# Abstract and Essay Student Competition Booklet
# ALIFE 2022

Organized by

**Olaf Witkowski**
**Alyssa Adams**
**Jitka Čejková**
**Julien Hubert**
**Keisuke Suzuki**
**Lisa Soros**
**Silvia Holler**

ALIFE
18-22 July 2022
La DOLCE vita

## List of contributors


| | |
|---|---|
| Louisa | Tierney |
| Musab | Kılıç |
| Andrea | Fanti |
| Federico | Pigozzi |
| Ryan | Wing |
| Shakiba | Shahbandegan |
| Anna | Catenacci |
| Stefan | Riegl |
| Zixuan | Liu |

# Complex Dynamical Systems and Networks

**Louisa Tierney**

Networks are essentially a representation of interactions in a system, revealing the larger structure of the system. Studying the dynamics of these complex networks can reveal an abundance of information, as structure always influences functions in a system (Strogatz, 2001). In one way or another, everything in life is composed of complex networks and interactions. The influential American biologist, Edward Osborne Wilson, once remarked that, "The greatest challenge today, not just in cell biology and ecology but in all of science, is the accurate and complete description of complex systems. Scientists have broken down many kinds of systems. They think they know most of the elements and forces. The next task is to reassemble them, at least in mathematical models that capture the key properties of the entire ensembles" (Wilson, 1999).

Network science helps scientists understand and find solutions to frequent problems in complex systems such as supply chain disruptions, power cuts, or financial network failures (e.g., Perera et al., 2017; Pagani & Aiello, 2013; Battiston et al., 2010). Networks of social interactions may also underpin the cycles of growth and innovations in cities (Bettencourt et al., 2007) and, more speculatively, planetary civilizations (Wong & Bartlett 2022). Complex networks play an important role not only in modern-day human-made issues, but also in all areas of science from microbiology to astronomy (e.g., Barbarasi & Oltavi, 2004; Jolley & Douglas 2012). Over the past several decades the relevance of complex networks along with their role in the natural world has become increasingly clear. The advancement of computational techniques and AI technology has allowed scientists to finally investigate complex systems and networks, giving them a bigger picture of a system's dynamics and innerworkings.

Of particular interest to me is the way that complex networks play a major role in astrobiology. Many scientists are implementing network science to study complex systems like chemical models built to investigate the complex network of reactions converting simple precursor molecules to the complex organics found in comets and meteorites (Jolley & Douglas, 2010) and the universal laws of life on Earth (e.g., West et al., 1997; West & Brown, 2004; Kim et al., 2019).

I have always been drawn to the concept of habitability. Throughout my first three years of high school, I researched exoplanet habitability and the current metrics we use to characterize habitability (e.g., Heller & Armstrong 2014; Schulze-Makuch et al., 2020). The more I researched, the more curious I became about the Earth-centric bias contained in the current definition of habitability, prompting me to investigate for a more representative and universal definition (Wong et al. in review). My interest in habitability combined with the advancement of AI network technology has led me to several potential projects using networks to better understand certain aspects of habitability.

Atmospheric networks contain a significant amount of information about a planet (Solé & Munteanu, 2004) and understanding how atmospheres play into the concept of habitability is a question that scientists are actively trying to answer today. My curiosity lies in the role semantic information, information that contributes to the stability of a system as it evolves in time (Kolchinsky & Wolpert, 2018), plays in atmospheres. Over the summer, I plan to use atmospheric chemical networks to further understand the dynamics of semantic information in atmospheres and its contribution to planetary habitability.

My passion for astrobiology and habitability led me to hear about the ALIFE conference and became overjoyed reading through the topics list, seeing that all my favorite areas of science would be discussed at this conference. Not only do I find genuine interest in the topics of the ALIFE conference, but also there are many areas mentioned that I would love to learn more about. Attending the ALIFE conference would alter my scientific journey, as it would be the first scientific conference I have attended. Hearing about the in-depth research conducted by other scientists in the AI and ALIFE community could open my eyes to more modern research and applications of ALIFE in all areas of science. Overall, the ALIFE conference would give me the opportunity to further my understanding of Artificial Intelligence as well as network science, transforming my scientific journey in the best way possible.

## References

Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101-113.

Battiston, S., Glattfelder, J. B., Garlaschelli, D., Lillo, F., & Caldarelli, G. (2010). The structure of financial networks. In *Network Science* (pp. 131-163). Springer, London.

Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences*, *104*(17), 7301-7306.

Heller, R., & Armstrong, J. (2014). Superhabitable worlds. *Astrobiology*, *14*(1), 50-66.

Jolley, C. C., & Douglas, T. (2010). A network-theoretical approach to understanding interstellar chemistry. *The Astrophysical Journal*, *722*(2), 1921.

Jolley, C., & Douglas, T. (2012). Topological biosignatures: large-scale structure of chemical networks from biology and astrochemistry. *Astrobiology*, *12*(1), 29-39.

Kim, H., Smith, H. B., Mathis, C., Raymond, J., & Walker, S. I. (2019). Universal scaling across biochemical networks on Earth. *Science advances*, *5*(1), eaau0149.

Kolchinsky, A., & Wolpert, D. H. (2018). Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface focus*, *8*(6), 20180041.

Pagani, G. A., & Aiello, M. (2013). The power grid as a complex network: a survey. *Physica A: Statistical Mechanics and its Applications*, *392*(11), 2688-2700.

Perera, S., Bell, M. G., & Bliemer, M. C. (2017). Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective. *Applied Network Science*, *2*(1), 1-25.

Schulze-Makuch, D., Heller, R., & Guinan, E. (2020). In search for a planet better than Earth: Top contenders for a superhabitable world. *Astrobiology*, *20*(12), 1394-1404.

Solé, R. V., & Munteanu, A. (2004). The large-scale organization of chemical reaction networks in astrophysics. *EPL Europhysics Letters*, *68*(2), 170.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*(6825), 268-276.

West, G. B., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, *276*(5309), 122-126.

West, G. B., & Brown, J. H. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology*, *208*(9), 1575-1592.

Wilson, E. O. (1999). *Consilience: The unity of knowledge* (Vol. 31). Vintage.

Wong, M. L., & Bartlett, S. (2022). Asymptotic burnout and homeostatic awakening: a possible solution to the Fermi paradox?. *Journal of the Royal Society Interface*, *19*(190), 20220029.

Wong, M. L., Bartlett, S., Chen, S., Tierney, L. (in review). Searching for life, mindful of lyfe's possibilities. *Life*.

# Artificial Biochemistry as an Alternative to Animal Testing

**Musab Kılıç**

**Abstract**

Human health is a serious matter. We want our clothes, meals, cosmetics, shampoos, and medicines not to have any side effects. Animal testing is used to keep us safe but what if there was another way? I argue that computer simulations for biochemical processes will take on animal testing and biological models will be trivial to implement and use with an ordinary personal computer.

The long-term and short-term effects of a product should be researched before that product is put to public use (CIOMS Working Group IV, 1998). Food and Drug Administration (FDA) approves drugs in the United States (Lipsky and Sharp, 2001) and European Medicines Agency (EMA) approves drugs in European Union (Pignatti et al., 2004).

Most of the prior research required often involves various forms of testing in vivo. In 2017, 1.89 million procedures were carried out for experimental purposes in Great Britain (Home Office, 2017). The total number of animals that research facilities used in regulated activities in the United States in 2017 were 792,168 (United States Department of Agriculture, 2017). This raises an ethical question on how much we are willing to harm, stress, and kill animals for our benefit.

