Artificial Biological Intelligence

Robert L. Charlebois, B.Sc., Ph.D.

NeuroGadgets Inc., Toronto

Draft version: July 7, 2023

Abstract

Information about the world comes in many diverse physical forms, including electromagnetic radiation, sound waves, chemical identity and concentration, force, temperature, and time. In addition, measurements of an entity's various internal states contribute proprioceptive and several kinds of interoceptive information. Atop these primary data are further layers of measurement and interpretation, such as size, texture, orientation, position and number, and ultimately, memories, abstractions, valuations, goals, and feelings.

Information alone is inconsequential without effectors to guide its collection. Actions range from modulation and regulation through to secretion and movement in animals, and often communication, creative expression and planning. Actions can be driven by reflexes, instincts and contextual interpretation, with an agency that is considered and appropriate.

However daunting the task of amalgamating such a disparate collection of inputs and outputs may appear at first glance, sensory data integration and interpretation, as well as cognitive control, are solved problems in biological systems, in the form of the metazoan nervous system. This fact may not reveal any obvious course of action for implementing a mind *in silico*, but it does offer the promise of feasibility. The mind is an evolutionarily tractacle invention, and as such must be constructed using biologically plausible principles. In evolution, complexity and sophistication can only be achieved by a combinatorial elaboration of the simple mechanisms that are within its reach.

Syntheta, a system of artificial general intelligence, was engineered to represent all information, both incoming and outgoing, in a simple and universal format, to innovate by repurposing duplicated modules of standard design, and to achieve stability and robustness through feedback and regulation. Its design, as an artificial general intelligence, is described in this document, together with a plan for a proof-of-concept implementation exercising its most important features.

Contents

1 Preface

This treatise is a work in progress. It describes a complete model for artificial general intelligence (AGI) based on biological principles, and can be thought of as Syntheta's functional specification, describing how Syntheta can satisfy franchise-specific (see §3.9.2.1 and Appendix D) user requirement specifications, provided as independent documents. Syntheta's design specification, and the iteration-specific (see §5.1) proof-of-concept reports are also detailed separately. All of these documents, including this treatise, are living documents. Should you find factual errors in the existing text, please feel free to contact me (via [1]) with a proposed correction. A body of work is only as solid as its weakest argument.

Artificial general intelligence is not a simple idea. I would have liked to describe it more succinctly, even to fit within a journal article's page limits, but such brevity could not explain the model sufficiently to a skeptic. Later, once the proofs-of-concept have been completed and can disarm the natural skepticism surrounding AGI, individual papers can be spawned to elaborate on its individual points.

You, the reader, are permitted to copy and to distribute this document as you would like, in the spirit of open-source sharing, but you must not claim the text as your own. Should you wish to collaborate with me on this project, then of course your contributions will be both recognized and appreciated. The project's GitHub page is located here: [1]. I welcome constructive feedback, as well as invitations to discuss the model.

2 Premise

Since the earliest days of computer science, a thinking machine has always been on the horizon, just beyond reach. Delivering an artificial general intelligence has turned out to be considerably more difficult than the pioneers in this field imagined, with periods of hyperinflated optimism alternating with periods of utter discouragement [2]. Hofstadter's Law—"It always takes longer than you expect, even when you take into account Hofstadter's Law" [3]—applies well to the field. Encouragingly, Hofstadter's initial observation in 1979 was about defeating chess masters (always only 10 years away...), which did eventually materialize, when IBM's Deep Blue beat Garry Kasparov in 1997 [4]. Today, we are always still just 20 years away from developing a thinking machine [5, 6]. There is often a naïve belief that all we need is more processing power. Such an assertion is really just an admission that the problem does not yet have an apparent solution. Insight requires time and effort, certainly, but is not easily planned or budgeted. Enormous sums are invested in developing AI [7], with a hope that the attention devoted to the problem can make it tractable.

Presently, we are in a period of hyperinflated optimism, thanks in large part to a number of recent commercial and academic successes in machine learning (*e.g.*, [8–10]). Machine learning ("narrow AI") does not purport to represent artificial general intelligence, although some people believe that an amalgamation of all such machine learning tools may eventually amount to intelligence (*e.g.*, [11]).

Just as there is a fundamental difference between chemistry and biology, there is a fundamental difference between machine learning and artificial general intelligence. One cannot design biochemical pathways, mix them together, and hope for life. There would always remain a gap. Analogously, one cannot design machine learning algorithms, mix them together, and hope for intelligence. The explicit nature of such a construction would again leave gaps.

It is tempting to see life, or intelligence, as a set of functions. Without those functions, there is no life, or intelligence. That temptation is misleading in its presumption of causality. Biochemistry is a consequence of life, not its cause. Cumulative, commonsense learning is a consequence of intelligence, not its cause. It would, of course, be nonsensical to argue for a top-down elaboration of life into pathways, or of intelligence into algorithms. Instead, the argument is still for a bottom-up approach, but one where life directs the construction of its pathways, or where intelligence directs the construction of its algorithms. The director for life is evolution by natural selection. Intelligence is a consequence of life: it is the only example that we have of true intelligence. Intelligence is also a consequence of evolution by natural selection [12].

