Ortus: an Emotion–Driven Approach to (artificial) Biological Intelligence

Andrew W.E. McDonald, Sean Grimes, and David Breen

Drexel Unversity, Philadelphia, PA 19104
awm32@cs.drexel.edu

Abstract

Ortus is a simple virtual organism that also serves as a framework for developing biologically based artificial intelligence. Born from a goal to create complex virtual intelligence and an initial attempt to model C. elegans, Ortus implements a number of mechanisms observed in organic nervous systems, and attempts to fill in unknowns based upon plausible biological implementations, psychological observations. Implemented mechanisms include excitatory and inhibitory chemical synapses, bidirectional gap junctions, Hebbian learning with its Stentian extension. We present an initial experiment that showcases Ortus' fundamental principles; specifically, a cyclic respiratory circuit, and emotionally–driven associative learning with respect to an input stimulus. Finally, we discuss the implications and future directions for Ortus and similar systems.

Introduction

While much work has been done to develop artificial intelligence (AI) systems that borrow principles from organic nervous systems, far less has been done that specifically targets the intersection of biology and artificial intelligence such that biological principles—rather than a specific applicability of the technology—are of primary concern, with the main goal being a virtual system that exhibits biological intelligence (BI). As our understanding of organic nervous systems and access to computation power both increase, widespread interest in systems that do exactly this is greatly increasing; as evidenced by DARPA's recent L2M project, in search of machines that learn throughout their lives (DARPA, 2017).

Researchers in the realm of computational biology and neuroscience have started making progress toward developing systems that model specific organisms or neural circuits, such as the nematode Caenorhabditis elegans (C. elegans) (Izquierdo and Beer, 2016), though these systems have the potential to require too much focus on organism–specific details to achieve proper functionality, shifting focus away from creating more generalized neurologically–inspired intelligent systems.

On the other hand more traditional (application focused) AI research has started taking more inspiration from human learning, such as developing an auto–encoder augmented by Hebbian learning, decreasing the need for an initial supervised–like learning period (Bowren et al., 2016). Further, Marblestone et al. (2016) discusses ways that artificial neural networks (ANNs) can more closely approximate neural functionality. In the context of biologically–inspired AI, the frameworks underlying these approaches may be too constraining for full exploration of the potential for that field of study.

Recent work at the intersection of these two areas includes Sinapayen et al. (2016), which investigates the applicability and biological plausibility of spiking neural networks learning by "stimulation avoidance". Perhaps the project most closely aligned to Ortus is a biologically inspired neural network modeled off of a honey bee's visual system, which merges biological mechanisms and neural networks (Roper et al., 2017).

Ortus is an initial implementation of, and framework for creating, virtual life aimed at approximating the intelligence of living organisms. Bourn from the study and analysis of C. elegans' connectome and behavior, it aims to strike a balance between biological abstraction, retention of biological fidelity, and computation scalability in order to as closely as possible approximate biological intelligence and learning. At its core, Ortus is a network of biologically–inspired, non–spiking neurons, capable of forming excitatory, inhibitory, and electrical synapses. Similar to the way the structure of C. elegans' 302 neuron nervous system is capable of complex behaviors including toxin avoidance, reflexively withdrawing from a "tap", and "remembering" the temperature it found food (Jarrell et al., 2012), Ortus' "connectome" (neural structure) enables its inherent functionality. Once running, Ortus refines its network—similar to the way organic nervous systems adjust themselves, based upon Ortus' intrinsic "understanding" that certain things are "good" and others are "bad", with regard to its own longevity. This understanding is derived from the structure of the nervous system it generates for itself from a set of input definitions given to it.

The remaining sections of this paper outline Ortus' design

and implementation, describe an initial experiment, discuss the implications of this framework, and analyze its shortcomings.

System Design

As Ortus aims to be a virtual analogy to intelligent life, we tried to only implement functions that either had a known analogous biological process, or which may have an analogous biological implementation that is unknown, but can be defended based off anecdotal evidence. Following each Ortus design element (ODE) below, is its biological rationale (BR):

ODE 1: The underlying source of "life" for Ortus is its respiratory circuit, which maintains a balance of O_2 and CO_2. O_2 is consumed, and CO_2 is generated, as a product of Ortus being "alive". As CO_2 rises and O_2 falls, motor neurons responsible for excitation and inhibition of the lung muscle are excited and inhibited (respectively), and the lung increases in activation, supplying O_2 and expelling CO_2.