There is criticism that the methodology and usefulness of animal testing haven't been researched enough. Reviews show that poor methodologies result in unnecessary animal experiments and a lack of gained knowledge on the side effects. Even if the experiments show a statistical advantage of a drug, it can't be deduced if humans will behave the same (Pound et al., 2004).

It has been argued that there are no other alternatives and it's therefore justified to use animals as human health is more important and urgent (Festing and Wilkinson, 2007). However, recent research shows in silico experiments are getting better and might replace in vivo in upcoming decades. Computer-Aided Drug Design(CADD) software is used to discover potential molecules. Machine learning algorithms are used to predict side effects (Galeano et al., 2020). Not only artificial biochemistry will save the lives of people and animals, but it is also much faster than doing in vivo experiments.

Coding a machine learning project used to require a huge amount of effort, academic knowledge, and programming skills. Nowadays, it takes under 50 lines of Python code and less than 3 minutes to get 97% accuracy on the MNIST database (Tensorflow, 2022). The computational power keeps growing every year, and I would argue computational biology will one day become as easy to use as the machine learning frameworks of today.

Recent developments in machine learning brought another ethical question with them: Is it safe to give everyone access to these powerful tools? Insufficient data might lead to incorrect results, or an evil party might use biased data for discrimination in purposes. I believe the same is true for computational biology; it's not a question of if, but when. Hence, artificial life needs a foundation of ethics for what is to come in future generations.

## References

CIOMS Working Group IV (1998). Benefit-risk balance for marketed drugs: Evaluating safety signals. `https://cioms.ch/wp-content/uploads/2017/01/benefit-risk.pdf`. Accessed: 2022-05-14.

Festing, S. and Wilkinson, R. (2007). The ethics of animal research: talking point on the use of animals in scientific research. *EMBO reports*, 8(6):526–530.

Galeano, D., Li, S., Gerstein, M., and Paccanaro, A. (2020). Predicting the frequencies of drug side effects. *Nature communications*, 11(1):1–14.

Home Office (2017). Annual statistics of scientific procedures on living animals great britain 2017. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/724611/annual-statistics-scientific-procedures-living-ani` `pdf`. Accessed: 2022-05-14.

Lipsky, M. S. and Sharp, L. K. (2001). From idea to market: the drug approval process. *The Journal of the American Board of Family Practice*, 14(5):362–367.

Pignatti, F., Boone, H., and Moulon, I. (2004). Overview of the european regulatory approval system. *The Journal of ambulatory care management*, 27(2):89–97.

Pound, P., Ebrahim, S., Sandercock, P., Bracken, M. B., and Roberts, I. (2004). Where is the evidence that animal research benefits humans? *BMJ*, 328(7438):514–517.

Tensorflow (2022). Training a neural network on mnist with keras. https://www.tensorflow.org/datasets/keras_example. Accessed: 2022-05-14.

United States Department of Agriculture (2017). Annual report animal usage by fiscal year 2017. https://www.aphis.usda.gov/animal_welfare/downloads/reports/Annual-Report-Animal-Usage-by-FY2017.pdf. Accessed: 2022-05-14.

# Vectorizing Cartesian Genetic Programming

**Andrea Fanti**

**Abstract**

Many Cartesian Genetic Programming (CGP) implementations exploit parallelism to achieve very fast execution times; however, none of these employ vectorization for the whole CGP evolutionary pipeline. Here a possible approach to fully vectorize the typical evolution step of multiple CGP populations is presented, assuming they share the same hyperparameters; the main application would be performing the same CGP experiment multiple times for statistical analysis. A minimal Python implementation of this approach is being developed at `https://github.com/rikifun-t/cgp-vec`; it uses the PyTorch library (Paszke et al., 2019), so that CPU and GPU devices can be easily used interchangeably.

## Introduction

*Cartesian Genetic Programming* (CGP) is a form of Genetic Programming first introduced by J. Miller in (Miller, 1999). Here a brief overview of the relevant aspects of the classical formulation of CGP from (Miller, 2011) is given. In CGP, a program is represented as a directed acyclic graph, where nodes are arranged in a grid of rows and columns, hence the "cartesian" adjective. The content of each node is the primitive function it performs, while the output edges represent from where the node gets its inputs. Since the grid dimensions are fixed, the genotype is of fixed length, and encodes, for each computational node, its function and its connections respectively in the *function gene* and the *connection genes*. A node can only get its input from program inputs or nodes in previous columns; moreover, the classical formulation also includes an optional parameter *levels back*, that specifies the maximum distance between the column of a node and the column of any of its connections. These specifications result in allelic constraints such that genes have different sets of valid alleles.

*Vectorization* is a form of parallel computation in which the same operations are performed on multiple inputs at once, also referred to as "array programming", since it allows to perform a single operation on entire arrays. Many programming languages support array operations, either as built–in functionality or through third–party libraries (Harris et al., 2020) (Paszke et al., 2019). These often revolve around $N$–dimensional arrays, sometimes called *tensors*. The dimensions of such an array are also referred to as its *shape*. The approach presented here assumes that the chosen backend for array programming already provides some common array operations, such as: arithmetic and boolean element–wise operators; extracting a pseudo–random array of integers in different intervals (this can be alternatively implemented using the more common pseudo–random extraction of floats in $[0, 1)$); *fancy indexing*, i.e. using other arrays or boolean masks to index a multi–dimensional array; *shape broadcasting*, i.e. reshaping two arrays with different shapes into a common shape, if appropriate conditions are met; finally, either top–$k$ selection, or sorting (stability may be required depending on the use case).

Note that all of the functions discussed here, excluding selection operators, don't need to distinguish between multiple populations and a single, combined population. From here on, it is assumed functions operate on a single population for the sake of clarity, with the only exception being selection operators.

## Genetic operators

Here the two fundamental genetic operations performed in CGP are considered: random initialization and mutation. Crossover is left out since it is not present in most CGP implementations, and several different variants exist.

In this section it is assumed the DNA is encoded with integers, and the CGP allelic constraints are represented with an array $n_a$ that stores the number of valid alleles for each *locus*, i.e. each location in the DNA. This means that the valid alleles for the gene at locus $i$ will be the integers in $[0, n_a[i])$; from here, the collection of all these intervals will simply be denoted as $[0, n_a)$. Note, however, that when the levels back parameter is specified, some genes would take values in 2 disjoint intervals; here it is assumed that mapping the single intervals $[0, n_a)$ to the appropriate disjoint intervals is performed during phenotype generation by adding a proper integer offset to affected genes.

With this representation, random initialization is straightforward, and simply involves extracting an array of random integers, although in different ranges. If not available from the array computing backend, this can be easily achieved by sampling floats uniformly in $[0, 1)$, rescaling by array multiplication with $n_a$, and finally truncating the decimals to obtain integers. The generated genes will then each be in the intervals $[0, n_a)$, as needed by the CGP allelic constraints.

Mutation is a bit trickier, since it requires to exclude the current genes values from the extraction of new valid alleles. First of all, a random boolean mask is generated, to select the genes to be modified. Then, instead of extracting integers in the $[0, n_a)$ intervals, they are sampled in $[0, n_a - 1)$. At this point, by adding 1 to all extracted integers that are greater or equal than the current alleles, obtaining random integers in $[0, a - 1] \cup [a + 1, n_a)$, where $a$ is the current allele at mutation locus. These can then be assigned to their selected locus as new, mutated alleles.

## Phenotype generation and evaluation

The evaluation of a CGP phenotype on an input may be vectorized in two independent ways: over multiple individuals, and over multiple inputs. The latter requires simply that primitives are themselves vectorized in the chosen backend, as is usually the case with arithmetic or boolean operations. Assuming the fitness function can also be vectorized (e.g.: regression losses), only one evaluation is needed to compute fitness values for all individuals at each generation. This makes it sensible to perform both the decoding of the genotype and evaluation of the phenotype in a single operation, without generating an intermediate phenotipic representation, which would only be used once for each individual.