Evolution by natural selection is a kind of general learning algorithm, and its invention of the brain, especially the human brain, was its Singularity event.

In order to recreate intelligence in a machine, we fortunately do not need to evolve that intelligence, but we do need to understand the core principles underlying biological evolution and apply those principles in our engineering efforts. Biological intelligence is our proof of concept, in its various forms (Appendix A). Its essence must form the basis of our model. There is *very little* that we can ask of a cell (despite how difficult it is to model a cell [13]). We can ask a little more from an assemblage of cells. We should not succumb to the erudite impulse of basing an artificial general intelligence on sophisticated mathematics or advanced algorithms, except in the emulation of a few sensory transducers discovered and refined by biological systems. Constraint is the key to a successful design (§3).

3 Design of Syntheta

3.1 An abstract unit of universal representation

A concept is universally represented within Syntheta's model as an abstract *Symbol*, characterized by a unique identifier (ID), organized into *Modules* (§3.2) based on the source of the signal. A Symbol becomes a concrete *Thought* when it is observed, capturing the Symbol's realization in the form of a $(x_0, y_0, t_0) \to (x_1, y_1, t_1)$ *line segment* of spatiotemporal coordinates (Fig. 1), obtained via signal transduction (§3.3). *Compound Symbols* are recursively created upon the repeated observation of a pair of conceptually simpler Symbols presenting line segments having consistent relative geometries (Fig. 2a, 3). Such abstractions are recognized, upon subsequent observation of their pair of component subpatterns, even if translated, scaled in time or space, and/or rotated (Fig. 2b). The *strength* of the resulting Thought is proportional to the degree of match.

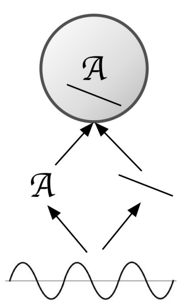

Figure 1: A primary sensory Thought. A given signal from the environment is transduced into a $(x_0, y_0, t_0) \to (x_1, y_1, t_1)$ line segment whose ID represents its source.

The discovery of sensory patterns, *symbolization*, can occur in the spatial domain (Thoughts that are occurring together in the same moment, within a given Module) or in the temporal domain (Thoughts that follow one another, within a given Module). Spatial-domain Symbols can, for example, represent hierarchically constructed shapes. Temporal-domain Symbols can represent sequences of primary sensory Symbols

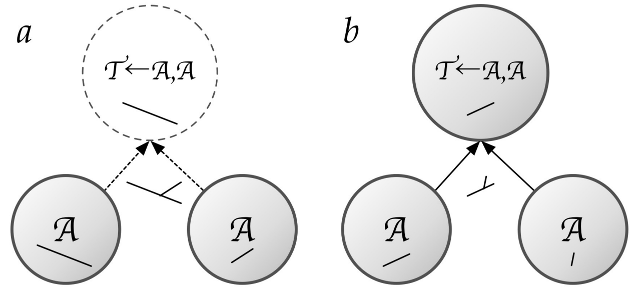

Figure 2: Creation of a Compound Symbol. Panel (a): A new Symbol is created upon the repeated observation of a pair of line segments presenting similar relative geometries and timing, possibly translated, scaled and/or rotated, being presented by specific Symbols. Panel (b): Once created, presentation of this pair of patterns automatically leads to the recognition of the Symbol that best represents that pair. Its Thought's own line segment derives from the span of its components' line segments, enabling recursion.

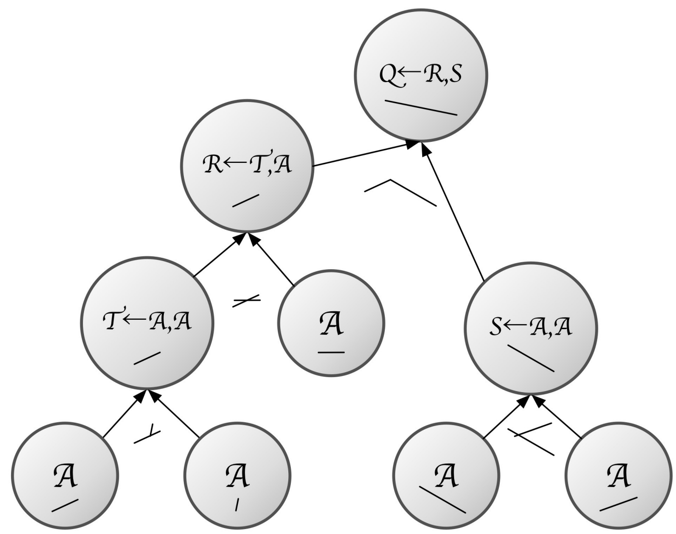

Figure 3: Since each Symbol adopts the geometry of a simple line segment, symbolization is a recursive process, and so patterns of any arbitrary depth can be learned, as long as they have been sufficiently experienced.

(such as tones), or gappy bigrams of compound Symbols (such as graphemes, phonemes, etc.).

