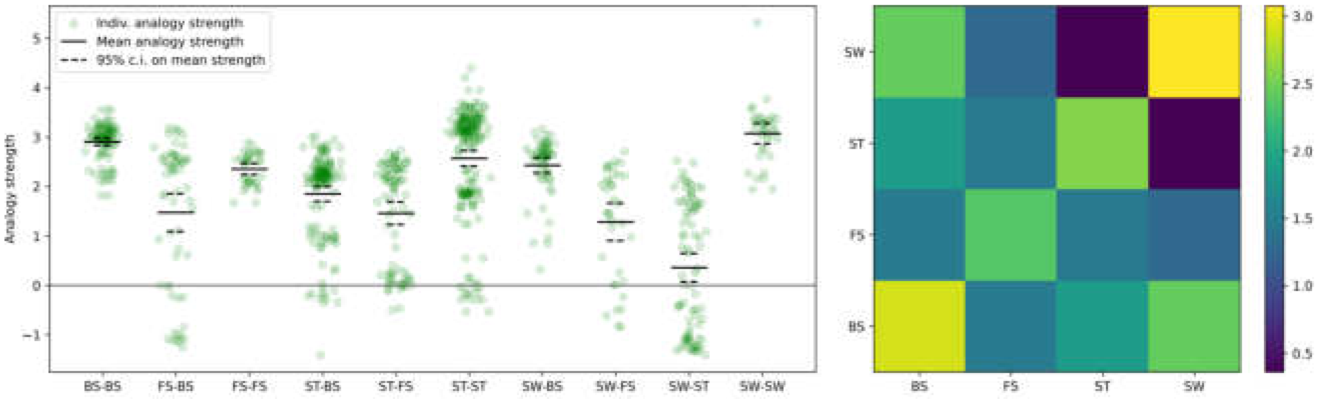


Figure 4: Analogy strength between pairs of circuits by dynamics class. Distribution, mean, and c.i. plot shown on left with same format as Fig. 3 and dynamics classes shown on x-axis; classwise heatmap shown on right with dynamics classes shown on x- and y-axes. Color legend shown at far right. Within-class analogies are generally stronger than cross-class analogies, with the exception of the very strong analogies between BS-switch and SW-switch classes.

Dynamics class and analogy

Our third experiment shows that Net-MATCH agrees with previous efforts to structure the space of legged walkers and that it reveals novel cross-architecture patterns.

To explain the highly heterogeneous results in the previous experiment, we applied Beer et al. (1999)'s classification of 3- and 4-neuron circuits generating the legged walking behavior to understand whether the insignificant or negative analogies resulted from cross-class comparisons. Splitting up 3-neuron circuits by which neuron timed each transition, we found 10 BS-switch and 6 FS-switch circuits as well as 4 unclassifiable circuits (one where FT played the role, two where FT and FS alternatively played the role, and one where BS and FS alternatively played the role). We excluded these circuits from the analysis. Splitting up 4-neuron circuits by dynamics class, we found 14 ST-switch and 6 SW-switch circuits. We then calculated the strengths of the pairwise analogies between all circuits of each of these 4 classes and the mean analogy strength across each pair of classes.

The results of this analysis (Fig. 4) are highly informative. Most classes invariably produce strong positive analogies among their members, with the exception of the analogies involving one outlier ST-switch circuit whose interneuron's off-on transition had been (misleadingly) automatically grouped with the motor neurons' in the T1 phase rather than on its own in the BS phase. With this exception, all negative analogies and the vast majority (216/220) of insignificant analogies among 3- and 4-neuron circuits came from the cross-class treatments. In order of mean analogy strength, the within- and cross-class analogies were SW-SW (3.08), BS-BS (2.91), ST-ST (2.57), SW-BS (2.43), FS-FS (2.36), ST-BS (1.85), FS-BS (1.47), ST-FS (1.46), SW-FS (1.28), and SW-ST (0.36).

It is particularly interesting that the functional traces of BS and SW circuits are generally strongly analogous to each other. On reflection, this result seemed natural to us. First, we discuss the underlying intuition for why this should be the case, and second, we analyze a single representative analogy from this population.

The intuition behind the analogy between SW-switch circuits and BS-switch circuits is that the interneuron (INT1) of a SW-switch circuit triggers FS to make the opposite transition to it, and triggers BS and FT to make the same transition as it. (That is, over a full loop, we see *INT1 off→BS, FT off; FS on→INT1 on→BS, FT on; FS off.*) This is precisely analogous to how in BS-switch circuits, BS triggers FS to make the opposite transition to it and FT to make the same transition as it. (That is, *BS off→FT off; FS on→BS on→FT on; FS off.*) To pre-empt any argument that one can contrive a perfectly good analogy for any pair of classes, we assert that this is a stronger analogy than can be made between ST-switch and FS-switch circuits: in ST-switch circuits, INT1 triggers FS to make the same transition and BS/FT to make the opposite transition; but in FS-switch circuits, FS triggers only BS/FT making the opposite transition. (Although we did not explore this possibility, there may be a stronger analogy between BS-switch circuit traces and timeshifted ST-switch circuit traces, as they have the same same/opposite and on/off signatures.)

The discussed analogy is illustrated in Figure 5, which depicts the analogy between traces of circuits 57 (SW) and 48 (BS). This analogy functions even though each stage has a different optimal alignment because these alignments are consistent with the overall trajectory discussed above: they are compatible when applied across stages. During the BS and FS stages, the interneuron (INT1) of circuit 57 aligns with the BS neuron of circuit 48. During the transition phases, the interneuron of circuit 57 goes unaligned and the BS and FT neurons align as a cluster with the BS and FT neurons of circuit 48.

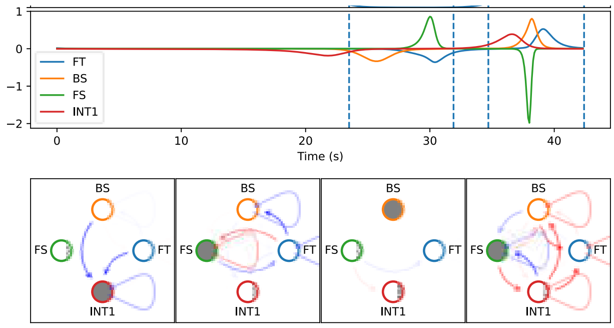


Figure 5: Evolved SW-switch circuit (seed 57) neural activity rate of change (row 1) and functional traces (row 2). Neuron communities correspond with those shown in Fig. 2, representing the common structure at each stage. These alignments are in fact compatible across stages.

Discussion

In this work, we found that analogical similarity as a metric in general and Net-MATCH in particular stand as powerful utilities for understanding the multiple realizability of a behavior. By drawing analogies between different neural circuits' functional traces, we were able to replicate in an emergent manner the dynamics classes devised by Beer et al. (1999). In addition, we were able to extend their analysis by drawing analogies between neural circuits of different sizes, using Net-MATCH to discover a natural analogy between BS-switch and SW-switch circuits.

As a tool for understanding the neural bases of behavior, Net-MATCH has different use cases, strengths, and weaknesses than analytic tools like dynamical modules.

While analogy and dynamical modules can both be applied to a system with unknown synaptic weights, dynamical modules provide the greatest benefit when used in concert with dynamical systems analyses like quasistatic approximation that assume access to the parameters of the system. In these cases, one can gain insight into the parameter space tradeoffs and relative parameter sensitivity in a neighborhood of the system parameters. In this regard, our work has significant limitations; our application of analogical comparison makes no predictions about behavior performance, and it also cannot be used to infer anything about the parameter space without generating new narrative sequences based on perturbed versions of the original circuit.

One additional limitation of our approach is that we needed to tailor our narrative stages specifically to this task, and it may be particularly difficult to generalize it to tasks that can be solved with different numbers of stages depending on the solution. In this situation, it may be necessary to impose additional structure on the timeline of the solution.

At the same time, analogical comparison has some advantages over dynamical modules analysis. It has fewer boundary cases, requiring no strict order of transitions or decision about which neurons belong to the same module at any given time. It can operate across different circuit sizes or architectures, and we have shown a case where it reveals shared structure across this boundary. Theoretically, it can even find analogies across different domains. We conclude that analogical comparison is a powerful new tool for the analysis of the space of circuits generating a behavior.

Future directions

We see three natural generalizations of this work.

First, our methodology can scale up to more complex problems and neural circuits. Because it is designed to identify groups of similar neurons, Net-MATCH can naturally scale up to reveal structure in the space of larger circuits that solve any task. However, as tasks become more complex, so does the problem of splitting up behavior into cycles and stages and aligning the resulting sequences of functional traces. In our present work, we made use of the previously characterized swing and stance phases of single-legged walkers. Moving towards more complex tasks, such as visual categorization, this problem will become significantly more trying, particularly for tasks where different valid solutions have different numbers of constituent stages. Resolving it will take careful consideration of the task's subtleties. However, in cases where we consider two directly analogous tasks, like two-legged and six-legged walking, we would hope to uncover similar patterns of neural dynamics due to the natural alignment between the time courses of both cycles. It will be interesting to see how the neurons of CPGs that control six legs will correspond with those in a CPG controlling two or only one leg.

Second, one can generalize the techniques we have applied in the neuroscience of artificial organisms to biological neuroscience. On the small-circuits level, one can apply analogical comparison to existing recordings of small subcircuits of the nervous systems of model organisms like *C. elegans* to understand (i) whether the same organism has analogous, but not identical, neural readings when repeating a behavior, and (ii) whether one can form analogies between the neural traces of two organisms that have engaged in the same behavior. One can alternatively go in a large-circuit direction and consider system-level fMRI data from people viewing similar or different scenes, from which functional traces can also be extracted and compared.

The third natural generalization of this work is towards multifunctional neural circuits. Can one use analogical comparison to detect which behavior a multifunctional circuit is currently generating? Would one find stronger analogies between a multifunctional circuit's traces during two behaviors, or between its trace in each behavior and that of a single-function circuit performing the same behavior? These are interesting and timely questions and we hope to explore them further soon.

References

Beer, R. D., Chiel, H. J., and Gallagher, J. C. (1999). Evolution and analysis of model cpgs for walking: Ii. general principles and individual variability. *Journal of computational neuroscience*, 7(2):119–147.

Beer, R. D. and Gallagher, J. C. (1992). Evolving dynamical neural networks for adaptive behavior. *Adaptive behavior*, 1(1):91–122.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:10008.

Breithaupt, K. and Leite, A. (2021). Persistent common structure as a measure of analogic similarity. ASIC 2021: Twentieth Annual Summer Interdisciplinary Conference.

Falkenhainer, B., Forbus, K. D., and Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, 41(1):1–63.

Funahashi, K.-i. and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks*, 6(6):801–806.

Gentner, D. and Forbus, K. D. (2011). Computational models of analogy. *Wiley interdisciplinary reviews: cognitive science*, 2(3):266–276.

Gentner, D., Holyoak, K. J., and Kokinov, B. N., editors (2001). *The analogical mind : perspectives from cognitive science*.

Hofstadter, D. R. and Sander, E. (2013). *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*.

Mitchell, M. (2021). Abstraction and analogy-making in artificial intelligence. *Annals of the New York Academy of Sciences*, 1505.

Zappacosta, S., Mannella, F., Mirolli, M., and Baldassarre, G. (2018). General differential hebbian learning: Capturing temporal relations between events in neural networks and the brain. *PLoS Computational Biology*, 14.

Navigating blind without a map: models of active wayfinding

Inman Harvey
Evolutionary and Adaptive Systems (EASy) Group
University of Sussex, Brighton BN1 9QH, UK
inmanh@gmail.com

Abstract

Maps are useful for navigation if (i) there is adequate known detail provided on the map, (ii) your present location on the map is known as is (iii) the location of your goal. Many natural examples of successful navigation, such as seasonal bird migration across continents and oceans, lack some or all of these. Success requires some strategy for continuously governing the direction of movement according to continuous sampling of available sensory cues. If sensory cues are strictly local, an instantaneous snapshot often gives insufficient guidance on the course to steer. Some form of *active perception* is needed, where motion gives rise to cues as to whether the current course needs to change. We illustrate with one old and two novel examples of increasing complexity: (A) Run-and-Tumble strategies, such as used by *e. Coli*, allowing the detection of local gradients. (B) Swerve-Zone, a novel artificial life model of bird migration showing how region-wide cues, even in the absence of discernible local gradients, can nevertheless still guide. And (C) Parting-the-Waves, a proposed strategy for exploiting the wave patterns underlying long-distance steering as used by Micronesian navigators, showing how the boat motion is essential for discriminating between swells. All three depend on some default motion; when stationary, you cannot sense where the goal lies. They exploit motion in different ways, but all simplify navigational search into tractable forms apparently amenable to evolution.

Introduction

All animals and humans have abilities to navigate locally in a familiar neighbourhood, but some go further. Seabirds can migrate between the Canadian Arctic and Madagascar in the Indian Ocean (Harrison et al., 2022). Monarch butterflies, weighing half a gram, can annually commute between Canada and Mexico (Taylor et al., 2020); it may take 3 or 4 generations for them collectively to complete the round trip of up to 5,000 kms. Traditional Polynesian and Micronesian navigators, sailing without tools or maps, crossed the Pacific Ocean between small remote islands with no land in sight for days or weeks. The routes taken may vary between trips, e.g. if disrupted by storms, but still arrive at their goals, both going out and returning back; awe-inspiring examples of cognition and regulation. 'Govern', like 'cybernetics', derives from the Greek word for steering.

Many different navigating techniques may be combined in such trips at certain stages, some may involve *direct sight* of the target direction, or of an intermediate waypoint. But here we focus on those long stages where the animal or human is *navigating blind*. Though there may be some global orientation indicator (maybe very approximate) to show e.g. global North or global South, or perhaps the wave direction of a longterm primary swell across an ocean, there is nothing visible from any single static location to directly indicate the target's direction or that of a waypoint. We shall show where *motion* of the agent turns out to be essential to guide its steering.

Wayfinding without knowing where you are

People use maps to accumulate, remember and share knowledge for aims that include navigation. Find the current (x, y) coordinates of the navigator, where you currently are on the map, and likewise the (x, y) coordinates of the destination. Look for plausible connecting routes. A route plan is a sequence of identifiable places and directions along identifiable routes between them.

Migrating birds, butterflies, eels, turtles are clearly mostly doing something very different. Though they have identifiable starting and ending places — let's simplify by calling this a N or S terminus since this typically fits in with seasonal migration — there are often no specific intermediate places they must pass through, and no identifiable tracks. The task is more basic: how can purely local cues offer guidance on a direction to steer, in egocentric polar coordinates?

We shall here ignore the sort of strategies that might well be used near the termini — including identifiable landmarks, dead reckoning, river-following — and focus on the central parts where the animal (or human navigator) has no source of information as to where it is. One can still navigate successfully if environmental cues suggest the right sort of direction to go in. What sort of cues, how might they work? And since they are shared across a species, how does that transmission work?

Here we discuss and analyse three different examples of navigating blind. First is (A) Run-and-Tumble, well-known in the biological literature. The second, and third examples are presented here for the first time as Artificial Life models illustrating novel potential strategies. (B) Swerve-Zones are for long range migration over land and water using regional cues that shape the trajectories taken. (C) Parting-the-Waves models wave navigation across oceans, based on swells reflected from islands.

(B) and (C) are inspired by real-life scenarios, respectively bird navigation and traditional Micronesian navigation in the Pacific. But we model these as Artificial Life scenarios in idealised simulations. We do not aim to explain any specific biological example, but by exploring the range of possible strategies we hope that this may offer new ideas to those who are trying to understand a specific example. In particular, this demonstrates how such navigation has no need for anything that looks like a map, and no need for an accurate sense of direction. Simple alternatives are available.

(A) Run and Tumble

As our first simplest example of navigating blind consider (Fig. 1) the case of chemotaxis where the target could be the denser centre of some distribution of sugar solution that is food for a bacterium. The figure shows us, the external analysts, a map overview of the big picture with gradient-climbing being an obvious strategy. But the map is not available to the bacterium; how can it detect and act on a

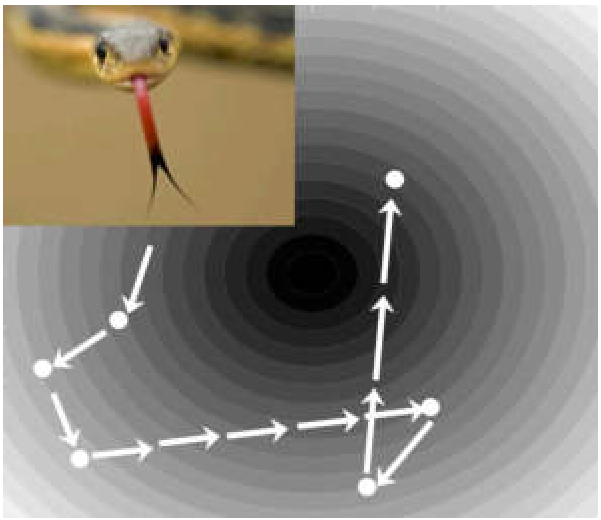


Figure 1. A snake with forked tongue can compare the cues sensed at each tip and detect a gradient. A tiny bacterium senses at a single point and can only detect a gradient after each movement. Increase or decrease of cue determines ‘Run’ (straight) or ‘Tumble’ (random turn, white spot), hence tending to stay in the denser regions.

gradient direction when it can only sense locally at its current position?

A snake with a forked tongue may be able to directly sense the gradient between chemo-sensors on the left and right tips, provided the cue strengths are different enough to be discriminated over that separation distance. The two sensors also need to be calibrated to match. A bacterium is too small to have two sensors separated thus, and hence has to trade time for space (Fig. 1). It discriminates between local cue strengths before and after a movement over some short distance. There is no need for left/right sensor calibration.

A Run-and-Tumble strategy may start anywhere, moving in any direction; it works in 2 or 3 dimensions. The strategy has two modes of action: a default Run mode of ‘continue as before’, and a Tumble mode of veering to a random new direction when triggered by a decrease in cue strength. Though the figure suggests a choice is made between these modes at regular intervals, this need not be so. An event-driven version would have the Run mode continuing indefinitely until a Tumble is triggered by a sensed decrease in cue. We may note that the Run mode does not require a perfectly straight course. So long as it is not chaotically changing, a roughly straight-ish course, even a veering bias to left or right, will still work.

This wayfinding method illustrates a key feature of *active perception*. The cue provides zero information as to the desirable direction to go in the absence of movement; it only becomes meaningful when the organism is actively in motion. Run-and-Tumble is easy to understand, and well-studied (Armitage and Schmitt, 1997; Egbert et al., 2010). There is no real distinction here between the signal used as a cue for navigation and the target goal of that navigating; the former is the local density, and the target need not be a unique goal, it is any denser region. Our next two studies on navigation, presented here for the first time, are more subtle and use signals not so directly related to the goal.

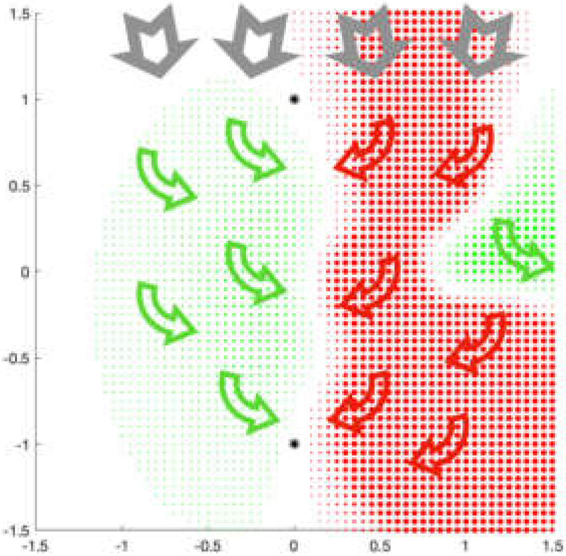


Figure 2. A tendency to migrate roughly south is modified by multiple overlapping regional cues of 2 classes: ‘green’ and ‘red’ tend to swerve the course to left or right respectively, thus channeling between the Swerve Zones.

(B) Swerve Zones

For our second example we propose a novel simple general class of strategies for long distance migration that relies on regional cues. Whilst this is inspired by real examples of bird and butterfly migration, we limit this study to Artificial Life simulations that may — or may not — correspond to one of the many ways that such creatures navigate. It is the job of a biologist to make such a determination, and these models are intended to expand the range of hypotheses they have available for analysing observed natural navigational skills.

Our minimal assumptions

Our world is for practical purposes 2D. Where e.g. a mountain range forces a bird to fly higher, this third dimension can provide a sensory cue ‘altitude’ in that 2D region. We assume the animal has no idea where it is -- if it knew yesterday, a

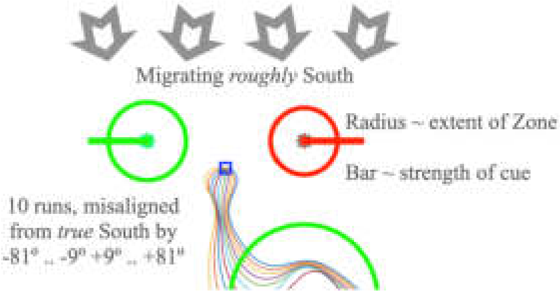


Figure 3. To ensure robustness to inaccurate directional sense, up to +/- 81° from true, 10 runs are tested. ‘Radii’ of Swerve Zones are indicated, modelled as hyperbolic secants. Bars represent (genetically specified) maximum perceived cue strength at the centre of a zone.

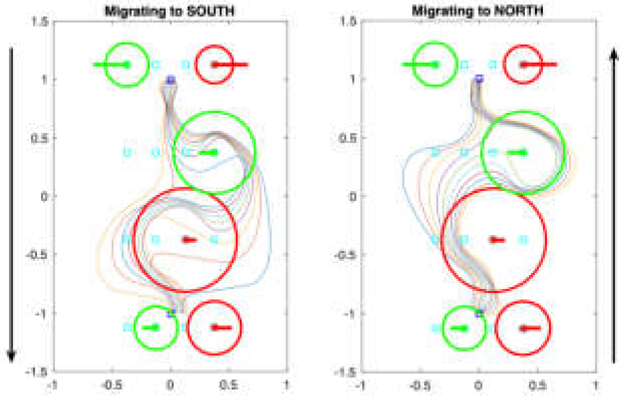


Figure 4. Example runs. Left side, migrating to South, 10 runs starting at (0,1) and aiming for (0,-1). Right side. Reversing the direction and also reversing the left/right sense of the Swerve Zones.

storm might have shifted it randomly today. We assume its strategy is reactive, its chosen direction of movement now depends solely on its sensory cues locally available now. We assume that these cues include some crude and approximate sense of global direction, such as might be provided by the earth's magnetic field, or some time-adjusted direction of the sun. In fact below we shall show below that it will be sufficient merely to be able to distinguish between North-ish and South-ish.

Apart from such globally available directional cues, we assume that all other cues are equivalent to regionally based 'pheromone' concentrations; strongest within some region and fading away to zero with distance. As an indication of scale, over a total range of 1000 to 2000 kms such regions may well extend for 100s of kms, and may multiply overlap each other. Flying over a mountain range can be treated as providing graded height cues; over the sea or over the land similar cues, here more sharply binary than graded. The air, or water, may contain chemical signatures over extended regions. We treat all such possibilities similarly. Only the local cue strength is available, not its gradient. We assume all these senses are noisy and unreliable.

We assume the direction taken is simply determined by the grossly inaccurate general sense of North-ish or South-ish, as modified by the strength of any currently perceived cues that induce deviations left or right (Fig. 2): swerves shaped by Swerve Zones (SZ). Thus an agent travelling South-ish in the absence of cues would be deflected to (e.g.) the left as it enters a 'pheromone' zone, the bigger the swerve for the stronger the cue; and then reverts to its South-ish course after it passes through the zone and moves out of range.

This is still sufficient to underwrite successful navigation strategies, provided there are sufficient regional sources spread around appropriately. We shall explore some minimal simulations. A strategy will be a set of rules that translate the currently sensed cues into the current direction to be followed.

Orientation and robustness

A guaranteed perfect global compass cue, and no storms on a straight line from one terminus to the other, is an idealised scenario that we must rule out. On a long journey even a small dead-reckoning deviation accumulates to missing the target.

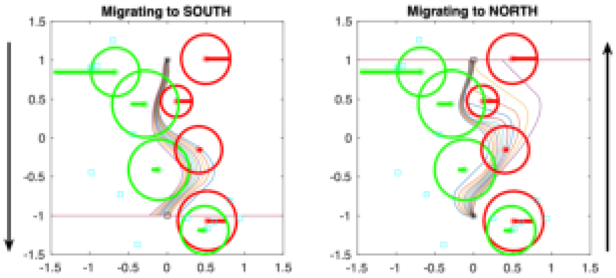


Figure 5. Further example runs. See text box for parameters.

Here we assume a global compass can afford to be up to 90º inaccurate and there is no accumulation of errors. We model the robustness that copes with this within our simulations by testing each strategy 10 times with the 'perceived South' varying (at 18º intervals) between +/- 81º discrepancy with 'True South' (Fig. 3).

One of our SZs does not provide such a sense of direction, but two (or more) can do something related. Suppose I am moving within sensing distance of different regions A and B, and I use one arbitrary parameter applied to the A-cue strength to shift my direction to the right, and another arbitrary parameter to the B-cue shifting me left. Multiple runs with multiple start positions and parameter choices will generate a phase portrait characterised by the gathering of a bundle of trajectories to pass between region A (on left) and B (on right). The robustness to parameter choices (and hence to sensor noise) is an example of *Rein Control* (Harvey, 2004), balancing opposing forces.

We can note that such an AB channel that directed southbound trajectories would do a similar job directing northbound if the swerves were reversed.This suggests that the same cues can be used both ways, subject to some simple internal switch. We now need to find an orientation strategy that uses such environmentally available cues, and decides which to ignore and which to attend to; are these left-shifting or right-shifting (subject to the seasonal north-south switch), and what relative strengths should be allotted to each cue?

How does the strategy get learnt?

In some biological species the strategy may be passed on through lifetime learning. Follow mum and dad on a round trip or two, remember what happened. This only defers the question of how their ancestors first learnt, and anyway for many species lifetime learning is not possible. Monarch butterflies in general do not individually complete a round trip of up to 5,000 km, it may take 3 generations to make the circuit. Eels may travel 7,000 km from the Sargasso Sea to the European stream their parents knew - but their parents are unavailable to show them the way, they are dead. So we assume the strategy is passed down the generations genetically, and must have been originally acquired and subsequently modified through evolution.

We can model this (Fig. 4) with an evolving population of agents in a 2D world with the North terminus at (0,1) the South at (0,-1). Success is measured by how close to x=0 the agent is when crossing the y=1 or -1 lines. The environment contains an assortment of 'Swerve-Zones' (SZ) that notionally emit 'pheromones' or cues. We here use SZs that are circularly

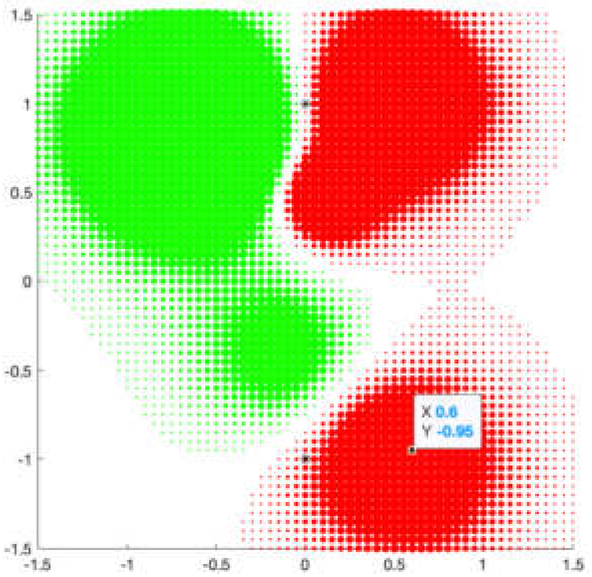


Figure 6. The summed swerves, as sensed by an agent producing the trajectories of Fig. 5.

symmetrical, either with a central circle of uniform cue strength, this tailing off to zero over an outer region, or using a hyperbolic secant function; but other SZ shapes can be considered.

Each agent in the population has a genotype that essentially specifies which of the SZs it is capable of sensing; and for each of these, whether it will be left- or right-shifting, and the factor by which the cue-strength is multiplied. An agent is first tested running N-S, and then the return S-N with the left/right shifting reversed. If we focus on the N-S run, the agent is modelled as having a highly inaccurate sense of where global south is by testing it 10 times over a wide range of 'perceived-south' values; success is required for all such variants. When out of range of any other cues, this perceived-south will determine the agent's direction. On coming into range of one or more cues, these will start to modify the perceived-south bias through the deflections they induce. The perceived-south bias was set at a relatively low level so as to only dominate when distant from any zone.

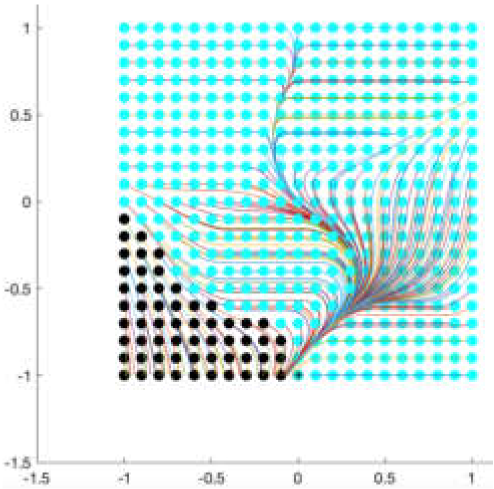


Figure 7. For a southbound agent that accurately perceives south, the swerves of Figs. 5, 6 induce a large basin of attraction (cyan dots) with trajectories leading towards the target area around (+/- 0.1,-1).