A recursive, vectorized single pass through all genotypes is sufficient for this operation, starting from all the output nodes of all individuals. Each type of node will be evaluated differently at each recursion step. Input nodes are the base cases, and return the program input components corresponding to their node numbers. Output nodes, instead, redirect their computation to the hidden or input node they are connected to. Finally, hidden nodes first require to gather their function and connection nodes; then, their connection genes are mapped to the numbers of the nodes they refer to, which are then evaluated recursively; finally, the outputs of the connections are fed to the primitives of the currently evaluated nodes, returning their outputs.

This is the only operation presented here that requires a non–constant number of vectorized operations. More specifically, it is proportional to the maximum depth of the phenotypes being evaluated, multiplied by the number of primitives. While the number of primitives is usually small, the maximum depth of the phenotypes is, in the worst case, the number of columns in the grid. Fortunately, CGP is known to be biased towards shorter phenotypes due to its "length bias" (Miller and Turner, 2015), which mitigates this problem.

Note that the use of GPUs and array computing for fitness evaluation in CGP was already treated in (Harding and Banzhaf, 2011); however, neither avoiding the intermediate phenotipic representation nor using vectorization to parallelize over multiple populations were considered. This also meant that the chosen approach was to only employ GPUs for multiple fitness evaluations of the same individual.

## Selection operators

While not specific to CGP, selection operators are needed to implement the Genetic Algorithms or Evolution Strategies that are usually used to evolve CGP populations. Here comma–selection, plus–selection, tournament selection and roulette–wheel selection are discussed. Note that these are the only operations that make a distinction between multiple individuals and multiple populations, since their selection process must be performed on each population independently. Hence, input arrays will be matrices of fitnesses, with each row (or column) representing a population; from here it is assumed populations are arranged in rows.

Comma–selection is trivial, as it is equivalent to top–$k$ selection, which is often present in array computing backend; if this is not the case, a simple alternative is sorting the rows of the array of individuals (using the rows of the fitness matrix) and then truncating it. Plus–selection is very similar, except for the cases where offspring need to be prioritized on ties, which is customary in CGP. In this case, the underlying selection or sorting function needs be stable, so that placing offspring at the front of the combined population has the desired effect.

Tournament selection can be achieved by first shuffling the rows, truncating to the first $n$ individuals, and finally applying usual top–$k$ selection. $n$ will thus be the tournament size. However, many array computing backends don't directly provide a way to shuffle rows independently; for this purpose, the `argsort` operator in conjunction with random sampling in $[0, 1)$ can be used, which are often available.

Roulette–wheel selection needs the rows of the fitness matrix to sum to 1. Cumulative sums can then be computed along each row, and compared with a matrix of random thresholds in $[0, 1)$; this should have as many rows as the number of roulette–wheel rounds to perform, and as many columns as populations. This means that, using broadcasting, a comparison with the cumulative sums will return a boolean 3–dimensional array. This array will contain, for each round, a boolean matrix with dimensions being the number of populations and the number of individuals in each population. When using $c < t$ as the comparison, with $c$ the cumulative sum and $t$ the random threshold, the matrix will contain rows composed of a series of `True`, followed by a series of `False`. This means that, by summing booleans as if they were 0 (`False`) and 1 (`True`), the sums of these rows will be the index of the individual to be selected in that population for that round. These indices can then be used to gather the selected individuals.

# References


Harding, S. L. and Banzhaf, W. (2011). Hardware Acceleration for CGP: Graphics Processing Units. In Miller, J. F., editor, *Cartesian Genetic Programming*, pages 231–253. Springer.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585:357–362.

Miller, J. and Turner, A. (2015). Cartesian Genetic Programming. In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, GECCO Companion '15, pages 179–198, New York, NY, USA. Association for Computing Machinery.

Miller, J. F. (1999). An empirical study of the efficiency of learning boolean functions using a Cartesian Genetic Programming approach. In *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation - Volume 2*, GECCO'99, pages 1135–1142. Morgan Kaufmann Publishers Inc.

Miller, J. F. (2011). Cartesian Genetic Programming. In Miller, J. F., editor, *Cartesian Genetic Programming*, pages 17–34. Springer.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Technical report, arXiv.

# Transformers as Reservoir Computers for Reinforcement Learning

**Federico Pigozzi**

### Abstract

Large Language Models (LLMs) have propelled breakthroughs in sequence modeling by pre-training on large web corpora and fine-tuning on downstream tasks. This project aims first at investigating how LLMs can be used as off-the-shelf reservoirs for model-free Reinforcement Learning (RL) tasks, thus optimizing much fewer parameters than usual and exploiting the common sense knowledge embedded within them, also considering the case of embodied RL agents.

## Introduction

Large Language Models (LLMs) (Devlin et al., 2018) have propelled breakthroughs in sequence modeling by pre-training on large web corpora, to the point of seemingly incorporating "universal" knowledge (Lu et al., 2021). But, to what extent is such knowledge universal remains an open question. In other words, do LLMs ground knowledge that is so universal to solve far-fetched downstream tasks? Most notably, Reinforcement Learning (RL) agents lag behind the computational abilities of animals and they might benefit from the common sense embedded in LLMs as if it was a *world model* (Ha and Schmidhuber, 2018). At the same time, to what extend the bodies—versus the brains—of *embodied* RL agents can aid in performing computation remains another open question.

This research project aims at answering the following research questions:

RQ1 Can we use off-the-shelf LLMs as reservoirs for model-free RL tasks?

RQ2 Can LLM reservoirs—juxtaposed with physical reservoirs—help in quantifying the amount of computation to be offloaded to the body in embodied RL agents?

The following subsections illustrate some background.

**Large Language Models** LLMs (Devlin et al., 2018) have propelled breakthroughs in sequence modeling by (1) pre–training on large web corpora; (2) fine-tuning on downstream tasks or even few-shot learning (Brown et al., 2020).

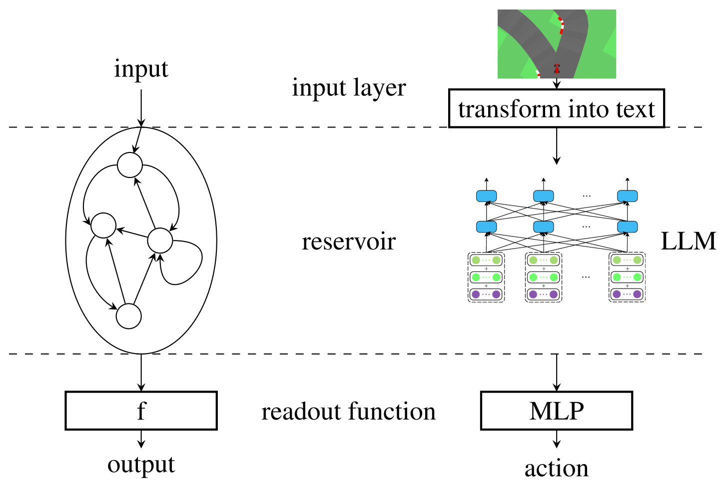


Figure 1: Outline of RC (left) and our proposed approach (right).

Recently, Chen et al. (2021) learned LLMs to predict action sequences in RL, while Huang et al. (2022) relied on frozen, pre-trained LLMs to generate action plans for model-based indoor environments.