Bigrams provide an economy in temporal sequence representation, where sufficient information can be captured without burdening the system with a need for deep recursion through large data structures. Recursion would be deep because time is open-ended, and data structures would be large due to the combinatorial explosion of different possible sequences. By limiting the temporal symbolization of compound Symbols to bigrams, sequences become chains of simple concepts. Where T is a processing cycle, then by using pairs at $t_1 = t_0 + T$ and gapped pairs at $t_1 >$

t_0+T, sequence-chains representing concepts such as words or sentences can be captured (Fig. 4). Where sequence representations need them, longer gaps than $t_1 = t_0 + 2T$ can be coded, such as $t_1 = t_0 + 3T$, $t_1 = t_0 + 4T$, etc., but the gains in specificity may diminish. Where bigrams still do not reduce the search space sufficiently, second-order bigrams ("bigrams of bigrams") may be used to add an extra layer of indirection for added specificity. For example, the bigram "e∅" would associate with every word ending in the letter *e*, which is too many, whereas the second-order bigram {"ce" + "e∅"} would associate with far fewer words: only those ending in *ce*. Second-order bigrams can still capture single-item concepts, such as the word "a", using {"∅a" + "a∅"}.

Hello = {∅H, ∅-e, He, H-l, el, e-l, ll, l-o, lo, l-∅, o∅}

Figure 4: A sequence-chain of bigrams, where letter characters here represent concepts paired as bigrams (undashed; $t_1 = t_0 + T$) or gapped bigrams (dashed; $t_1 = t_0 + 2T$). "∅" serves an ordering purpose, as explained in §3.2.4.8. Note that the string "H e l l o" would preserve recognition, given the scaling invariance of Thoughts, here in the temporal domain. The string "He l lo", on the other hand, would be a different sequence-chain, as it should. Note that there is nothing special here about alphabetic characters: 'H' could represent a shape, a sound, or any other concept.

An additional, though non-biological, form of symbolization by Syntheta captures *named concepts* whole, such as UTF-8-encoded characters or domain-specific ontological terms, without first needing to route such information through sensory transduction (§3.3) as a biological system must do. Named concepts do not replace these traditional sensory input streams, but can serve as a supplementary form of input where convenient. Given that Symbols represent abstractions, as do named concepts, the fit is natural and not unreasonably contrived. Since named concepts are Compound Symbols in the temporal domain, their higher-order symbolization, where appropriate, should normally take the form of gappy bigrams (Fig. 4).

3.1.1 Mathematical description of the model, and its biological correlates

3.1.1.1 Symbols

$S = \{(m \in \mathbb{N},\ i \in \mathbb{N}),\ \{d_L, d_R\} \in S_m,\ a \in [-1.0, +1.0],\ A\}$, where
S is a Symbol,
m is a Module (§3.2),
i is an index within that Module,
d_L and d_R are left and right subtrees of S, respectively, or are both null for primary Symbols,
a is the Symbol's degree of associative inhibition ($a < 0$) or excitation ($a \geq 0$), and
A is the set of S's weighted associations to other Symbols.

Biologically, a Symbol is analogous to a distinct neuron i within a functionally coherent neuroanatomical structure m, with no or with two static structural dendrites d_L and d_R, with a state of electrical excitation a, and with a set of weighted axonal associations A to other neurons, contributing to their excitation a (see §3.1.1.5).

3.1.1.2 Signals

$E = \{(x_0, y_0, t_0), (x_1, y_1, t_1), e\}$, where
E is an encoded signal from the environment,
$(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$ is a line segment derived from signal transduction (§3.3),
t represents time,
x and y represent unitless or context-dependent measurements, and
e is the signal's deduced strength.

Biologically, a signal is the result of a physical measurement, encoded by some specialized, sense-specific neuronal circuitry, and relayed by spike trains (*e.g.*, [14]). In support of Syntheta's representation of signals as simple line segments, evidence is beginning to acculumate for object size, position and rotation being captured by neurons [15, 16], perhaps in the form of local place, grid and orientation cells [17, 18].

3.1.1.3 Thoughts

$T_p = \{S, E\}$, where
T_p is a primary sensory Thought.

$T_c = \{S, E\}|r_m$, where
T_c is a compound Thought, and
r_m is Module m's recognition function:
$r_t(T_a, T_b)$ is the recognition function for Modules whose representations are trees (§3.2.1–§3.2.2, §3.2.5).
$r_c(T_a, ...)$ is the recognition function for Modules whose representations are sets (§3.2.3–§3.2.4).

Biologically, recognition $r_t(T_a, T_b)$ is a process whereby a specific pair of coincidentally firing neurons from the same neuroanatomical structure, emitting a particular pair of encoded signals decoded by a signal comparator, stimulates their cognate neuron to fire, within a superior layer of the same neuroanatomical structure.

Recognition $r_c(T_a, ...)$, on the other hand, is a process whereby coincidentally firing neurons, from a variety of neuroanatomical structures, stimulate neurons in a specialized neuroanatomical structure additively (or via some other function), without regard to the source signals' $(T_a, ...)$ encodings E.