BR 1: Maintaining a given system concentration of CO_2 is fundamental to mammalian life, and is very strongly linked to the mammalian fear response (e.g., you get scared if you can't breathe). The ability to regulate one's level of CO_2 appears to at least be part of the basis for defining what is a "good" and "bad" event/stimulus in the mammalian brain.

ODE 2: Currently, Ortus has two "emotions", fear and pleasure, represented by emotional interneurons, `eFEAR` and `ePLEASURE`, which are both tied into the respiratory circuit. When CO_2 rises, `eFEAR` rises, and Ortus would then be in a fearful state. As CO_2 falls, its contribution to Ortus' fearful state falls. The interaction between O_2 and `ePLEASURE` is the same. In this way, any stimulus presented in combination with either increased CO_2 or increased O_2 will become known as either a desirable (good) or undesirable (bad) stimulus. In this way, emotions are the driving force behind associative learning in Ortus. In Ortus, the idea of "emotions", are simply the rise and fall in activation of different neurons or groups of neurons, tied to very fundamental behaviors—such as "breathing". The concepts of "good" and "bad" sensations or emotions only carry meaning to us because of their associations to circuits that are either fundamentally desirable or undesirable from a longevity/survival perspective.

BR 2: Clearly, mammals don't normally get scared when they exhale, nor do they feel a measurable increase in pleasure upon inhalation, however the relations do exist on some level. As stated by Verma et al. (2015), "emotions, motivations, and reinforcement are a closely related, evolutionarily—conserved phenomena maintaining the integrity of an individual and promoting survival in a natural environment". Along with Gore et al. (2015), which suggests that associative learning is funneled through innate behavior circuits to assign positive or negative emotions to neutral sensory stimuli, it seems that building a virtual organism driven by emotional states is a fairly sound approach.

While at the human level, the emotional "part" of the brain is quite complex, it is not unreasonable to assume that as organismic complexity (and thereby intelligence) decreases, the complexity of emotions decreases. Numerous experiments done on rodents, such as those described by Weiner et al. (2015), show that the major structures of the brain can be examined by lesioning portions of the brain. One can infer then, that representing regions of the brain by single neurons would enable a rough approximation of the region's functionality. If one follows this line of thought, the possibility emerges that organisms like C. elegans may, in fact, be driven by "emotions" as well. For example, C. elegans is capable of toxin avoidance, a tap–withdrawal response, as well as learning that it found food at a certain temperature **(citations?)** – one must ask how this can be. There is nothing external that assists it in differentiating good from bad, yet it wants to avoid certain things, while is attracted to others. In Ortus, we make the assumption that that these behaviors are a result of a *very* simple emotional subsystem that forms the basis for C. elegans' behavior.

ODE 3: Ortus employs four different classes of interneurons at the moment. First are "Sensory Extension Interneurons" (SEIs). These take input directly from sensory neurons, and pass the input along to the second class of interneurons, "Sensory Consolidatory Interneurons" (SCIs). SCIs take anywhere from 1 to the number of sensory inputs as chemical synapses, with incoming synaptic weights equal to *1 / (# sensory inputs)*. The idea behind SCIs is to enable different types of sensors to combine their input, and trigger emotions, effectively as a "new" sensory input, thereby forming associations between two stimuli. We defer the descriptions of the last two interneurons to **ODE 5**.