**Reservoir Computing** Reservoir Computing (RC) (Nakajima and Fischer, 2021) consists of a learning machine that maps input signals to a higher dimensional space through a non-linear reservoir; the embedding produced by the reservoir is then fed to a readout function for the final output. In RC, the reservoir is a fixed "black-box" and we learn only the readout, thus reducing the complexity of the learning problem and exploiting the computational power of an off-the-shelf reservoir. Figure 1 (left) is a schematic view of RC.

**Physical reservoirs** There is evidence that physical bodies can operate as reservoir computers. In particular, Nakajima et al. (2015) highlighted how soft bodies, given their infinite degrees of freedom, are ideal physical reservoirs. Indeed, the embodied cognition paradigm (Pfeifer and Bongard, 2006) posits that bodies (of embodied agents) can carry on computation alongside brains, as cognition arises from the interaction between the brain and the body. Still, how to quantify the amount of computation to be offloaded to the body (and, thus, the amount to be retained by the brain) remains an open issue (Nolfi, 2021).

RQ2 tackles that open issue. Considering embodied agents, we juxtapose a physical, *wet* reservoir (ideally, the soft body) with a neural, *dry* reservoir; compared to other dry reservoirs, LLMs are ideal as their complexity is the sole to rival that of physical bodies.

## Methods

Figure 1 (right) outlines our proposed approach. Our research questions give rise the following issues.

**Model out-sourcing** The project relies on frozen, pre-trained LLMs as reservoirs. Training LLMs from scratch is no easy endeavor and we highly advise against it; the solution is then to out-source it.

Several open-source implementations of LLMs do exist. As a matter of example, we consider Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), since it has the least parameters, suits the needs of the project, and is open-source, but others are indeed possible. BERT employs an encoder-decoder architecture, with an encoder that distills input sentences (i.e., a sequence of text tokens) into a latent representation that is then fed to a decoder that reconstructs the original input. In the original implementation, Devlin et al. (2018) trained on two different tasks at the same time: masked language modeling (predicting missing words from the context sentence) and next sentence prediction (predicting the likelihood of a sentence given the previous one). As a result, BERT represents contextual embeddings for words. Once we freeze the weights, BERT can be seen as a reservoir whose encoded embeddings are numerical vectors to be fed to any readout function to perform a downstream task.

**Input representation** LLMs accept inputs in the form of sequences of text. Considering that the vast majority of RL tasks present inputs (i.e., observations) in different modalities (e.g., images for an Atari video-game), we must investigate what is the appropriate input representation for an LLM reservoir, or, in other words, how to effectively translate inputs from disparate modalities into text.

Albeit intricate, the issue presents solutions. As a matter of example, consider `CartPole`, whose observation is the concatenation of four real numbers: the cart position, cart velocity, pole angle, and pole velocity at tip. An input representation consists of concatenating the corresponding text representations (i.e., “4” becomes “four”), possibly with syntactic information (e.g.,“cart position is four”). As a more advanced example, consider the `CarRacing` environment (Klimov, 2016), whose observation is the screen image tensor. An input representation consists of extracting low-level features from the image tensor, like distances from nearest roadsides, together with the angle and the velocity of the car.

**Readout** We must investigate how to optimize a readout outputting the action for an RL agent and taking as input the LLM embedding—we remark the LLM stays frozen and we never update it. As seen in other fine-tuning settings (Pigozzi et al., 2022), the readout can be a neural network with few parameters, to be optimized with any standard RL procedure.

**LLMs and embodiment** For the sake of RQ2, we juxtapose a wet and a dry reservoir and learn a readout that dynamically assigns confidence scores to one or the other reservoir outputs (with, e.g., an attention model or a mixture of experts). By inspecting such confidence scores, we would gain insights into the balance between computation performed by the body and by the brain.

## Limitations

There is evidence that LLMs incorporate the same bias of the corpora they are trained on. The standard RL benchmarks envisioned so far provide safe environments. At the same time, the project treats LLMs as reservoirs, akin to feature extractors that output no sequences of text; as such, the risk of doing harm to human beings is very low. Still, we acknowledge that evidence.

## Relevance to ALife

Proving that LLMs ground knowledge that is so universal to solve model-free RL tasks would be a relevant discovery. As a consequence, being the LLM reservoir frozen, we can solve tasks by optimizing much fewer parameters than usual.

Consequences can also be far-fetched. Within the artificial intelligence and ALife communities, there is a divergence between works on connectivist deep learning and works on the embodied cognition side of the fence and some have argued in favor of combining the two fields (Risi, 2021). Our project might indeed bridge that divergence by exposing what is the optimal balance—in an embodied agent—between the computation to be retained by the brain and the computation to be offloaded to the body. As outlined in Introduction, describing such balance remains an open issue: highlighting how wet and dry reservoirs can complement each other would be a stepping stone towards a unified understanding of cognition.

We envision open-sourcing all the code and the models to reproduce and replicate the experiments.

## References


Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., and Mordatch, I. (2021). Decision transformer: Reinforcement learning via sequence modeling. *Advances in neural information processing systems*, 34.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ha, D. and Schmidhuber, J. (2018). Recurrent world models facilitate policy evolution. *Advances in neural information processing systems*, 31.

Huang, W., Abbeel, P., Pathak, D., and Mordatch, I. (2022). Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. *arXiv preprint arXiv:2201.07207*.

Klimov, O. (2016). Carracing-v0. *URL https://gym. openai. com/envs/CarRacing-v0*.

Lu, K., Grover, A., Abbeel, P., and Mordatch, I. (2021). Pretrained transformers as universal computation engines. *arXiv preprint arXiv:2103.05247*.

Nakajima, K. and Fischer, I. (2021). *Reservoir Computing*. Springer.

Nakajima, K., Hauser, H., Li, T., and Pfeifer, R. (2015). Information processing via physical soft body. *Scientific reports*, 5(1):1–11.

Nolfi, S. (2021). *Behavioral and cognitive robotics: an adaptive perspective*. Stefano Nolfi.

Pfeifer, R. and Bongard, J. (2006). *How the body shapes the way we think: a new view of intelligence*. MIT press.

Pigozzi, F., Tang, Y., Medvet, E., and Ha, D. (2022). Evolving modular soft robots without explicit inter-module communication using local self-attention. *arXiv preprint arXiv:2204.06481*.

Risi, S. (2021). The future of artificial intelligence is self-organizing and self-assembling. *sebastianrisi.com*.

# Consciousness and Artificial Life: The Inevitable Collision

**Ryan Wing**

## Introduction

Artificial life research, and humanity more broadly, are on an inevitable collision course with a reality for which we are woefully unprepared. We will develop the ability to create conscious life. Millions of people are already working on projects across multiple domains, funded by billions in annual investment that are collectively pulling this inevitability forward in time with exponential decay. Will it be wetware? Robotic? Hybrid? Artificial General Intelligence? A new technology yet to be developed? The existing candidates are increasingly accessible and have proliferated throughout the world. Who will be the first to create conscious artificial life? A scientist? A corporation? A government? A teenager with access to YouTube? The identity of the instantiator, and thus the initial knowledge holder, will matter, a lot. Humanity's collective abilities to discern 'how' to move from idea to reality are now increasing exponentially. We have, however, largely avoided reckoning with the adage 'With great power comes great responsibility.' Unintended consequences are unintended. 'Externalities' are ramifications of a decision, yet are external to it. Sometimes there's 'collateral damage.' We can't allow that abnegation of responsibility to happen with empowering ourselves to create conscious artificial life. Understanding how consciousness arises and how it can be reliably identified should be a top priority for anyone interested in the creation of "life as it could be."