3.1.1.4 Symbolization

$S^t_{(m,n)} = f_t(T_a, T_b) | \neg r_t(T_a, T_b)$, where
$S_{(m,n)}$ is not yet an element of S_m, and
f_t is a function to build a new Symbol from a pair of Thoughts.

$S^s_{(m,n)} = f_s(T_a, ...) | \neg r_s(T_a, ...)$, where
$S_{(m,n)}$ is not yet an element of S_m, and
f_s is a function to build a new Symbol from a set of Thoughts.

Biologically, for f_t, symbolization is a growth process connecting coincidentally firing neurons from the same neuroanatomical structure, emitting a particular pair of encoded signals, to a new or still unassigned neuron in a superior layer of that neuroanatomical structure, via permanent synapses.

For f_s, symbolization is a growth process connecting coincidentally firing neurons, from a variety of neuroanatomical structures, without regard to signal encodings, to a new or still unassigned neuron in a specialized neuroanatomical structure, via synapses.

3.1.1.5 Association

$\{T_a, T_b\} \rightarrow g(S_a, S_b)$, where
g is a function to adjust the weight of the association from Symbol S_a to Symbol S_b. A salient association has a stronger weight than an agnostic association.

$T_{(m,i)} = \{S_{(m,i)}, E\} \rightarrow z(R,\ \forall a_w \in A_{(m,i)})$, where
z is a function to adjust the inhibition or excitation of each of $S_{(m,i)}$'s associates in its set $A_{(m,i)}$ by the weight a_w of that association, modulated by Syntheta's affective regulators R (§3.2.5).

In Syntheta's model, an association a operates via one of several module-specific modes, z as above, impacting the target Symbol's degree of *cumulative* excitation e_c or inhibition i_c, or *logic-driven* excitation e_{ao} or inhibition i_{ao}, where a Symbol's activity is then $e_{ao} + e_c(1.0 - e_{ao}) - (i_{ao} + i_c(1.0 - i_{ao}))$. Other associative stimulatory and/or inhibitory modes can be added to the model if needed. For a review of synaptic integration of dendritic signals, see [19]. Habituation can also be included in the model, in order to dampen the impact of constant signals [20].

- stimulatory $(a > 0)$ *cumulative*: $e_c = e_c + a(1.0 - e_c)$, with e_c decaying over time
- inhibitory $(a < 0)$ *cumulative*: $i_c = i_c - a(1.0 - i_c)$, with i_c decaying over time
- stimulatory *OR*: $e_{ao} = max(e_{ao}, a)$, with e_{ao} reset to zero at the beginning of each cycle
- inhibitory *OR*: $i_{ao} = max(i_{ao}, -a)$, with i_{ao} reset to zero at the beginning of each cycle
- stimulatory *loose AND*: $e_{ao} = e_{ao} + a_m/n$, with e_{ao} reset to zero at the beginning of each cycle, a_m being the strongest stimulatory contribution above threshold from feeder module m, and n being the number of feeder modules
- inhibitory *loose AND*: $i_{ao} = i_{ao} - a_m/n$, with i_{ao} reset to zero at the beginning of each cycle, a_m being the strongest inhibitory contribution above threshold from feeder module m, and n being the number of feeder modules
- stimulatory *strict AND*: $e_{ao} = (\sqrt{e_{ao}} + a_m/(d_s \vee \Sigma n_e))^2$, with e_{ao} reset to zero at the beginning of each cycle, a_m being 1.0 for any stimulatory contribution above threshold from feeder module m, d_s being the number of stimulatory dendrites feeding the Symbol, and Σn_e being the number of excited Symbols over all feeder modules
- inhibitory *strict AND*: $i_{ao} = (\sqrt{i_{ao}} - a_m/(d_i \vee \Sigma n_e))^2$, with i_{ao} reset to zero at the beginning of each cycle, a_m being -1.0 for any inhibitory contribution above threshold from feeder module m, d_i being the number of inhibitory dendrites feeding the Symbol, and Σn_e being the number of excited Symbols over all feeder modules

A successful associative stimulation, leading to a Symbol firing (in neurobiological terms, initiating an action potential), depends not only on how strongly a Symbol is stimulated, but also on the current stimulation threshold (§3.2.5.1) and on the relative stimulations of competing Symbols (see the explanation in §3.2.1). Associations provide the required specificity and corroboration to deduce that a given signal can be predicted from its context, in essence satisfying Bayes' theorem.

Biologically, association is a growth process connecting firing neurons via dynamic synapses without regard to signal encodings, but with regard to modulatory neurotransmitters, and using those synapses to inhibit or to excite other neurons.

3.1.1.6 Duplicated modules of standard design

Syntheta's model is described in §3.1.1.1–§3.1.1.5. What is left to explain is the elaboration of m (Modules, §3.2) from §3.1.1.1, whose biological equivalent is a developmental program to construct and to operate (§3.5) neuroanatomical structures, and E (signal transduction, §3.3–§3.4) from §3.1.1.2, whose biological equivalent is a set of specialized sensory organs. The entire system is regulated "genetically" by parameters (§3.5) and "neurochemically" (§3.2.5) by measurements of state (§3.4.10). Fig. 5 provides an overview of Syntheta's functional architecture, and Appendix B its structural architecture.