We can display an individual's performance through the 10 southbound tracks and the 10 northbound tracks, against an indication of the size, placing and perceived handedness (left/right) of the SZs emitting cues. The overall evaluation of fitness is taken as minimising, over these 20 tracks, the average distance from target when crossing the respective finishing lines. This fitness was used within a simple Microbial Genetic Algorithm (Harvey, 2011).

Artificial evolution turned out to be remarkably easy, with successful runs appearing from the start. With a population size of 30, the equivalent of around 30 generations was typically sufficient to produce successful strategies, provided there were enough randomly provided SZs to be exploited. Figures 5, 6 and 7 illustrate one scenario where 8 out of 20 available SZs were to be selected as either green or red. The parameters used are given in the text box.

Successful runs were achieved with many variants, but here is one example set of parameters used in Figs 5, 6, 7.

20 potential SZs were randomly positioned with centres (x, y) within $-1.0 < x < 1.0$, $-1.5 < y < 1.5$, and with nominal radii varying in the range $0.1 < rad < 0.3$. The strength of each such cue diminished as a hyperbolic secant of the distance from its centre, sech(5*distance/rad). This implied a strength of 1.0 at the centre, 0.0135 at the nominal radius as shown in Figures.

The strategy of an individual agent was determined by a genotype that specified just 8 cues for SZs (from the potential 20) that the agent could sense; and specified for each of these a weighting factor in the range [-20.0,+20.0]. These perceived weighted cues were summed to give a total Swerve (measured in radians), capped within the range +/- $\pi/2$ radians, i.e.+/- 90°.

Each agent was tested 10 times on southbound migrations, with 'Perceived South' varying between +/- 81° discrepancy from 'True South'. The direction moved by the agent is then Perceived South + Swerve; likewise (reversed) for 10 northbound migrations.

The robustness is apparent. It may be noted that some tracks, with a highly skewed perception of 'global South', make significant detours yet still manage to home in on the ultimate target. Runs (not shown here) where the agents were midway given large random shifts East or West — the equivalent of a storm disruption — were similarly robust. The results were so much better than anticipated that some explanation, with a reality check, was felt necessary.

Why might Swerve Zones be so effective?

Although it sounds very difficult to achieve pin-point accuracy in finding a destination after 1000s of kms, this is actually an overstatement of what is required. We are here only considering getting within some 'ballpark' distance after 1000s of kms of going 'vaguely' in the right direction. Let us put some scale on this, and illustrate with real examples.

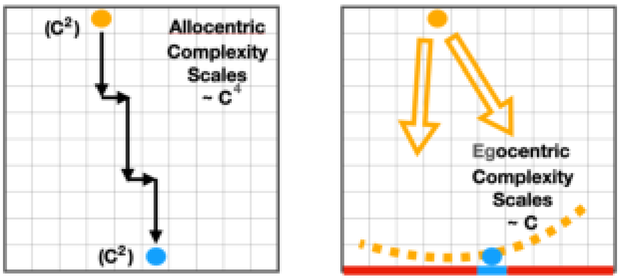


Figure 8. Complexity or specificity of map based navigation needing both start and end co-ordinates scales around O(C^4). Egocentric navigation can scale around O(C).

We need a scale S of size, eg in kms, and a scale C of complexity. Very crudely we can take S as the size across of the target destination -- the region wherein we assume some recognition of 'place' kicks in and different navigational strategies are available. We compare this S to the size across (eg diameter of an idealised circle) for the whole area available for potential migratory routes, and use the ratio as the scale measure S. So a chessboard might have a size measure S= 2cms (across each cell) and complexity measure C=8 (cells per side).

Traditional Micronesian navigators of outrigger canoes across an area of Pacific ocean 1000 km across have the concept of *etak* of arrival, entering the region where they can detect their destination atoll, through perhaps seeing a change in illumination of distant clouds, or through atoll-based birds flying to their daily fishing grounds. This can easily be 100 km across; S=100 km, C=10.

Many Monarch butterflies overwinter within an area of Mexico perhaps 120 km across (Taylor et al., 2020), and we might estimate their potential range across N. America as 4000 km; implying S=120km, C=30 to 40.

A map-based strategy for finding one's way to a destination might assume the difficulty of search scales something like C^4. On a chessboard of side C cells, discriminating where your start-cell is scales by C^2, and likewise for the end-cell. But the class of strategy we are considering is nothing like as complex: start anywhere on the North side of a chessboard, head very roughly south and see where you arrive on the south edge. Out of C possibilities, just one is your target. If we assume that arrival within the neighbourhood of the target will be recognised, whereupon some different short-range navigation strategies will operate, then the difficulty of the long-range navigation scales with C, not C^4 (Fig. 8).

So consider the example of Fig. 5, where some 20 dispersed potential Swerve Zones were available to be exploited. If we crudely consider the evolutionary choice for each SZ to be ternary, between ignored/green/red (neglecting the cue strengths) then the evolutionary search space is (underestimated at) size 3^{20}. But the great majority of these strategies will channel the noisy tracks of migrating agents into crossing the 'finishing line' within some fairly restricted region of x-values — since the Swerve Zones inevitably tend to converge such tracks *somewhere.* And the proportion of these converged tracks that are within range of the target (e.g. for Fig.7 within +/- 0.1) is going to be closer to 1 in 100 than 1 in 3^{20}.

There is the need for the northbound strategies to work as well as the southbound; but the natural symmetries make this relatively easy. With the benefit of hindsight, the framing of this navigational problem in terms of Swerve Zones made the search space computationally simple, and the immediate success of the evolutionary simulations is unsurprising.

(C) Parting the Waves

Our third example of 'navigating blind' is drawn from people. Among the greatest feats of human navigation, as with the migrating birds achieved without instruments or maps, has been traditional Micronesian and Polynesian wayfaring in small sailing outrigger boats across the Pacific ocean. Near to islands there are visible signs, perhaps clouds forming over land, or birds on their daily flights out at dawn to their fishing grounds and back at dusk to their land base. But far from land there are no such signs, perhaps for many days voyaging. At night with any clear sky, the stars provide directional cues for pilots without a magnetic compass; in cloudy weather even this is unavailable. One of the key navigation strategies used, distinctive to this Pacific voyaging, was the exploitation of patterns of interacting swells, as they are reflected and refracted around islands.

Such wave navigation skills were lost in Polynesia long ago, but lasted longer in Micronesia (Hutchins and Hinton, 1984). One of the last traditionally trained navigators in the Marshall Islands, Captain Korent Joel, before he died in 2017 passed on some of his experience to Western scientists (Genz et al, 2009; Genz, 2016; Huth, 2013; Huth, 2016). There remain several mysteries as to what wave patterns are being exploited, and how they are sensed by the pilot. The dominant primary swell provides a background against which secondary swells – reflections of the primary from individual islands – may play a key role (Fig. 9). A previous study (Harvey, 2018a, 2018b) offered speculations as to large-scale invariant pathways stretching all the way from the origin island to destination island, tracks potentially visible from satellite photos. In this study we take instead a purely local focus. A typical voyage is long distance between small islands, mostly with no land in sight – but what wave based cues are available locally to a pilot navigating blind on an isolated boat?

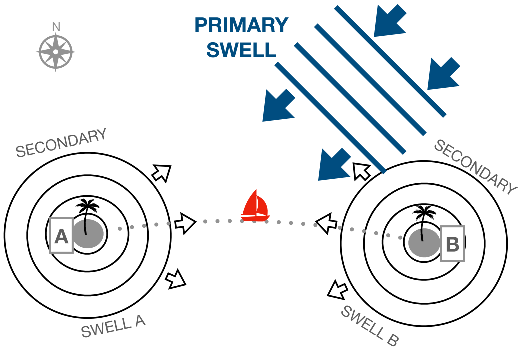


Figure 9. An idealised picture of a boat midway between two islands A and B, perhaps 100km apart. A Primary swell is shown from the NE. Weaker reflected Secondary swells of the same period radiate out from A, B.

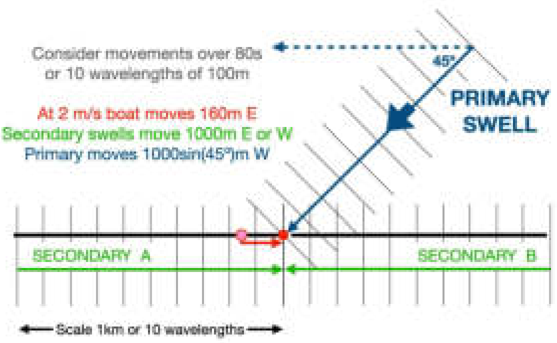

Consider movements over 80s
or 10 wavelengths of 100m
45°
PRIMARY
SWELL
At 2 m/s boat moves 160m E
Secondary swells move 1000m E or W
Primary moves 1000sin(45°)m W
SECONDARY A
SECONDARY B
Scale 1km or 10 wavelengths

Figure 10. The Primary and two Secondary swells, with the same periods, cannot be separately distinguished at a stationary location. But when the boat is moving (here E at 2m/s), the perceived frequencies separate and can be distinguished apart. Swell is here 0.125Hz, period 8s, wavelength 100m.

Roll, pitch and heave

We assume that the waves in the open ocean are dominated by a long slow primary swell that is fairly consistent in period and direction over several days. Such swells are often seasonal and predictable; or may perhaps arise from some severe weather event 1000 kms distant. Mixed with this will be any number of lesser swells and local wind waves, of different periods, amplitudes and directions; including crucially some secondary swells reflected from islands. We assume that, apart from some approximate idea of the direction of the primary swell relative to the boat, the only cues available are from the felt motion of the boat dancing on the waves. We can conceptualise this in terms of a smartphone with an internal accelerometer fixed flat on the deck aligned with the fore-aft axis. This can detect roll, around this fore-aft axis; pitch, around the port-starboard axis; and heave, the vertical acceleration as the boat rises and falls on the water. A smartphone (and a smart navigator) may also sense yaw (rotation about a vertical axis), surge and sway (linear acceleration along fore-aft and port-starboard axes); but first we show how far we can go with just roll, pitch and heave.

Whilst the boat is stationary on the ocean, the frequency of the primary swell should clearly stand out in the heave records sampled over a few minutes. For our examples in simulation we shall take this to be a sine wave at 0.125 Hz, i.e. with an 8 second period. An essential principle for deep water waves of period P (applicable when the water depth is more than half the wavelength W; Barber and Ghey, 1969) is:

$$W = P^2 g / 2\pi$$

where g is gravitational acceleration (9.8m/s^2). This means our primary swell with an 8s period has a wavelength of (nearly exactly) 100m, and a speed of 12.5 m/s.

A smartphone with its heave sensor could readily use a Fast Fourier Transform (FFT) to identify a peak in the frequency spectrum identifying this primary swell, and thus its wavelength. It is unlikely that the human brain directly implements an FFT algorithm, but we know of brain mechanisms that have somewhat equivalent functionality in the auditory domain. The cochlea in the inner ear is a spiral bony tube filled with fluid: as auditory vibrations pass along this, different regions resonate with different frequencies, and sensitive cells placed along the cochlea can identify where this occurs, effectively filtering as a mechanical FFT.

Ocean swell frequencies (0.02Hz to 0.2Hz) are on a very different scale to the auditory frequencies (20Hz to 20kHz) that the cochlea samples, but we assume that somehow they can be detected and recognised. We need to know the direction of a swell as well as its frequency, which is why pitch and roll must also be used. If we have heave, pitch and roll — filtered at the Primary swell frequency – in synchrony with each other, then the relative phase of pitch and roll indicates the Primary wave direction, If the boat rises to an approaching crest with bow rising together with the port side, then the wave is clearly coming onto the port bow (in Figs. 9 and 10, from the NE quadrant). All 3 sensors, roll, pitch and heave, are needed to make this discrimination.

In an ideal world, with no noise and a circularly symmetrical boat, comparing the amplitudes of pitch and roll would give a fine resolution for the direction within that quadrant. But the world is not ideal, pitch and roll operate differently because the shape of the hull affords different angular momentum about the different pitch and roll axes. The finest discrimination of swell direction, relative to the boat, is going to be when one axis of symmetry, e.g. the fore-aft axis, is facing directly into the swell; then the relevant roll should be zero, with any small deviation left or right showing up in some roll amplitude, in phase or anti-phase with the pitch.

Splitting the frequencies

What the navigator really wants to know is not the direction of the Primary swell, but rather the directions of much weaker Secondary swells; definitely from the destination island as the target to be aimed for, and perhaps also from the starting island so as to assess the track being made good and hence any possible side currents. These Secondary swells are weaker than the Primary as energy is lost in reflection, and they further dissipate with distance as they expand radially from their source island. The task will be to distinguish these very weak Secondary signals from the Primary. At the boat, the combination of Primary swell with such weaker secondary swells – with the same periods but different phases and

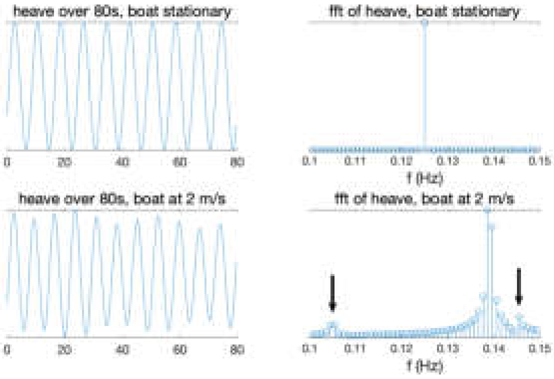


Figure 11. Top: heave experienced from combined swells at a stationary boat, and a Fast Fourier Transform showing the single peak at 0.125Hz. Bottom: when the boat moves E at 2m/s, the FFT now splits the frequencies, with the weaker secondaries at *FAM*=0.105Hz, *FBM*=0.145Hz.

directions – will (if they can be treated as sine waves) add up to one single wave with the same period. The sum will be not too different from the dominant constituent, the Primary swell (Fig. 11, Top left). From local sensors for roll, pitch and heave located at a single location, it is in principle impossible to disentangle the component swells from their sum.

A wide sensor array, spanning several 100m wavelengths, could in principle collectively discriminate between the different component swells. This is not feasible on a small boat of a few metres long. But just as the bacterium, with its Run-and-Tumble, uses motion to compensate for its lack of separated sensors, the moving boat can do likewise.

Fig. 9 presented a snapshot of a stationary boat amidst primary and secondary swells. Fig. 10 shows the difference sensed where the boat is moving, here at 2 m/s E over 80 seconds, which corresponds to 10 wavelengths. The boat will, having moved 160m, pass 10 - 1.6 = 8.4 waves of secondary swell A, and 10 + 1.6 = 11.6 waves from B in those 80s; hence registering their sensed frequencies at 0.105Hz and 0.145Hz respectively. Given that the primary swell is coming from the NE obliquely at 45°, the boat will pass 10 + 1.6*cos(45°) = 11.1314 primary waves in those 80s, for a sensed frequency of 0.1391Hz.

Thus the frequencies of Primary and Secondary swells are, as perceived from the moving boat, now split apart and distinguishable (Fig. 11) via differing Doppler effects. If one filters out the now altered Primary frequency, the Secondary swell frequencies for heave — and also pitch and roll — are made clear (Fig. 12). Any detectable roll will indicate that the boat's heading is not aligned with the desired course.

The Navigator's set of Procedures

So the algorithm to be performed, by smartphone program or navigator's trained intuition, is:

- With the boat stationary for several minutes, estimate the period of the Primary swell. In our example this is 8s, equating to a Primary-Frequency-Static *FPS*=0.125Hz, a wavelength of 100m, a wave speed of 12.5 m/s.
- Note which quadrant the primary swell comes from, and roughly estimate theta, the angle relative to boat direction.
- Then sail in the current direction as steadily as possible, assessing the experienced swells. Explicitly or implicitly do the equivalent of an FFT analysis, looking for frequency peaks each side of *FPS*. Identify the new observed Frequency of the Primary-in-Motion, here shifted by a Doppler effect to *FPM*=0.1391Hz.
- The shift implies that in 8s the component of the boat's motion in the primary direction is an extra (*FPM-FPS*)/*FPS* of the 100m wavelength. The boat's speed is unknown, but 8*speed*cos(theta) = 100*(*FPM-FPS*)/*FPS*, and hence speed*cos(theta) = 1.400
- We can use this to roughly estimate the speed, dependent on how well we have estimated theta. This in turn points to where in the frequency domain we can seek Frequency spikes (as perceived from the boat in Motion) for secondary swells A (*FAM*=0.105Hz) and B (*FBM*=0.145Hz).
- Once *FAM* and *FBM* are identified, they mutually confirm the earlier rough estimates, making them more precise. We note that (*FAM*+*FBM*)/2=0.125Hz=*FPS*, and (*FBM-FPS*)=0.020 multiplied by the wavelength of 100m gives us an accurate estimate of 2 m/s for the boat's speed.
- Hence from speed*cos(theta) = 1.400 we derive cos(theta) = 0.700 and thus theta=45.5°. Close enough to the true value of 45°, given the precision used in the intermediate steps.

Once the key frequencies *FAM* and *FBM* for the secondary swells (as perceived from the boat in motion) are identified, the remaining task is straightforward. Take either one, e.g. *FBM* for the secondary swell from island B, and filter out the pitch and roll signals around that frequency. Cross correlate these filtered pitch and roll signals, and positive or negative correlation indicates that the bow of the boat is left or right of heading directly into that oncoming swell.

Such an algorithm could be explicitly programmed into a smartphone secured to the deck of a boat. Did Micronesian navigators use comparable strategies? — a key indicator is that this strategy requires the observation from the *moving* boat over several minutes to 'Part the Waves', to allow the frequencies to be separately distinguished.

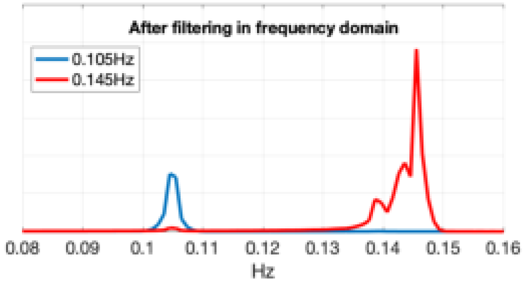


Figure 12. The boat's motion allows the (now differing) Primary frequency to be filtered out. The perceived Secondary frequencies can be clearly identified.

Similarities between the Examples

We can characterise what is common between these three examples – (A) Run-and-Tumble in bacteria, (B) Swerve Zones in models of large-scale seasonal bird migration, (C) Parting the Waves in proposals for explaining Micronesian wave-navigation – in that they illustrate strategies for wayfinding towards distant goals that are out of sight, using strictly local sensing.

Ambiguous location, resolved through motion

In all these examples, the locally available cues give little or no clues about the agent's current location — and this is not needed. Local cues also give (A) no clue as to the target location or (B, C) little clue, beyond the vaguest global sense of direction. The target's location only becomes eventually identified through motion towards the attractor of the navigational dynamics.

No dead reckoning

Any dead-reckoning strategy will break down if there is a single break in the chain of updates. In contrast, the strategies here are viable without any such need for a reliable *history*. They only need a basic relationship between *currently* available cues and *current* course to steer.

Proximal sensing

They may perhaps use a crude and approximate global sense of direction — where North and South lie (for migrating birds) or the rough direction of the primary swell (for

Micronesian pilots) — but none of these examples use distal sensing of the target such as vision provides; hence we describe this as 'navigating blind'. In the absence of motion, there are either no cues as to the target direction (bacteria and birds) or insufficient cues (Micronesians).

Move and orient, not orient and move

Often a navigation task is initially framed as a static problem, typified by the use of a map: first the appropriate direction from the current position, taken as static, is estimated. Only then is the next move initiated.

Our three examples all reverse this procedure, analysis is based on some initial default motion. In case (A), Run-and-Tumble, this can be in any random direction. In cases (B) and (C), such default motions can be imprecisely based on a general sense of direction, North or South for migrating birds, orientation to a primary swell for the wave navigator.

Cues for the wrong directions, not for the right one

In the absence of any cues, all these three example strategies direct a course of 'carry on as before'. It is only the *presence* of a cue that stimulates a change of direction: Tumble or Swerve or respond to the Roll. Such cue signals are not acting as *beacons*, as waypoints to aim for. Instead they act as *anti-beacons* to avoid: (A) decreasing cue strengths, (B) swerve zones or (C) roll cues at the identified secondary frequency.

Cue responses can be Sloppy, crude and basic …

Responses are limited to swerving left or right — and in the Tumble case it does not even matter which choice is made. The size of the swerve may fluctuate with weightings of the cues, but the simulations demonstrate that the strategies are tolerant to a wide range of such parameters. A slight nudge for as long as is needed will be as good as a strong brief nudge. Suppose an agent is consistently deviating by 45º off-course. Rather than heading directly for the target, it gently spirals in towards it. It might take some 41% longer along such a spiral, but safe arrival is still achieved and that is the main criterion.

… Which implies Easily Evolvable

Fragile complex systems pose a challenge for evolution, whether natural or artificial. In contrast the inherent sloppiness in all these three examples offers easy incremental evolutionary phylogenetic trajectories from even simpler precursors. (A) is already basic. With (B) just a couple of Swerve Zones offers navigation better than random, and complexity scales incrementally as extra zones are co-opted. Our simulations demonstrated easy evolvability. With (C), the secondary swells radiating from a destination island are weak from afar but easily discernible when close in. This suggests a learning trajectory for both evolution and a novice navigator.

Differences between the Examples

In the bacterial example (A), there is not a single target to be aimed for. A distributed layout of 'food' can take any form, with any number of dense areas; a Run-and-Tumble strategy is indiscriminate in its tendency to climb any local density gradient. No global sense of direction is needed. In contrast, both the bird migration (B) and the wave navigation (C) examples have identifiable point targets, with choices to be made: migrating N or S, navigating from island A to B or vice versa. In both cases some global sense of direction is required, whether from the sun or the known direction of the primary swell. But it appears that such a global sense of direction can be very approximate, since its role is little more than breaking symmetry between migrating NS and SN, between navigating AB and BA.

Both examples (B) and (C) have the effect of setting a direction to steer for, but in allocentric terms they differ in significance. With (C) this direction is indeed directly towards the destination island, or directly away from the island of departure; ideally the whole route would be a straight line. However with (B) this is not the case; as can be seen in Figures 4 and 5, multiple twists and turns may be expected.

Conclusions

Migrating birds and navigating sailors have plenty of different strategies in their toolbox, with different ranges of application. Our examples focus on just one subset, appropriate for long range navigation where both the location of the target and the location of the agent itself is not known, and the link between proximal sensing and steering choice can afford to be sloppy since it is self-correcting.

Though inspired by such natural examples, we have restricted ourselves here to Artificial Life models in abstract settings. The aim is to expand the range of possible explanations available to scientists studying these natural phenomena, and indeed offer new tools for robotic navigational studies. The principles underlying (B) Swerve-Zones and (C) Parting-the-Waves appear to be novel and are presented here for the first time.

A main theme is that navigational problems using only proximal cues may seem difficult or indeed intractable when viewed in a static framework. But adding in the motion of the navigator and then analysing their moving perceptions can transform these problems and make them radically simpler.

A key observation was how surprisingly easy it was to quickly evolve effective strategies for Swerve Zones -- at least in our idealized toy models. The real world is messier, so we should consider how this approach might scale. As the environment changes, e.g. through climate change, evolution must adapt to suit. The selection pressure on migration is severe, basically arrive-or-die, but fortunately it looks like this class of strategy scales really well.

The Parting-the-Waves scenario is at first sight a niche topic associated with ocean swells. But our natural senses of vision and hearing are based on wave phenomena. The principles behind trading time for space, afforded by all kinds of active perception, extend across these senses also. We have focussed here on 'navigating blind' in the sense of using strictly proximal sensory cues. But even distal sensing such as vision and hearing is ultimately based on local sensing within the eyes and ears; aided by legs and wings and sails.

References

Armitage J and Schmitt R (1997). Bacterial chemotaxis: *Rhodobacter sphaeroides* and *Sinorhizobium meliloti*: variations on a theme? *Microbiology* 143: 3671–3682.

Barber, M. F. and Ghey, G. (1969). *Water Wave*s. Wykeham Publications, London.

Egbert, M. D., Barandiaran, X. E. and Di Paolo, E. A. (2010). A Minimal Model of Metabolism-Based Chemotaxis. *PLoS Comput. Biol.* 6(12): e1001004. https://doi.org/10.1371/journal.pcbi.1001004

Genz, J., Aucan, J., Merrifield, M., Finney, B., Joel, K. and Kelen, A. (2009). Wave navigation in the Marshall Islands: Comparing indigenous and Western scientific knowledge of the ocean. *Oceanography* 22(2):234–245.

Genz, J.H. (2016). Resolving ambivalence in Marshallese navigation: Relearning, reinterpreting, and reviving the “stick chart” wave models. *Structure and Dynamics* 9(1). https://escholarship.org/uc/item/43h1d0d7

Harrison, A.-L., Woodard, P. F., Mallory, M. L. and Rausch, J. (2022). Sympatrically breeding congeneric seabirds (Stercorarius spp.) from Arctic Canada migrate to four oceans. *Ecology and Evolution* 12(1). https://doi.org/10.1002/ece3.8451

Harvey, I., (2004). Homeostasis and Rein Control: From Daisyworld to Active Perception. In Pollack, J., Bedau, M., Husbands, P., Ikegami, T., and Watson, R.A., editors, *Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems, ALIFE9*, pp. 309-314. MIT Press, Cambridge MA.

Harvey, I., Di Paolo, E., Wood, R., Quinn, M. and Tuci, E. A. (2005). Evolutionary Robotics: a new scientific tool for studying cognition. *Artificial Life*, 11(1-2):79-98.

Harvey, I. (2011), The Microbial Genetic Algorithm. In Kampis, G., Karsai, I., Szathmáry, E., editors, *Advances in Artificial Life: ECAL 2009*, Part II, LNCS 5778, pp. 126-133. Springer, Heidelberg. ISBN-13: 978-3642212826.

Harvey, I. (2018a). Parallel paths across the Pacific: a speculative explanation for the *dilep* in Marshallese navigation. arXiv: 1802.09151 [physics.pop-ph]

Harvey, I. (2018b). An Artificial Life perspective on traditional Pacific wave navigation. In Ikegami, T., Virgo, N., Witkowski, O., Oka, M., Suzuki, R. and Iizuka, H., editors, *Proceedings of the 2018 Conference on Artificial Life*. MIT Press, Cambridge MA: 359-360. https://doi.org/10.1162/isal_a_00067

Hutchins, E. and Hinton, G. E. (1984). Why the islands move. *Perception*, 13:629-632.

Huth, J. E. (2013). *The lost art of finding our way.* Harvard University Press.

Huth, J. (2016). Conclusions: a cross-disciplinary journey through spatial orientation. *Structure and Dynamics* 9(1):154-178.

Taylor, O. R., Pleasants, J. M., Grundel, R., Pecoraro, S. D., Lovett, J. P. and Ryan, A. (2020). Evaluating the Migration Mortality Hypothesis Using Monarch Tagging Data. *Front. Ecol. Evol.,* 07 August 2020. https://doi.org/10.3389/fevo.2020.00264

En route for implanting a minimal chemical perceptron into artificial cells

Pasquale Stano[1*], Pier Luigi Gentili[2], Giordano Rampioni[3], Andrea Roli[4,5], Luisa Damiano[6]

[1]Department of Biological and Environmental Sciences and Technologies, University of Salento; Lecce, Italy
[2]Department of Chemistry, Biology and Biotechnology, Università degli Studi di Perugia; Perugia, Italy
[3]Sciences Department, University of RomaTre; Rome, Italy
[4]Department of Computer Science and Engineering, Campus of Cesena, Università di Bologna; Cesena, Italy
[5]European Centre for Living Technology, Venice, Italy
[6]Università IULM; Milan, Italy

*pasquale.stano@unisalento.it

Abstract

This paper describes a potentially rewarding research program aimed at designing, modeling, analyzing and experimentally realizing artificial cells in the wetware domain endowed with a 'neural network'-like module for achieving minimal perception. In particular, we present a possible implementation based on bacterial phosphorylation signaling networks (dubbed as "phospho-neural network" by Hellingwerf and collaborators in 1995). At this initial stage only preliminary discussions are possible. The scenario we devise minimizes unrealistic assumptions and it is based on the state-of-the-art of contemporary artificial cell technology. This contribution is intended as a plan to forster the construction and the theoretical analysis of next-generation artificial cells.

Wetware Artificial Cells

Tremendous advancements in artificial cell (AC) research have been reported in the past years, thanks to the joint efforts of several communities (for example, the MaxSynBio, Build-a-Cell, BaSyC, FabriCell initiatives), attracted by the novel idea of constructing cell-like systems by assemblage processes. The latter differ from conventional synthetic biology because a "bottom-up" approach is employed. Complex cell-like structures are literally built from scratch by employing molecules such as DNA, RNA, ribosomes, enzymes, lipids, etc., or allegedly primitive molecules (fatty acids, ribozymes, short peptides, …), or even completely artificial molecules (amphiphilic molecules and/or polymers, *ad hoc* designed transition metal-based catalysts, organocatalysts, etc.). A distinctive element that characterizes bottom-up approaches to ACs is the focus on basic scientific questions (e.g., what is the minimal complexity for life? What is the minimal genetic information required to sustain cellular self-reproduction? Can we build a minimal cognitive chemical system?). Because AC technology does not resemble any other existing technology, the entire field is experiencing a momentum and a wide variety of results are continuously reported. It is possible to bet about the central role that AC technology could play in future, if properly developed in theoretical and applied science arenas.