## Why Progress is Critical

If humans were not consciousness or reciprocally recognized as conscious by others, humanity could not have developed the concepts of empathy, compassion, or justice. Absent these, we would not in-turn have ethics, morality, or the legal systems largely built upon them. It seems unavoidable that humanity will use these frameworks to address the creation of artificial life. However, if they ultimately rely on a reasonable certainty of consciousness, we currently lack a universal method of measurement or identification. Given the social and political intensity around issues of 'sanctity of life' vis-a-vi abortion rights, it is reasonable to expect strongly held, opposing opinions surrounding artificial life. While a certain degree of emotionality is likely unavoidable, a more complete understanding of and universal ability to measure consciousness prior to the first 'creation event' might enable proactive action on some of these issues to help avoid prolonged disruption of ongoing research. Knowledge of consciousness' boundaries or necessary conditions would offer greater decision making clarity. Enabling proactive action on this inevitable collision requires highly prioritizing consciousness research.

## How can we identify consciousness?

**From Definition to Identification** Thomas Nagel claimed that "An organism has conscious mental states if and only if there is something it is like to be that organism." Nagel (1974) This often cited definition seems to accurately point to a specific, recognizable, shared experience of consciousness: "I" am an individual, "there is something that it is like to be" me, and that is different than what it is like to be anyone else. Only I seem to have access to my subjective experiences, my thoughts, and my emotions, just as others only have access to their own. While helpful, the abstract nature of such a definition highlights the need for more precise and measurable indicators of consciousness such as its source/s or cause/s, if it is binary or a gradient, a possible minimum threshold to arise, and if it is substrate independent. Theoretically, each of these should be able to be measured, computed, or modeled.

**Global States** Global states ('levels of consciousness') refer to the overall, subjective experience of a conscious being. Examples include wakefulness, dreaming, sedation, minimally conscious, and (more debatably) the psychedelic state. [Seth and Bayne (2022)] Consciousness research mostly addresses subjective, wakeful experience, but all global states require reliable measurement and identification.

**Local States** Local states are specific, subjective experiences such as moods, emotions, the sense of willpower and choice about its use, the sense of existence within a body, and an ongoing autobiographical narrative. Local states help construct a sense-of-self. It's unknown why a sense-of-self exists when the brain could more simply, unconsciously process sensory input while still allowing for methodological navigation, manipulation of an environment, and learning.

**Beyond Humans** All mammals are widely presumed to be conscious in some form. Consciousness in other animals is more debated, but birds are gaining more broad acceptance. The cephalopod nervous system differs vastly from mammals' and likely evolved along a separate evolutionary path. This may show there is not one path to consciousness, but potentially many. [Godfrey-Smith (2016)]

## Theories of Consciousness

A theory of consciousness (TOC) offers an explanatory and/or predictive mechanisms for conscious experience. If a TOC is not measurable or testable, it is not able to be verified or disproven and cannot be studied scientifically. To be comprehensive, a TOC should address functional *and* phenomenological components, causal mechanisms, forms, states, and offer predictive abilities. TOCs also need to account for: **Local States:** affect, temporality, volition, thought; **Global States:** dreaming, meditation, disorders of consciousness, the psychedelic state; **Conscious Beings:** human infants, non-human animals, artificial life [Seth and Bayne (2022)] The TOCs *very* briefly described below have gained prominence in modern research.

**Higher Order Theories** Propose that consciousness is a recursive process of higher-level representations (meta-representations) referring back to another, lower-level representation. Some states are incapable of reaching consciousness as they cannot be targeted by meta-representations while others are inherently conscious due to a necessary pairing of meta and lower-level representations. [Seth and Bayne (2022)]

**Global Workspace Theories** Mental states are conscious if they are 'globally available' to specialized information processing systems such as attention, evaluation, memory, and verbal report. Lower-level sensory information is 'broadcast' via an 'ignition' mechanism to higher-level cortical areas where it becomes conscious and can flexibly guide behavior and cognition in context dependent ways. [Seth and Bayne (2022)]

**Integrated Information Theory** IIT starts with phenomenology and then seeks possible explanations in physics, not visa versa, as properties of the external world are only known to us via interpretations by consciousness. A set of axioms is used to derive postulates that, in principle, can determine if any system in a particular state is conscious, how conscious, and which type of consciousness it possesses. A numerical measure of a system's integrated information, referred to as $\Phi$ (Phi), is used to make consciousness determinations by identifying $\max(\Phi)$ in the system. Local states of consciousness arise when conceptual structures form specific shapes in high-dimensional space resulting from a system's cause-effect structure. [Tononi (2015)]

## Strategies for Progress

**Confront the measurement problem** Quantum physics is not alone in having a measurement problem; Consciousness research also has a non-trivial challenge in trying to identify measurable, reliable markers of consciousness. Directly confronting it will yield more progress, faster than purely abstract or conceptual TOC conjecture. Theoretical TOCs should focus on proposals that can be modeled or are demonstrably testable.

**Prioritize cross-disciplinary efforts** Neuroscience, while serving a critical role, is unlikely to unilaterally produce answers to the numerous, diverse aspects of consciousness research. Multidisciplinary efforts will produce the most comprehensive TOCs. Finding answers relies on formulating questions that aim effort in the correct directions. Collaboration between multiple, diverse minds will create the most optimal solutions.

**Beyond Neural Correlates of Consciousness** Researchers should avoid becoming hyper-focused on neural correlates of consciousness. Each part of the brain is not a complete, independent system. The brain is only one component of the nervous system. We often draw imaginary boundaries around parts of systems to study them as individual subjects, but consciousness might result from the parts, in context, as a complete system.

### Be Lean

Not everything should be run like a startup or a business, but insights from the "Lean Startup Method" [Reis (2011)] may be valuable if applied to consciousness research.

**Minimum Viable Products, Fail Fast, and Fail Forward** A minimum viable product (MVP) is the quickest, fastest, most basic version of an idea that can be created (while remaining representational). Sharing MVP TOCs (with a disclaimer) to gather feedback and make refinements would provide opportunities to "fail fast" and "fail forward." Discovering your proposal isn't viable when it's only a basic framework rather than a heavily invested in, complete project is extremely valuable.

**Pivot** Just like evolution, when ideas are culled sooner rather than later, minds can pivot to new ideas for solutions with greater fitness. Every failed idea provides valuable information for future efforts, which should be celebrated, not shamed. [Reis (2011)]

## References


Godfrey-Smith, P. (2016). *Other minds: The octopus, the sea, and the deep origins of consciousness*. Farrar, Straus and Giroux.

Nagel, T. (1974). What is it like to be a bat. *Readings in philosophy of psychology*, 1:159–168.

Reis, E. (2011). The lean startup. *New York: Crown Business*, 27:2016–2020.

Seth, A. K. and Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, pages 1–14.

Tononi, G. (2015). Integrated information theory. *Scholarpedia*, 10(1):4164. revision #150725.

# Handle with Care

**Shakiba Shahbandegan**

We live in a mass-produced existence, not only as beings but also as ways we can express ourselves as individuals. When there is no direct line of communication between people, there should be a platform that translates nuances, give them means of expression, and keeps the reactions unchanged.

Our relationships are mediated by technology. As a result, the philosopher Walter Benjamin's most significant points start to come out clear; as time goes on, communicating our experiences as people seems to become more and more difficult. Because of the new and sometimes brutal technology that has mediated their experience of conflicts, people who have a lot to say are having difficulties communicating any of it. There is always a pause, searching for the right turn of phrase to give someone who wasn't there, access to their experience. Language and communication have atrophied to such an extent that it will always be impossible for them to make others comprehend it.