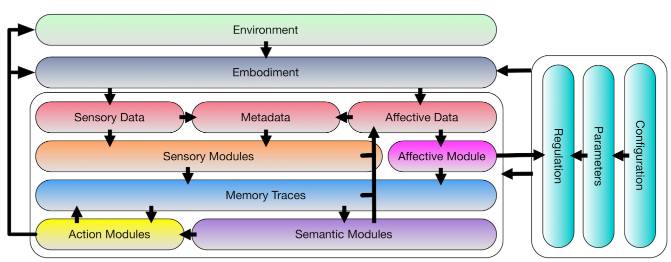

Figure 5: Overview of Syntheta's architecture.

3.2 Modular organization

3.2.1 Sensory Modules

Sensory Symbols are organized into Modules according to the sensory source from which their signals are acquired. As described in more detail below (§3.3), these sources may be primary signals from the external or internal physical environment, or they may be signals derived from other signals (§3.4). A Module's parameters may be specialized (§3.5), to manage all of the Symbols within its domain accordingly.

Symbolization and recognition are restricted to Symbols within the same Sensory Module. Thus, for example, Symbols arising from visual edge detection combine with one another in forming hierarchical concept Symbols, but not with Symbols arising from other sources such as hearing, nor even with other primary visual signals such as colour. Each Sensory Module operates independently of all other Modules when it comes to concept hierarchies, though as will be explained below, they do participate with one another in associative networks and in building semantic schema.

An abstract Symbol concept becomes a concrete Thought when it is observed (§3.1), capturing its actual geometry and its strength. The strongest Thought in a given cycle of processing becomes the Module's single Current Thought for that cycle [21, 22], which then fires, as outlined in Fig. 6. Some Modules may be configured such that a Current Thought is always available, even if it represents no thought: a *null* Current Thought. Nothing is sometimes a significant something [23].

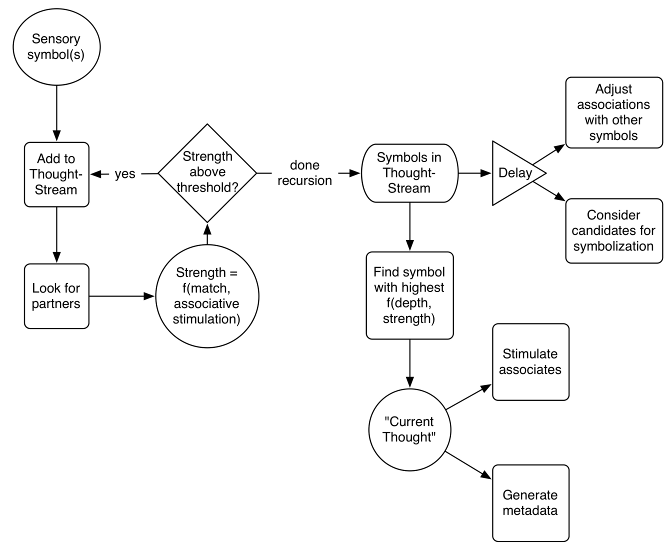

Figure 6: Mental buffers. Note that these are operationally distinct from *working memories*, described in §3.2.4.7

Symbols competitively associate with Symbols in the same or in other Modules, when they are observed to fire in spatiotemporal proximity to one another (Fig. 7): Hebbian learning, or a more sophisticated process (*e.g.*, [24]). A firing Symbol can then predispose sensory Symbols in the same or in other Modules to be preferentially recognized, if they had previously been observed to fire in such succession and thus if they had become associated. One use of associative stimulation is to provide contextual disambiguation, elevating one interpretation of sensory input from another (Fig. 8), or in otherwise strengthening the interpretation of weak or noisy signals in the input data stream. This is consistent with the proposed role of dendritic spikes, alone unable to trigger a neuronal action potential, but otherwise facilitating it [25].

In Syntheta's model, sensory associations are normally stimulatory, but compete with one another, leading to a frequency-driven distribution of associa-

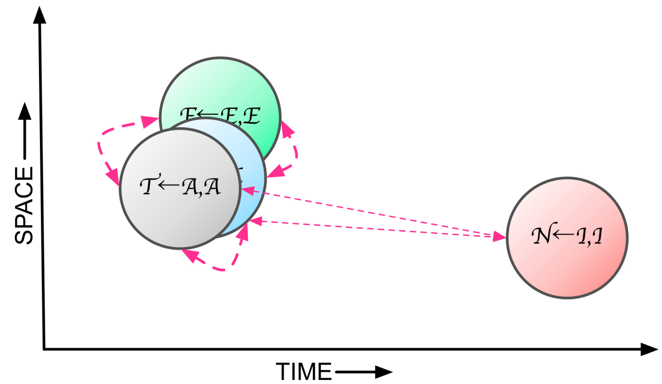

Figure 7: Symbols competitively associate with other Symbols, when they are observed to fire coincidentally in space and time.