BR 3: Admittedly, SEIs may not be necessary. We included these so that we could give sensory neurons a functionality that was separate from interneurons if the need arose. With regard to SCIs, it is suggested by Xie et al. (2016) that neurons in the brain are organized according to the idea that if there are N neurons, then the brain has the ability to represent all $2^N - 1$ possible combinations. Clearly, for anything but the most simple organisms, having one neuron that collects the input for each of the $2^N - 1$ possibilities is unrealistic (as the authors note); however, the authors suggest there may be additional combining of neuron inputs to decrease the computational and spacial complexity (in organic brains, that is). It also possible that mammalian brains are not quite as connected as their owners are led to believe, and a great deal of what amounts to interpolation is the reality. Regardless, we employed the idea of $2^N - 1$ SCIs because losing sensory resolution with such few neurons (that only have one neuron per sensory input) doesn't make sense. This will have to be reassessed. There is, however, a far stronger argument for the strength of the SCI in-

puts: it seems that synaptic strength scales inversely with the number of connections, "K as $\frac{1}{\sqrt{k}}$" (Barral and D Reyes, 2016). This makes intuitive sense; if neuron A synapses onto B with a weight of 1, and C, D, E all synapse onto F, the weight of each of the latter three connections must be $\frac{1}{3}$ in order to maintain equivalence between B and F. In addition, there is evidence that neighboring neurons in the same "layer" are connected in C. elegans (Azulay et al., 2016), which suggests that a certain amount of information consolidation may occur.

ODE 4: Ortus currently implements both Hebbian learning and the Stentian extension to Hebbian learning. Specifically, for each chemical synapse, on every timestep, if activity is sufficiently synchronous or sufficiently asynchronous, the synapse strengthens or weakens (respectively) according to its *mutability index* (*MI*). The *MI* determines a synapses' potential to be modified. Currently, *MI*s are static, though in the future we plan to vary them with synaptic age, among other things.

BR 4: Hebbian and Stentian learning, relating to the correlation (or lack thereof) between presynaptic and postsynaptic pairs is a proven learning paradigm in neuroscience, as described by Kutsarova et al. (2016).

ODE 5: The third class of interneurons is comprised of "Emotion Extension Interneurons" (EEIs), which are essentially auxiliary emotion neurons. EEIs are specific to each emotion, and receive input from SCIs via chemical synapses, and pass that input on to their respective primary emotion neurons via gap junctions. Emotional learning is achieved by strengthening synapses (via Hebbian learning) of interneurons that form synapses between SCIs (consolidated sensory information), and EEIs. The GJ connection enables a given emotional state, incurred by any stimulus, to permeate through Ortus—resulting in an "emotional state" of the system (gap junctions in Ortus have no activation threshold). The effect of this, is that if a certain sensory input causes a given emotion to increase in activation, the introduction of another sensory input will cause the synapses at the junction of the newly introduced sensory input and elevated emotion to strengthen. (graph here?). Further, as each SCI synapses onto one EEI per emotional state, each EEI has a "backlink" chemical synapses to their parent SCI. The idea behind the backlinks is to allow an emotional state to "trigger", or cause Ortus to "remember" a stimulus that previously invoked (and thus become associated with) that emotional state. In this case, "remembering" is defined as measurably increased activation in the SCI that the EEI backlinks into. In practice, we have observed a very slight difference in Ortus as a result of the backlink connections; currently they exist more as a concept that doesn't break the system. A diagram of this layout may be seen in **DIAGRAM!**.

BR 5: Specifically with regard to the mammalian dopamine circuit, there exists a small locus of dopaminergic cells, but the they receive inputs from diverse sources, as well as project to divers parts of the brain, as discussed by **?**. From a psychological perspective, it is clear that when one enters an emotional state, the emotion is encompassing, not localized. It is also clear that a stimulus or activity is likley to be "colored" by the emotional state one was in when the stimulus occurred. The approach we took with Ortus is a simplification of the dopamine circuit, and a generalization across emotional states; it will likely need refinement.

ODE 6: Certain emotional states preside over (dominate) others. In Ortus, fear dominates pleasure.

BR 6: The idea of a hierarchical system, where certain emotional states dominate others, is supported by the inhibition of fear in mice, in favor of searching for food when hungry (blood glucose levels falling causing the release of hormones), but greater concern for safety when not hungry (Verma et al., 2015). Further, Leknes and Tracey (2008) discusses the "Motivation–Decision Model", which suggests that anything that is more important for survival than pain should inhibit the feeling of pain.