We don't just communicate verbally— we have various options or channels for communication. One way is to have a manual, a guidebook that, like any object we purchase from the store, will tell us the details about something, showing the parts, how they work, how to interpret weird sounds and alarms, and how to fix it.

In this essay, I am going to write about a project that uses technology to find a new means of communication between people and gives them a chance to express themselves and their personalities in a unique way. This project, called "Handle with Care," visualizes one's feelings and hard-to-express emotions in the form of an object with a distinctive shape.

Handle with Care is proposed by Parisa Ghaderi[1], an assistant professor of graphic design at Michigan State University (MSU) in the department of Art, Art history and Design. She collaborated with Dr. Parisa Kordjamshidi[2], an assistant professor at MSU in the department of Computer Science and Engineering, and Sania Sinha[3], an undergraduate researcher at the Heterogeneous Learning and Reasoning Lab at MSU.

Parisa came up with an idea to visualize these difficult emotions as a tangible object and a manual that can be referred to in times of trauma and conflict.

For the first step of this project, a two-part survey was designed. The first part [1] consists of 54 questions ranging from personal traits to lifestyle choices. The second part [2], through 20 questions, asks for a description of an object of comfort. This could be an imaginary object one would envision in their mind, that could be held dear to orient them in times of conflict or emotional trauma.

The survey was conducted in December 2021, using a curated crowd of 18 people from different walks of life, single, divorced, newlywed, about to divorce, in a happy relationship and young love. The survey was conducted in a way to make sure these people were able to provide enough details about their past, present, and even future self without feeling vulnerable and exploited.

The results of this survey were then processed using Natural Language Processing (NLP) techniques and visualized into forms of urban furniture with unique shapes.

This was done by using Sentence Transformer to obtain phrase embeddings of the answers of a person. Subsequently, Principal Component Analysis (PCA), a dimensionality reduction technique, is used to obtain the top three components from the embeddings. These are used as x, y, z coordinates to represent the phrase in a 3D space. Additionally, descriptions of color are tracked and converted into standard RGB format and finally, an average RGB value is kept as the resultant color (blue is default). Finally, a convex hull is used on the 3D points representing the phrases to obtain a solid shape and the resultant color is used as the color of the shape. This shape, generated as a 3D mesh is subsequently converted into a STL format, which is standard for 3D printing purposes.

The furniture can be thought of as the translation of the descriptions provided by the subjects of the survey into physical reality. More specifically, 14 different pieces of varying sizes were created. The sizes ranged from 15x10 inches (the smallest object) to 30x35 inches (the tallest object). The color hues varied based on the responses from the participants who indicated their sense of color for the imaginary comfort object described in the second part of the survey. The objects had soft edges and corners to make it safe for people to sit, step on, walk by, or interact with them. Figure 1 shows one of these objects made by a 3D printer as a sample.

Parisa, the creator of Handle with Care, sees the built objects of this project as city furniture, thinking they can continue to exist to become a participatory project in the future. For instance, some questions and answers could be engraved on these benches to engage the audience while sitting or passing by so they can see themselves as a collective subject. Parisa believes a great potential space to exhibit this furniture could be the "Blue Bridge" in Grand Rapids, Michigan since the bridge is known for being a metaphor for connection.

[1] ghaderip@msu.edu

[2] kordjams@msu.edu

[3] sinhasa3@msu.edu

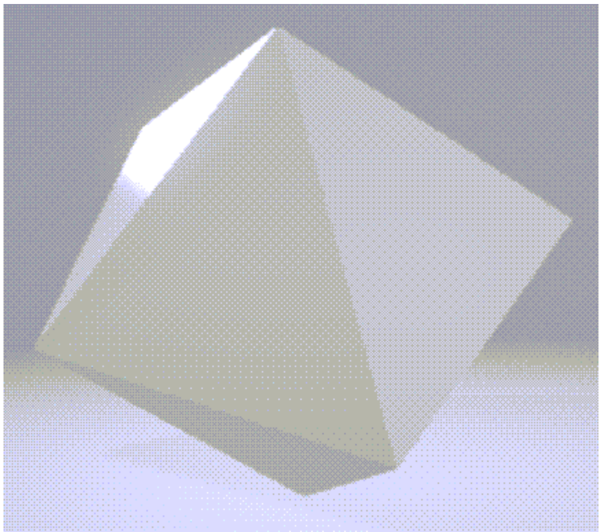

Figure 1. A sample simulated object based on the results of the surveys

Parisa believes in an ideal world in which people could hand each other a manual that functions as an introduction or a guide to how they truly feel in different stages of their lives. She believes there will be less miscommunication, tension, loneliness, and sense of loss as people can understand each other by referring to or reading about one another. Indeed, communication brings people closer to each other emotionally in this ideal world.

## References


[1]https://docs.google.com/forms/d/e/1FAIpQLSeafbujlr1VUVsnunMV8dTTMcwM6JhZtA4OBxZYm2gVNObjDA/viewform?usp=sf_link

[2]https://docs.google.com/forms/d/e/1FAIpQLSejRrlQvi7Lr_3LwIZQgyDVXqBobnBc7B_O1DHuNp59V2sfLA/viewform?usp=sf_link

# AI in Agriculture: ALIFE 2022 Proceedings

Anna Catenacci

### Abstract

In a world that's constantly changing, agriculture is a sector that will need to be prepared to rapidly shift with it if we are to maintain our population. The industry will need to become more sustainable in many ways; both the farming and distribution of food will become increasingly complex due to climate change, resource scarcity, and a growing population (Ayed and Hanana, 2021). Traditional farming methods may be challenged by the variable and less predictable climates that a warming planet brings, and food security may be imbalanced by threats like our own socio-economic systems, which is particularly relevant, as we have all recently seen these systems challenged by the outbreak of covid-19. With the effects of crises like these in mind, it becomes obvious that the process of farming and distributing food should become more efficient in order to increase food security across the globe.

## Artificial Intelligence and Food Security

The precision and forecasting the artificial intelligence (AI) offers a valuable tool in rapidly adapting farming systems to our environment. AI can be applied on farms to make rapid decisions about the care of produce in rapidly changing local environments.

### The Source: Farming

Using AI in farming means using algorithms to track and interpret patterns around the farm. It could allow farmers to keep up with weather patterns and pest infestations in their area with 98% accuracy by adjusting automated machines that care for and track the crop growth. (Ayed and Hanana, 2021). AI is the system that could use deep learning to analyze patterns collected in data to help increase crop yield by suggesting solutions to improve the growing environment and to diagnose disease and pests. Systems like smart irrigation, drones, and satellite guided technology can also be set up to allow the AI to directly care for the plants. Drones and similar technology have the ability to fertilize, water, and apply herbicide or pesticides to crops, using data obtained from sensors (Ayed and Hanana, 2021). Satellite guided technology is useful for diagnosing disease, pests, and the most fertile plots of land. The images and data obtained from this technology can also be used to guide drones and irrigation systems to give the plants the nutrition and water they need.

### Distribution: Supply Chain

Additionally, AI could aid the distribution of produce, as it has the ability to conduct rapid decision-making based on large datasets. In supply chain, AI could help to mitigate unexpected errors instantly (Toorajipour, et al. 2021). In making the supply chain more efficient, the agriculture industry would ultimately have less food wasted while increasing food security for our globe's population. Data collected regarding the traceability of crops can be utilized in a very similar way as previously discussed. In the context of the supply chain, produce would be tracked while it is being transported to its final destination. Ideally, this information could reduce food waste by not providing more food than will be used to stores and restaurants.