Following the pioneering phase of the early 2000s, recent efforts have shown that it is possible to design and construct ACs that perform a variety of cell-like functions. It is beyond the scope of this abstract to summarize the state-of-the-art (interested readers can find more information in excellent recently published reviews [1-4]). For the present discussion it is enough to say that today it is possible to reconstruct several basic cellular functions, including those related to crucial mechanisms such as gene expression, DNA replication, signaling, transport, bioenergy generation, morphological transformations, small enzymatic pathways, transmembrane protein functions. Until now, most of these processes have been demonstrated one at a time, while one of the next challenging goals refers to the integration of these different "modules" into a whole, in order to reach higher complexity levels.

In this contribution, we will make one step forward, and describe a new research goal in AC research. We base our discussion on realistic expectations about what it will be possible to build in the near future, under the hypothesis that conditions will be found to allow the different coexisting "modules" of this hypothetical system work smoothly together.

A Wetware Embodied AI?

A fruitful scientific field for developing the present research plan is "Embodied AI": the area of AI that focuses on the role played by the body in cognitive processes, and thus uses, as synthetic models of natural cognitive systems, "complete agents" – i.e., agents that, differently from computers, have a body, and, through this body, can perceive and react to their environments, accomplishing cognitive tasks [5,6]. Since its birth in the late 80s, Embodied AI has been working on modeling cognitive embodied agents mainly through hardware models – electromechanical robots. Intrigued by the opportunity of developing Embodied AI in the *wetware* domain, we therefore asked whether, and to what extent, ACs can constitute a platform for the theoretical and experimental investigation of a "Chemical Embodied AI".

In this respect, there are different interesting modeling options. The first one derives from the fact that ACs are expected to be developed within systems-level frameworks, e.g., autopoiesis [7]. Accordingly, if ACs were autonomous, they would display a cognitive dimension [8,9], and then would represent very interesting models for Embodied AI. The problem with this approach is that current ACs do not yet realize the *organizational closure*, the prerequisite for autonomy [10]. A second compelling option can be conceived in terms of ACs that do not have an organizational closure in itself, but their modular organization is thought as a part of a larger organizational closure (whose construction must be approached stepwisely).

Here we consider the implantation of an upstream AI "module" to control gene expression inside ACs. In particular, Neural Networks (NNs) draw our attention. In AI, NNs are generally implemented in the logical domain of software [11]. Our main research question becomes: is it possible to devise chemical NNs (*chemical perceptrons*) implemented in the wetware domain and in particular in the field of Embodied AI? And, by means of modeling, what is the minimal NN complexity that would generate interesting behavior?

Phospho-Neural Networks

A brief review on the attempts of constructing chemical NNs can be found in [12,13]. Here we will move straight to the point that we consider of relevance for AC technology. Hellingwerf and collaborators, in 1995, have put forward a lucid discussion about interpreting the bacterial signaling systems (the so-called two-component systems, TCSs) as chemical NNs [14]. These authors actually called them "phospho-NNs", because their functioning is based on molecular phosphorylation.

TCSs enable bacteria to sense, respond, and adapt to their environments, letting the cell perceive chemical signals present in their surroundings. In a typical TCS, a membrane protein (sensor, S) with histidine kinase activity catalyzes its autophosphorylation in the presence of a stimulus. Next, the sensor is capable of transferring the phosphoryl group to a response regulator (R), which can then affect cellular physiology by regulating gene expression or by modulating protein activity (Figure 1a). This series of reactions can be interpreted as a transmission of information, from outside to the genome (and, in turn, to the profile of proteins present in the cell).

Generally each TCS transmits information in a linear way, i.e., the *i*-th signal is sensed by the *i*-th cognate sensor, which self-phosphorylates and in turn phosphorilates, later, the *i*-th cognate regulator, affecting one or more gene expression mechanism(s). To function properly, however, some TCSs involve convergent or divergent branched pathways. Moreover, and this is a relevant observation for our proposal, *crosstalk* between TCSs is also possible, at the level of sensors and/or regulators, leading to a NN-like system (Figure 1b). Normally, discussions about sensing and the molecular biology of well-functioning TCSs emphasize in which conditions crosstalk is reduced or eliminated. In the new perspective of implanting artificial TCSs inside ACs (Figure 1c), however, crosstalk becomes essential, and thus the interest goes to conditions for favoring and *controlling* its occurrence [15].

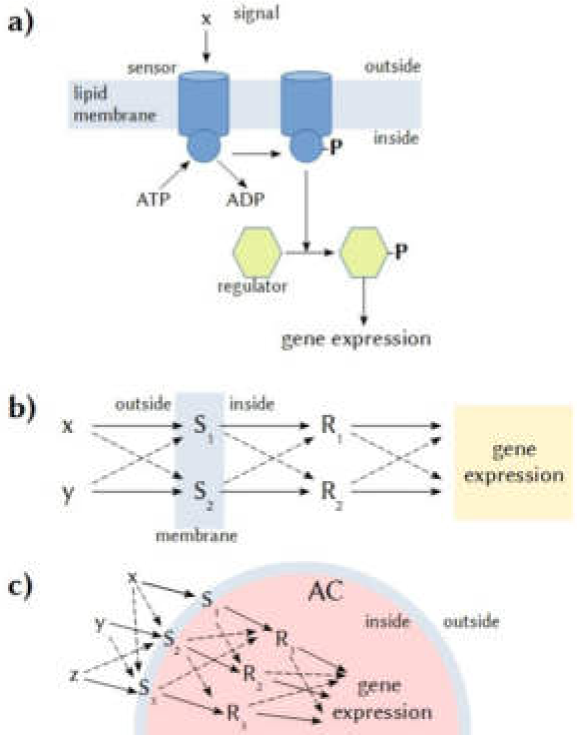


Figure 1. Phospho-NNs inside ACs. (a) The structure of a TCS; P represents the phosphate group. (b) When two TCSs crosstalk (shown dashed arrows), a NN-like system is generated. (c) Schematic drawing of a phospho-NN, based on three TCSs, implanted into an AC.

At this initial stage we identify two relevant goals, summarized as it follows. (1) Analysing the feasibility of intra-AC phospho-NNs, which will depend both on the right selection of TCSs and on the technical capability of building ACs endowed with the required molecular components. Critical elements are the transmembrane histidine kinase sensors, which should be reconstructed in functional form and in the desired orientation. Recent results on transmembrane protein reconstitution in ACs [16-18] constitute the starting point for design and experimentation. (2) With respect to numerical models, recognizing that chemical embodiment implies an intrinsic heterogeneity of AC physicochemical microenvironment. In turn, this generates the coexistence of many conformations (and activities) for the macromolecules constituting the phospho-NNs. The suggestion is that chemical perceptrons should be modeled by placing side-by-side both binary and not binary (fuzzy [19]) input/outputs.

References

[1] Robinson, A. O., Venero, O. M. and Adamala, K. P. (2021). Toward Synthetic Life: Biomimetic Synthetic Cell Communication. *Curr. Opin. Chem. Biol.*, 64:165–173.

[2] Olivi, L., Berger, M., Creyghton, R. N. P., De Franceschi, N., Dekker, C., Mulder, B. M., Claassens, N. J., Ten Wolde, P. R. and van der Oost, J. (2021). Towards a Synthetic Cell Cycle. *Nat. Commun.*, 12:4531.

[3] Shim, J., Zhou, C., Gong, T., Iserlis, D. A., Linjawi, H. A., Wong, M., Pan, T. and Tan, C. (2021). Building Protein Networks in Synthetic Systems from the Bottom-Up. *Biotechnol. Adv.*, 49:107753.

[4] Elani, Y. (2021). Interfacing Living and Synthetic Cells as an Emerging Frontier in Synthetic Biology. *Angew. Chem. Int. Ed. Engl.*, 60:5602–5611.

[5] Brooks, R. (1991). Intelligence without Representation. *Artificial Intelligence*, 47:139–159.

[6] Pfeifer, R. and Scheier, C. (1999). *Understanding Intelligence*. MIT Press, Cambridge MA.

[7] Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Reidel Publishing Company, Dordrecht.

[8] Ceruti, M. and Damiano, L. (2018). Plural Embodiment(s) of Mind. Genealogy and Guidelines for a Radically Embodied Approach to Mind and Consciousness. *Front. Psychol.*, 9:2204.

[9] Varela, F. J. (1979). *Principles of Biological Autonomy*. Elsevier North-Holland, Inc., New York.

[10] Damiano, L. and Stano, P. (2018). Synthetic Biology and Artificial Intelligence. Grounding a Cross-Disciplinary Approach to the Synthetic Exploration of (Embodied) Cognition. *Complex Systems*, 27:199–228.

[11] Rumelhart, D.E. and James McClelland (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA.

[12] Blount, D., Banda, P., Teuscher, C. and Stefanovic, D. (2017). Feedforward Chemical Neural Network: An In Silico Chemical System That Learns XOR. *Artif. Life*, 23:295-–317.

[13] Stano, P. (in press). Chemical Neural Networks and Synthetic Cell Biotechnology: Preludes to Chemical AI. In Chicco, D., Facchiano, A. and Mutarelli, M., editors, *Proceeding of the CIBB 2021 Computational Intelligence Methods for Bioinformatics and Biostatistics*. Lecture Notes in Bioinformatics, Springer.

[14] Hellingwerf, K. J., Postma, P. W., Tommassen, J. and Westerhoff, H. V. (1995). Signal transduction in bacteria: phospho-neural network(s) in Escherichia coli? *FEMS Microbiol. Rev.*, 16: 309–321.

[15] Agrawal, R., Sahoo, B. K. and Saini, D. K. (2016). Cross-Talk and Specificity in Two-Component Signal Transduction Pathways. *Future Microbiol.*, 11:685–697.

[16] Yanagisawa, M., Iwamoto, M., Kato, A., Yoshikawa, K. and Oiki, S. (2011). Oriented Reconstitution of a Membrane Protein in a Giant Unilamellar Vesicle: Experimental Verification with the Potassium Channel KcsA. *J. Am. Chem. Soc.*, 133:11774–11779.

[17] Altamura, E., Milano, F., Tangorra, R. R., Trotta, M., Omar, O. H., Stano, P. and Mavelli, F. (2017). Highly Oriented Photosynthetic Reaction Centers Generate a Proton Gradient in Synthetic Protocells. *Proc. Natl. Acad. Sci. U.S.A.*, 114:3837–3842.

[18] Amati, A. M., Graf, S., Deutschmann. S., Dolder, N. and von Ballmoos, C. (2020). Current problems and future avenues in proteoliposome research. *Biochem. Soc. Trans.*, 48:1473–1492.

[19] Gentili, P. L. (2018). The Fuzziness of the Molecular World and Its Perspectives. *Molecules*, 23:2074.

Modelling a Common Cognitive Bias and a Simple Heuristic to Overcome it

Michael Vogrin[1], Guilherme Wood[1] and Thomas Schmickl[1]

[1]University of Graz
michael.vogrin@uni-graz.at

Abstract

We emulated an experiment that shows the Einstellung-effect by building an agent-based model and developed a new heuristic that helps to overcome the effect.

Introduction

The "Einstellung-effect", also called the "Expertise Reversal Effect", is a phenomenon where a high degree of knowledge or expertise can decrease performance (Luchins, 1942; Bilalić et al., 2009; Kalyuga et al., 2012). This seems counterintuitive, as expertise in a domain usually increases performance in that domain (Vicente and Wang, 1998). The Einstellung-effect was first demonstrated in a series of experiments involving puzzles about measuring water using differently sized jugs (Luchins, 1942). Participants learned a specific method to solve problems, but when confronted with a similar problem that could not be solved using the specific method, they were unable to solve it. Naive participants, who did not go through the learning phase, were able to find the solution at a higher rate. Trained participants overlooked the solution because they had a certain mindset (German: "Einstellung") that prevented them from seeing it. Since then, the Einstellung-effect has been replicated in many different contexts (Levitt and Zuckerman, 1959; McKelvie, 1985, 1990), such as anagrams (Ellis and Reingold, 2014) and even magic tricks (Thomas et al., 2018), albeit the most studied context is arguably within the game of chess (Bilalić et al., 2010; Sheridan and Reingold, 2013).

One instance of the Einstellung-effect was shown in an experiment that we replicate for the study at hand (Goldstein and Gigerenzer, 1999). German and American participants where quizzed about which of two American cities had the higher population. It was observed, that German participants outperformed American participants when asked about American cities. The explanation for this surprising finding was that Germans often only knew one of the two American cities that were shown. They then picked the city that was familiar to them. Such reasoning is called "recognition heuristic" (Gigerenzer and Goldstein, 2011), and worked in this case. In contrast, Americans often knew both cities, were not able to use this heuristic, and consequently performed a little worse.

Our study aims at replicating this experiment in an agent-based model. We also present a simple heuristic, the "first contact heuristic" (FCH), that can mitigate this effect. We suggest that, given sufficient meta-cognition, Americans could have guessed the correct city by using this heuristic.

Self-Organization in Evolution

Methods

We designed an agent-based model that emulates the experiment done by Goldstein and Gigerenzer in 1999. Thus, we use terms such as "Germans" and "Americans" or "cities" to refer to "agents" and "objects that can be known". However, the model can be generalized to other domains.

One American and one German agent are generated and function independently from and analogously to each other. There are 25 German and 25 American cities. At each time step, agents have a 90% chance to learn about a city corresponding to their nationality (i.e. German agents learn about German cities, analogously for American agents), as well as a 10% chance to learn about other cities. Agents learn by saving the cities' ID in a list that serves as their memory. The cities have exponentially distributed random populations, and cities with higher populations are more likely to be learned by agents. After 100 repetitions of this process, agents are asked 25 questions about American and German cities each. Questions consist of two randomly drawn cities (from the same country) and for a correct answer agents must pick the city with the higher population. Agents can be in one of two groups: either *not* using the first contact heuristic (groups G1 and A1) or *using* the first contact heuristic (G2 and A2). G1 and A1 guess randomly if they have both or neither of the cities mentioned in a given question in their memory. If they have exactly one of the cities in their memory, they pick this city, using the recognition heuristic. G2 and A2 function exactly as G1 and A1, except that they do not guess randomly in cases in which they have both of the cities mentioned in a given question in their memory. In-

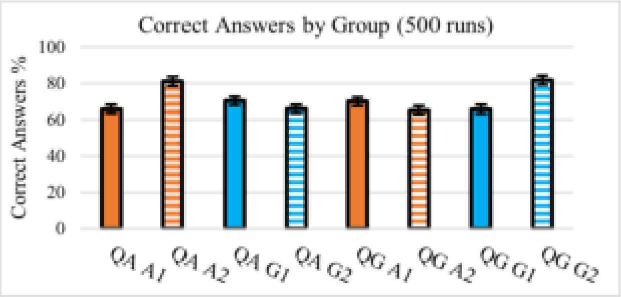


Figure 1: On a quiz about American cities (QA, left half) Germans (G1, orange) have higher scores than Americans (A1, blue). In agents using the FCH (striped), Americans (A2) have higher scores than Germans (G2). The results are reversed in the quiz about German cities (QG, right half).

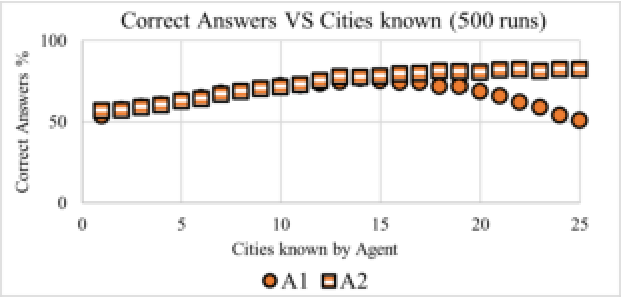


Figure 2: Group A1 (circles, no FCH) shows an optimum of correct answers at around 14 known cities, while group A2 (squares, using FCH) has increasing correct answers with increasing known cities before plateauing at 18 known cities.

stead, they pick the city which they learned about first, using the first contact heuristic.

Results

We report mean correct answers of German (G) and American (A) agents, without (1) and with the first contact heuristic (2) after 500 runs. In the first group (G1 and A1), Germans score higher than Americans in the quiz about American cities (QA), and vice versa. Agents using the first contact heuristic (G2 and A2) have higher scores, see figure 1. Mentioned differences were found to be statistically significant ($p < .00001$) using a Mann-Whitney U Test. Further, we report correct answers and known cities of American agents without (A1) and with (A2) the first contact heuristic, see figure 2. A1 shows a maximum of correct answers at 14 known cities, and a decreasing percentage of correct answers with further increasing known cities. A2 has increasing correct answers with increasing known cities. Groups G1 and G2 are not shown here but behave very similarly.

Discussion

In our study, a surplus of information prevents agents to use the recognition heuristic. Consequently, agents with less knowledge perform better, as they can overcompensate their relative lack of knowledge by using this heuristic. This is shown in figure 1, in which German agents have better scores on a quiz about American cities and vice versa. Beyond a certain point, more knowledge does not lead to more correct answers, see figure 2. Agents often recognize both cities in questions about their own country, but only one of the cities in questions about the foreign country. In the former case, they cannot use the recognition heuristic, since they recognize both. In the latter case, they choose the city that they recognize. The heuristic works well because the likelihood of hearing about a city is, in the real world as well as in the model, linked to its size: Larger cities usually are more well known. Our results replicate the “Less-is-More”-effect, shown by Goldstein and Gigerenzer 1999.

Given that, one can analyze situations where the likelihood of knowing a given entity (e.g. a city) is related to a property (e.g. the population size) as trade-off situations. A high amount of knowledge is beneficial, but comes with a cost: a lesser ability to use the recognition heuristic. In cases as the study presented here, the cost of knowledge beyond a certain point is higher than the benefit. To counteract this effect, we propose a new heuristic that needs no further external information but instead uses meta-cognition: agents using our “first contact heuristic” (A2 and G2) pick the city they heard about first. The recognition heuristic implicitly assumes that one is more likely to hear about larger cities. Analogously, our “first contact heuristic” assumes that one likely hears about larger cities *first*, and only later about smaller cities. In contrast to the recognition heuristic however, it can be used even if both cities are known. This is shown by the higher scores of A2 compared to A1 in the QA condition, as well as the higher scores of G2 compared to G1 in the QG condition, shown in figure 1.

The recognition heuristic can be useful, but fails when the probability of recognition is inversely related, or not connected at all, to the property in question: if one would be asked about which of two cities has the *smaller* population, and only recognizes one out of two options, then one should *not* pick the recognized option. Our results suggest that the recognition heuristic can be an effective method to draw inferences when used in the right situation. However, when it cannot be applied, the “first contact heuristic” can fulfill a similar role and is a valuable piece of the cognitive toolkit of simulated as well as real persons alike.

Acknowledgements

This study is supported by the Field of Excellence COLIBRI (Complexity of Life in Basic Research and Innovation) of the Karl-Franzens University of Graz.

References

Bilalić, M., McLeod, P., and Gobet, F. (2010). The mechanism of the einstellung (set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science*, 19(2):111–115.

Bilalić, M., McLeod, P., and Gobet, F. (2009). Specialization effect and its influence on memory and problem solving in expert chess players. *Cognitive Science*, 33(6):1117–1143.

Ellis, J. J. and Reingold, E. M. (2014). The einstellung effect in anagram problem solving: evidence from eye movements. *Frontiers in Psychology*, 5:679.

Gigerenzer, G. and Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, 6(1):100–121.

Goldstein, D. G. and Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In *Simple heuristics that make us smart*, pages 37–58. Oxford University Press, Oxford, UK.

Kalyuga, S., Rikers, R., and Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, 24(2):313–337.

Levitt, E. E. and Zuckerman, M. (1959). The water-jar test revisited: The replication of a review. *Psychological Reports*, 5(3):365–380.

Luchins, A. S. (1942). Mechanization in problem solving: The effect of einstellung. *Psychological Monographs*, 54(6):1–95.

McKelvie, S. J. (1985). Einstellung: Still alive and well. *The Journal of General Psychology*, 112(3):313–315.

McKelvie, S. J. (1990). Einstellung: Luchins' effect lives on. *Journal of Social Behavior and Personality*, 5(4):105.

Sheridan, H. and Reingold, E. M. (2013). The mechanisms and boundary conditions of the einstellung effect in chess: evidence from eye movements. *PloS One*, 8(10):e75796.

Thomas, C., Didierjean, A., and Kuhn, G. (2018). It is magic! how impossible solutions prevent the discovery of obvious ones? *Quarterly Journal of Experimental Psychology*, 71(12):2481–2487.

Vicente, K. J. and Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, 105(1):33.

Modeling the Cell as a Network of Parallel Processes—a New Approach

Margareta Segerståhl[1] and Boris Segerståhl[1,2]

[1]Bisari Research Institute, Kirkkonummi, Finland
[2]University of Oulu, Oulu, Finland
margareta.segerstahl@gmail.com

Abstract

This study addresses the problem of combining insights from artificial life, artificial intelligence, and biology in an efficient way to form a holistic unified view of life and living systems. Today, the study of biological life has a common *root object* – the cell – although lacking a formal definition of it. The theory of artificial life and artificial intelligence lacks this type of a root object. Here, we present a generalized model of the real biological cell in terms of a framework that is derived from theoretical studies of life. The framework is conceptualized generally as the MIC framework (Metabolism, Information, Compartment). The result is an autopoietic model with generic systemic properties and a network structure that can be examined further from a formal system-theoretic perspective. This study introduces a new way of describing the cell, providing new kind of access to existing biological knowledge of life. It may provide new tools for more efficient utilization of biological data and knowledge in the design and study of artificial life.

Introduction

It is widely acknowledged that the study of Artificial life (ALife) is not only about mimicking the design principles of biological life. The science of ALife is already moving in many new directions fully in its own right. Many of the new approaches are not directly tied to the basic principles, constraints, or to the material realm of biological objects. Research has proceeded well also in the absence of a formal definition or a basic concept that would link the applied field of ALife directly to biology and thereby to the natural foundation of all organismic life on earth.

However, as the result of the latest technological and computational advances, ALife is entering a new era. There is significant current interest for developing new kinds of life-like, physical machines that can interact intelligently and adaptively not only with humans, but also with many other kinds of biological life-forms. There is real need for new kind of understanding of biological life from an engineering perspective. This should preferably involve a formal integration of the wider perspective of living systems to the general picture in such a way that it can cover all instances (biological and artificial).

There is today no universal way of studying life using general formal approaches or models that would capture the essence of what a living system is. There is no general way of describing life by using a general methodology based on systems theory or mathematical methods that could offer a common perspective for scientists with different backgrounds. A general model should be applicable at least in the natural sciences and engineering. Abstracting biological behavior as computational behavior (Regev & Shapiro, 2002), reverse engineering of biological systems (Villaverde & Banga, 2014) or using principles of theoretical physics (Grenfell et al., 2006) have been suggested as possible ways forward.

A major theoretical problem is linked to the fact, that the biological sciences do not have a consistent formalism for describing the life of living cells or organisms in an abstract way. This is not a major problem among most biologists, but it hinders the possibility of using the full potential of currently existing biological knowledge in ALife research.

Theories of life

Many different theories and definitions try to capture the essence of life from a scientific perspective (for reviews see Cornish-Bowden & Cárdenas, 2020; Letelier et al., 2011). Explanations may derive from chemistry, mathematical study of life as abstract systems of organization and relations, cybernetics, molecular biology and more recently systems biology. Several of the theories focus on the circularity of metabolism and how it leads to metabolic closure at whole-system level. Notably, these theories have been developed independently, but there are similarities between them. Especially three of them attracted our attention and inspired us to conduct this study. The theory of *autopoiesis* (Maturana & Varela, 1980) Focuses on the ability of the living system to synthesize the components of which it consists of, while setting aside the question about the precise nature of the realizing mechanisms. Instead, it does highlight the importance of a physical separation of the system from its environment and how the living system can influence the properties of the environment (known as structural coupling). The *Chemoton* model was introduced by Tibor Gánti as an abstract model of a minimal chemical system that can be considered to be alive

(Gánti, 1975, 2003). It features three interconnected, cyclic processes that support metabolism, membrane synthesis and information processing. These are according to Gánti the three main properties that all living cells have in common. The *metabolism-repair (M,R)* systems approach was developed by Robert Rosen. It belongs to a field of biomathematics that is known as *relational biology,* established by Nicolas Rashevsky (Rashevsky, 1954). Rosen developed his theories over several decades, during which some of his views also changed as reflected in books that he wrote about life (Rosen 1985, 1991, 2012; reviewed in Pattee, 2007). The focus was on examining how a living system, that consists of components that have limited lifespans, needs to be internally organized in order for it to be able to (re)produce copies of all kinds of components that it consists of, thereby managing to repair and maintain itself against detrimental forces that would otherwise lead to the death and decay of the organismic entity.

Rosen emphasized the purely formal nature of his approach and did not connect it to any real-life examples. (*M,R*) systems have been contrasted against autopoietic systems and they are considered to form a more general class of systems that includes autopoietic systems as a subset (Letelier et al., 2003). Another study (Cornish-Bowden, 2015) has compared them against the Chemoton model. It ended with a general plea for the scientific community to form a holistic general synthesis of as many living systems theories as possible.

A new approach

In this study, we have tackled the problem of modeling life and living organisms conceptually by formulating a description of the real biological cell from a general systems perspective. We took as our starting point the one thing for which there seems to be a wide consensus—that *the biological cell is the basic unit of natural life*. We focused specifically on the question of how to describe the cell in a generic manner so that broad classes are included, taking advantage of existing theoretical views to life (reviewed above). This resulted in a general model of cellular biomolecular organization that is presented in the form a tailored network formalism. Our model provides an orderly foundation for the development of more rigorous mathematical formalism and understanding of the cell's biomolecular-level system-organizing properties, while remaining connected to the biological reality.

Overview of the MIC Framework

In the study of Alife and minimal modeling of living cells, there exists a certain general consensus, that for a cell or a cell-like system to be considered alive it must at least exhibit the following properties: *metabolism, information and compartmentalization* (e.g., Banzhaf & Yamamoto, 2015; Fellermann et al., 2007; Gánti, 1975; Rasmussen et al., 2016; Solé, 2009; Solé et al., 2007). We conceptualized these properties collectively for practical modeling purposes of this study as the *MIC properties of life*.

We used this conceptualization as a framework for considering what are the most relevant aspects of cell biology to be focused on. These topics form the main body of textbook literature on molecular cell biology and many details are known at the atomic scale of resolution (e.g., Alberts et al., 2014; Stryer, 1988). The outcome of this survey was, that despite the many differences that exist between the three main types of biological cells (archaeal, eubacterial, eukaryotic), one may postulate that there is a universal biomolecular foundation on which the cellular realization of the MIC properties is based on:

(1) Compartmentalization is provided by a cell membrane that separates the cell from the outside environment. The two main types of components found in all biological cell membranes are phospholipids and membrane-bound proteins. Structural assembly of the lipid fraction of the cell membrane depends universally on the physical behavior of phospholipid molecules in liquid water environment. Signal recognition particle (SRP) mediated mechanisms attach cell membrane proteins to the lipid fraction of the membrane. These mechanisms have an ancient basis and include protein and RNA components that are universal for all cells (Nagai et al., 2003).

(2) Cellular metabolism produces complex chemical compounds from simple raw materials. In all modern cells this is predominantly based on chemical reaction pathways and networks that are catalyzed by protein enzymes. All cellular protein strands are produced by the translation step of gene expression, involving ribosomes (complexes of rRNA molecules and ribosomal proteins), mRNA molecules that provide the genetic message to the site of protein chain synthesis, and tRNA molecules that carry amino acids to the site of protein chain synthesis.

(3) Chromosomal double-stranded DNA molecules are the universal physical carriers of genetic information in all living cells. Cells synthesize new DNA predominantly through the well-characterized mechanisms of semi-conservative DNA replication.

We set out to model this system from a holistic self-production perspective. The MIC conceptualization of life was included by considering these properties in a theoretical way as abstract components of a modeling space, in terms of which the most relevant aspects of cell biology for realizing these universal properties of cellular life could be described in a generic way. We divided this system-conceptual modeling space into thematic *partitions*, as described in table 1. Note, that the table lists four partitions. We extended the MIC paradigm by adding a partition named *Embodiment* to the bigger picture for practical modeling purposes. It accommodates the description of the main events of cell membrane assembly that all living cells have in common.

Furthermore, we adopted guidelines from the autopoietic theory of life (Maturana, 2002; Maturana and Varela, 1980) and set out to organize the model from the perspective of cellular material production pathways. The autopoietic theory describes living systems as self-producing entities that use simple raw materials that they acquire from the outside environment to synthesize the complex components from which they consist of. It recognizes the biomolecules of the cellular MIC properties as the main material foundation of living cells (Maturana, 1980). While the autopoietic theory is well recognized in ALife research (e.g., Beer, 2004, 2015; Ikegami and Suzuki, 2008; Masumori et al., 2020; Suzuki and Ikegami, 2009), it is not part of the current description paradigm of molecular cell biology in the biological literature.