ODE 7: Ortus builds its nervous system by reading in a *.ort* file, using our "*Ortus Development Rules*" language. First, "elements" (neurons and muscles) must be specified, with attributes such as the type of element, its affect, and activation threshold. Then, relationships between elements may be specified. Currently, there are four relationships:

- `+A` **causes** `+B`

  Where the "+" indicates an increase (and a "-" a decrease) in the presynaptic neuron causes an increase in the postsynaptic neuron. This translates into a chemical synapse.

- `A` **correlated** `B`

  Where A and B are correlated. This translates into a gap junction.

- `A` **opposes** `B`

  Where A and B inhibit each other.

- `A` **dominates** `B`

  Where A inhibits B.

Along with relationship–level attributes such as: "mutability", "weight", "polarity", and "age". At this time, most relationship attributes need not be specified as Ortus uses default values. Once Ortus reads these instructions in, it creates interneurons and connections as described by the **ODE**s above to satisfy all constraints imposed, and to allow associative learning between all emotional states and sensory stimuli (both individual and combined) as described in **ODE 5**.

BR 7:

Our goal in creating the *Ortus Development Rules* was to approximate, in a greatly simplified manner, the way a brain grows itself, based upon genetic instructions that cause circuits to form in certain ways via gene expression (Weiner

et al., 2015). Ortus' neural structure is essentially wired to force it to behave in certain ways, similarly to the development of the mammals, and how genes . Further, Schröter et al. (2017) provides evidence for the existence of organizational "motifs" found in C. elegans that may underly more complex networks in larger brains; this also lends credence to Ortus' rule–based development approach.

Implementation Details

Ortus is written in C++ and OpenCL. Each iteration of the OpenCL kernel constitutes one timestep, during which, each neuron sums incoming (positive or negative) "activation" from presynaptic cells via chemical synapses (CSes) and gap junctions (GJs). **NOTE: maybe this is a good spot for the equations!**

The chemical synapse and gap junction activation transfer equations were simplified from those described by Wic (1996) to ignore physical properties of neurons (such as neuron length).

Ortus' neurons are based upon C. elegans' neurons, and are non–spiking, but do not transmit any "activation" (footnote about why activation is used instead of voltage? or use 'potential?) below a given threshold, as in (Graubard, 1978). The equations for CS and GJ synapses were derived (simplified) from Wicks' 1996 paper (cite), and appear to have been used in recent C. elegans' models (cite Beer's 2013).

Implementation Details - C++, OpenCL - each neuron operates on its incoming connections, so it only does "postsynaptic" work, leaving its "presynaptic" work to neurons that it is presynaptic to (see "shortcomings") - sensory consolidation causes resolution loss, currently maximum resolution for any sensor array (e.g., not just O2) has a maximum resolution of 2 for intra–sensor consolidation, and 3 for inter–sensor consolidation. Single neuron sensors have no minimum for either. (NOTE: sensor array not implemented yet... take this out if it doesn't get implemented)

Experimental Design and Results

Experiment. To test our initial implementation of Ortus, we implemented the configuration described above, with the addition of an H_2O sensory neuron. We then "exposed" Ortus to water via H_2O sensory stimulation four times in bursts spaced 100 timesteps apart, in order to allow injected activation to naturally decay. During each burst of H_2O exposure, we prevented Ortus from exhaling CO_2 or inhaling O_2, thereby inducing an enhanced state of fear. After repeating the conditioning exposure four times, we allowed Ortus 200 timesteps to let any injected activation to decay. Finally, we exposed Ortus to water, without inducing a fearful state, to determine if it had learned to be fearful of water.

Results. Since CO_2 activation causes fear, and H_2O was presented during this time, we expected Ortus to become progressively more and more "scared" of water each time it was presented while it was in a state of elevated fear. As can be seen in **FIGURE**, during each exposure to H_2O, activation of `eFEAR` increased when compared to the previous exposure. Further, after the four rounds of conditioning, we observed a large spike in fear. Compared to simply exposing Ortus to H_2O without prior fear conditioning, in which a very slight raise was observed, it is clear that we were able to successfully use classical conditioning to instill a fear of water in Ortus that was previously almost non–existent.

.... should probably show graph of what happens when O2 isn't inhibited...

Discussion

While the implementation presented is quite simple, the initial results presented pave the way for quick advances to Ortus' capability with regard to "behaviors" resulting from emotional amalgamation as well as the introduction of more complex sensory input.