### Data Collection

To collect accurate data about these various aspects that improve crop yield, use of the Internet of Things (IoT) has been proposed for use to collect the data that AI will use to make tactical decisions. (Phasinam, et all. 2022). While crops are growing, camera recognition tools can capture images of plants, soil, and pests to recognize problems and disease (Ayed and Hanana, 2021). To identify weather patterns, probes and sensors that detect humidity, wind speed, soil moisture and fertility, temperature, and water quality are effective ways that IoT networks can collect a variety of data about the environment (Phasinam, et al. 2022). This approach would allow the farmer to tailor his techniques to trends in his specific area, as well as track the change in the climate over time. With climate change pressuring our ability to rapidly adapt to new trends in the environment we grow food in, this pattern recognition will become incredibly important with time.

## Long-term application of AI and ML

In the context of long-term data collection, with the use of machine learning, this process of using AI to make decisions in farming would help farmers to adjust their practices to be the most sustainable and the most efficient that they could be in their own, very specific area. The data and decisions that AI collects over time would be useful to tailoring a farmer's practices to his own land by using machine learning (ML) to analyze large sets of data pertaining to water, dirt, pests, etc. When a farmer sets out to combat pests, disease, or to make the farm environment more hospitable to their crops, the same tools can be used to track the treatments the farmer tries. The results of this analysis can show quantitatively which treatments worked the best. After some time identifying these problems that a farmer wants to solve, and their most effective solutions, the AI could begin suggesting treatment plans to ailments in the future are known to have worked and been helpful to the farmer in the past.

## References


Ben Ayed, R., & Hanana, M. (2021). Artificial intelligence to improve the food and agriculture sector. *Journal of Food Quality*, *2021*.

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research*, *122*, 502-517.

Phasinam, K., Kassanuk, T., & Shabaz, M. (2022). Applicability of internet of things in smart farming. *Journal of Food Quality*, *2022*.

Bannerjee, G., Sarkar, U., Das, S., & Ghosh, I. (2018). Artificial intelligence in agriculture: A literature survey. *International Journal of Scientific Research in Computer Science Applications and Management Studies*, *7*(3), 1-6.

# Computational post-cognitivism? How simulations of artificial life support theory-building

**Stefan Riegl**

**Abstract**

Philosophers recently argued that non-representational forms of computationalism might be compatible with certain forms of post-cognitivism, such as autopoietic theory (Villalobos and Dewhurst, 2017). Questions arise: Given a system of the real world that is autopoietic, can we look at it from a computational perspective? How to do that? What can we expect? Answering those questions is difficult and requires time and effort. Simulating computational models of artificial life support answering those questions. How computational models can help answering those questions is explored here.

## Introduction

In the middle of the 20th century the cognitivist paradigm emerged, declaring symbol manipulation as central aspect of cognition. This harmonized well with the computational theory of mind a.k.a. computationalism, which claims that at its essence the mind processes information and cognition means performing computations. However, shortcomings of cognitivism motivated new research program, which often rejected central aspects such as symbol representation, symbol processing, or possibly a form of metaphysical realism. One such theory is autopoiesis (Maturana, 1975), which later got adopted into modern post-cognitivist theories and frameworks, such as autopoietic enactivism. Summarized briefly, autopoietic theory defines living organisms not by a somewhat arbitrary selection of biological functions (like reproduction or metabolism), but by the dynamics that allow the organisms to produce and maintain itself.

Post-cognitivist streams try to solve problematic shortcomings of preceding theories at the cost of becoming increasingly incompatible with with those theories. As a consequence, insights from those modern theories are hard to bring in line with the arguably useful results from a vast body of existing, more mainstream research. Hence, philosophers re-visited some assumptions in recent years and are now challenging the general incompatibility between post-cognitivism and (non-representational) computationalism (Villalobos and Dewhurst, 2017).

Autopoiesis can be used as one exemplary instance of post-cognitivism with relatively modest additional assumptions. We can formulate research questions that may help us to investigate, whether blending it with computationalism might be worthwhile. The questions are:

1. When is it possible to apply a computationalist perspective at all?

2. What do we actually need to do to integrate a computationalist perspective?

3. Can we expect to learn anything interesting from applying a computationalist perspective (and if so what)?

The problem with those questions is: If they were easy to answer, they would not be unresolved issues anymore. Specific post-cognitivist and computational theories usually are quite comprehensive on their own and a mix of instances of either can bear a lot of complexity, which makes further inquiries effortful and time-consuming. However, there may be tools that may advance the discourse and bring us closer to answer the above questions.

This is where simulation of artificial life can help. Autopoiesis states concrete constraints on what should be understood as living organism, which inspired simulations of computational models in the past (McMullin, 2004). Turning theories and isms into computational models allows us to test predictions, evaluate complexity, and importantly explore implications by running the model, which would be too complicated to work out analytically.[1]

With the value of artificial life simulations in mind, we can consider whether such models can help answering the questions above.

## Can we do it?

Stated in full, the question becomes: **How can computational models help figuring out, *whether we can* apply computationalism to an autopoietic system?**

[1] Note that the "computational" in computational model and computational theory means different things. A model of cognition can be computational in both senses, in either, or neither.

The naive answer is: If it works, we can. That is: When we set out to implement a model of ”computational autopoiesis”, we are already assuming that it is possible to apply computationalist principles to an autopoietic system. If both theories are in fact work well with each other and the attempt to combine both in a computational model succeeds, we know that it is possible. However, we might not learn much if the attempt fails (we might e.g. just have made a methodological, avoidable error during modeling).

A more graded approach can be more useful. Autopoiesis and any specific form of computational theory of mind, respectively, consist of hypotheses and assumptions of which some are more essential to the theory or more important in a specific context than others. Experimentation can be started with a prototype implementing the most relevant parts that gets extended iteratively, which produces intermediary results that allow better guesses on feasibility much earlier.

## How do we do it?

Stated in full, the question becomes: **How can computational models help figuring out, *how to* apply computationalism to an autopoietic system?**

Models of reality do not always need to be theory-driven, such as the Lotka-Volterra-model for predator-prey-pupulations or curve-fitting in machine learning (Weisberg, 2012; Frigg and Hartmann, 2020). However, purely data-driven models, such as the otherwise highly successful deep artificial networks of recent years, may suffer from various problems, such as nontransparent processes, being data-hungry or hard to create, have biased results[2], lack explainability, or generalize badly (Lapuschkin et al., 2019).

Because of such and similar issues, some researchers argue that in order to build better models we do not need more data of the same kind, but better theories (van Rooij and Baggio, 2020). A theory-guided guided approach is no guarantee for flawless models, but it can help avoiding certain pitfalls. For example, one specific strategy is to turn verbal theories into formal theories, which then already provides a starting point for a computational model. In practice going from verbal theory to formal theory to model are often not consecutive steps, but rather alternating phases that specify and constrain each other mutually (Guest and Martin, 2021).

Specifically for the case of a computational take on autopoiesis, it was suggested to adopt Piccinini’s notion of mechanistic computation (Villalobos and Dewhurst, 2017) as a specific instance for a computationalist theory, as it does not imply a notion of representation like the cognitivism does. The next step then is the development of a formal theory of the verbal theories of autopoiesis and mechanistic computation, in parallel with the implementation of the corresponding computational model.

[2]When data as the single source of truth leads to problematic results, the models are often informally summarized with the phrase ”garbage in, garbage out”.

## What can we expect?

Stated in full, the question becomes: **How can computational models help figuring out, *what to expect when we* apply computationalism to an autopoietic system?**

While we as researchers might be interested in a combination of autopoietic and computationalist principles and can imagine it to be worthwhile, we don’t actually know whether it is worth the effort. Can we expect our method of modeling to bring about any useful results?