Results

Our model, shown in figure 1, provides a general way of describing the biological cell from a holistic living systems perspective. The model refers to general aspects of molecular cell biology that all living cells have in common. It provides a visual representation of the type of system-level network connectivity that organizes the production of the material components on which the cellular MIC properties are based on.

The layout of the network is a particularly important aspect of our model and clarity of the visual representation was of specific interest to us (Polančič and Cegnar, 2017). We have arranged the model so that it is organized as a process flowchart. This gives it an appearance that is structurally very consistent even though the biochemistry that it describes is very complex. The network contains substance nodes (rounded) and process nodes (rectangular) that are connected by directed arrows. The nodes were defined and organized relative to each other so that there is a regular alternation of the two types. Solid arrows indicate a formal direction of process flow from the perspective of complexity increase of the material products.

Going through the network in the direction of the arrows, increasingly more complex organizational states of matter occur. Each type of product is formed from components that were produced by the events of the previous process step(s) of the network. The type of products that are formed have functional roles in the network. They are needed as components of the machineries through which the process node events are operated in real cells. Based on this, the model includes dotted directed arrows that describe how the core types of components that are formed by cells (blue nodes) effectively form a system-wide control network organization via the type of products that they form. Connection to real biology is maintained by labeling nodes using biological or biochemical terminology. Nodes are also numbered to enable easy referencing of specific parts of the model.

The abstract modeling space, which is divided into four system-conceptual partitions, is an important constituent part of our model. It makes it possible to assign higher-level system-related attributes to the characterization of the type of material components of cell biology that form the network. Each network components has a location in the system space relative to the partitions that provide system-level coordinates for how the components are distributed across the abstract system-space that represents the MIC properties of life. For example, the cell membrane (node 22) is formally localized to the C-partition as the main type of substance of molecular cell biology that provides a physical barrier between the cell and its environment. At the same time we know, however, that the membrane also has many channels (formed by protein molecules) and these are needed for the membrane to be able to realize its system-level function of providing raw materials from the environment to the living cell (node 2). In the model, node 2 is also a higher-level connecting element that joins the C and M-partitions together. In a similar manner, the 20 different kinds of natural amino acids are referred to by one substance node (node 7). On one hand, the node represents products that are formed by cellular metabolic reaction events (node 4) and that are used by the cell to synthesize protein strands (node 10). On the other hand, node 7 has a systemic role in our theoretical abstraction of the cell. It is an important connecting element at the description level of the abstract partitions, with a specific location in the intermediate system space between the M and the I-partition of the cell model.

Thematic partition	Heuristic description
Metabolism (M)	The biochemical reactions that distribute energy and matter in the cell, producing biomolecular chemical substances.
Information management (I)	The material components that contain the genetic hereditary information of the cell, and the cellular components and events that process it for cellular usage and storage.
Compartment (C)	A compartment defines the cell as an object in space, establishing a barrier between the cell and its physical environment, influencing the flow of energy and matter into and out of the cell.
Embodiment (E)	The mechanisms through which the cell's biomolecular components come to share a system-wide interface, that connects them to the external environment and gives them a holistic existence as an integrated cell-level entity.

Table 1: Theoretical partitioning of molecular cell biology into operating units, that together provide the kind of organization for the cell that formally (in the light of the autopoietic theory and the MIC properties of life) makes it a living system entity.

Our model also clearly shows, that real molecular cell biology of the I-partition is organized as an autocatalytic subset of network functions and process events: DNA is used as the molecular template both for the synthesis of DNA during DNA replication and for the synthesis of RNA strands during gene expression; RNA molecules participate in DNA synthesis as so-called primers and in protein synthesis as mRNA, tRNA and rRNA molecules; Proteins are needed as components of the biomolecular complexes that form the cellular machinery for the synthesis of new DNA, RNA, and protein strands. These functional connections, formally operating within the theoretical I-partition of the model, are visualized by directed dotted arrows. Autocatalytic sets (Kauffman, 1993) are an important theory for the modeling of gene regulatory networks and chemical-level reaction networks of cell metabolism. This interesting network motif emerges here at a high level of abstraction. It is a real organizational property of molecular cell biology that can be clearly seen in our model where chemical reaction events and many intermediate steps are abstracted away.

It is possible to also consider other kinds of ways of describing our network model. Figure 2a shows a version of the model where the system-wide network organization, including the

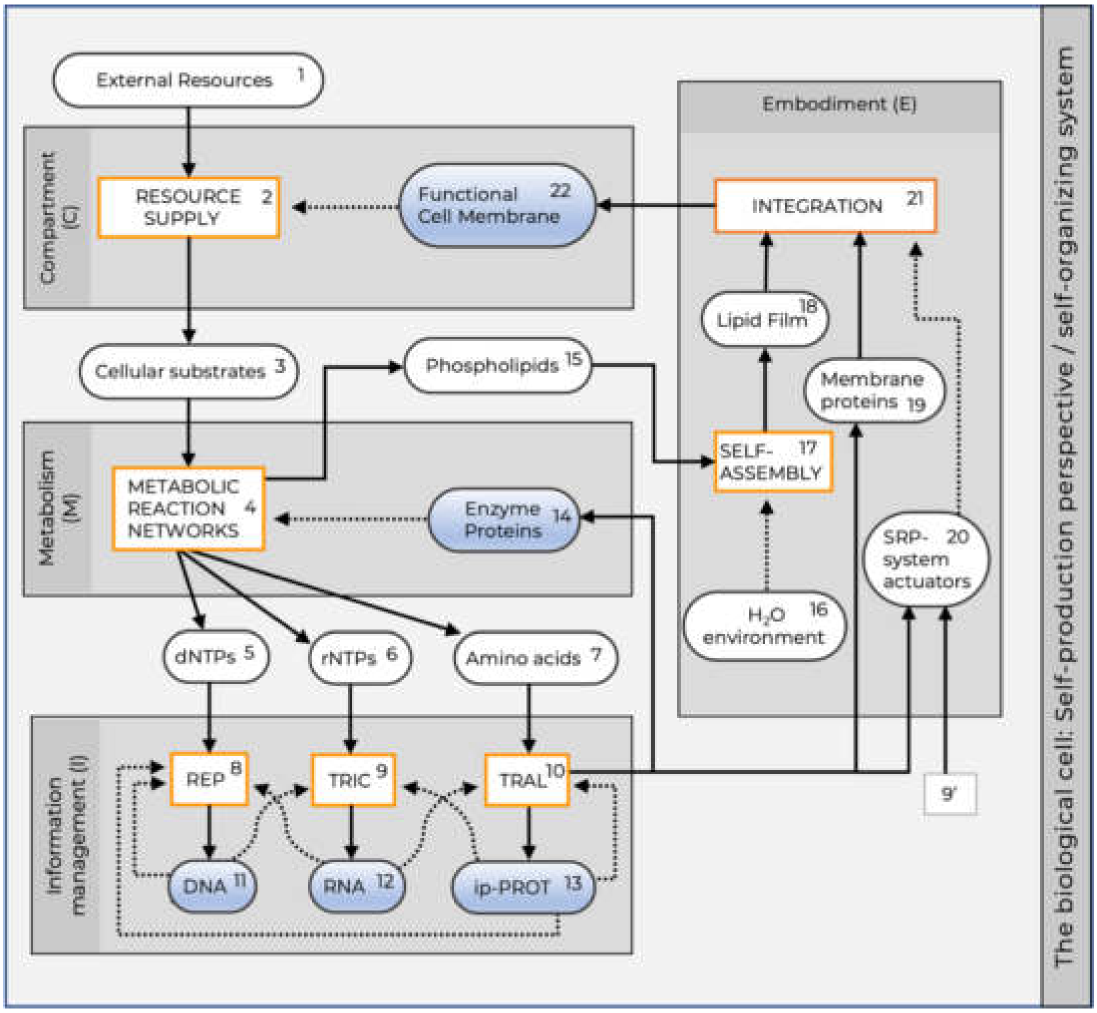


Figure 1: Diagrammatic model of the cell that describes it from a holistic perspective as a network of parallel processes. The description is given from an unconventional material production perspective, motivated by the autopoietic systems view that living cells produce the complex biological material components from which they consist of. The grey areas in the background describe an abstract modeling space and its system-theoretic partitioning into four parts (see table 1). The network describes knowledge that can be found in current textbooks of molecular cell biology. It features key aspects of molecular cell biology that all living cells have in common regarding the production of the main types of material components on which the cell's physical existence formally as a MIC (Metabolism, Information, Compartment) system entity is based on. The model is given in the form of a process flowchart with regularly alternating substance nodes (rounded) and process nodes (rectangular). Blue nodes indicate the main types of complex biomaterial products. Additional network elements are solid arrows that show the direction of process flow and how the different types of products and their system-wide distribution across the abstract modeling space of the thematic partitions. Dotted arrows indicate functional contribution coming from the type of material components that are formed for the execution of the target node events. Nodes are connected to the biological reality by the labeling. Numbering (1-22) enables easy referencing of specific parts of the model. dNTPs = the four types of deoxyribonucleotides, the monomeric substrates for DNA strand synthesis. rNTPs = the four types of deoxyribonucleotides that are the monomeric substrates of RNA strand synthesis. Node 7 refers to the 20 natural amino acids. REP = the process of semiconservative DNA replication. TRIC = the process of RNA synthesis during transcription step of gene expression. TRAL = the process of ribosomal protein synthesis during the translation step of gene expression. ip-PROT = **i**nformation **p**artition proteins, the set of proteins that are involved in the molecular mechanisms of the process node events of the I-partition. 9' is a duplicate of node 9 (replacing a solid arrow to enhance visual clarity of the diagram). The network components of the I-partition (nodes 8-13) form an autocatalytic subset in real biological cells. The dotted link 12→10 covers the contribution of mRNA, tRNA as well as rRNA molecules to the mechanisms of cellular protein synthesis. SRP = signal recognition particle dependent system for attaching cell membrane proteins to the lipid fraction of the membrane (see Nagai et al. 2003 for additional information). H_2O = reference to liquid water as the universal medium in which cells exist as physical entities, and as the main driving force behind biological self-assembly of lipid films.

self-referential aspects, can be seen more clearly. Figure 2b provides another kind of version of the model, where only the system-wide network connectivity that emerges at the highest possible description level of the MIC partitions is shown.

Discussion

Our model describes general knowledge of molecular cell biology in a new way by reorganizing it according to existing theoretical viewpoints to life and living systems. The conventional way of describing the cell from a holistic perspective is by using images that reflect the visual appearance that the cells have under the microscope as concrete physical objects. There is currently no universal model in general scientific use, that would describe the cell in a generic way. Thus, our model is a significant attempt to change this situation and provide new kind of access to biological knowledge about the system-level aspects of the natural biomolecular organization of the cell—the basic unit and (using our terminology) the formal *root object* of natural life.

An important design principle of graphical network representations is to keep the number of network components (especially the links) as low as possible (Polančič and Cegnar 2017). This provided a modeling constraint that influenced the selection of appropriate levels of abstraction and the choice of terminology that was used for labeling the nodes especially in those parts of the system where the amount of detailed biochemical knowledge is particularly high. Each node in our model captures a level of description that is at the same time sufficiently detailed as well and as general enough for the modeling purposis of this study (judged by the overall purpose to form a holistic, yet reasonably simple general model for the description of natural biomolecular organization of living cells from the MIC system perspective).

Our model integrates several viewpoints that exist for the theoretical study and modeling of life and living organisms. It describes the cell as an autopoietic system (that can synthesize the complex components from which it consists of), organized as a MIC system entity (a reflection of the Chemoton theory), and one part of the model is organized as an autocatalytic set. The (*M,R*) system formalism is another framework that is of interest to ALife research.

The model that we have provided features interactions and parallel mappings of process events and the resulting network has a system-level functional role in connecting distinct parts of the abstract cell-system modeling space (the partitions shown in Figure 1). These properties connect it to the theoretical domain defined by the (*M,R*) systems and relational biology. (*M,R*) systems focus on the functional relationships of components and products that together form a living system that can (re)produce the type of components that it consists of. Based on Letelier et al. (2003), (*M,R*) systems in deal with concepts such as *input* materials, that are transformed into *output* materials, that include *catalysts* that operate process events of material production, *components that select* for the synthesis of (biologically) *meaningful products*, and an agent that is referred to as the *efficient material cause* because it defines what products are meaningful for the organism. Two specific problems have been identified regarding the modeling of cells from the (*M,R*) systems perspective: (1) (*M,R*) systems do not account for the cell membrane, yet it is unrealistic to imagine real cells that have no cell membrane and (2) (*M,R*) systems seem to lack the ability to capture information processing aspects of living cells (Cornish-Bowden, 2015). However, Rosen has described an outline for dealing with properties of cellular information. He defined three categories of information—genomic, phenotypic and environmental—that are not equivalent which has consequences for the possibility of extrapolating results from formal models into the concrete realm of physical realizations (Rosen, 1986).

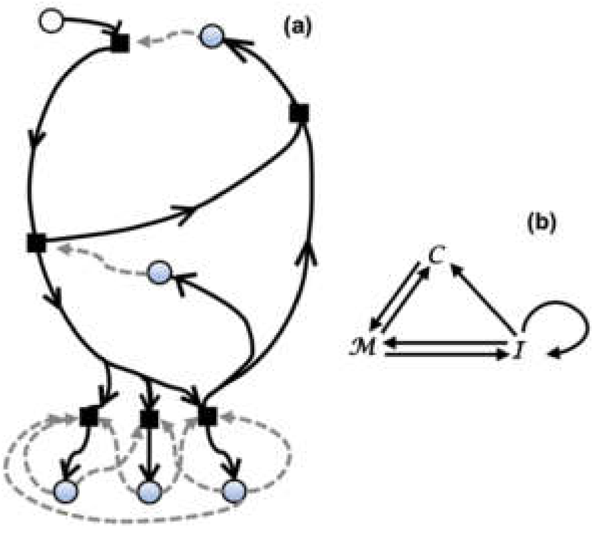


Figure 2: Two variations of our system-theoretic holistic cell model (see figure 1 for the full description). (a) A short-hand version of the model that highlights the system-level process flow aspects for the synthesis of the complex material products from which the cell consists of. Circles represent substance nodes and squares represent process nodes. The blue nodes correspond to the blue nodes of figure 1. The autocatalytic set of the information management (I-partition) components forms the bottom part of the graph. (b) Description of system-level network closure that arises at the abstraction level of the systemic cell partitions (see table 1). C, M, I = Compartment, Metabolism, Information management. This is an ultimate reduction of the model into a form that can be analyzed, for instance, in the light of network motifs of complex networks (Milo et al. 2002). The diagram shows how the three partitions support each other, in addition to which the information management partition also produces the types of material components from which it consists of (shown by the self-referential link I→I).

Based on our model, we divide the problem of applying the (*M,R*) system formalism to the modeling of living cells into two parts. *First* we focus only on the information management part of the model (the I-partition) and consider how (*M,R*) system viewpoints can be connected to the knowledge of molecular components and process dynamics of molecular cell biology that are associated with this part of the network. For modeling purposes, the rest of the cell can then be viewed relative to I-partition as byproducts. The result is a bipartite view of cellular organization. *Then* attention is shifted to the whole-cell level and to the reciprocal mutual inter-dependences that exist both between as well as within these two theoretical parts of cellular system organization (the I-partition versus the rest of the cell, see figures 1 and 2a). This is an even higher-level systemic view to the organization of the living cell than the one provided by the four partitions (figure 1). It assigns a certain kind of asymmetry to the biological organization of the living cell, where the rest of the cell can be considered to constitute an acquired environment for the components and processes of the I-partition.

But all four parts (compartment, metabolism, information management and embodiment) are needed to form a real living cell. Each of them depends on the other parts for the supply of input materials as well components of the molecular machinery that is needed for realizing the process node events assigned to them. They support themselves by supporting each other, and all this together supports the life and survival of the cell as a holistic, autopoietic MIC system entity.

Our modeling formalism combines theoretical viewpoints and descriptions of reality in a way that may allow expanding the use of formal methods of systems biology (Machado et al., 2011; Szallasi et al., 2006) to the whole-cell level. The (*M,R*) systems theory has already been brought to the attention of systems biologists (Gatherer & Galpin, 2013). By examining from the holistic perspective of living systems theories what is already known about natural cells and molecular cell biology, important general systemic properties of higher-level organizational aspects of living systems may be discovered and properly characterized.

A particularly interesting topic for future study is to conduct a system-theoretic analysis of the higher-level structural and dynamic properties of the material components and interaction mechanisms that form the autocatalytic set of the I-partition in real cells. This is a nonconventional way of studying the mechanism and events of cell biology that form the foundation for the cell's information related properties and genetic inheritance.

Our model contains feedforward and feedback loops and the general structure that we have presented in this study resembles the network structures of communication networks and control systems (Machado et al., 2011; Milo et al., 2002). These aspects can be studied further from the perspective of complex networks (see, e.g., Basler et al., 2016; Liu & Barabási, 2016). The system dynamics of this network in the concrete physical realm are influenced by the recycling and repurposing of the type of materials that are formed. A related issue is to find good ways of presenting energetic and thermodynamic aspects of real cellular life in relation to our model. We aim to extend our modeling approach with a method(ology) for including systems biological knowledge of gene regulatory networks and metabolic pathways. An eco-evolutionary framework needs to be developed for studying evolution of life from the living systems perspective using our model as a connecting element between theory and reality of life's complexity evolution—After all, living cells are evolutionary adaptive entities and can only fully be understood in relation to their living environment. We also need a way to describe cellular reproduction as a key operation that the components perform together as an organismic whole. We continue our search for practical ways of linking further living systems viewpoints, such as symbiosis (King, 1977), cooperation (Stewart 2019) and hypercycles (Eigen & Schuster, 1977, 1978a, 1978b), to our model.

Conclusion

The MIC model describes parallel processes and flows in a generic cell. It combines several existing theories of life. It is possible to extend the model to a broad class of biological and life-like systems.

Acknowledgements

We are grateful to Kimmo Kaski, Wolfgang Banzhaf, Robin Dunbar and Juan Carlos Letelier for their influential scientific discussions with MS that encouraged her to take on this modeling task, to Sinikka Segerståhl for sharing her knowledge of the complexity of language and its usage and to Katarina Segerståhl for many discussions on the complexity of human organizations. JCL made important comments to an early version of figure 1 and suggested adding a distinct fourth part to the description of the abstract modeling space to accommodate the process events of cellular self-encapsulation through cell membrane assembly. KS designed the visual details of figure 1.

References

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2014) *Molecular Biology of the Cell,* 6th edn, Garland Science, New York, NY.

Banzhaf, W., & Yamamoto, L. (2015). *Artificial Chemistries*. MIT Press.

Basler, G., Nikoloski, Z., Larhlimi, A., Barabási, A. L., & Liu, Y. Y. (2016). Control of fluxes in metabolic networks. *Genome research*, *26*(7), 956-968.

Beer, R. D. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, 10(3), 309-326.

Beer, R. D. (2015). Characterizing autopoiesis in the game of life. *Artificial Life*, 21(1), 1-19.

Cornish-Bowden, A. (2015). Tibor Gánti and Robert Rosen: contrasting approaches to the same problem. *Journal of Theoretical Biology*, *381*, 6-10.

Cornish-Bowden, A., & Cárdenas, M. L. (2020). Contrasting theories of life: Historical context, current theories. In search of an ideal theory. *BioSystems,* 188, 104063.

Eigen, M., & Schuster, P. (1977). The hypercycle: a principle of self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, *64*(11), 541-565.

Eigen, M., & Schuster, P. (1978a). The hypercycle: a principle of self-organization. Part B: The abstract hypercycle. *Naturwissenschaften*, *65*(1), 7-41.

Eigen, M. & Schuster, P. (1978b). The hypercycle: a principle of self-organization. Part C: The realistic hypercycle. *Naturwissenschaften*, *65*, 341-369.

Fellerman, H., Rasmussen, S., Ziock, H. J., & Solé, R. V. (2007). Life cycle of a minimal protocell—a dissipative particle dynamics study. *Artificial Life*, *13*(4), 319-345.

Gánti, T. (1975). Organization of chemical reactions into dividing and metabolizing units: the chemotons. *BioSystems, 7*(1), 15-21.

Gánti, T. (2003). In Szathmáry, E., Griesemer, J. (Eds.) *The principles of life*. Oxford University Press, Oxford.

Gatherer, D., & Galpin, V. (2013). Rosen's (M, R) system in process algebra. *BMC systems biology*, *7*(1), 1-11.

Grenfell, B. T., Williams, C. S., Björnstad, O. N., & Banavar, J. R. (2006). Simplifying biological complexity. *Nature Physics*, *2*(4), 212-214.

Ikegami, T., & Suzuki, K. (2008). From a homeostatic to a homeodynamic self. *BioSystems, 91*(2), 388-400.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA.

King, G. A. M. (1977). Symbiosis and the evolution of prokaryotes. *BioSystems*, *9*(1), 35-42.

Letelier, J. C, Marín, G., and Mpodozis, J. (2003). Autopoietic and (M, R) systems. *Journal of Theoretical Biology, 222*(2), 261-272.

Letelier, J. C., Cárdenas, M. L., and Cornish-Bowden, A. (2011). From L'Homme Machine to metabolic closure: steps towards understanding life. *Journal of Theoretical Biology*, 286, 100-113.

Liu, Y. Y., & Barabási, A. L. (2016). Control principles of complex systems. *Reviews of Modern Physics*, *88*(3), 035006.

Machado, D., Costa, R., Rocha, M., Ferreira, E., Tidor, B. and Rocha, I. (2011). Modeling formalisms in systems biology. *AMB Express*, 1(1), 1-14.

Masumori, A., Sinapayen, L., Maruyama, N., Mita, T., Bakkum, D., Frey, U., Takahaski, H. and Ikegami, T. (2020). Neural autopoiesis: Organizing self-boundaries by stimulus avoidance in biological and artificial neural networks. *Artificial Life, 26*(1), 130-151.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing, 9*(3-4), 5-34.

Maturana, H. (1980) The introduction in: Maturana, H. R. and Varela, F. J. (1980) Autopoiesis and Cognition, *Boston studies in the Philosophy of Science*, Vol. 42, pp. ix – xxx, especially the figure on page x.

Maturana, H. and Varela, F. (1980) Autopoiesis and Cognition. *Boston Studies in the Philosophy of Science*, Vol. 42

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science, 298*(5594), 824-827.

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., and Jovine, L. (2003). Structure, function and evolution of the signal recognition particle. *The EMBO Journal, 22*(14), 3479-3485.

Pattee, H. H. (2007). Laws, constraints, and the modeling relation–history and interpretations. *Chemistry & Biodiversity*, *4*(10), 2272-2295.

Polančič, G., & Cegnar, B. (2017). Complexity metrics for process models–A systematic literature review. *Computer Standards & Interfa*ces, 51, 104-117.

Rashevsky, N. (1954). Topology and life: in search of general mathematical principles in biology and sociology. *The bulletin of mathematical biophysics*, *16*(4), 317-348.

Rasmussen, S., Constantinescu, A., and Svaneborg, C. (2016). Generating minimal living systems from non-living materials and increasing their evolutionary abilities. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1701), 20150440.

Regev, A. & Shapiro, E. (2002). Cellular abstractions: Cells as computation. *Nature, 419*(6905), 343-343.

Rosen, R. (1985). *Anticipatory systems,* Pergamon Press, Oxford, New York, Paris.

Rosen, R. (1986). On information and complexity. In Casti, L.J. & Karlqvist, A. (Eds.) *Complexity, language, and life: Mathematical approaches* (pp. 174-196). Springer, Berlin, Heidelberg.

Rosen, R. (1991). *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press.

Rosen, R. (2012). *Anticipatory systems,* Springer, New York, NY. second edition.

Solé, R. V. (2009). Evolution and self-assembly of protocells. *The International Journal of Biochemistry & Cell Biology*, *41*(2), 274-284.

Solé, R. V., Munteanu, A., Rodriguez-Caso, C., & Macía, J. (2007). Synthetic protocell biology: from reproduction to computation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1486), 1727-1739.

Stewart, J. E. (2019). The origins of life: the managed-metabolism hypothesis. *Foundations of Science*, *24*(1), 171-195.

Stryer, L. (1988). *Biochemistry*. Freeman, New York, NY, third edition.

Suzuki, K., & Ikegami, T. (2009). Shapes and self-movement in protocell systems. Artificial Life, 15(1), 59-70.

Szallasi, Z., Stelling, J., & Periwal, V. (Eds.) (2006). *Systems Modeling in Cell Biology, From Concepts to Nuts and Bolts*. The MIT Press.

Villaverde, A. & Banga, J. (2014). Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of the Royal Society Interface*, 11(91), 20130505.

Inside looking out

Fernando Rodriguez

University of Sussex
fr97@sussex.ac.uk

Abstract

One of the defining, foundational axes of enactivism was its emphasis on the necessary relation between cognition and phenomenological experience, assumedly rooted on the particular, organizationally recursive nature of autonomous systems. However, in spite of many advances, there is no conclusive understanding about the emergence of this experiential dimension yet; a conundrum that has lead to contrasting positions within the framework. In this context, we suggest that an enactive, not fully committed interpretation of ideas from the Integrated Information Theory of Consciousness (IIT) may result fruitful; In particular, the formal notions of intrinsic information and integration as indicative of an intrinsic perspective and emergence respectively.

Autonomy and phenomenology

Maturana and Varela (1973) claim that, given their recursive nature, autopoietic systems possess an intrinsic *identity* underlied by the autonomous subordination of structural changes to the preservation of their organization; determining an independent and self-contained (biological) phenomenological subdomain (Maturana and Varela, 1973, p.69,p.110). In later work, and in order to account for a general notion of *autonomy* beyond cellular specificities, Varela (1979) introduced the formal concept of *organizational closure*, characterizing autonomous behavior in terms of operations within a self-referential space of transformations.

In *The Embodied Mind* (Varela et al., 1991) autonomy is characterized in enactive terms, in light of an embodied and embedded view on cognition. This is illustrated with Bittorio, a minimally autonomous cellular automata described as capable of *bringing forth a domain of significance* by selectively enacting external regularities (hence, displaying cognitive behavior), but considered obviously devoid of experience (e.g. in contrast to color perception) (Varela et al., 1991, p.150-157). This gap is the crux of the matter; why/how some varieties of autonomy would entail phenomenological experience and not others? While Varela (1997) claims that is the *autonomy of living systems* that provides a referential perspective for meaning and intentionality in phenomenological terms. And that experience, as for us, results from the particular nested and sensorimotor nature of the autonomy of the nervous system; there is still no decisive justification for this assumption. This has lead to, very broadly speaking, *traditional* positions (Di Paolo, 2005; Egbert and Barandiaran, 2014; Barandiaran, 2017; Di Paolo et al., 2017; Froese and Taguchi, 2019), more or less *radical* positions (Hutto and Myin, 2012, 2017; Abramova and Villalobos, 2015; Villalobos and Silverman, 2018), and other approaches leaning towards more representational explanatory stances (Clark, 2016; Seth and Tsakiris, 2018).

Formalizing autonomy

For an autonomous system, its dynamic interactional selectivity will determine different state transitions, mediated by environmental circumstances, so that $(st_i, e_k) \mapsto st_x$ (where st_i and st_x are states of the system, and e_k is an environmental state). Because every system is influenced by environmental interactions, such mappings are normally many-to-one; therefore, whenever an enaction takes place, there is a categorization determined by the particular nature of the autonomy, which implicitly specifies a certain structural instantiation, with some inherent sensitivity and possibilities for action; It is in this sense that an enaction is simultaneously a distinction and an action. For a deterministic case, we can define an *enaction set* as the set of all environments enacted in the same way, a related function that maps pairs of states to these sets, and sets of *enaction categories* containing the valid transitions available to a system:

$$es(st_i, st_x) = \{e_k : (st_i, e_k) \to st_x\} \tag{1}$$

$$f_{es}(st_i, st_x) := \{e_k : (st_i, e_k) \to st_x\}. \tag{2}$$

$$ec(st_i, st_x) := (st_i, st_x, f_{ec}(st_i, st_x)). \tag{3}$$

In spite of their simplicity, these definitions allow us to determine relevant properties of a (relatively simple) autonomous system, such as its organization, selectivity, or structural degeneracy. An exhaustive (and wonderful) illustration of the dynamic properties of autonomy, in the context of autonomous patterns in the Game of Life, can be found in (Beer, 2004, 2014, 2020a,b).

Integrated information?

Even if powerful, our current formal descriptions of autonomous systems seem unable to capture some important qualities associated to a phenomenological dimension; in particular, the above mentioned notions of phenomenological (operational) domain and the emergence of a global coherence. In this context, to explore the possibility of incorporating concepts from the Integrated Information Theory of Consciousness (IIT), such as *intrinsic information* and *integration*, probably by making an enactive reinterpretation, may result fruitful.

The IIT considers selectivity to be a requirement for *intrinsic information*, as it underlies the capacity of a mechanism to constraint past and future system's states: *its cause-effect power*. Cause-effect information is conceived as a measure of how relevant a change is, from the *perspective of the system itself* (Oizumi et al., 2014) which is given by the minimum (shared) between cause (ci) and effect (ei) information. For simplicity's sake, we will make use of the example system presented in Oizumi et al. (2014) (fig. 1), where, given a system of mechanisms A,B and C; cause and effect information can be obtained from:

$$ci = D(p(ABC^p \mid A^c = 1) \mid\mid p^{uc}(ABC^p)) \qquad (4)$$

$$ei = D(p(ABC^f \mid A^c = 1) \mid\mid p^{uc}(ABC^f)) \qquad (5)$$

Here, $p(ABC^p \mid A^c = 1)$ and $p(ABC^f \mid A^c = 1)$, called cause and effect repertoires, are the constrained probability distributions for past and future system's states, given that $A = 1$; the superscripts ^{p},^{c} and ^{f} stand for past, current and future, and ^{uc} denotes unconstrained distributions; whereas D, represents the distance between distributions, measured using the earth mover's distance (EMD) method.