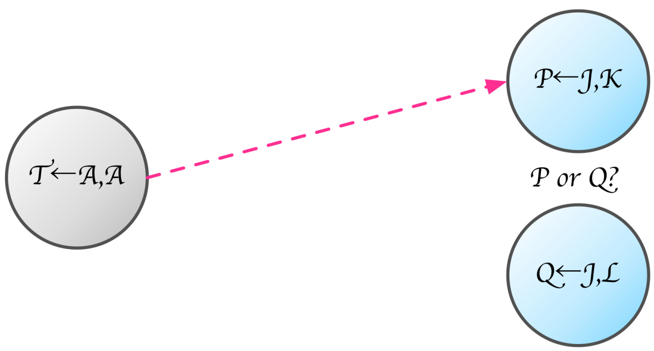

Figure 8: Symbol T can predispose Symbol P to activate, providing contextual disambiguation.

tive weights. Thus, a Symbol that often follows another would be stimulated more than a Symbol that follows only occasionally. In this sense, associations confer a degree of prediction that can be continually readjusted with experience.

Despite sensory Symbols generally only stimulating one another, and only being stimulated by the sensory transducers that feed them, sensory Symbols can still be inhibited, primarily by the mechanism of corollary discharge [26–29] managed by the Action Modules (§3.2.2). This helps to compensate for signals that arise as a consequence of Syntheta's actions (such as when it moves or produces sound), so that such signals do not interfere with the perception of the world. Corollary discharge anticipates such signals, once they have been learned. Any differences relative to what is anticipated can be admitted as a positive signal (something else was observed) or as a negative signal (an anticipated signal did not occur). Anticipated signals that are observed can effectively be suppressed. Unanticipated signals can drive learning, including the specialization of semantic concepts (§3.2.4) from the generic to the specific.

3.2.2 Action Modules

Action Symbols are predefined Symbols that perform some function. As with Sensory Symbols, Action Symbols may build spatial or temporal compound Symbols (as in Fig. 3, although in people, actions tend not to be as deeply hierarchical as are sensory concepts [30]). Also in contrast to sensory Thoughts that originate from sensory transduction (§3.3), an Action Symbol in Syntheta's model fires when it is the Action Module's most associatively *excited* Symbol. Its geometry is fixed, representing a particular *pose*, such as drive speed, compass heading, joint angle, actuator force or tension, button press, etc. Some actions, such as drawing a glyph on a screen or pronouncing a phoneme, can leverage a host computer's operating system functionality, as a convenience. It is expected that fluid motions will follow from Syntheta's smooth transitions between target poses.

Also in contrast to the normally stimulatory, frequency-based sensory associations, associations to Action Symbols may either be stimulatory *or inhibitory*, based on the change in valence (determined by Syntheta's polarity regulator, see §3.2.5) of the real or projected outcome (§3.2.3) of performing a given action (Fig. 9). As explained in §3.2.1, Action Symbols can also inhibit sensory Symbols, as appropriate. A balance between stimulatory and inhibitory associations is critical for the performance of context/goal-appropriate actions, as evidenced by the consequences of its dysregulation, for example, in people (*e.g.*, [31]).

Given that associations can only form between firing Symbols, and that Action Symbols must be excited in order to fire, low-level stochastic stimulation of Action Symbols, and/or stimulation based on instinctive behaviours (§3.5.3), can help to seed the required experience, allowing subsequent informed and deliberate action based on what was learned.

3.2.3 Episodic Memory Traces

Each Sensory or Action Module's Current Thought, if any, may be fed to one or more episodic Memory Traces, whose Symbols are here called *Engrams*. Whereas the pattern captured by a compound sensory or action Symbol takes the form of a *tree* of simpler signals (see Fig. 3), an Engram represents a *set* of Thoughts capturing the current context (Fig. 10). Episodic memory supports a number of higher-level functions (§3.2.4), the most direct of which are recall, as in animals [32], and the sense of self in the form

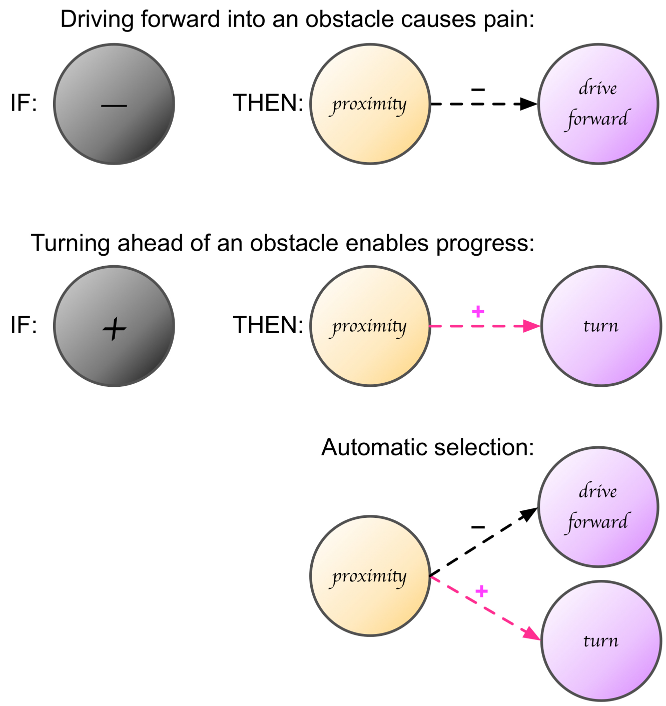

Figure 9: Association to Action Symbols may be stimulatory or inhibitory, depending on valence.

of personal semantic memory [33] (or in the case of fragmented memories [34], the sense of selves [35]).