Clearly more emotions can be added, and can be combined in various ways, meaning there is the potential for Ortus to experience quite complex emotional states.

While Ortus' eventual goal is to develop complex "neural" functionality akin to that observed in complex systems, we

Kutsarova et al. (2016) indicates that, relative to synaptic strengthening and weakening, axonal branch tips emerge to form new synapses. synchronous activity stabilizes synapses and prolongs axonal branches. This is one way for Ortus to alter its structure in addition to synaptic weights.

? suggests that in mice, while synchronous activity (correlation) is necessary for synaptic plasticity, it is not sufficient; neuromodulatory signaling is also required. This could be a way to control some of the unstable behavior seen as a result of the purely correlation–based learning rules implemented.

Since disabling different parts of a given circuit can have slightly different effects (e.g. males will seek males, instead of females) Weiner et al. (2015), it should be feasible to grow portions of Ortus where more nuanced behavior is desired.

However one current issue is if we have SCIs for sensory neurons `A`, and `B`, we will have `iA`, and `iB`, and `iAB`. If `iAB` is activated, that means that `iA` and `iB` will also be activated. A solution might be for more complex SCIs to inhibit less complex ones, though that would perhaps unnecessarily add to computational complexity.

Sensory stimulation is sensed by sensory neurons that are topographically arranged, similar to those in organic systems. A recent study suggesting that the brain is organized based upon an n = 2something... and something about permutations lends credence to this approach (cite n something 2 permutation paper).

Talk about the combinatorial explosion, but then discuss how that can be pruned away... similar to babies. The $2^n - 1$

paper suggested reducing the resolution of inputs, so, instead of each sensor being able to connect with every other sensor, funnel 3 sensors (for example) into 1, to create a "functional connectivity module" (FCM). That approach might not be bad, however, it could be severely limiting. Perhaps this can be cut down by looking at relationships between sensors, and getting rid of things that don't matter?

Choices are: 1) an array of 9 sensors, if 0,1,2 ¿ A, 2,3,4 ¿ B, 4,5,6 ¿ C, 6,7,8 ¿ D, then sum nCr(k,n) for n from 1 to k is $2^n - 1$. It is clear that to get a fully connected network, capable of expressing all possible combinations, even from a living organism perspective, things become unfeasible.

2) Rather, it is probably the case that our brains aren't quite as connected as they appear to be. (e.g., within a certain radius, two places on your skin feel like the same location. visual system either might be an exception, however it is well known that visual information gets interpolated.) There must be some sort of neural processing shortcuts ... approximation. e.g., if you are quickly reading, you might read "New York", when the only word there was "York"

Weiner et al. (2015) talks about topographic organization. . . useful for sensors in array (e.g. visual)

Need to implement habituation

From an artificial intelligence standpoint, the

1. less connections
2. ability to specify initial behavior
3. don't need a massive training set
4. 3 types of interactions: inhibitory CS, excitatory CS, and GJ.

Together, these afford artificial neural networks far greater flexibility and thereby power. It is possible that this will enable more nuanced behavior.

From an artificial life standpoint: people are modeling networks with less complexity than this to attempt to recreate C. elegans' behavior from its connectome (cite openworm). This work pivoted from that approach, because in attempting to model the entire connectome, it became clear that there were too many unknowns. However, it also became clear that there was a certain logic to the way the neurons were wired; a notion backed up by (cite c elegans wiring works... have some in mendeley).

A simple idea is, how can C. elegans ¡DO SOMETHING¿ – point here is to describe the requirement that certain things inhibit while others excite, in order to allow a certain behavior to happen.

Schröter et al. (2017) suggests that C. elegans' neurons may have multiplexed functions, meaning that one neuron may contribute to more than one behavior. This is another possible way to decrease the number of necessary neurons in Ortus.

floor so that learned stuff can't be unlearned beyond a certain point – e.g., as a relationship ages, it's mutability should tend toward 0.

Shortcomings and Issues. There is no temporal differentiation between CS and GJ transmission. On each timestep, neurons collect all incoming "activation" from.

There is minimal presynaptic processing – (NOTE:how true is this?)