In turn, integration implies that a system able to support emergent experience must be an irreducible whole, not far from the enactive premise of an emergent identity, even if the specific sense of emergence (Chalmers, 2011) is not always unequivocal. Unfortunately, as described by Mediano et al. (2018), different methods for measuring the integration of a system yield inconsistent results, which is obviously problematic for conclusive interpretations. Although less specific, we could consider using a more reliable measure for integration like that from De Rosas et al. (2020):

$$\Psi_{t,t'}(V) := I(V_t; V_{t'}) - \sum_j I(X_t^j; V_{t'}) \qquad (6)$$

While both approaches consider systems to exist intrinsically and to be intrinsically determined (even if susceptible to external perturbations) by means of their organization (causal structure); practically, this can be implemented in different manners. Indeed, IIT's repertoires are evaluated by fixing the environment and defining a new transition matrix whenever is needed; this however, if we conceive system's transitions in terms of enaction, would be obscuring the fact that such transitions entail (syntactic, but phenomenologically relevant) distinctions. More concretely, if, for instance, we change the environment (element D=0/1) from figure 1, we could certainly specify two transition matrices, but isolated these are unable to reflect the full range of available enactions, which are closer to be the actual distinctions from the perspective of the system (see table 1).

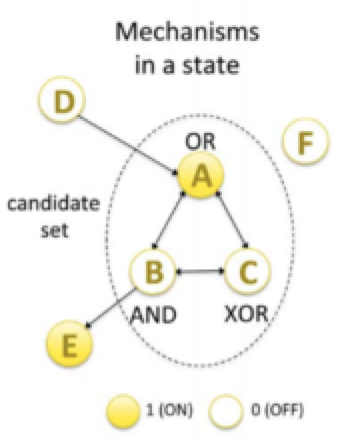


Figure 1: Minimal example system introduced by Oizumi et al. (2014) in the last (3.0) version of IIT.

ec	st_i	env_k	st_x
e1	000	$\{env_1\}$	000
e2	000	$\{env_2\}$	100
e3	001	$\{env_1, env_2\}$	100
e4	010	$\{env_1, env_2\}$	101
e5	011	$\{env_1, env_2\}$	101
e6	100	$\{env_1\}$	001
e7	100	$\{env_1\}$	101
e8	101	$\{env_1, env_2\}$	111
e9	110	$\{env_1, env_2\}$	100
e10	111	$\{env_1, env_2\}$	110

Table 1: The superset of enaction categories available to the system ABC after taking both environmental conditions (ABCD, for D=0 and D=1) into consideration

Thus, we could define a slightly different cause and effect repertoires as distributions of the enaction categories that have led to-, or will inform of affordances with respect to the current state of a mechanism. Of course, in order to ensure congruence when quantifying information, the unconstrained distributions must change accordingly.

References

Abramova, K. and Villalobos, M. (2015). The apparent (ur-)intentionality of living beings and the game of content. *Philosophia*, 43:651–668.

Barandiaran, X. (2017). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, (36):409–430.

Beer, R. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, (10):309–326.

Beer, R. (2014). The cognitive domain of glider in the game of life. *Artificial Life*, 20:183–206.

Beer, R. (2020a). Bittorio revisited: Structural coupling in the game of life. *Adaptive Behavior*, 28(4):197–212.

Beer, R. (2020b). An integrated perspective on the constitutive and interactive dimensions of autonomy. *Proceedings of the ALIFE 2020: The 2020 Conference on Artificial Life*, July 13-18:202–209.

Chalmers, D. (2011). Strong and weak emergence. *In: The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion. Oxford University Press.*, pages 244–256.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action and the Embodied Mind.* Oxford University Press.

De Rosas, F., Mediano, P., Jensen, H., Seth, A., Barret, A., and Carthart-Harris, R. (2020). Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLOS. Computational Biology*, 16(12).

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, (4):429–452.

Di Paolo, E., Burhmann, T., and Barandarian, X. (2017). *Sensorimotor Life: An enactive proposal.* Oxford University Press.

Egbert, M. and Barandiaran, X. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Frontiers in Human Neuroscience*, 8(590):1–15.

Froese, T. and Taguchi, S. (2019). The problem of meaning in ai and robotics: Still with us after all these years. *Philosophies*, 4(2).

Hutto, D. and Myin, E. (2012). *Radicalizing Enactivism. Basic minds without content.* MIT Press.

Hutto, D. and Myin, E. (2017). *Evolving Enactivism. Basic Minds Meet Content.* MIT Press.

Maturana, H. and Varela, F. (1973). *Autopoiesis: the organization of the living. [De maquinas y seres vivos. Autopoiesis: la organizacion de lo vivo]. 7th edition from 1994.* Editorial Universitaria.

Mediano, P., Seth, A., and Barret, A. (2018). Measuring integrated information: Comparison of candidate measures in theory and simulation. *Entropy*, 21(1):17.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLOS Computational Biology*, 10(5).

Seth, A. and Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11):969–981.

Varela, F. (1979). *Principles of Biological Autonomy.* North Holland.

Varela, F. (1997). Patterns of life: Intertwining identity and cognition. *Brain cognition*, (34):72–87.

Varela, F., Rosch, E., and Thompson, E. (1991). *The embodied mind: Cognitive science and human experience.* The MIT Press.

Villalobos, M. and Silverman, D. (2018). Extended functionalism, radical enactivism and the autopoietic theory of cognition: prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences*, 17:719–739.

Towards Hierarchical Hybrid Architectures for Human-Swarm Interaction

Jonas D. Rockbach[1,2], Luka-Franziska Bluhm[1], and Maren Bennewitz[2]

[1]Human-Machine Systems, Fraunhofer FKIE, Wachtberg, Germany
[2]Humanoid Robots Lab, Computer Science VI, University of Bonn, Germany
jonas.rockbach@fkie.fraunhofer.de

Abstract

This contribution summarizes an integrated view on human-swarm interaction which investigates how human cognition should be joined with the distributed intelligence of robot swarms. From our perspective, a capable human-swarm hybrid that is embedded in the world can be formalized as nested agent interaction matrices that are hierarchically organized.

Human-Swarm Interaction

A joint human-swarm loop (JHSL) is a hybrid agent where humans are joined with a robot swarm via interfaces to solve particular tasks in the world (Hasbach and Bennewitz, 2021). While swarm engineering is the discipline that focuses on the design of robot swarms (Dorigo et al., 2021), human-swarm interaction (HSI) investigates how the cognitive decision making capabilities of humans can be merged with the distributed intelligence of robot swarms (Hasbach and Bennewitz, 2021; Kolling et al., 2016).

In general, a JHSL is made of at least three interdependent facets; humans, swarm robots, and interfaces. Recently, we have proposed a view on HSI that aims at integrating these different facets, or views, of the JHSL (Hasbach and Bennewitz, 2021). From this integrated perspective, humans, robot swarm and interfaces are considered to be the building blocks for designing an intelligent hybrid agent that is situated in the world. We therefore proposed that HSI could also be interpreted as (hybrid) human-swarm *intelligence*. Fig. 1 shows a simplified summary of the human-swarm intelligence design space.

In this contribution, we summarize an aspect of our previous work by elaborating on how a JHSL can be formalized as nested interaction matrices that are hierarchically organized and what this means for the design of HSI.

Hierarchical Hybrid Architecture

Intelligent Hybrid Agents

A JHSL $L = \{H, S, C, I\}$ can be described as a set of humans H, swarm robots S, components $C \subseteq \{H, S\}$, and local interaction matrices I. A component c_i is a group of locally interconnected human and swarm agents that must interact to achieve subgoals as a collective. The agents may participate in multiple components simultaneously. The set of local interaction matrices I defines the information architecture of the JHSL, i.e., the (possible) local interactions inside and between agent components C (Hasbach and Bennewitz, 2021; Heylighen, 2001; Simon, 1982). For example, a subswarm engaged in a collective decision making task (Hamann, 2018) in a local area can be considered a component c_i that is determined by the locally dense agent interaction matrix i_i that in turn is a subset of the overall agent interaction possibilities I.

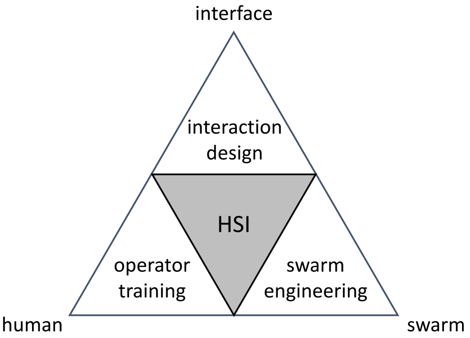


Figure 1: The triad of human-swarm intelligence summarizes its main facets and design dimensions: human (operator training), swarm (swarm engineering), and interface (interaction design, e.g., user experience). Adapted from Rockbach and Bennewitz (2021).

In general, an intelligent JHSL L must deal with the dynamics in the world E by implementing sensory-action rules that enhance the probability of reaching a desired target state $e_{goal} \in E$ (Russell and Norvig, 2016; Wooldridge, 2009). Therefore, the interaction possibilities I between humans and the robot swarm should not be the design goal itself but rather the means in order to maximize the capabilities of the JHSL that gets the tasks done in particular situations (Hollnagel and Woods, 2005). More specifically, the complexity of the interaction dynamics should be kept as minimal as possible while the capability of the JHSL to reach goal states should be as maximal as possible (Hasbach and Bennewitz, 2021). This design principle is vividly demonstrated by the sparsely connected interaction matrix of the human nervous system (Genç et al., 2018).

In sum, three layers of nested interaction networks are defined in our framework:

1. Layer 1: The sensor-decide-action loops *in* each participating agent that determines the behaviour of a single human or robot agent.

2. Layer 2: The sensor-decide-action loops *combined* by the participating agents *in* each component that determines the behaviour of a functional subgroup of interacting human and robot agents.

3. Layer 3: The sensor-decide-action loops *combined* by the components that determines the behaviour of the *overall* JHSL.

Thus, layer 1 is nested in layer 2 which in turn is nested in layer 3. Note also that the layers refer to modelled abstractions of the JHSL, rather than to actual physical differences.

Hierarchical Hybrid Architecture

Consider a JHSL that is part of a search-and-rescue scenario (Hasbach and Bennewitz, 2021; Murphy, 2014). The main purpose of such a hybrid disaster response team is to locate, evacuate and treat a maximum number of victims while minimizing the risk for the own participating agents.

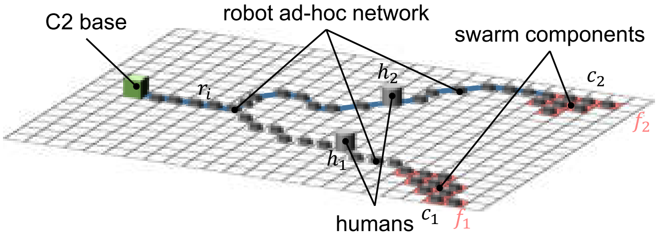


Figure 2: Example of a JHSL situated in a world with a search-and-rescue scenario. The C2 base is coupled with relevant task features of the environment, such as the health state of victims, via the robot ad-hoc network where humans and robots form local components that act in the environment. A possible path between C2 and the task feature f_2 via the JHSL components is shown in blue.

Fig. 2 shows an example of a search-and-rescue JHSL in a simplified $\mathbb{N}^2$ environment. A socio-technical command and control (C2) base is placed at a save location in the west with the role of coordinating different assets (Bluhm et al., 2021). Similar to the control architecture of the nervous system (Albus, 1981; Hasbach and Bennewitz, 2021; Hohwy, 2013), C2 constitutes a higher-order structure that has a broad view of the situation but also needs time to decide. On the other hand, the two deployed search-and-rescue operators that can be seen in the centre have a more narrow-detailed view of the situation while they can act directly, and therefore faster, in the environment. The robot swarm subsets in turn are located directly at relevant but hostile parts of the environment F, $F \subseteq E$, in the east and have once again a narrower view with shorter reaction times. The robots participating in c_1 and c_2 form two dense interaction submatrices that render them a component, or a team, nested inside the larger hybrid interaction matrix.

To conclude, the swarm S can take the role of interfacing higher-order human agents with relevant task features in order to extend the human sensory-motor range (Hasbach and Bennewitz, 2021; Rockbach and Bennewitz, 2021; Sheridan, 1992). This hierarchical hybrid architecture of the JHSL is shown in Fig. 3.

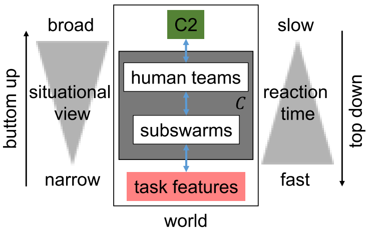


Figure 3: Model of layer 3 that shows the hierarchical hybrid architecture for HSI with narrow-fast loops at the front-end to relevant task features and slow-broad loops at the backend of the organizational hierarchy.

As adaptation is considered the underlying principle of decision making in complex worlds (Ashby, 1960; Hasbach and Witte, 2021; Hollnagel and Woods, 2005), the JHSL must be able to dynamically adjust its interaction matrices to different situations. For example, imagine the situation where a search-and-rescue team locates possible victims that it cannot treat by its own capacity. How should the hybrid interaction matrix of layer 2 and layer 3 adapt? If available, the local team could call for other assets in the area that may either allocate by themselves or are guided by the C2 element. When these allocated assets join the first-responder team, they will start working together (Coucke et al., 2020; Salas et al., 2008), i.e., form a new enhanced component[1] in terms of a locally denser interaction matrix compared to the rest of the system.

Importantly, it should be remembered that the organizational structure of the JHSL (Fig. 2) should be adapted to the current situation. The design of HSI must take such considerations of hybrid intelligence into account.

[1] The enhancement of a subsystem by joining other agents is a simplification, see Hamann and Reina (2021).

References

Albus, J. S. (1981). *Brains, behavior, and robotics*. BYTE Books, Peterborough.

Ashby, W. R. (1960). *Design For A Brain: The origin of adaptive behaviour*. Chapman & Hall Ltd, New York, second edition.

Bluhm, L.-F., Lassen, C., Keiser, L., and Hasbach, J. (2021). Swarm View: Situation Awareness of Swarms in Battle Management Systems. In *STO-MP-SCI-341*.

Coucke, N., Heinrich, M. K., Cleeremans, A., and Dorigo, M. (2020). Hugos: A multi-user virtual environment for studying human–human swarm intelligence. In *International Conference on Swarm Intelligence*, pages 161–175. Springer.

Dorigo, M., Theraulaz, G., and Trianni, V. (2021). Swarm Robotics: Past, Present, and Future [Point of View]. *Proceedings of the IEEE*, 109(7):1152–1165.

Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., Ling, J. M., Güntürkün, O., and Jung, R. E. (2018). Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. *Nature communications*, 9(1):1–11.

Hamann, H. (2018). *Swarm robotics: A formal approach*. Springer, Heidelberg.

Hamann, H. and Reina, A. (2021). Scalability in computing and robotics. *IEEE Transactions on Computers*.

Hasbach, J. D. and Bennewitz, M. (2021). The design of self-organizing human–swarm intelligence. *Adaptive Behavior*.

Hasbach, J. D. and Witte, T. E. (2021). Human-Machine Intelligence: Frigates are Intelligent Organisms. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*.

Heylighen, F. (2001). The science of self-organization and adaptivity. *The encyclopedia of life support systems*, 5(3):253–280.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press, Oxford.

Hollnagel, E. and Woods, D. D. (2005). *Joint Cognitive Systems. Foundations of Cognitive Systems Engineering*. CRC Press, Boca Raton, FL.

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., and Lewis, M. (2016). Human Interaction with Robot Swarms: A Survey. *IEEE Transactions on Human-Machine Systems*, 46(1):9–26.

Murphy, R. R. (2014). *Disaster robotics*. MIT press, Cambridge, MA.

Rockbach, J. D. and Bennewitz, M. (2021). Robot Swarms as Embodied Extensions of Humans. In *Proceeding of the 2021 International Workshop on Embodied Intelligence. In press*.

Russell, S. and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson, 3rd edition.

Salas, E., Cooke, N. J., and Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, 50(3):540–547.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT press, Cambridge, MA.

Simon, H. (1982). *The Sciences of the Artificial*. MIT press, Cambridge, MA.

Wooldridge, M. (2009). *An introduction to multiagent systems*. John wiley & sons, New York.

Minimal Models for Spatially Resolved Population Dynamics – Applications to Coexistence in Multi – Trait Models

Rudolf M. Füchslin[1,2], Pius Krütli[3], Thomas Ott[4], Stephan Scheidegger[1], Johannes J. Schneider[1], Marko Seric[1], Timo Smieszek[5], Mathias S. Weyland[1]

[1] Zurich University of Applied Sciences, School of Engineering, Winterthur, Switzerland,
[2] Europan Centre for Living Technology, Venice, Italy
[3] Transdisciplinarity Lab, ETH Zürich, Switzerland
[4] Zurich University of Applied Sciences, School of Life Sciences, Wädenswil, Switzerland
[5]IntiQuan GmbH, Basel, Switzerland
rudolf.fuechslin@zhaw.ch

Abstract

Spatial resolution is relevant for many processes in population dynamics because it may give rise to heterogeneity. Simulating the effect of space in two or three dimensions is computationally costly. Furthermore, in Euclidean space, the notion of heterogeneity is complemented by neighbourhood correlations. In this paper, we use an infinite-dimensional simplex as a minimal model of space in which heterogeneity is realized, but neighbourhood is trivial and study the coexistence of viral traits in a SIRS – model. As a function of the migration parameter, multiple regimes are observed. We further discuss the relevance of minimal models for decision support.

Spatial Resolution in Population Dynamics

It is well known that population dynamics in spatially resolved systems show features not observed in homogeneous systems (Sun et al., 2021). Spatially structured systems enable microenvironments that give rise to local "symmetry breaking" or spatial heterogeneity. Not all microenvironments have to be in the same state, even if the fundamental laws governing the local dynamics are the same everywhere. This spatial heterogeneity may result from stochastic effects and/or reaction-diffusion processes, e.g. shown in (Turing, 1990) or the complex dynamics emerging in seemingly simple bacteria (Govindarajan et al., 2012; Shapiro et al., 2009).

A broad class of processes combine colocalization of individuals with the transfer of an attribute from an individual with this attribute to one without it. This transfer can conserve the attribute or replicate it. The former case is relevant in the study of conserved quantities in physics or economics, whereas the latter represents processes that one can understand as infections or, regarding information in societies, as knowledge transfer or teaching processes.

As application, we study a minimal model of spatial resolution to a variant of the SIRS – model with two traits:

$$\begin{aligned} S + I_k &\xrightarrow{\alpha_k} 2I_k \\ I_k &\xrightarrow{\rho_k} R_k \\ R_k + I_{\bar{k}} &\xrightarrow{\xi_{\bar{k}}} 2I_{\bar{k}} \\ R_k &\xrightarrow{\nu_k} S \end{aligned} \qquad (1)$$

As usual, the variable S represents susceptibles, I_k infected, and R_k recovered. The index $k \in \{1,2\}$ represents the multiple traits. We set $k = 1,2 \Rightarrow \bar{k} = 2,1$. The parameter α_k models the infection, ρ_k recovery, ν_k waning immunity and ξ_k cross-infection. We emphasize that models as given in eqs. (1) are not restricted to diseases but can be transferred to, e.g., the spread of cultural innovations (Walker et al., 2021).

In eqs. (1), the infection processes are modelled by a single parameter α_k, ξ_k respectively. As a motivation for the presented minimal model relevant, these parameters combine (at least) three variables: The infectivity of the I, the susceptibility of S, R_k and the contact rate of the infected and the susceptibles.

In a conventional SIRS model, there is no easy way to disentangle physiological parameters (infectivity, susceptibility) from the influence of the contact rate. One could think that doubling the contact rate can be represented by a twofold increased infection rate α_k. This holds for low physiological infection/susceptibility parameters, but saturation effects kick in for higher values. This can easily be understood if one considers that it is not the number of contacts alone but also the time of exposure that influences the risk of infection. Transmission processes which require proximity and time of exposure (thereby limiting the number of potential sources of infection) eventually reach a saturation level for the infectivity.

Minimal Models: Epidemiology on an Infinite Dimensional Simplex

Besides understanding a specific situation, we claim that there is an interest in studying generic phenomena resulting from spatial resolution. Whereas a study that aims to understand the details of a specific epidemiological development should map the real world as precisely as possible (complete models), a study focusing on generic properties should work with a space as simple as possible.

The notion of "space" combines a variety of mathematical structures; first, the concept of space implies that one can distinguish between here and there. Furthermore, spaces such as the three-dimensional Euclidean space allow to quantify the "theres" (means "non-heres") by a notion of distance and thereby invoking the notion of neighbourhood and neighbourhood correlation. Studying such spaces gives detailed insight into the processes taking place in them but is computationally expensive.

Probably the simplest structure that allows some form of spatial heterogeneity but with only a trivial notion of neighbourhood is an infinite-dimensional simplex. For our purposes, a simplex is a set of discrete locations or nodes that are all mutually connected. If the number of these locations goes to infinity, one speaks of an infinite-dimensional simplex. In our investigations, a location contains two sites, see Fig. 1. Each site is occupied by a representative of the five species S, I_k, R_k or empty (occupied by a V). As discussed in (Füchslin et al., 2019; McCaskill et al., 2001), the key point of such a simplex is that it enables to implement a mean-field formulation of a dynamics as given in eq. (1). The fundamental observation underlying this is that since all locations experience the same neighbourhood (all locations are mutually connected), the probability of being in a specific state is equal for all locations. Influx of a representative X of one of the species into a location is determined by a mobility parameter m, the number of empty sites on the location and the average $\overline{X}$ on all other locations, see Fig. 1.

More formally, if U denotes the set of all allowed states $u = (s, i_1, i_2, r_1, r_2)$ and $x(u)$ gives the number of x in a state $u \in U$, we must calculate the (time-dependent) probabilities $P(s, i_1, i_2, r_1, r_2; t)$. Because we have two sites per location, it must hold $0 \le s + i_1 + i_2 + r_1 + r_2 \le 2$. The probabilities $P(s, i_1, i_2, r_1, r_2; t)$ are combined into a vector $\vec{P}(t)$, and one writes the dynamics of the system as:

$$\frac{d\vec{P}(t)}{dt} = A(\vec{P}(t))\vec{P}(t) \tag{2}$$

Here $A(\vec{P}(t))$ is a matrix that depends on $\vec{P}(t)$. To illustrate the construction of $A(\vec{P}(t))$, we analyze the dynamics of the state $s = 1, i_1 = 1, i_2 = 0, r_1 = 0, r_2 = 0$. This state can be reached or left either by internal epidemiological dynamics (eqs. (1)) or by influx from some other site.

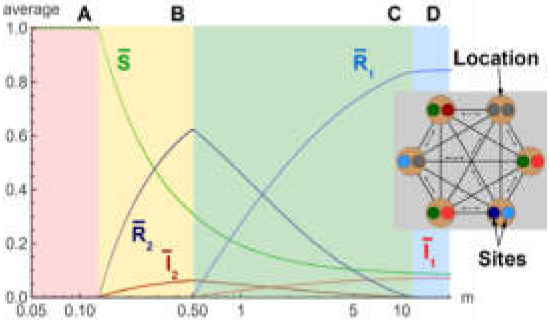


Figure 1: Longterm Averages as a function of the migration parameter m. The other parameters are: $\alpha_1 = 3.0, \alpha_2 = 2.0, \xi_1 = 0.03, \xi_2 = 0.04,$ $\rho_1 = 0.12, \rho_2 = 0.04, \nu_1 = \nu_2 = 0.01$. Time is given in days. The inlet shows a simplex of six locations with two sites at each location.

For the internal dynamics, we get:

$$\left.\frac{dP(1,1,0,0,0;t)}{dt}\right|_{epi} = -\alpha_1 P(1,1,0,0,0;t) + \nu_1 P(0,1,0,1,0;t) + \nu_2 P(0,1,0,0,1;t) \tag{3}$$

For state changes induced by migration, we get:

$$\left.\frac{dP(1,1,0,0,0;t)}{dt}\right|_{mig} = m\overline{S}P(0,1,0,0,0;t) + m\overline{I}_1 P(0,1,0,0,0;t) - 2m\overline{V}P(1,1,0,0,0;t) \tag{4}$$

The averages $\overline{X}$ are (with $\overline{V} = 2 - \overline{S} - \overline{I}_1 - \overline{I}_2 - \overline{R}_1 - \overline{R}_2$):

$$\overline{X} = \sum_{u \in U} x(u) P(u;t) \tag{5}$$

As an example, we use this formalism to study an essential effect of spatial heterogeneity, namely the coexistence of different traits in a population. In recent times, questions concerning the coexistence of different traits of viruses gained attention in epidemiology (Ackleh et al., 2016; Guo & Wang, 2022; Roberts et al., 2015). As it turns out, the coexistence of traits appears in our model, see Figure 1, where the averages the system approaches after long time are shown as a function of m. We distinguish regions A, B, C, and D in which either no, one or both traits exist. The simple model we present illustrates an effect which is relevant with respect to public health: In case of co-existing traits, reducing m may decrease one trait, but lead to an increase of the other trait. This phenomenon is no surprise but needs to be considered, if the two traits differ in the severity of the disease they cause.

Discussion

It is clear that the presented model is not suited for predicting the course of an actual pandemic. The model is much too simple (no age dependency, et cetera), and the spatial structure does not reflect any actual geography. However, it still teaches us some important lessons: First, the coexistence of different viral traits is potentially possible but depends on the migration rate m (equivalent to the contact rate). Second, and more important for decision support, a decrease in the contact rate may suppress one trait that is dominant at higher contact rates. The decrease may, however, give room for a trait with different properties.

The SARS-CoV-2 pandemic resulted in a massive increase of interest in mathematical and model-based epidemiology. In a recent review (Gnanvi et al., 2021), the authors compare different simulation techniques for modelling the dynamics of the SARS-CoV-2 pandemics (Compartment models of the SIR- or SEIR type (46%). Only 1.3% of the studies used agent-based models). Most of these studies focused on a particular case in a specific geographic setting. This approach is sensible, particularly because the contact structure of the populations in different countries is known with a resolution concerning age and type of activity (Fumanelli et al., 2012; Prem et al., 2017). These are "complete" models in that they try to include as much of reality as possible. This comes at a price: The models contain many parameters and, in consequence, are difficult to calibrate. In contrast, minimal models may give only qualitative insight into the processes they analyze. However, this comes with the advantage that it is often easy to relate cause and effect. More concretely, concerning epidemiology, even if minimal models are not suited as tools for prediction, they justify a detailed scrutiny of variants of minor importance. This is because these variants may become relevant after changing system parameters, e.g. via non-pharmaceutical interventions.

Giving a justification for (expensive) observations is highly relevant in decision support; it is an example of the value and importance of modelling for politics and society and helps strengthen the role of science and artificial life, in particular.

Acknowledgments. This work has been partially supported by the EU – funded project ACDC (Grant Agreement No. 824060) and a project of the FOPH (Federal Office of Public Health) in Switzerland (contract: 142003923 / 332.11-78/5).

References

Ackleh, A. S., Deng, K., & Wu, Y. (2016). Competitive exclusion and coexistence in a two-strain pathogen model with diffusion. *Mathematical Biosciences & Engineering*, *13*(1), 1.

Füchslin, R. M., Schneider, J. J., Ott, T., & Walker, R. (2019). Simplified modeling of the evolution of skills in a spatially resolved environment. The 2019 Conference on Artificial Life, Newcastle, United Kingdom, 29 July-2 August 2019,

Fumanelli, L., Ajelli, M., Manfredi, P., Vespignani, A., & Merler, S. (2012). Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread.

Gnanvi, J., Salako, K. V., Kotanmi, B., & Kakaï, R. G. (2021). On the reliability of predictions on Covid-19 dynamics: A systematic and critical review of modelling techniques. *Infectious Disease Modelling*.

Govindarajan, S., Nevo-Dinur, K., & Amster-Choder, O. (2012). Compartmentalization and spatiotemporal organization of macromolecules in bacteria. *FEMS microbiology reviews*, *36*(5), 1005-1022.

Guo, J., & Wang, S.-M. (2022). Threshold dynamics of a time-periodic two-strain SIRS epidemic model with distributed delay. *AIMS Mathematics*, *7*(4), 6331-6355.

McCaskill, J. S., Füchslin, R. M., & Altmeyer, S. (2001). The stochastic evolution of catalysts in spatially resolved molecular systems.

Prem, K., Cook, A. R., & Jit, M. (2017). Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, *13*(9), e1005697.

Roberts, M., Andreasen, V., Lloyd, A., & Pellis, L. (2015). Nine challenges for deterministic epidemic models. *Epidemics*, *10*, 49-53.

Shapiro, L., McAdams, H. H., & Losick, R. (2009). Why and how bacteria localize proteins. *Science*, *326*(5957), 1225-1228.