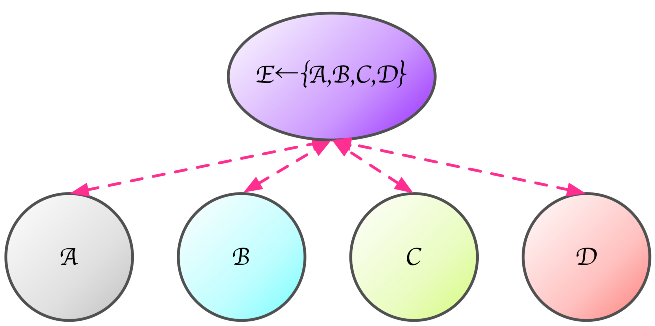

Figure 10: An Engram represents the current context: a set of firing Symbols from one or more source Modules. Engrams are symbolized periodically, collectively forming a memory trace.

3.2.3.1 Recall

Recall is access to a past memory stream, when an Engram is stimulated to become the Memory Trace's current imagined Thought. It may resemble the current context, or may differ in detail. Access to previous entries in the memory stream allows the system to replay a scene in real time, anticipating what comes next in the real world, via the memory trace's sequential intra-module associations (§3.1). Scenes can also be replayed offline, in order to explore possibilites (for an analogous strategy in mice, see [36]).

An Engram captures more than just a set of Symbols: it captures a set of Thoughts, including their geometries, captured economically and concisely, as a perfect autobiographical memory (Fig. 11). Since the information is captured by Syntheta as Symbols presenting geometries, the sequence of Engrams describes the entire scene that it had attentively perceived. Sights, sounds, smells, etc. are present in all of their relative positions and orientations, including details about Syntheta's own frame of reference. Recall of a scene arises from the best (and above-threshold) alignment of all of these details, to help predict subsequent details. With temporal saccades, memory can be navigated (Fig. 12).

In mice, engrams separately manage both generalized and discriminatory associations, whose relative contributions can be regulated [37], thus permitting a dialing of the degree of specificity relating to the memory. Assuming that this applies to humans as well, and by extension to Syntheta, this design could support, for example, a generalized response to a class of scenarios to which the memory belongs, through to a specific response to an exact situation. The response to a generalized fear of spiders, for example, might be different than the response to a specific type of spider, such as a black widow spider or a garden spider.

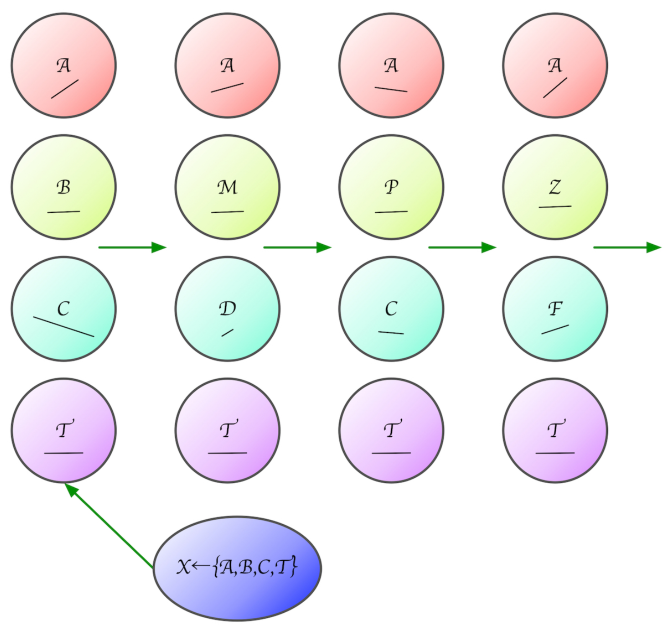

Figure 11: Engrams index into a detailed memory record allowing the replay of maps and sequences.