In principle, some problems in research only become apparent ”when we get there” like a very complicated game of chess. The more complex the research topic, the harder it becomes to predict every consequence, implication, and limitation by just abstract reflection. Moreover, research can be influenced by implicit assumptions or biases. The hands-on approach of computational modeling forces us to make those hidden influencing factors visible, which helps determining the feasibility of a research objective.

As mentioned above, the continuous development and testing of a computational model (especially when complemented by theory-building) provides intermediary results that provide a first glimpse of how the final results may look like, without requiring too big of a commitment in the beginning. Such early results are valuable, as they can provide indications that course-correction might be necessary, allowing to shape final results with on-line feedback during the process.

## Conclusion

While computational models often lead to an implementation in computer program code, there are other practices from the field of software-development that can be beneficial for scientific practice. Theories (can) make predictions and when compared to actual results they can be off, just like program code may fail expected results in assertions and tests. ”Through writing code, we debug our scientific thinking” (Guest and Martin, 2021). Loosely speaking, computational models allow us to write unit tests for theories and ”fail fast”[3], which is valuable feedback to see how models and theory can be fixed or abandoned in favor of a more useful method or even research question.

Simulating computational models of artificial life can guide future research. Such models can make it clearer at an early stage whether investing energy is possible and hence support our decision whether future research make sense, how to do it, and what to expect.


## Acknowledgements

The author would like to thank `[Name redacted]` who suggested an initial version of the questions discussed here.


[3]”Fail fast” is a phrase not uncommon in the area of software engineering and other fields.

# References


Frigg, R. and Hartmann, S. (2020). Models in Science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2020 edition.

Guest, O. and Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, page 1745691620970585.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., and Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, 10(1):1096.

Maturana, H. R. (1975). The organization of the living: A theory of the living organization. *International journal of man-machine studies*, 7(3):313–332.

McMullin, B. (2004). Thirty Years of Computational Autopoiesis: A Review. *Artificial Life*, 10(3):277–295.

van Rooij, I. and Baggio, G. (2020). Theory Development Requires an Epistemological Sea Change. *Psychological Inquiry*, 31(4):321–325.

Villalobos, M. and Dewhurst, J. (2017). Why post-cognitivism does not (necessarily) entail anti-computationalism. *Adaptive Behavior*, 25(3):117–128.

Weisberg, M. (2012). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

# Benefit of Communication Constraints

**Zixuan Liu**


### Abstract

We analysed several recent studies in this essay where multi-agent or multi-robot systems can benefit from communication constraints, including constraints of communication range, connectivity and frequency of the social interaction process of social learning. We then conclude that reasonable constraints make the system more adaptable to environmental noise and dynamics.


## Introduction

Swarm robotics is widely considered as a potential approach to use simple structure robots to solve problems of complicated environments collectively. One of the most studied potential applications is monitoring and detecting environments. For example, in a search and rescue area, a team of drones is sent off to searching a area for casualties. Such problems are widely studied under a concept of social learning. In social learning models, the individuals of the swarm are able to both exploring the environment independently and gain information through social interactions with other fellows. There are controllable factors of the system, such as the communication range of each individual, the network connectivity of the communication, as well as the frequency of communication. In other words, we can control the frequency and connectivity of the social interactions by control how far a other agent they are able to communicate with, who they can speak to, and how often they speak to others. The question is, do we really gain anything from the constraints? There is a widespread belief that better communication connectivity always improves the information exchange in a network. However, some evidence that supports the opposite could be sometimes true has been dig out recently.

## Communication Range Constraints improve adaptivity

To effectively perform collective learning/monitoring the targeted environments, robots swarms need to be able to cope noisy information, and in dynamic environments to adapt the latest information and spread them to other as quickly as they can. A recent study of Talamali et al. (2021) suggests that the swarm achieves better adaption to dynamic environment if the individuals have shorter rather than longer communication range. For a time-varying environment, it could be better if information is constrained to prevent the swarm from reaching consensus too fast to adapt the new changes of the environment.

## Network Connectivity Constraints

More counter-intuitively, Crosscombe and Lawry (2022) show that lower connectivity of a network results in lower error than fully connected networks and a moderate connectivity is better than lower or higher connectivity when the exploration to the environment is both infrequent and noisy. This study also suggests that high regularity in the network improves the ability to tolerate environment noise in decentralised systems. In other words, individuals are better to interact with several fixed partners rather than random others.

We can both benefit from shorter communication range and lower network connectivity. The question then will be, why not just stick with the shorter and lowers? Is there any thing gained by more social interactions?

## Social Interaction Speed up Consensus

Valentini et al. (2016) studied the speed versus accuracy trade-off and in their model the main factor affecting the trade-off is the individuals' "neighbourhood size" which is the number of other individuals that one share information with. The study has shown that the more interaction do help in terms of speed. The consensus reaching can be speed up with individuals' more social interactions while their study also showed the swarm can not adapt to the environmental changes once the consensus has been reached. Social interactions helps

## The Frequency of Interaction Matters

Besides studying the range constraints and the connectivity of social interactions, researches also focused on how to best balance the relative frequency of interaction and exploration. Kwa et al. (2021) reviews recent the studies on the exploration-exploitation dilemma in many multi-agent or multi-robot studies where the exploration and exploitation are mutually independent actions. Liu et al. (2021) looked into a specific problem where both gathering new information and making use of those currently available by interactions are probabilistic. The results showed that the system tend to converge at a level equal to the environmental noise level when the probability of exploration is relatively higher than that of interactions. The system failed to correct the errors by the interactions between agents. From this perspective, the probability of interaction is less significant when the noise of the environment is lower. On the other hand, the system tend to reach consensus at a random state when the interaction probability is greater. In other words, the consensus is reached without enough environmental information and therefore the randomness increases. The system can only perform well at a particular range of relative probability of exploration and interaction.

## Conclusions and Future Works

In general, reasonable communication constraints improve the adaptivity of a multi-agent or multi-robot system to noisy or dynamic environments while more social interactions can drive consensus among the population faster. The constraints of communication range can help the ystem adapt time-varying environments and keep the system from being stuck in consensus. The totally connected network is always not the best option in a noisy environment. Also the best frequency of social interactions is dependent on the frequency of exploration, in other words, it is usually not the highest possible frequency.

So what stories are these studies telling? We should be cautious dealing with the social interactions in social learning contexts. Individuals' communication strategies can significantly influence the group learning performance. Furthermore, it will be interesting the combine those different constraints to see if a noisy and dynamic environment can be adapted just by keep individuals from talking to others too much. It is the time to give up the stereotype that communication strength is very significant.

## References


Crosscombe, M. and Lawry, J. (2022). The impact of network connectivity on collective learning. In Matsuno, F., Azuma, S.-i., and Yamamoto, M., editors, *Distributed Autonomous Robotic Systems*, pages 82–94, Cham. Springer International Publishing.

Kwa, H. L., Kit, J. L., and Bouffanais, R. (2021). Balancing collective exploration and exploitation in multi-agent and multi-robot systems: A review. *Frontiers in Robotics and AI*, 8.

Liu, Z., Crosscombe, M., and Lawry, J. (2021). Imprecise fusion operators for collective learning. In *ALIFE 2021: The 2021 Conference on Artificial Life*. MIT Press.

Talamali, M. S., Saha, A., Marshall, J. A. R., and Reina, A. (2021). When less is more: Robot swarms adapt better to changes with constrained communication. *Science Robotics*, 6(56):eabf1416.

Valentini, G., Ferrante, E., Hamann, H., and Dorigo, M. (2016). Collective decision with 100 kilobots: Speed versus accuracy in binary discrimination problems. *Autonomous agents and multi-agent systems*, 30(3):553–580.