Sun, G.-Q., Zhang, H.-T., Wang, J.-S., Li, J., Wang, Y., Li, L., Wu, Y.-P., Feng, G.-L., & Jin, Z. (2021). Mathematical modeling and mechanisms of pattern formation in ecological systems: a review. *Nonlinear Dynamics*, *104*(2), 1677-1696.

Turing, A. M. (1990). The chemical basis of morphogenesis. *Bulletin of mathematical biology*, *52*(1), 153-197.

Walker, R., Eriksson, A., Ruiz, C., Newton, T. H., & Casalegno, F. (2021). Stabilization of cultural innovations depends on population density: Testing an epidemiological model of cultural evolution against a global dataset of rock art sites and climate-based estimates of ancient population densities. *PloS one*, *16*(3), e0247973.

AgTech that doesn't cost the Earth: Creating sustainable, ethical and effective agricultural technology that enhances its social and ecological contexts

Alan Dorin[1*], Alexandra S. Penn[2] and Jesús M. Siqueiros[3]
[1] Computational and Collective Intelligence, Dept. of Data Science & AI, Faculty of Info. Tech., Monash University, Australia
[2] Centre for the Evaluation of Complexity Across the Nexus, Centre for Research in Social Simulation, Dept. of Sociology, University of Surrey, UK
[3] Unidad Académica del IIMAS en el Estado de Yucatán, IIMAS, UNAM, México
*alan.dorin@monash.edu

Abstract

To feed the growing human population we require increased food production and security, while using less land and causing less environmental damage. Significant changes in agriculture are needed to meet these demands. One widely touted solution is smart, AI-enhanced Agricultural Technology. In this article we argue that improved technology is insufficient to address the needs of many farmers, but that by taking a whole-of-system approach native to Artificial Life we can shift towards creating sustainable, ethical and effective AgTech. This can innovate industrial agriculture in developed nations and benefit small landholders from vulnerable communities, whilst reducing the environmental impacts of food production globally.

Introduction and background

The rapidly increasing human population demands more food, and more reliable food production, but significant changes in agricultural practices are needed to meet these demands under urban intensification, in an increasingly unpredictable climate, whilst reducing poverty, resource consumption and environmental damage [1, 2]. The majority of the world's farmers operate very small plots in developing nations [3]. Agricultural Technology (AgTech) can *potentially* reduce their poverty, increase their well-being and food security [4], and drive economic development [5], but it hasn't yet, and it won't, unless changes are made in how it is created, applied and socially integrated. "AgTech" encompasses electronics and algorithms for monitoring, managing and harvesting crops and livestock, agrochemicals and biotech for growing and breeding robust nutritious food, even processes for marketing and distributing food [6, 7]. The data-driven digital technological aspects of AgTech are being prominently disrupted and improved by Artificial Intelligence (AI) [8-10]. However, the impacts of AgTech on food security and poverty are exceedingly complex, context-specific, and have little to do with the technological components of the situation that AI addresses most directly [4, 11]. Consequently, increased data availability and naïve AgTech enhancements via AI miss key factors relating to trust, reliability and social integration required especially for adoption by small-holder and subsistence farmers [8, 11]. They can generate substantial environmental footprints when used by large scale industry that may be forced upon developing nations (although, see [12]). AI and data-enhanced AgTech is therefore no silver bullet. A new approach is needed to lessen social and technological division. Relevant factors that must be incorporated into AgTech research include cultural, social, political and economic concerns that determine the relationship between humans, food, and environment [13]. Improved technology is only one issue for sustainable and ethically feeding people (e.g. [14, 15]).

Here, we link ideas from within the field of Artificial Life (AL) to sustainable, ethical and effective AgTech. These ideas aren't unique to AL, some are explored in fields addressing agriculture's social and environmental aspects. However, the multi/inter/trans-disciplinarity of AL links these aspects *and* technology, and this is specific to the field. Below we classify and discuss AL's potential contributions to AgTech.

Why *should* Artificial Life focus on AgTech?

Artificial Life is a scientific research field that studies natural living phenomena (e.g. organisms, ecosystems, social systems) through research and experimentation with artificial processes, often synthesised in software [16, 17]. Through its concerns with dynamics, synthesis, interactions and complex adaptive systems, AL offers broad, powerful perspectives for AgTech beyond AI-enhanced smart technology (e.g. [10]) and attainment of quantifiable engineering goals. AL principles point AgTech towards ethical food production situated within a whole-of-system context. We classify its contributions as: (i) Understanding and perspective; (ii) Simulation and modelling; (iii) Design and innovation; (iv) Intervention. This categorisation is counter to our intuition as AL researchers that linearity is appropriate, but it is selected for alignment with product lifecycle stages: requirements solicitation, design, development and testing, manufacture, sales, deployment, and customer relationship management.

(i) Understanding and perspective. As noted above, a key point brought home by AL's attention to complex adaptive systems, is that AgTech operates within a hybrid social, economic, technological, agricultural, climatic, ecological, biological system. If overall system behaviour is poorly understood, technological intervention will be unpredictable and failure prone. AL, as a "systems thinking" native, explores cascades of causal effects of technology beyond a myopic focus, e.g. on profit or crop yield, to encompass users' changes of practice and thinking as they interact with AgTech within extended social and environmental feedback loops. An AL perspective implies transdisciplinary co-design and a requirement to understand AgTech's consequences for people, land and environment [13] by unravelling the system's hypothetical trajectories – even if this approach is not (yet) widely adopted within the field itself. This can be achieved via participatory modelling strategies (e.g. (Evolved) Double Diamond [18], Companion Modelling [19]) that link relevant simulations (see (ii) below), stakeholders and researchers in design processes to aid social learning and innovation and to generate new, sustainable and just social, technological, environmental processes.

(ii) Simulation and modelling. AL often adopts interactive computer models and simulations to acquire a preliminary understanding of complex adaptive system behaviour [17, 20]. Only when this has been achieved is it sensible to decide what real challenges should be addressed, what system components or processes are suited to intervention, and what tools are best for intervention. This last point is important because AgTech is only one potential element of food security. Other options include improved land management processes [21, 22] social technologies [23-25], even the avoidance or removal of counter-productive or toxic technology [26, 27].

AL's diverse simulation techniques (e.g. agent-based models, cellular automata, L-systems, network-models) have been applied to bee-pollination [28], land use [29], plant/herbivore interactions [30] and livestock movement [31]. Spatial ecological interactions, information exchange, trade and cooperation in social systems and technology innovation, also form relevant streams in AL simulation research [32-35]. Such approaches can shift AgTech to explore new ideas, promote engagement, and anticipate implementation outcomes.

(iii) Design and innovation. The design, development and sale of current AgTech falls largely within engineering and technology industries. Examples include data platforms, smart farm tractor control and coordination, UAVs, greenhouses, water / nutrient supply systems, monitoring tools, and robotics [8, fig.4]. If tech-focused design processes dominate, manufacturers may entice farmers into ongoing, potentially unsustainable, complex relationships of dependence, such as sowing crops engineered for resistance to glyphosate [36]. They are also likely to develop technology assuming farms aim for profit and increased yield, underlying assumptions that can further divide largescale industrial wealth from subsistence farming poverty [37]. AL simulations can explore how the design and diffusion of technology alters unfolding societal technological dependencies, economies and ecosystems [38].

In software industries *software-as-a-service* contracts with individual users may support companies to maintain products and customer satisfaction [39]. Iterative user experience surveys can help technology address customer requirements within changing application landscapes [40]. AgTech can potentially extend these approaches to sustain interactions between technology and local, dynamic, social and ecological environments after roll-out. However, the diversity of stakeholders, from rural subsistence farmers to multinational corporations, makes this a costly proposition since with it comes vast differences in literacy, digital connectivity and communication preferences or constraints. If these hurdles are not leapt, AgTech risks increasing social division by ignoring the voices of the majority of farmers, who operate plots of less than 2 ha. [3, 41], and forcing them into unsustainable practices. This eventually has ramifications felt by industrial agriculture through international cascades of environmental destruction, famine, war and mass migration (exacerbated by climate change) – scenarios that AL simulations can explore [42, 43]. The preferred scenario is an AgTech that doesn't impose a-contextual technology's limitations on local practices. By investing in co-design and flexible (perhaps simulation-informed) implementation at a local scale, industry facilitates the shift of farmers from users to "active agents" (a prominent AL topic) who learn as they design relevant, sustainable, culturally coherent agrisystems [44]. This ethical approach empowers farmers to appropriate technology on their terms, adds to its longevity and immeasurably improves its value [45].

(iv) Intervention. If stages (i-iii) lead a co-design team to agree on technological intervention, this doesn't imply a "set and forget" mentality will succeed. Beyond initial intervention too, the principles of complex systems guidance familiar to AL researchers suggest that the farmer must continue to participate [46, throughout] in continuous iterative monitoring and adaptation of the dynamical system with novel technology as an untried component. User engagement and empowerment are key to managing this responsibility. Even if the system is understood prior to intervention, no model can predict behaviour subject to as many external influences, feedback loops and degrees of freedom as technology situated within agriculture; an exemplar of complexity. We also expect adaptation and evolution, of both biological and social components to occur. Stakeholders must continually learn and renegotiate the role of AgTech within the agrisystem – the system must be tweaked interactively from inside, a requirement with which many AL researchers are familiar – if food security is to be achieved. AL could further contribute experience with evolving complex systems to developing new adaptive management processes and tools.

Conclusion

Embedding approaches (i-iv) into participatory adaptive management processes [46], enhanced by technology, with stakeholders as full partners, would be a revolutionary vision of people empowered by AgTech to adaptively manage and steer their complex agricultural systems. Artificial Life's approaches have the potential to establish the paradigms to achieve this. We challenge the community to embrace participatory approaches and combine them with our unique technological expertise to make this vision a practicable reality.

References

1. Tallis, H.M., et al., *An attainable global vision for conservation and human well-being.* Frontiers in Ecology and the Environment, 2018. 16(10): p. 563-570.
2. Vågsholm, I., N.S. Arzoomand, and S. Boqvist, *Food Security, Safety, and Sustainability—Getting the Trade-Offs Right.* Frontiers in Sustainable Food Systems, 2020. 4(16): p. 1-14.
3. Lowder, S.K., M.V. Sánchez, and R. Bertini, *Which farms feed the world and has farmland become more concentrated?* World Development, 2021. 142(105455).
4. DeJanvry, A. and E. Sadoulet, *World poverty and the role of agricultural technology: direct and indirect effects.* Journal of Development Studies, 2002. 38(4): p. 1-26.
5. Tripp, R., *Technology and Its Contribution to Pro-Poor Agricultural Development*. 2005, Agriculture and Natural Resources Team - UK Department for International Development. Available from: https://www.fao.org/sustainable-food-value-chains/library/details/en/c/267205/.
6. Watz, E., *Digital farming attracts cash to agtech startups.* Nature Biotechnology, 2017. 35: p. 397–398.
7. Murase, H., *Special issue on Artificial intelligence in agriculture.* Computers and Electronics in Agriculture, 2000. 29(1-2): p. 1-178.
8. Southwood, R. and K. Wong, *Building a Data Ecosystem for Food Security and Sustainability in AgTech V3.0*. 2021, International Center for Tropical Agriculture (CIAT) / CGIAR: Cali, Colombia. p. 1-35. Available from: https://hdl.handle.net/10568/111666.
9. Spanaki, K., et al., *Disruptive technologies in agricultural operations: a systematic review of AI-driven AgriTech research.* Annals of Operations Research, 2022. 308: p. 491-524.
10. Smith, M.J., *Getting value from artificial intelligence in agriculture.* Animal Production Science, 2018. 60(1): p. 46-54.
11. Victor Galaz, et al., *Artificial intelligence, systemic risks, and sustainability.* Technology in Society, 2021. 67: p. 101741.
12. Hasegawa, T., et al., *Risk of increased food insecurity under stringent global climate change mitigation policy.* Nature Climate Change, 2018. 8(8): p. 699-703.
13. Marin, A., A. Ely, and P.v. Zwanenberg, *Co-design with aligned and non-aligned knowledge partners: implications for research and coproduction of sustainable food systems.* Current Opinion in Environmental Sustainability, 2016. 20: p. 93-98.
14. Marco V. Sánchez Cantillo, et al., *The State of Food and Agriculture 2019: Moving forward on food loss and waste reduction*, in *The State of Food and Agriculture (SOFA)*. 2019, Food and Agriculture Organization of the United Nations: Rome, Italy. p. 182.
15. Gustavsson, J., et al., *Global food losses and food waste: Extent, causes and prevention*. 2011, Food and Agriculture Organisation of the United Nations: Rome. p. 37.
16. Boden, M.A., *Artificial Life*, in *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*, R.A. Wilson and F.C. Keil, Editors. 2001, MIT Press: MA, USA / London, England. p. 37-39.
17. Bedau, M.A., *Artificial Life*, in *The Cambridge Handbook of Artificial Intelligence*, K. Frankish and W.M. Ramsey, Editors. 2014, Cambridge University Press: Cambridge, United Kingdom. p. 296-315.
18. Pyykkö, H., M. Suoheimo, and S. Walter, *Approaching Sustainability Transition in Supply Chains as a Wicked Problem: Systematic Literature Review in Light of the Evolved Double Diamond Design Process Model.* Processes, 2021. 9(12): p. 2135.
19. Barreteau, O., *Our Companion Modelling Approach.* Journal of Artificial Societies and Social Simulation, 2003. 6(1).
20. Bromwich, B., et al., *Systems Analysis for Water Resources. Report to Defra, The Department for Environment, Food and Rural Affairs, UK Government 2020.* 2020. Available from: http://randd.defra.gov.uk/Document.aspx?Document=14947_WT15121.FinalReport.pdf.
21. Kamalongo, D.M. and N.D. Cannon, *Advantages of bi-cropping field beans (Vicia faba) and wheat (Triticum aestivum) on cereal forage yield and quality.* Biological Agriculture & Horticulture, 2020. 36(4): p. 213-229.
22. Baum, C., W. El-Tohamy, and N. Gruda, *Increasing the productivity and product quality of vegetable crops using arbuscular mycorrhizal fungi: a review.* Scientia horticulturae, 2015. 187: p. 131-141.
23. Tortia, E.C., V.L. Valentinov, and C. Iliopoulos, *Agricultural cooperatives.* Journal of Entrepreneurial and Organizational Diversity, 2013. 2(1): p. 23-36.
24. Khandker, S.R. and G.B. Koolwal, *How has microcredit supported agriculture? Evidence using panel data from Bangladesh.* Agricultural Economics, 2016. 47(2): p. 157-168.
25. Andersen, R., et al., *Community seed banks: sharing experiences from North and South.* 2018: Paris. p. 44. Available from: https://cgspace.cgiar.org/handle/10568/92510.
26. Carson, R., *Silent Spring*. 1962, Great Britain: Penguin Books.
27. Gleadow, R., J. Hanan, and A. Dorin, *Averting robo-bees: why free-flying robotic bees are a bad idea.* Emerging Topics in Life Sciences, 2019. 3(6): p. 723-729.
28. Dorin, A., et al., *Simulation-governed design and tuning of greenhouses for successful bee pollination*, in *Artificial Life*, T. Ikegami, et al., Editors. 2018, MIT Press: Tokyo, Japan. p. 171-178.
29. Gomes, E., et al., *Modelling future land use scenarios based on farmers' intentions and a cellular automata approach.* Land Use Policy, 2019. 85: p. 142-154.

30. Hanan, J., et al., *Simulation of insect movement with respect to plant architecture and morphogenesis.* Computers and Electronics in Agriculture, 2002. 35(2-3): p. 255-269.
31. Stricklin, W., et al., *Artificial pigs in space: using artificial intelligence and artificial life techniques to design animal housing.* Journal of Animal Science, 1998. 76(10): p. 2609-2613.
32. Dobbie, S., et al., *Agent-based modelling to assess community food security and sustainable livelihoods.* Journal of Artificial Societies and Social Simulation, 2018. 21(1): p. 1-23.
33. Lansing, J.S. and J.N. Kremer. *Emergent properties of balinese water temple networks: co-adaption on a rugged fitness landscape*. in *Artificial Life III. Studies in the Sciences of Complexity*. 1994. Addison Wesley p. 201-223.
34. Schlüter, M., B. Müller, and K. Frank, *The potential of models and modeling for social-ecological systems research: the reference frame ModSES.* Ecology and Society, 2019. 24(1): p. 1-31.
35. Rebaudo, F. and O. Dangles, *Adaptive management in crop pest control in the face of climate variability: An agent-based modeling approach.* Ecology and Society, 2015. 20(2): p. 1-18.
36. Powles, S.B., *Evolved glyphosate-resistant weeds around the world: lessons to be learnt.* Pest Management Science, 2008. 64(4): p. 360-365.
37. Bronson, K., *Looking through a responsible innovation lens at uneven engagements with digital farming.* NJAS - Wageningen Journal of Life Sciences, 2019. 90-91: p. 100294.
38. Berger, T., *Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis.* Agricultural Economics, 2001. 25(2-3): p. 245-260.
39. Choudhary, V., *Comparison of Software Quality Under Perpetual Licensing and Software as a Service.* Journal of Management Information Systems, 2014. 24(2): p. 141-165.
40. Kujala, S., et al., *UX Curve: A method for evaluating long-term user experience.* Interacting with Computers, 2011. 23(5): p. 473-483.
41. Lowder, S.K., J. Skoet, and T. Raney, *The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide.* World Development, 2016. 87: p. 16-29.
42. Shults, F.L., et al. *Artificial Societies in the Anthropocene: Challenges and Opportunities for Modeling Climate, Conflict, and Cooperation*. in *Winter Simulation Conference (WSC)*. 2021. IEEE p. 1-12.DOI: doi: 10.1109/WSC52266.2021.9715391.
43. Berger, T. and C. Troost, *Agent-based Modelling of Climate Adaptation and Mitigation Options in Agriculture.* Journal of Agricultural Economics, 2014. 65(2): p. 323-348.
44. Marin, A., P.V. Zwanenberg, and A. Cremaschi, *Bioleft: A collaborative, open source seed breeding initiative for sustainable agriculture*, in *Transformative Pathways to Sustainability: Learning Across Disciplines, Cultures and Contexts*, A. Ely, Editor. 2021, Routledge, Taylor and Francis: London. p. 90-108.
45. Van Zwanenberg, P., et al., *Seeking unconventional alliances and bridging innovations in spaces for transformative change: the seed sector and agricultural sustainability in Argentina.* Ecology and Society, 2018. 23(3): p. 1-11.
46. Roling, N.G. and M.A.E. Wagemakers, eds. *Facilitating sustainable agriculture: participatory learning and adaptive management in times of environmental uncertainty*. 2000, Cambridge University Press: Cambridge, United Kingdom. 348.

A Participatory Complex Systems Modelling Approach Towards Rewilding in the UK

Imran Khan[1] and Christopher Sandom[2]

[1] School of Physics, Engineering, and Computer Science, University of Hertfordshire, Hatfield, UK, AL10 9AB
[2] School of Life Sciences, University of Sussex, Brighton, UK, BN1 9RH
i.khan9@herts.ac.uk

Introduction

The year 2021 marked the start of the United Nation's Decade of Ecosystem Restoration (Fischer et al., 2021). Nature recovery efforts aim to bring back the diversity of life to areas that have been degraded by humans: to change how people interact with the environment, and to allow more plants and animals to thrive alongside humans. For land owners/managers (i.e. those responsible for managing land resources), there is an increasing interest in adopting a more nature-friendly land management approach (Suding, 2011).

Numerous options to nature recovery exist, including "regenerative agriculture" (farming approaches that aim to have lower or net-positive environmental and social impacts) (Newton et al., 2020) to the more recent approach of "rewilding", which aims to *"restore self-sustaining and complex ecosystems with interlinked ecological processes that promote and support one another while gradually minimising human intervention"* (Perino et al., 2019). One approach to rewilding, called "trophic rewilding", looks to achieve this through the (re)introduction of missing "keystone" species: species that, despite having a low abundance, play a crucial role in improving ecosystem connectivity, increasing trophic complexity and biodiversity, and restoring the autonomy of natural processes (Perino et al., 2019).

For land managers and restoration professionals interested in rewilding, there remains uncertainty around the best-suited rewilding options for a particular ecological context. While practitioners are considering taking more data-driven decisions towards nature restoration, the underlying science for understanding precisely who the keystone species are, what populations are required, and what effects these changes may have on their wider ecosystem, remains difficult to access, understand and, therefore, apply.

Researchers and practitioners in Artificial Life (i.e. ALifers) are well-positioned to address this issue. While nature recovery is fundamentally a complex, ecological problem, (the science underpinning) rewilding efforts can be approached through the lens of complex systems modelling. The question of "What are the effects of species X on ecosystem Y given its properties Z?" can, with sufficient information about the constituent components, be modelled as a complex system. These modelling approaches can be coupled with local, expert ecosystem knowledge, to provide a starting point in tackling one of the "grand ecological challenges" of systems ecology (Martinez, 2020).

Current Work

We describe ongoing work where we look to approach this challenge. Working with ecologists and real-world stakeholders, we have (co-)developed an interactive tool to assist land managers or owners in their rewilding efforts: by helping them explore and engage with their rewilding options and opportunities, as well as to understand some of the potential effects that the (re)introducion of (several) species may have on their local ecosystem.

Taking a participatory approach towards the development of this tool through regular stakeholder engagement and participatory workshops, we place real-world stakeholders and local community experts at the heart of our approach. As part of this project, we aim to address issues around the accessibility of underlying science, the competency in engaging with, and interpreting, complex systems models and to improve the autonomy of end-users for making data-driven decisions when approaching ecological problems.

Our intention is **not**, however, to develop a strictly prescriptive or predictive tool. Moving from the *"myth of the technological fix"* (Oelschlaeger, 1979), the tool aims to complement, not replace, the knowledge of real-world stakeholders—local ecologists, restoration experts, and land owners—by synthesising knowledge from on-the-ground experts with ALife-inspired modelling approaches, in an attempt to improve ecosystem recovery efforts.

Tool Development though Participatory Modelling and Design

Through a series of workshops and ongoing engagement between January-March 2022 with stakeholders—who include experts in conservation ecology and nature restoration (via Rewilding Britain Network, who work closely with land managers in the UK)—we identified three core requirements that the proposed tool sought to meet. Specifically, these stakeholders need a solution to help them understand: (1) which species might be suitable for (re)introduction given the properties (size, net primary productivity, vegetation type) of their land, (2) the possible and appropriate population densities for each available species, and (3) the possible resultant interactions between species and on the wider ecosystem resulting from (re)introducing (several) species. These form the starting point of the tool development, which can be broken down into two distinct, (concurrently-developed) parts: the construction of the complex system (i.e. the underlying mathematical models), and the user interface, to provide stakeholders with the ability to interact directly with the model to run their own simulations, and to provide comprehensible outputs of the results.

The system was modelled through collaborative efforts with one key stakeholder—a restoration ecologist (CS) working as an advisor to land managers in the UK—who also provided access to the necessary data to construct the model. Properties of system components (i.e., different carnivore and herbivore species) were modelled using a range of existing datasets (Middleton et al., 2021; Lundgren et al., 2021; Kissling et al., 2014; Sandom et al., 2017; Faurby et al., 2018; Santini et al., 2018; Middleton, 2021) which included data related to a species' dietary requirements, its natural density distributions, and physiological data (e.g. mass, metabolic rates). Additional datasets related to habitat suitability and additional dietary preferences of species, were compiled in collaboration with CS. Interactions between system components (such as how herbivore traits affect different types of vegetation) were derived through relevant ecological literature (Carbone and Gittleman, 2002), and constructed as mathematical models between system components. Regular evaluations of the system was performed by CS, with the intention of both enhancing the stakeholders' understanding of the underlying system to improve the confidence in its final results (Penn et al., 2013), but also to identify (and rectify) any (potential) erroneous modelling that may have occurred.

Similarly, the (co-)design of the user interface as well as the outputs of the simulation were developed through a combination of previously-established user requirements (provided by CS), as well as engagement with end-users/stakeholders during workshops: with further feedback provided by an expert in human-computer interaction and participatory design. Taking feedback from these workshops, the user interface underwent numerous iterations, with respect to both design and usability, as the complexity of the tool evolved. Involving key stakeholders throughout the course of the interface development ensures that those issues related to accessibility and comprehension of the underlying data are addressed, allowing these end-users to derive the intended value from the tool. A prototype of this tool is available at rewildingdemo.imytk.co.uk.

Concluding Remarks

The UN's Decade of Ecosystem Restoration has reignited nature recovery efforts, with rewilding efforts being considered as an option towards driving desirable outcomes for ecosystems. While the science of nature restoration is slowly improving, it still remains difficult to access and understand for practitioners. Ecosystems are remarkably complex. These complexities and uncertainties around the longer-term outcomes of rewilding interventions may make it difficult for restoration experts to fully realise and explore their options based (solely) on their expert knowledge.

The nature of such a complex (dynamical) system makes it near-impossible to construct a complete model that accurately accounts for all of its components and interactions. These incomplete systems may give rise to potentially-attractive phenomena for ALifers, such as chaos or emergence. However, it is these same, inexplicable outcomes that may cause some end-users (particularly those engaging with real-world problems and whose decisions have significant ecological and financial consequences) to approach these models with caution, or may hinder their buy-in altogether. Central to our approach, therefore, is ensuring that the expert knowledge provided by stakeholders remains a key component in the development of the model. At the same time, we maintain transparency of the (current) incompleteness of our system, instead allowing end-user expertise to complement, and even supersede, its recommendations.

Our present approach does not aim to be *prescriptive* in its recommendations, nor does it currently attempt to *predict* the long-term ecological effects of rewilding interventions. Rather, it provides a snapshot in time: immediate insights into "what if?" questions which, when coupled with the contextual knowledge of a local ecosystem, can empower real-world stakeholders to make data-driven decisions, by being able to use data and knowledge that has previously been inaccessible to them. In this way, ALifers are able to lend their "ALife tools"—those that have allowed us to instantiate our own thought experiments—to other disciplines and beyond.

Far from providing technological solutions to socio-ecological problems, ALifers interested in tackling these issues may instead benefit through collaborative efforts and participatory approaches with other disciplines and real-world stakeholders. With due consideration for the expertise provided by both academic and non-academic experts in other fields: as we demonstrate, ALifers may be able to bridge the gap for real-world stakeholders into better accessing, understanding, and engaging with the underlying science that may facilitate action. Our ALife-inspired methods and perspectives, then, can be used as a catalyst in finding solutions for some of the grand societal and ecological challenges that concern us at this present moment.

Acknowledgements

This work was funded by the Natural Environment Research Council (NERC) Discipline Hopping Grant through the University of Sussex, UK.

References

Carbone, C. and Gittleman, J. L. (2002). A common rule for the scaling of carnivore density. *Science*, 295(5563):2273–2276.

Faurby, S., Davis, M., Pedersen, R. , Schowanek, S. D., Antonelli1, A., and Svenning, J.-C. (2018). Phylacine 1.2: The phylogenetic atlas of mammal macroecology. *Ecology*, 99(11):2626–2626.

Fischer, J., Riechers, M., Loos, J., Martin-Lopez, B., and Temperton, V. M. (2021). Making the un decade on ecosystem restoration a social-ecological endeavour. *Trends in Ecology Evolution*, 36(1):20–28.

Kissling, W. D., Dalby, L., Fløjgaard, C., Lenoir, J., Sandel, B., Sandom, C., Trøjelsgaard, K., and Svenning, J.-C. (2014). Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecology and Evolution*, 4(14):2913–2930.

Lundgren, E. J., Schowanek, S. D., Rowan, J., Middleton, O., Pedersen, R. Ø., Wallach, A. D., Ramp, D., Davis, M., Sandom, C. J., and Svenning, J.-C. (2021). Functional traits of the world's late quaternary large-bodied avian and mammalian herbivores. *Scientific data*, 8(1):a17 1–21.

Martinez, N. D. (2020). Allometric trophic networks from individuals to socio-ecosystems: Consumer–resource theory of the ecological elephant in the room. *Frontiers in Ecology and Evolution*, 8.

Middleton, O. (2021). CarniTRAIT: A dataset of carnivore traits.

Middleton, O., Svensson, H., Scharlemann, J. P., Faurby, S., and Sandom, C. (2021). Carnidiet 1.0: A database of terrestrial carnivorous mammal diets. *Global Ecology and Biogeography*, 30(6):1175–1182.

Newton, P., Civita, N., Frankel-Goldwater, L., Bartel, K., and Johns, C. (2020). What is regenerative agriculture? a review of scholar and practitioner definitions based on processes and outcomes. *Frontiers in Sustainable Food Systems*, 4:194.

Oelschlaeger, M. (1979). The myth of the technological fix. *Southwestern Journal of Philosophy*, 10(1):43–53.

Penn, A. S., Knight, C. J., Lloyd, D. J., Avitabile, D., Kok, K., Schiller, F., Woodward, A., Druckman, A., and Basson, L. (2013). Participatory development and analysis of a fuzzy cognitive map of the establishment of a bio-based economy in the humber region. *PloS one*, 8(11):e78319.

Perino, A., Pereira, H. M., Navarro, L. M., Fernández, N., Bullock, J. M., Ceaușu, S., Cortés-Avizanda, A., van Klink, R., Kuemmerle, T., Lomba, A., et al. (2019). Rewilding complex ecosystems. *Science*, 364(6438):eaav5570.

Sandom, C. J., Williams, J., Burnham, D., Dickman, A. J., Hinks, A. E., Macdonald, E. A., and Macdonald, D. W. (2017). Deconstructed cat communities: quantifying the threat to felids from prey defaunation. *Diversity and Distributions*, 23(6):667–679.