No location–based axon growth, as is seen in, relative to hebbian plasticity (cite)

Conclusion

We have presented Ortus, an initial approach to a framework that exhibits basic emotionally–based learning (fear conditioning), self–sustaining inherent behavior (cyclic breathing), and a topographical approach to sensory processing that enables inter-sensory associative learning, with or without an emotional component.

Ortus also has a system to map sensory inputs such as "breathe" to the stimulation of O2 sensors.

Acknowledgements

This work was supported by NSF grant No. ??????????????????.

References

(1996). A Dynamic Network Simulation of the Nematode Tap Withdrawal Circuit : Predictions Concerning Behavioral Criteria of the Nematode Tap Withdrawal Synaptic Function Using. 16(12):4017–4031.

Azulay, A., Itskovits, E., and Zaslaver, A. (2016). The C. elegans Connectome Consists of Homogenous Circuits with Defined Functional Roles. *PLOS Computational Biology*, 12(9):e1005021.

Barral, J. and D Reyes, A. (2016). Synaptic scaling rule preserves excitatoryinhibitory balance and salient neuronal network dynamics. *Nature Neuroscience*, 19(12):1690–1696.

Bowren, J., Pugh, J., and Stanley, K. (2016). Fully Autonomous Real-Time Autoencoder-Augmented Hebbian Learning through the Collection of Novel Experiences. *Proceedings of the Artificial Life Conference 2016*, (Alife 15):382–389.

DARPA (2017). Toward machines that improve with experience.

Gore, F., Schwartz, E. C., Brangers, B. C., Aladi, S., Stujenske, J. M., Likhtik, E., Russo, M. J., Gordon, J. A., Salzman, C. D., and Axel, R. (2015). Neural Representations of Unconditioned Stimuli in Basolateral Amygdala Mediate Innate and Learned Responses. *Cell*, 162(1):134–145.

Graubard, K. (1978). Synaptic transmission without action potentials: input-output properties of a nonspiking presynaptic neuron. *Journal of Neurophysiology*, 41(4):1014–1025.

Izquierdo, E. J. and Beer, R. D. (2016). The whole worm: Brain-body-environment models of C. elegans. *Current Opinion in Neurobiology*, 40:23–30.

Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., Albertson, D. G., Hall, D. H., and Emmons, S. W. (2012). The Connectome of a Decision-Making Neural Network. *Science*, 337(July):437–444.

Kutsarova, E., Munz, M., and Ruthazer, E. S. (2016). Rules for shaping neural connections in the developing brain. *Frontiers in Neural Circuits*, 10(January):111.

Leknes, S. and Tracey, I. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience*, 9(4):314–320.

Marblestone, A. H., Wayne, G., and Kording, K. P. (2016). Towards an integration of deep learning and neuroscience. *bioRxiv*, 10(September):1–61.

Roper, M., Fernando, C., and Chittka, L. (2017). Insect Bio-inspired Neural Network Provides New Evidence on How Simple Feature Detectors Can Enable Complex Visual Generalization and Stimulus Location Invariance in the Miniature Brain of Honeybees. *PLOS Computational Biology*, 13(2):e1005333.

Schröter, M., Paulsen, O., and Bullmore, E. T. (2017). Micro-connectomics: probing the organization of neuronal networks at the cellular scale. *Nature Reviews Neuroscience*, 18(3):131–146.

Sinapayen, L., Masumori, A., and Ikegami, T. (2016). Learning by Stimulation Avoidance: A Principle to Control Spiking Neural Networks Dynamics. page 17.

Verma, D., Wood, J., Lach, G., Herzog, H., Sperk, G., and Tasan, R. (2015). Hunger Promotes Fear Extinction by Activation of an Amygdala Microcircuit. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 41(2):431–439.

Weiner, J. A., Burgess, R. W., and Jontes, J. (2015). *Mechanisms of Neural Circuit Formation*.

Xie, K., Fox, G. E., Liu, J., Lyu, C., Lee, J. C., Kuang, H., Jacobs, S., Li, M., Liu, T., Song, S., and Tsien, J. Z. (2016). Brain Computation Is Organized via Power-of-Two-Based Permutation Logic. *Frontiers in Systems Neuroscience*, 10(November):95.