Santini, L., Isaac, N. J. B., and Ficetola, G. F. (2018). Tetradensity: A database of population density estimates in terrestrial vertebrates. *Global Ecology and Biogeography*, 27(7):787–791.

Suding, K. N. (2011). Toward an era of restoration in ecology: successes, failures, and opportunities ahead. *Annual review of ecology, evolution, and systematics*, 42:465–487.

Detecting New Phase Transition Points in Large-Scale Numerical Simulations of an Adaptive Social Network Model

Hiroki Sayama[1,2]

[1]Center for Collective Dynamics of Complex Systems, Binghamton University, State University of New York, USA

[2]Waseda Innovation Lab, Waseda University, Japan

sayama@binghamton.edu

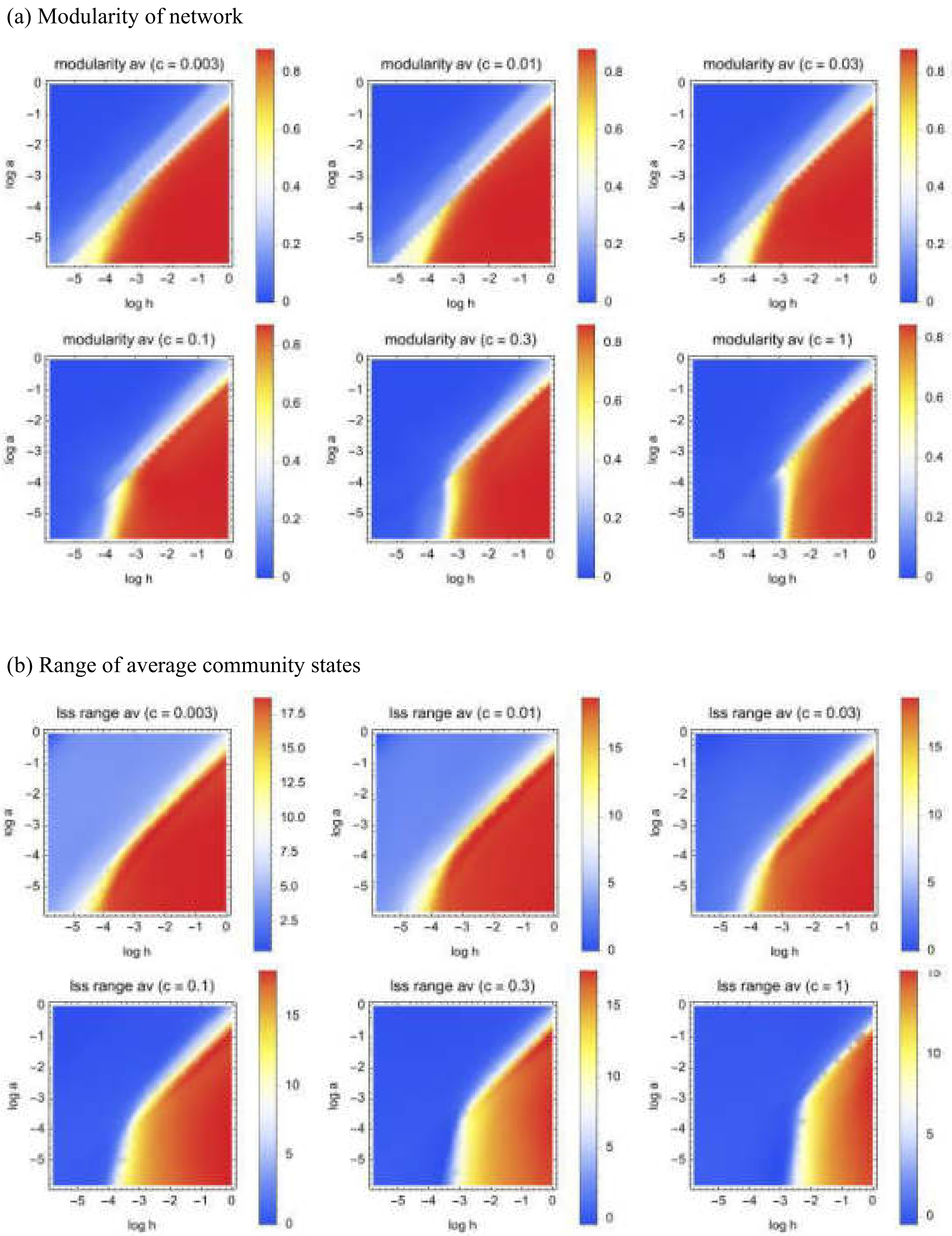
Understanding social fragmentation transition, i.e., transition of social states between many disconnected communities with distinct ideas and a well-connected single network with homogeneous ideas, is a timely research topic with high relevance to various current societal issues (Blex & Yasseri, 2020; Kozma & Barrat, 2008; Levin et al., 2021; Sasahara et al., 2021). We had previously studied this problem using numerical simulations of adaptive social network models (Sayama, 2020) and found that two individual behavioral traits, *homophily* (i.e., tendency to strengthen connections to similar agents and weaken those to dissimilar ones) and *attention to novelty* (i.e., tendency to strengthen connections to agents whose opinions stand out compared to others), among others, had the most significant impact on the outcomes of social network evolution. Specifically, when homophily was strong, the social network evolved into fragmented states of many disconnected clusters with diverse opinions, but when attention to novelty was strong, the social network evolved to well-connected yet informationally homogeneous states. However, the previous study was rather limited in terms of the range of parameter values examined, and possible interactions between multiple behavioral traits were largely ignored, especially about the other behavioral trait, *social conformity* (i.e., how strongly agents assimilate themselves to social neighbors).

In the present study, we have examined a broader spectrum of social network behaviors through a larger-scale numerical experiment with expanded parameter sweep ranges by an order of magnitude in each parameter dimension. Specifically, each of the five model parameters (c for conformity, h for homophily, a for attention to novelty, and two other threshold parameters; see (Sayama, 2020) for details) was varied over {0.003, 0.01, 0.03, 0.1, 0.3, 1} and each parameter value combination was simulated five times with independently generated random initial conditions. The whole set of simulations was conducted for three network sizes (n = 30, 100, 300). This resulted in a total of 116,640 simulation runs, taking a substantial amount of computational time and resource. The simulations were conducted in Python.

In order to capture nontrivial, nonlinear interactions among behavioral parameters, we have modeled and visualized the outcome dependence on parameters using neural networks using Wolfram Research Mathematica 12's neural network predictor. The following five outcome measures were obtained and modeled from the final configuration of each simulation: average edge weight, number of communities, modularity, range of average community states, and standard deviation of average community states (Sayama, 2020). The first three outcome measures captured the network topology, while the last two did the opinion states of the social network.

Results are shown in Figure 1 for modularity and range of average community states. The competition between homophily (h) and attention to novelty (a) is still observed as the primary determinant of social fragmentation in a low-conformity (c) regime (top rows in Fig. 1a and 1b), seen as the diagonal phase transition line in the plots. However, another vertical transition line emerges at an intermediate homophily level in a high-conformity regime (bottom rows in Fig. 1a and 1b), which was not previously known. This new result shows that, when agent's social conformity is sufficiently strong, social homogenization can occur even without attention to novelty. This also implies a previously unrecognized competition between social conformity (c) and homophily (h) in low-a regions (near the bottom edge of each plot in Fig. 1). Namely, when c is low, social fragmentation dominates, but when c is high, social homogenization emerges for sufficiently small values of h.

This study has detected new phase transition points in the adaptive social network model and demonstrated nontrivial, nonlinear interactions among the multiple behavioral mechanisms. In particular, the competition between social conformity and homophily (observed in the absence of attention to novelty) is intriguing because both behaviors have very similar effects at an individual agent level (i.e., they both make ego and alter similar to each other) and their differences are often vague and undetectable in empirical social network studies (Shalizi & Thomas, 2011). Meanwhile, those two behaviors are mathematically distinct since social conformity is about node dynamics while homophily is about edge dynamics. The finding that their competition may lead to very different societal outcomes down the road, offers a lot of implications for how we should rethink/redesign our social interactions in this highly interconnected world.



(a) Modularity of network

(b) Range of average community states

Figure 1: Phase diagrams of adaptive social network evolution. Results of large-scale parameter-sweep numerical simulations were modeled and visualized using neural networks. Each plot shows outcome dependence on homophily (h, horizontal axis), attention to novelty (a, vertical axis) and conformity (c, varied from top-left to bottom-right). (a) How final network topology (modularity) depends on h, a and c. (b) How final opinion diversity (range of average community states) depends on h, a and c. Red and blue regions correspond to fragmented and homogenized network states, respectively. Number of nodes n = 300. Similar patterns were observed for other outcome measures and network sizes.

Acknowledgments

This work was supported in part by JSPS KAKENHI Grant Number 19K21571.

References

Sayama, H. (2020). In *ALIFE 2020: Artificial Life Conference Proceedings*, pages 113-120. MIT Press, Cambridge, MA.

Blex, C., & Yasseri, T. (2020). Positive algorithmic bias cannot stop fragmentation in homophilic networks. *The Journal of Mathematical Sociology*, 1-18.

Kozma, B., & Barrat, A. (2008). Consensus formation on adaptive networks. *Physical Review E*, *77*(1), 016102.

Levin, S. A., Milner, H. V., & Perrings, C. (2021). The dynamics of political polarization. *Proceedings of the National Academy of Sciences*, *118*(50).

Sasahara, K., Chen, W., Peng, H., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2021). Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, *4*(1), 381-402.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, *40*(2), 211-239.

Innovation and informal knowledge exchanges between firms

Juste Raimbault[1,2,3,4]

[1]LASTIG, Univ Gustave Eiffel, IGN-ENSG, Saint-Mandé, France
[2]Centre for Advanced Spatial Analysis, UCL, London, United Kingdom
[3]UPS CNRS 3611 ISC-PIF, Paris, France
[4]UMR CNRS 8504 Géographie-cités, Paris, France

juste.raimbault@polytechnique.edu

Abstract

Firm clusters are seen as having a positive effect on innovations, what can be interpreted as economies of scale or knowledge spillovers. The processes underlying the success of these clusters remain difficult to isolate. We propose in this paper a stylised agent-based model to test the role of geographical proximity and informal knowledge exchanges between firms on the emergence of innovations. The model is run on synthetic firm clusters. Sensitivity analysis and systematic model exploration unveil a strong impact of interaction distance on innovations, with a qualitative shift when spatial interactions are more intense. Model bi-objective optimisation shows a compromise between innovation and product diversity, suggesting trade-offs for clusters in practice. This model provides thus a first basis to systematically explore the interplay between firm cluster geography and innovation, from an evolutionary perspective.

Introduction

Innovation is a central process of evolution, from biological evolution to social, cultural (Mesoudi and Thornton, 2018) and technological evolution (Sood and Tellis, 2005). Understanding the drivers of technological innovation is in that context crucial from a theoretical perspective for insights into evolutionary theories of social change and evolutionary economics among others, and from a practical perspective for sustainable planning and management of societies. Technological innovation may indeed be an essential aspect of transitions towards sustainability (Adams et al., 2016), although it should not be their sole driver at the detriment of other dimensions of transitions such as social change.

Geographical proximity, or in practice the implementation of firm clusters, is thought to have a positive impact on innovation capabilities (Bittencourt et al., 2019). In that context, the role of local informal interactions between innovation agents has been suggested as important for breakthrough innovations by empirical and theoretical studies. In the context of firm cluster, Gnyawali and Srivastava (2013) propose the intensity of social interaction as a key factor alongside cluster competition intensity to determine potential future innovations. Clusters are understood as enablers of tacit knowledge exchanges between inventors from different firms (Arikan, 2009). Furthermore, the mobility of employees between firms in the same area may be a support for the transfer of competences and tacit knowledge (Almeida and Kogut, 1999). Firms benefit from a stronger connection in local social networks (Kemeny et al., 2016). The idea of firms as innovation incubators in which ideas evolve can be linked to an evolutionary approach to social systems which has been widely studied by the Artificial Life community (Marriott et al., 2018).

From an evolutionary perspective, the concept of market niche has been used to explain technological change (Schot and Geels, 2007). The firm in that context acts as the primary space where evolution of ideas occurs, and knowledge flows between firms can be understood in analogy with gene flows between isolated geographical areas in biological innovation. The transfer of concepts from biology to economic geography remains however valid only to a certain extent (Schamp, 2010), and a precise definition of genomes, species, evolution and co-evolution in social systems is not straightforward (Raimbault, 2019). Regarding innovation, multiple scales from firms to cities can be for example considered (Raimbault, 2020). We choose here to focus on the microscopic scale of innovation ecosystems, more precisely how research and development employees of firms act as carriers of ideas leading to the emergence of breakthrough innovations (Song, 2016).

One privileged tool to study and simulate the emergence of innovations from this microscopic perspective are agent-based models (ABM). Various ABMs have been proposed for the diffusion of innovation (Kiesling et al., 2012). Sayama and Dionne (2015) use ABMs to simulate an ecology of ideas and study collective decision making and creativity. Lopolito et al. (2013) combine knowledge exchange, expectations of agents and learning as core mechanisms to simulate innovation niches. Dosi et al. (2021) introduce an ABM to investigate the role of patenting on the innovativity of firms competing on a set of submarkets, including consumer demand. Chen and Chie (2006) focus on functional modularity of products and use a genetic programming formalism to evolve technologies. Ma and Nakamori (2005) describe an ABM of technological change which takes into account both intrinsic fitness selection pressure and environment selection pressure, the latest being determined by the interaction with customers. The role of space is studied by Vermeulen and Pyka (2018) in a multi-level approach combining interregional knowledge networks and knowledge diversity within regions. Diverse aspects of firm clusters have been studied by means of ABMs, such as firm competitiveness, local networks, and policy-making, among others (Fioretti et al., 2005).

These previous modeling efforts however do not specifically tackle the particular question of informal knowledge flows within firm clusters. It remains still an important dimension, with implications in urban planning and concrete aspects of firm cluster implementation among others. We propose in this paper to study this issue by developing a stylised agent-based model of technological change within firm clusters. Following the approach of artificial societies (Epstein and Axtell, 1997), we do not aim at providing a highly realistic or data-driven model, but rather a simple tool to explore the interplay between basic mechanisms in the emergence of a macroscopic phenomenon (innovation within firms in our case). More precisely, our contribution relies on the following points: (i) we provide a simple ABM linking innovation within firms and informal knowledge flows within firm clusters, based on an evolutionary model for innovation and exhibiting a strong analogy with biogeography optimisation algorithms (Simon, 2008); (ii) the model is implemented with a specific instance for the genotype-phenotype mapping, using a generalised Rastrigin function as synthetic fitness landscape; (iii) the model is systematically explored, using global sensitivity analysis and a genetic algorithm optimisation to unveil various patterns of innovation in firm clusters.

The rest of this paper is organised as follows: we first describe and formalise the agent-based model; we then describe results of various numerical experiments; and finally discuss the implications of these results and diverse potential model developments and applications.

Agent-based model

Rationale

The main structure of the model corresponds to a set of firms, each composed by a set of employees. An employee is represented by some ideas, and these are mixed through evolution crossover within firms, but also mutated at the scale of each individuals. Indeed, empirical evidence using patents as a proxy of innovation suggest that inventions are produced by the superposition of exploration (recombination of existing technologies, which would correspond to a crossover in the genome) and exploitation (small incremental changes to existing combinations, captured by a local gene mutation) processes (Youn et al., 2015).

The main feature of the model is an additional crossover between firms, which captures the process of informal knowledge flows. In practice, employees of different firms in the same sector living in the same geographical area will share connections through social networks, meet intentionally or unintentionally, and share ideas. Although in many cases professional secrecy is strictly observed, a tension with knowledge sharing exists (Rouyre and Fernandez, 2019). Therefore, informal knowledge is still exchanged, on non-sensitive subjects such as work or management practices, or technical subjects unrelated to the company's core business. We model interactions between employees of different firms through spatial interaction modeling (Wilson, 1975), which has already been used to study innovation and knowledge spillovers (LeSage et al., 2007). In practice, adding this geographical component makes our model closer to biogeography optimisation algorithms (Simon, 2008).

At the scale of intra-firm innovation, we need to introduce an evolution model. Therefore, it is necessary to specify a selection process linked to some mapping between the genotype of inventions and their phenotype, in other words a fitness function. Ma and Nakamori (2005) use a linear mapping obtained by applying a constant matrix to the genome, inspired from the model of Kauffman and Macready (1995). The concept of fitness landscape is applied in different streams of complexity economics (Khraisha, 2020). We choose to work with a similar heuristic, using a complicated fitness landscape obtained from the genome. Our model is in practice applicable with any optimisation landscape, but for the sake of simplicity we will work below with a generalised Rastrigin function which is a classical difficult optimisation problem used as a benchmark for optimisation algorithms.

Model description

The core element of the agent-based model are firms f_k with $1 \leq k \leq N_f$. These are located in space by coordinates $(x_k, y_k) \in \mathbb{R}^2$. Each are composed by a set of employees e_{ki} with $1 \leq i \leq S_k$ where S_k is the size of the firm. In principle, number of firms, locations and sizes can take any value, but we will parametrise them with realistic values as detailed later. An employee is summarised by a set of ideas, captured by a real genome of fixed size: $e_{ki} = (x_j^{(ki)})(t) \in \mathbb{R}^G$ where G is the genome size. These ideas will evolve as time step t changes. We do not include more detailed employee characteristics such as home location, assuming that spatial interaction modeling captures microscopic interactions around firms. We also do not include competences or field, as the innovation model is a simple genetic algorithm without detailed economic structure. At a given time step, a firm is also characterised by its current product $p_k(t) = (p_{kj})(t) \in \mathbb{R}^G$ and the corresponding fitness value $y_k(t)$ (which can be interpreted as a societal value of the innovation, or as the turnover potential of the product for the firm).

Starting from an initial state at $t = 0$, the model proceeds iteratively to evolve and exchange ideas, and to innovate within firms, for a fixed number of time steps until final time t_f. At each time step, the following actions are taken in order.

1. Ideas are exchanged within firms, corresponding to the core of the genetic algorithm capturing innovation. Each employee has a fixed probability p_C of realising a crossover with another one. Following a random draw, if this is realised, one other employee is selected at random, and the current employee copies a fixed share s_C of the other genome (obtained in practice with a random draw of probability s_C for each gene). The update is done synchronously so that no propagation can occur within one time step. The exchange is not symmetric (reflecting the asymmetry of idea exchanges in real life), but each employee gets to renew its ideas. Genomes are then mutated with a probability p_M (at the gene level for all employees), and mutations correspond to a uniform random increment $m \in [-x_M/2; x_M/2]$ where x_M is a parameter giving the amplitude of the mutation.

2. New ideas are tried by employees, in other words the fitness function y is evaluated for each genome $y_{ki} = y(e_{ki})$. Within each company, the next product is selected as the one maximising the fitness: $p_k(t) = \text{argmax}_i y_{ki}$ and the corresponding fitness value is taken as $y_k(t)$. At this stage, the analogy with the genetic algorithm is slightly modified to reflect actual research processes within firms: a fixed share of employees s_P of the firm is randomly chosen to work on the product during the next cycle, and thus update their genome as the product genome $e_{ik} = p_k$. This leads to a loss of diversity which may be detrimental to a genetic algorithm with the sole aim of optimisation, but this is not the purpose of our model which is to capture actual innovation processes in a stylised way.

3. Informal knowledge flows occur between firms, in practice being carried by local social networks of employees and their daily activities within the geographical area of the cluster. We assume informal flows within firms are indistinguishable from formal work exchanges accounted for at the first step, and this step only captures inter-firms interactions. For any pair of employees $(e_{k_i i}, e_{k_j j})$ from distinct firms k_i, k_j, a probability of interacting is given by
$$p_{ij} = p_E \cdot \exp\left(-d(k_i, k_j)/d_E\right)$$
where p_E is a parameter giving the local intensity of informal exchanges (which will for example capture the difference between a rural, periurban and urban cluster), $d(k_i, k_j)$ is the geographical distance between firms and d_E is the characteristic distance of interactions. Taking the average on employees and aggregating by firms gives the expected number of interactions between firms as
$$I_{kl} = S_k \cdot S_l \cdot \exp\left(-d(k,l)/d_E\right)$$
which corresponds to a classic spatial interaction model (Wilson, 1975). Two interacting employees will act as for the internal crossover: the first agent copies a random part of the genome using the s_C parameter.

Model indicators

We study various aspect to quantify model dynamics and outcomes. First, innovation utility is measured through fitness values. At each time step, we consider (i) $b(t) = \max_k y_k(t)$ the best fitness value across all firms, and (ii) $\bar{f}(t)$ the average fitness value across firms. We then capture economic inequality between firms, both through the relative fitness difference Δf between the best and the worst performing company, and with the entropy $\mathcal{E}_f$ of the distribution of fitness. Finally, we capture product diversity at the genotype level (which is complementary to previous inequality indicators at the phenotype level), using an average dissimilarity index obtained with cosine similarity:

$$d(t) = \frac{1}{2 \cdot N_f \cdot (N_f - 1)} \sum_{k \neq l} \left(1 - \frac{p_k(t) \cdot p_l(t)}{||p_k(t)|| \cdot ||p_l(t)||}\right)$$

The more diverse phenotypes are, the higher this diversity index will be.

Model setup

A certain number of parameters can be parametrised to match realistic configurations. The size of firms can in practice be approximated by a power law for the largest sizes of the distribution (Growiec et al., 2008). Therefore, we distribute S_k following a rank-size law $S_k = S_0 \cdot k^{-\alpha_S}$ where indices are in decreasing size order, S_0 is the size of the largest firm and α_S the level of hierarchy. Locations of firms are taken randomly in $[0; 100]^2$, such that the order of magnitude for the d_E is in the same interval (corresponding to realistic sizes of urban regions where clusters are generally located). Initial employee genomes are initialised randomly in $[-10; 10]$ and the initial product and corresponding fitness are chosen randomly among employees.

Regarding the fitness landscape, we work on an particular implementation using a function difficult to optimise. We work with a generalised Rastrigin function, that we define here as

$$y(\vec{x}) = -\sum_{i,j} m_{ij} \left[x_i^2 - 10\cos\left(2\pi x_i\right)\right]$$

where m_{ij} is a random uniform static matrix of size $G \times G$ and with coefficient in $[0; 1]$, capturing the random fitness landscape used by Ma and Nakamori (2005), and the rest is the classic Rastrigin function.

We furthermore fix a certain number of parameters which can reasonably correspond to real world values. We take medium-sized companies by taking $S_0 = 100$ and a number of firms $N_f = 10$, corresponding to a medium-sized cluster, such as in the case of several start-ups working in digital services. We fix the genome size at $G = 10$ to avoid exploring too large dimensional spaces. We run the model with $t_f = 100$, corresponding to a magnitude of 10 years if one time steps is roughly one month. The rest of the parameters are left free and will be explored in the numerical experiments.

Results

The model is implemented in `scala` for performance purposes. Simulations and design of experiments are achieved using the software OpenMOLE for model exploration and validation (Reuillon et al., 2013), which provides seamless model embedding, simple access to high performance computing infrastructures and state-of-the-art model sensitivity analysis and validation techniques. Model source code is open and available on the git repository of the project at `https://github.com/JusteRaimbault/InnovationInformal`. Simulation results used in the paper are available on the dataverse repository at `https://doi.org/10.7910/DVN/X8PWPF`.

Explored parameter space corresponds to, when not specified otherwise, the following parameters and ranges: firm size hierarchy $\alpha_S \in [0.1; 2.0]$, crossover probability $p_C \in [0; 1]$, crossover share $s_C \in [0; 1]$, mutation probability $p_M \in [0; 1]$, mutation amplitude $x_M \in [0; 2]$, product work share $s_P \in [0; 1]$, interaction probability $p_E \in [0; 10^{-4}]$ (this highest bound gives already a considerable mixing of ideas leading to a total uniformity of products), and distance decay $d_E \in [1; 100]$.

Statistical convergence

We first proceed to an internal validation experiment, aimed at testing whether model outputs are robust to stochasticity, or in other words if they exhibit good statistical convergence properties. We sample 100 parameter points using a Latin Hypercube Sampling (Giunta et al., 2003), and run 1000 replications of the model for each parameter point. This large number of replications is first necessary to estimate the statistical properties of indicators.

We first look at the variability of indicators, looking at Sharpe ratios defined as $\mu[I] / \sigma[I]$ which μ and σ estimators of mean (resp. variance) for the indicator I. Most indicators, including best and average fitness, fitness entropy and diversity, exhibit a low variability with the first quartile of Sharpe ratios larger than 4 across all parameter points. Fitness relative difference is more stochastic with a median of 1.48. Altogether, indicators have thus a low variability.

We then investigate how to discriminate two estimated average indicator values. A relative distance between averages is given by $2 \cdot \frac{\mu[I]+\mu[J]}{\sigma[I]+\sigma[J]}$, for indicators I, J and estimated across all pairs of parameter points. This value is high for best and average fitness (first quartile at 6.8 and 9.9), low for inequalities (median at 2.5 for relative difference and 0.56 for entropy), and relatively high for diversity (first quartile at 2.8). In the case of normal distributions, a confidence interval of size $\sigma/2$ is obtained with 64 repetitions (as confidence interval size is $2 \cdot 1.96 \cdot \sigma/\sqrt{n}$), so we run our experiments with $n = 100$ to ensure a proper separation of indicator values.

Global sensitivity analysis

We then proceed to a global sensitivity analysis to investigate the respective role of parameters in terms of indicators variance. This technique described by Saltelli et al. (2008) provides aggregate measures of parameter relative importance, both at the first order (all other parameters being fixed) and also capturing interaction effects with other parameters (total order). We take $N = 10,000$ design points for the estimation. Results of sensitivity indices estimation are shown in Table 1. We find that the size distribution of firms influences fitness inequality and diversity, but not performance. Crossover parameters mostly influence the entropy of fitness. The parameter with most influence overall is mutation probability, with 3 indicators being significantly changed. We will fix this parameter in the following to ensure a refined exploration, focusing on second order effects. Share of product adoption within the firm s_P has a significant total order influence on best fitness, what may correspond to the fact that this parameter sometimes induces innovation locks through the loss of diversity. Spatial interaction parameters p_E and d_E influence strongly inequality but not diversity, although some diversity reduction could have been expected from exchanges. Finally, we confirm the low sensitivity to stochastic noise as all indicators have low indices with respect to the random seed. Altogether, this global sensitivity analysis confirms that all mechanisms play a role and that they interact in a complex manner.

Parameter space exploration

We now turn to a more targeted exploration of the parameter space to discuss model behavior. We choose to fix mutation parameters in order to focus on the role of exchanges and geography. We therefore take $p_M = 0.01, x_M = 1$. We also fix current product share, as it is similarly a specific parameter of the genetic algorithm, and in terms of thematic interpretation is internal to firms. We take an intermediate value of $s_P = 0.5$. The crossover parameter s_C is in contrary involved in informal knowledge exchanges. We explore a coarse grid for $s_C \in \{0.25; 0.5\}$, for $p_C \in \{0.25; 0.5\}$ and for $\alpha_S \in \{0.1 : 1.0 : 2.0\}$, combined to a more refined grid for exchange parameters: $\log(p_E) \in \{-7; -6; -5; -4\}$ and

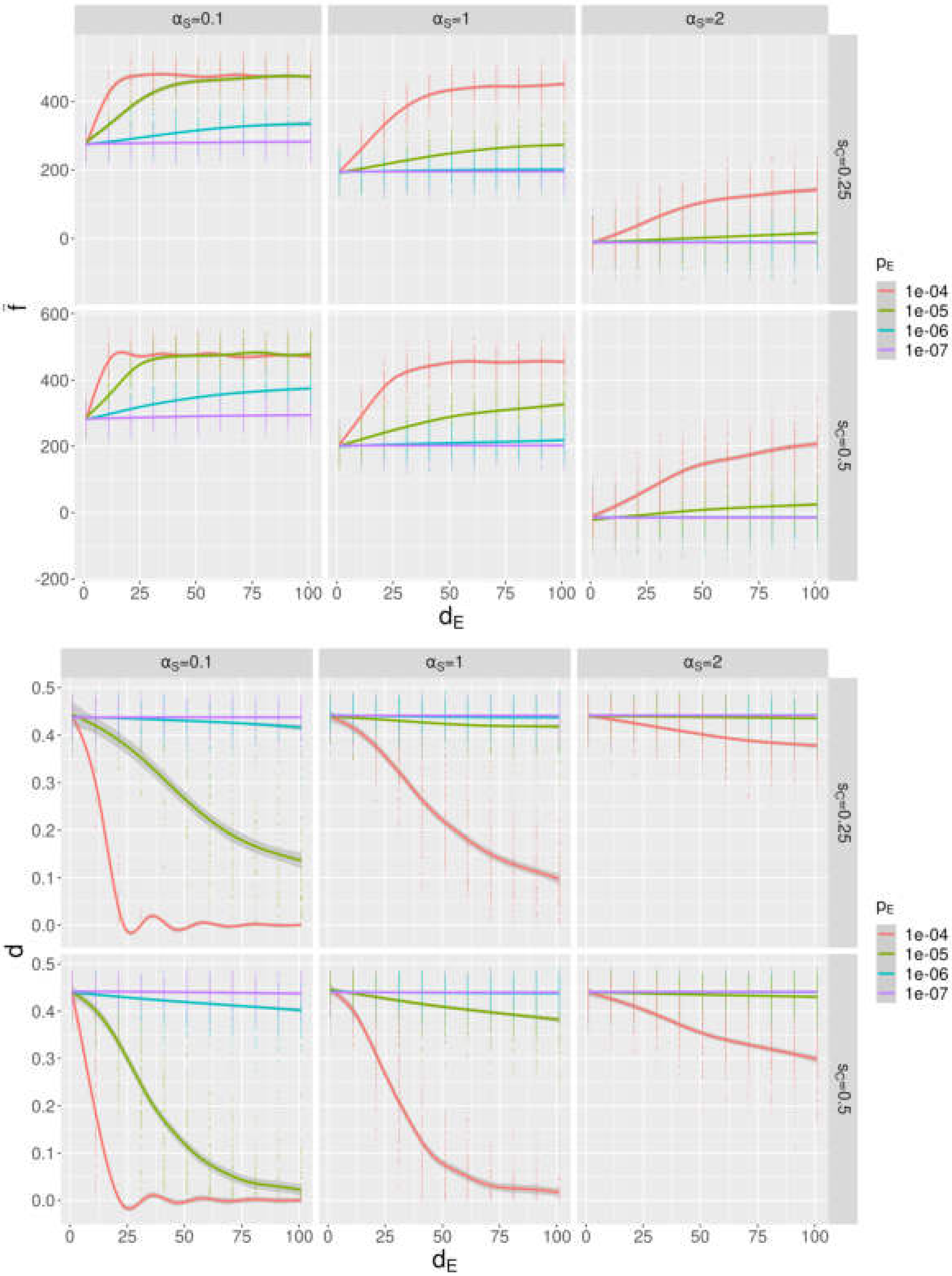


Figure 1: Behavior of average fitness (top plot) and product diversity (bottom plot), as a function of distance decay d_E. We plot raw replication points and smoothed average values, for different values of interaction probability (color scale), and for varying firm size hierarchy α_S (columns) and crossover share s_C (rows). Both plots are shown for $p_C = 0.5$, with no significant qualitative change for $p_C = 0.25$.

Table 1: Saltelli sensitivity indices, for indicators at t_f in rows and parameters in columns. We give for each pair the first order index (F) and the total order index (T). Non-significant values were assimilated to 0.

	α_S		p_C		s_C		p_M		x_M		s_P		p_E		d_E		seed	
	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T
b	0.001	0.002	0.001	0.003	$9 \cdot 10^{-4}$	0.002	0.41	0.75	0.17	0.52	0.03	0.13	$5 \cdot 10^{-4}$	0.002	$9 \cdot 10^{-4}$	0.002	0.003	0.007
$\bar{f}$	0.02	0.07	$6 \cdot 10^{-4}$	0.002	0.0	0.003	0.36	0.69	0.21	0.55	0.02	0.008	0.0	0.004	$4 \cdot 10^{-4}$	0.004	$8 \cdot 10^{-4}$	0.007
Δf	$7 \cdot 10^{-4}$	0.56	0.0	0.9	0.0	0.0	0.003	0.0	0.0	0.24	0.0	0.48	0.0	0.18	0.0	0.17	0.0	0.0
$\mathcal{E}_f$	0.14	0.64	0.0	0.44	0.27	0.36	0.48	0.84	0.014	0.35	0.23	0.41	0.16	0.39	0.0	0.40	0.05	0.46
d	0.007	0.13	0.001	0.04	0.01	0.1	0.45	0.7	0.21	0.42	0.0	0.1	0.003	0.09	0.006	0.09	0.006	0.05

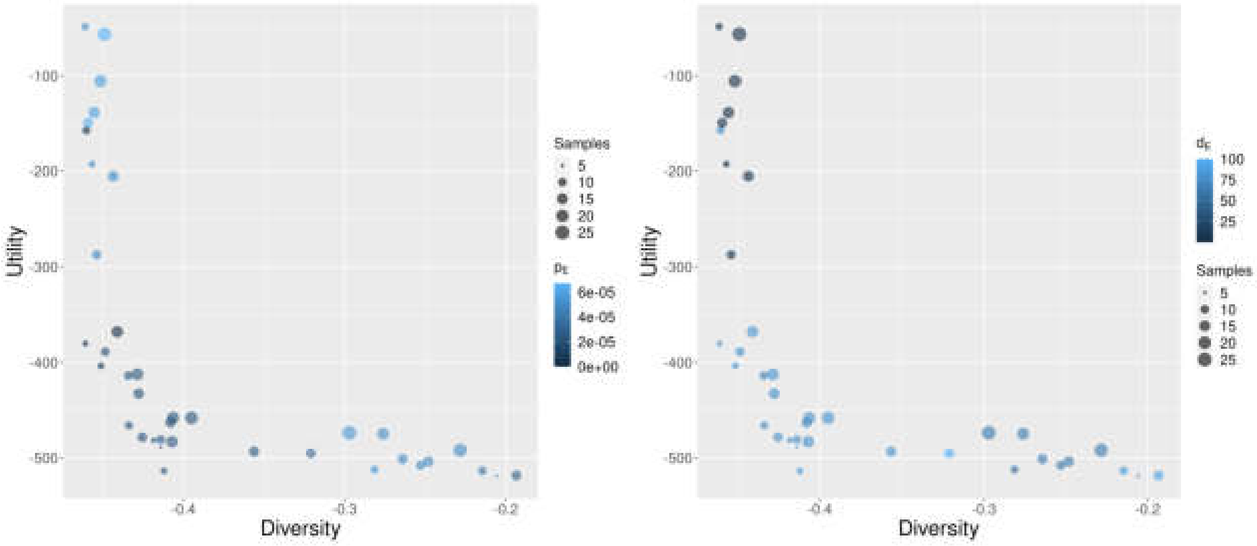


Figure 2: Pareto fronts between average fitness and diversity, obtained with a NSGA2 bi-objective optimisation algorithm. Point size gives the number of stochastic samples, while point color gives interaction probability p_E for the left plot and distance decay d_E for the right plot.

d_E varying from 1 to 101 with a step of 10. We run 100 model replications for each parameter point, corresponding to a total of 52,800 model runs.

We show the behavior of main indicators in Fig. 1. Both average fitness and product diversity exhibit a similar behavior. When distance decays increases, i.e. when the integration of the firm cluster in terms of informal knowledge is strengthened, innovation is improved as the average fitness increases significantly. This effect disappears for low interaction probabilities, recalling that the informal knowledge flow is the outcome of these two processes of social intensity and spatial interactions. The increase in fitness is at the detriment of product diversity. When comparing columns with varying α_S, we find that unequal firm sizes (larger α_S values) are non optimal, as highest fitness values are obtained for the lowest hierarchy. The crossover share s_C (plot rows) does not change much indicator behavior, except for diversity at high interaction probabilities (middle column of bottom plot).

It is interesting to note that changing p_E leads to a qualitative change in model regime: low values mostly imply a steady increase of fitness (decrease of diversity), while intense interaction lead to a sharp increase followed by a plateau. This implies a change in the way to conceive clusters depending on their geographical situation: proximity will be beneficial more quickly in urban environments compared to rural settings for example.

Optimisation

We finally run a bi-objective optimisation algorithm, to investigate the potential compromise between innovation in terms of fitness and product diversity. Indeed, diversity is crucial to maintain for a longer term robustness and resilience of the socio-technical system (Reinmoeller and Van Baardwijk, 2005). We use a NSGA2 genetic algorithm with two optimisation objectives to be maximised: diversity and average fitness. We use the OpenMOLE implementation of a steady state NSGA2, with a population of 200 individuals, for 10,000 generations. In practice, this implementation minimises objectives, so we use opposites $-\bar{f}$ and $-d$ as optimisation objectives. The number of stochastic samples for each parameter point is determined through an embedding strategy, adding this number as an additional optimisation objective. In the final population, we filter points with less than 5 repetitions.

We show the optimisation results in Fig. 2. We find a Pareto front between the two objectives, confirming the compromise between global performance and diversity. Both extremities of the front are rather steep/flat, meaning that a reduced number of points provide an effective compromise. Investigating the values of some parameters, we find some kind of U-shaped behavior for interaction probability p_E: high values for this parameter put the points on extremities of the front. This is an interesting behavior as quite unexpected from an intuitive point of view: increasing interactions should mix more ideas and decrease diversity - which is true, but also includes the opposite, i.e. optimal diversity when interaction are high. The explanation may rely on the fact that these points (top-left extremity of the front) correspond to low values of distance decay d_E, as seen on the right plot of Fig. 2. The localised regime impedes the effect of interactions in that case. We observe a similar behavior with product share s_P, but with different underlying processes: a too high share could have been expected to induce technological locks. This shows altogether the complexity of interacting processes within firm clusters, leading to the emergence of innovations.

We can investigate the part of the Pareto front which would constitute some "reasonable" compromise, i.e. where trade-offs between the two objectives are of similar amplitude. We therefore filter the points such that $-\bar{f} < -400$ and $-d < -0.4$, obtaining 13 compromise points. Interestingly, the parameter values for these points are rather localised. Their values with standard deviations are: $\alpha_S = 0.13 \pm 0.04$, $p_C = 0.94 \pm 0.05$, $s_C = 0.23 \pm 0.05$, $p_M = 0.03 \pm 0.008$, $x_M = 1.26 \pm 0.34$, $s_P = 0.12 \pm 0.05$, $p_E = 2 \cdot 10^{-5} \pm 5 \cdot 10^{-6}$ and $d_E = 77 \pm 7.8$. This corresponds to equal firm sizes, frequent crossovers of a quarter of the genome, very low mutations (as fixed in the grid sampling experiment), a small but not negligible product share (keeping diversity within companies is thus important for the compromise), and a very low interaction probability but at a long range. In practice, this would be interpreted a regional firm system with few but important informal idea exchanges between firms.

Discussion

We explored a novel model for innovation diffusion within and between firms from an evolutionary perspective. One important contribution of this work compared to previous literature is the stylised realistic parametrisation, coupling paradigms from evolutionary computation and economic geography. The main takeovers drawn from our simulations are (i) a strong effect of informal knowledge exchanges on innovation fitness, but which is rapidly plateauing; (ii) a more optimal configurations in terms of fitness when firms are close in size, compared to highly hierarchical firm systems; (iii) a compromise between innovation fitness and diversity, with the trade-off region of similar amplitude being characterised by an equal-size regional firm system. The second point relates with the idea of modular systems being a favourable context for innovation and creativity as found by Dionne et al. (2019). In our model, spatial clusters corresponding to firms play a crucial role. The third point suggests that these clusters are balanced and geographically distributed in the compromise configuration. More generally, the configuration of spatial niches may play an important role in evolutionary systems.

Our stylised results can furthermore be linked with documented empirical facts. The innovation success of a firm cluster relies on a strong interplay between local interactions and global integration (Fitjar and Rodríguez-Pose, 2014). Put in another way, local interactions are not sufficient to drive innovation. However, given the sharp fitness increase exhibited by our model when increasing local knowledge flows, we can suggest that these may be necessary, and that a firm in complete isolation would have difficulties to innovate (whether knowledge flows can occur without being local or informal is another question out of the scope of our work). Furthermore, although empirical stylised facts are not unanimously agreed on, there exists evidence that cluster size may lead to "agglomeration diseconomies", in other words that clustering becomes detrimental above a certain size (Folta et al., 2006). We do not obtain this aspect, since the effect of distance and interaction probability are always increasing in terms of fitness. However, the opposite effects on diversity may be interpreted as detrimental as maintaining diversity is important for the resilience of complex systems (Fraccascia et al., 2018). Finally, in relation with cluster size, we find that firm size hierarchy α_S is to be minimal (firm of equal sizes) to obtain a higher fitness. This implies that clusters should not be dominated by large companies for a better innovation performance.

Numerous extensions and applications are open at this point. More advanced model validation procedures would bring further knowledge on its complex behavior: behavior search algorithms provide a feasible output space (Chérel et al., 2015), while the Calibration Profile algorithm can be applied to conditional optimisation along discrete axis for parameters of interest (Reuillon et al., 2015). The several stylised facts contained within the conceptual model introduced by Gnyawali and Srivastava (2013) may be the basis for more general models which would need to reproduce these facts. The combination of this model with urban innovation diffusion models such as (Raimbault, 2020) could provide a multiscale model of innovation clusters. Exploring other instances of fitness landscapes is also crucial to assess the robustness of our results and to be able to generalise. Regarding aspects that were not taken into account, teleworking can significantly change the role of informal knowledge exchanges and the geography of clusters. It was shown recently to influence the productivity of firms (Bergeaud et al., 2022) and is one component of what Duranton (1999) calls the "tyranny of proximity": face-to-face contacts have a novel importance in that context. Furthermore, this model could be parametrised and calibrated on real world data, including patent data for innovation and real cluster case studies. Finally, this model could have potential policy applications, to plan and manage company clusters to foster innovation in the context of sustainability.

To conclude, we have introduced and explored a simple instance of an innovation diffusion model, focused on informal knowledge flows and the geography of firm clusters. The model was explored for a particular instance of fitness landscape. We find a strong effect of these flows on innovation performance, and a compromise between diversity and innovation corresponding to regional firm systems. These results and the model can be the basis of future empirical, theoretical and modeling research, in link with policy applications.

References

Adams, R., Jeanrenaud, S., Bessant, J., Denyer, D., and Overy, P. (2016). Sustainability-oriented innovation: A systematic review. *International Journal of Management Reviews*, 18(2):180–205.

Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917.

Arikan, A. T. (2009). Interfirm knowledge exchanges and the knowledge creation capability of clusters. *Academy of management review*, 34(4):658–676.

Bergeaud, A., Cette, G., and Drapala, S. (2022). Telework and productivity: Insights from a new survey. *Available at SSRN 4015066*.

Bittencourt, B. A., Galuk, M. B., Daniel, V. M., and Zen, A. C. (2019). Cluster innovation capability: a systematic review. *International Journal of Innovation*, 7(1):26–44.

Chen, S.-H. and Chie, B.-T. (2006). A functional modularity approach to agent-based modeling of the evolution of technology. In *The complex networks of economic interactions*, pages 165–178. Springer.

Chérel, G., Cottineau, C., and Reuillon, R. (2015). Beyond corroboration: Strengthening model validation by looking for unexpected patterns. *PloS one*, 10(9):e0138212.

Dionne, S. D., Sayama, H., and Yammarino, F. J. (2019). Diversity and social network structure in collective decision making: evolutionary perspectives with agent-based simulations. *Complexity*, 2019.

Dosi, G., Palagi, E., Roventini, A., and Russo, E. (2021). Do patents really foster innovation in the pharmaceutical sector? results from an evolutionary, agent-based model. Technical report, LEM Working Paper Series.

Duranton, G. (1999). Distance, land, and proximity: economic analysis and the evolution of cities. *Environment and Planning a*, 31(12):2169–2188.

Epstein, J. M. and Axtell, R. (1997). Artificial societies and generative social science. *Artificial Life and Robotics*, 1(1):33–34.

Fioretti, G. et al. (2005). Agent-based models of industrial clusters and districts. *Contemporary issues in urban and regional economics*.

Fitjar, R. D. and Rodríguez-Pose, A. (2014). When local interaction does not suffice: sources of firm innovation in urban norway. In *Regional development and proximity relations*. Edward Elgar Publishing.

Folta, T. B., Cooper, A. C., and Baik, Y.-s. (2006). Geographic cluster size and firm performance. *Journal of business venturing*, 21(2):217–242.

Fraccascia, L., Giannoccaro, I., and Albino, V. (2018). Resilience of complex systems: State of the art and directions for future research. *Complexity*, 2018.

Giunta, A., Wojtkiewicz, S., and Eldred, M. (2003). Overview of modern design of experiments methods for computational simulations. In *41st Aerospace Sciences Meeting and Exhibit*, page 649.

Gnyawali, D. R. and Srivastava, M. K. (2013). Complementary effects of clusters and networks on firm innovation: A conceptual model. *Journal of Engineering and Technology Management*, 30(1):1–20.

Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2008). On the size distribution of business firms. *Economics Letters*, 98(2):207–212.

Kauffman, S. and Macready, W. (1995). Technological evolution and adaptive organizations. *Complexity*, 1(2):26–43.

Kemeny, T., Feldman, M., Ethridge, F., and Zoller, T. (2016). The economic value of local social networks. *Journal of Economic Geography*, 16(5):1101–1122.

Khraisha, T. (2020). Complex economic problems and fitness landscapes: Assessment and methodological perspectives. *Structural Change and Economic Dynamics*, 52:390–407.

Kiesling, E., Günther, M., Stummer, C., and Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, 20(2):183–230.

LeSage, J. P., Fischer, M. M., and Scherngell, T. (2007). Knowledge spillovers across europe: Evidence from a poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86(3):393–421.

Lopolito, A., Morone, P., and Taylor, R. (2013). Emerging innovation niches: An agent based model. *Research Policy*, 42(6-7):1225–1238.

Ma, T. and Nakamori, Y. (2005). Agent-based modeling on technological innovation as an evolutionary process. *European Journal of Operational Research*, 166(3):741–755.

Marriott, C., Borg, J. M., Andras, P., and Smaldino, P. E. (2018). Social learning and cultural evolution in artificial life. *Artificial life*, 24(1):5–9.

Mesoudi, A. and Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B*, 285(1880):20180712.

Raimbault, J. (2019). Modeling interactions between transportation networks and territories: a co-evolution approach. *arXiv preprint arXiv:1902.04802*.

Raimbault, J. (2020). A model of urban evolution based on innovation diffusion. In *The 2020 Conference on Artificial Life*, pages 500–508. MIT Press.

Reinmoeller, P. and Van Baardwijk, N. (2005). The link between diversity and resilience. *MIT Sloan management review*, 46(4):61.

Reuillon, R., Leclaire, M., and Rey-Coyrehourcq, S. (2013). Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems*, 29(8):1981–1990.

Reuillon, R., Schmitt, C., De Aldama, R., and Mouret, J.-B. (2015). A new method to evaluate simulation models: the calibration profile (cp) algorithm. *Journal of Artificial Societies and Social Simulation*, 18(1):12.

Rouyre, A. and Fernandez, A.-S. (2019). Managing knowledge sharing-protecting tensions in coupled innovation projects among several competitors. *California Management Review*, 62(1):95–120.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Sayama, H. and Dionne, S. D. (2015). Studying collective human decision making and creativity with evolutionary computation. *Artificial Life*, 21(3):379–393.

Schamp, E. W. (2010). On the notion of co-evolution in economic geography. In *The handbook of evolutionary economic geography*. Edward Elgar Publishing.

Schot, J. and Geels, F. W. (2007). Niches in evolutionary theories of technical change. *Journal of Evolutionary Economics*, 17(5):605–622.

Simon, D. (2008). Biogeography-based optimization. *IEEE transactions on evolutionary computation*, 12(6):702–713.

Song, J. (2016). Innovation ecosystem: impact of interactive patterns, member location and member heterogeneity on cooperative innovation performance. *Innovation*, 18(1):13–29.

Sood, A. and Tellis, G. J. (2005). Technological evolution and radical innovation. *Journal of marketing*, 69(3):152–168.

Vermeulen, B. and Pyka, A. (2018). The role of network topology and the spatial distribution and structure of knowledge in regional innovation policy: A calibrated agent-based model study. *Computational Economics*, 52(3):773–808.

Wilson, A. G. (1975). Some new forms of spatial interaction model: A review. *Transportation Research*, 9(2-3):167–179.

Youn, H., Strumsky, D., Bettencourt, L. M., and Lobo, J. (2015). Invention as a combinatorial process: evidence from us patents. *Journal of the Royal Society interface*, 12(106):20150272.

Network-Based Phase Space Analysis of the El Farol Bar Problem

Shane St. Luce[1,2] and **Hiroki Sayama**[1,2,3]

[1]Department of Systems Science and Industrial Engineering, Binghamton University, SUNY, USA
[2]Center for Collective Dynamics of Complex Systems, Binghamton University, SUNY, USA
[3]Waseda Innovation Lab, Waseda University, Japan
sstluce1@binghamton.edu

Abstract

The El Farol Bar problem highlights the issue of bounded rationality through a coordination problem where agents must decide individually whether or not to attend a bar without prior communication. Each agent is provided a set of attendance predictors (or decision-making strategies) and uses the previous bar attendances to guess bar attendance for a given week to determine if the bar is worth attending. We previously showed how the distribution of used strategies among the population settles into an attractor by using a spatial phase space. However, this approach was limited as it required $N - 1$ dimensions to fully visualize the phase space of the problem, where N is the number of strategies available.

Here we propose a new approach to phase space visualization and analysis by converting the strategy dynamics into a state transition network centered on strategy distributions. The resulting weighted, directed network gives a clearer representation of the strategy dynamics once we define an attractor of the strategy phase space as a sink-strongly connected component. This enables us to study the resulting network to draw conclusions about the performance of the different strategies. We find that this approach not only is applicable to the El Farol Bar problem, but also addresses the dimensionality issue and is theoretically applicable to a wide variety of discretized complex systems.

A comprehensive conceptual and computational dynamics framework for Autonomous Regeneration Systems

Tran Nguyen Minh-Thai[1,3], **Sandhya Samarasinghe**[1] and Michael Levin[2]

[1] Complex Systems, Big Data and Informatics Initiative (CSBII), Lincoln University, New Zealand
[2] Allen Discovery Center at Tufts University, Medford MA, USA
[3] College of Information and Communication Technology, Can Tho University, Vietnam

Sandhya.Samarasinghe@lincoln.ac.nz

Abstract

Many biological organisms regenerate structure and function after damage. Hydra (tiny aquatic animals) and planarians (free-living flatworms) are widely used as models of adaptive regeneration. After being injured, cells in any amputated fraction of these animals fully restore tissues and organs to their previous anatomical and physiological state; this phenomenon is known as body-wide immortality. The regeneration of organs is widespread in other animals such as snails, axolotls (an amphibian known as a 'walking fish'), and zebrafish. Regeneration is a process of collective action, through cellular computing in living systems. Despite the long history of research on molecular mechanisms involved in regeneration, many questions remain about algorithms by which cells can cooperate toward the same invariant morphogenetic outcomes. Therefore, conceptual frameworks are needed not only for motivating hypotheses for advancing the understanding of regeneration processes in living organisms, but also for regenerative medicine and synthetic biology. Inspired by the body-wide immortality in planarian regeneration, this study offers a novel generic conceptual framework that hypothesizes mechanisms and algorithms by which cell collectives may internally represent an anatomical target morphology toward which they build after damage. Further, the framework contributes a novel nature-inspired computing method for self-repair in engineering and robotics. Our framework, based on past *in vivo* and *in silico* studies on planarian regeneration, hypothesizes efficient novel mechanisms and algorithms to achieve complete and accurate regeneration of a simple *in silico* flatworm-like organism from any damage, much like the body-wide immortality of planaria, with minimal information and algorithmic complexity. In regeneration, stem cells regenerate new tissues with the help of existing tissue cells and we propose mechanisms for this collaboration. Our framework extends our previous circular tissue repair model and integrates two levels of organization: tissue and organism. In Level 1, three individuals *in silico* tissues (head, body and tail- each with a large number of tissue cells and a single stem cell at the centre) repair themselves through efficient local communications. Here, the contribution is extending our circular tissue model to other shapes and investing them with tissue-wide immortality through an 'information field' that holds the minimum body plan. In Level 2, individual tissues combine to form a simple organism. Specifically, the three stem cells form a network that coordinates organism-wide regeneration with the help of Level 1. Here we contribute novel concepts for collective decision making by stem cells for stem cell regeneration and large-scale recovery. Both levels (tissue cells and stem cells) represent networks that perform simple neural computations and form a feedback control system. Specifically, tissue cells are represented by a network of perceptron (threshold) neurons with local communication and stem cells are represented by a network of linear neurons. With simple and limited cellular computations, our framework minimises computation and algorithmic complexity to achieve complete recovery. We report results from computer simulations of the framework to demonstrate its robustness in recovering the organism after any injury. This comprehensive hypothetical framework that significantly extends the existing biological regeneration models offers a new way to conceptualise the information-processing aspects of regeneration, which may also help design living and non-living self-repairing agents.

The Impossibility of Automating Ambiguity

Abeba Birhane

Mozilla Foundation, USA and University College Dublin, Ireland
abeba@mozillafoundation.org

Abstract

On the one hand, complexity science, and enactive and embodied science approaches emphasize that people, as complex adaptive systems, are ambiguous, indeterminable, and inherently unpredictable. Strong emphasis is placed on the inherently indeterminable nature of the person and the inextricably entangled relationships between person, other, and technology. These traditions have challenged Cartesian ambitions that neatly delineate human behaviour and actions into dichotomies, instead embracing fluidity. Uncertainty, ambiguity, and fluidity, not static dichotomies, exemplify human beings and their interactions (McGann and De Jaegher, 2009). People, far from being static Cartesian selves, are active, dynamic, and continually moving. We are fully embedded and enmeshed with our designed surroundings and we critically depend on this embeddedness to sustain ourselves. Furthermore, our historical paths and the moral and political values that we are embedded in constitute crucial components that contribute to who we are. The idea of defining the person once and for all, drawing clear categories, and making accurate predictions of future behaviours thus appears a seemingly futile endeavour. In complexity science terms, human beings and their behaviour are complex adaptive phenomena whose precise pathway is simply unpredictable (Juarrero, 2000).

On the other hand, Machine Learning (ML) systems that automate, datafy and claim to predict human behaviour are becoming ubiquitous in all spheres of social life. Automation is something that is achieved once a given process is complete, that is, it is understood and discrete, such that it can be implemented from a set beginning to a set finish reliably. People and social systems, however, are partially-open, always becoming, and inherently unfinalizable (Bakhtin, 1984). Automation as complete understanding, therefore, stands at odds with human behaviour which is inherently incomplete making machine categorization, classification and prediction futile. Given the open and incomplete nature of human beings and social systems, automating sensible ambiguity (as opposed to automating nonsense and random) and indeterminability is ill-conceived. A machine capable of grasping humanity, by definition, is capable of grasping open-endedness, incompleteness, fluidity, and ambiguity. Alas, this becomes something other than machines or automation as we know them.

I contend that ubiquitous Artificial Intelligence (AI) and ML systems that sort, categorize, classify the world, and forecast the future are not only scientifically unsound but also ethically dubious. Through the practice of clustering, sorting, and predicting human behaviour and action, these systems impose order, equilibrium, and stability to the active, fluid, messy, and unpredictable nature of human behaviour and the social world at large. The social world is messy and fluctuating but also inundated with persistent social norms, power asymmetries, and historical injustice (Benjamin, 2019). Historical norms and traditions are often unkind and unjust to individuals and groups at the margins of society, and accordingly, attempts to find stable patterns to sort and categorize the social world encode these deeply ingrained norms and injustices. When ML systems pick up patterns and clusters, this often amounts to identifying historically and socially held norms, conventions, and stereotypes. This is not an argument to abandon all ML modelling of complex social systems, but a call for cautions and modest approaches. Given the non-determinable and incompressible nature of complex systems, a single model may only be able to capture a snapshot of the systems at best (Cilliers, 2002). ML approaches that acknowledge the fundamental limitations of attempting to capture complex social systems in data and models, and are modest in their claims pave the way to scientifically rigorous modelling. Furthermore, given existing power asymmetries, historical injustices, and disproportionate negative impacts of ML systems on the most marginalized in society, responsible and just ML requires centreing the needs, welfare and benefit of the most marginalzed.

References

Bakhtin, M. (1984). Problems of dostoevsky's poetics. 1929. *and trans. Caryl Emerson. Minneapolis: U of Minnesota P.*

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code*. John Wiley & Sons.

Cilliers, P. (2002). *Complexity and postmodernism: Understanding complex systems*. Routledge.

Juarrero, A. (2000). Dynamics in action: Intentional behavior as a complex system. *Emergence*, 2(2):24–57.

McGann, M. and De Jaegher, H. (2009). Self–other contingencies: Enacting social perception. *Phenomenology and the Cognitive Sciences*, 8(4):417–437.

Life Worth Mentioning: Complexity in Life-Like Cellular Automata

Eric Peña[1,2] and Hiroki Sayama[1,2,3]

[1]Binghamton University, SUNY, Department of Systems Science and Industrial Engineering
[2]Center for Collective Dynamics of Complex Systems
[3]Waseda University, Waseda Innovation Lab
eric.pena@binghamton.edu

Abstract

Cellular automata (CA) have been lauded for their ability to generate complex global patterns from simple local rules. The late English mathematician, John Horton Conway, developed his illustrious Game of Life (Life) CA in 1970, which has since remained one of the most quintessential CA constructions—capable of producing a myriad of complex dynamic patterns and computational universality. Life and several other Life-like rules have been classified in the same group of aesthetically and dynamically interesting CA rules characterized by their complex behaviors. However, a rigorous quantitative comparison among similarly classified Life-like rules has not yet been fully established. Here we show that Life is capable of maintaining as much complexity as similar rules while remaining the most parsimonious. In other words, Life contains a consistent amount of complexity throughout its evolution, with the least number of rule conditions compared to other Life-like rules. We also found that the complexity of higher density Life-like rules, which themselves contain the Life rule as a subset, form a distinct concave density-complexity relationship whereby an optimal complexity candidate is proposed. Our results also support the notion that Life functions as the basic ingredient for cultivating the balance between structure and randomness to maintain complexity in 2D CA for low- and high-density regimes, especially over many iterations. This work highlights the genius of John Horton Conway and serves as a testament to his timeless marvel, which is referred to simply as: Life.