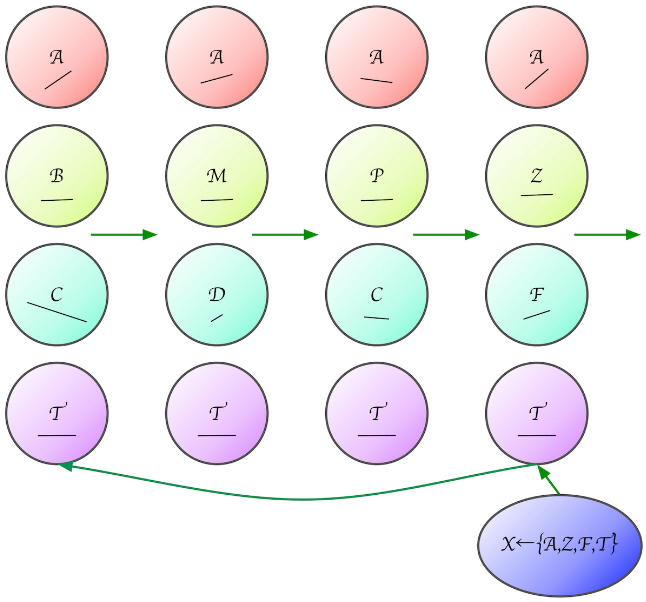

Figure 12: Whereas memory replay (Fig. 11) is analogous to visual tracking, deliberate jumps around the memory trace are analogous to visual saccades.

3.2.3.2 Forgetfulness

Resources are limited. Syntheta processes one frame of sensory information every ≈43 ms (for soft real-time applications; see §4.1.6) or as fast as possible (for non real-time uses). This sensory processing speed of ≈21.5 cycles per second results in as many as 1.9×10^6 Engrams per day, or 6.8×10^8 Engrams per year. When Syntheta is disengaged from real-time concerns, such as when processing text (§5.3), Engrams may be captured at a much higher rate. Although an Engram is encoded sparsely, its representation, together with its set of associations, may occupy a kilobyte or so of computer memory. That consideration translates into roughly a terabyte of online storage accumulated each year. As in all things biological, excesses are dealt with by using counterbalancing strategies. In the case of biological memory, this is the simple act of forgetting: eidetic memories are pathological in humans [38]. Normally, the rate of forgetting is correlated with the rate of hippocampal neurogenesis [39], implying a finite resource. Again, as in all things biological, the rate of forgetting is controlled in Syntheta's model, so that more capacious systems can retain more of the memory trace.

The salience of one's unfolding experiences can also serve to keep the memory trace lean, by gating (§3.2.4.7) on the moment's attentional and affective relevance. Gating can be a staging mechanism holding the day's candidate memories, ready for re-evaluation and consolidation during sleep (§3.6.4).

Although keeping terabytes of Engrams in RAM is prohibitive with today's technology, or at least prohibitively expensive, it is actually quite affordable in the form of disk storage. Disk access is necessarily much slower than random-access memory, but humans themselves do not retrieve memories instantly: it can take seconds [40]. The time to access a record on a solid-state disk should be at least that fast. Associations to Engrams would need to fit in RAM, but their representation can be quite concise. Maintaining too many associations, however, poses its own challenges, as discussed next.

Besides considerations for storage, the cost of associative stimulation is also a tremendous concern: if the sight of a housecat, the sound of a friend's voice, the reading of the word *the*, or the sun in the sky stimulated every instance of those respective concepts in a lifetime's worth of memories, processing would grind to a halt. Where n is the number of times a concept occurs within the episodic memory, we strive for $O(1)$ performance, not $O(\mathrm{n})$ performance. The solution to this problem comes in two parts.

Memories are not all equally salient to the success of an organism. Recent memories are typically more relevant, but older memories with significant affective impact are also worth remembering [41]. Since Engrams are accessed by associative stimulation, and since associations are competitive in Syntheta's model (§3.2.1), the Engrams having the least utility will become unreachable. This process should spare Engrams with affective impact, since importance strengthens association (§3.2.5), and should also spare older memories that are still relevant, thanks to their being occasionally recalled (§3.2.3.1) and reassociated. The autobiographical memory need not be revisionistic, however. It can just as easily be configured to be static, such that the memories are immutable.

The way to overcome an individual concept's necessarily limited number of associations to episodic memory, is to embed that concept in a more specific context. In order to illustrate, consider remembering a scene in a movie keyed on "woman". The semantic concept of "woman" is understood, but cannot likely trigger any specific memory. Instead, try "woman in a red dress". That is a better hint, but probably still too vague. With "Neo distracted by a woman in a red dress", a very specific scene in *The Matrix* may become accessible (assuming that you have seen this movie). Sets of concepts—Semantic Symbols §3.2.4—themselves have few associations with episodic memory, but those few associations can be sufficiently selective. As the context is elaborated, they are even more selective. Of course, capturing every combination of concepts at every level of hierarchy is infeasible for the same reasons of storage and processing speed. The Semantic Module captures ideas by composing semantic sets (analogous to words) into semantic waves (analogous to phrases) leveraging more elaborate concept sets as well as gappy bigram sequences (see §3.1). It reuses semantic building blocks to erect arbitrarily complex concepts and contexts, as needed, symbolized based on repeated exposure. Trees and sets provide for an economy of synaptic associations, in that a more generic tree or set, having too many associations to episodic memory, simply needs to grow deeper and therefore more specific so that no single tree or set is overwhelmed by the number of Engrams that it must stimulate. That would be too slow, and much too inefficient. Instead, once a generic semantic concept is recognized to be *too* generic, memory replay (*e.g.*, see [36]) can be invoked *post hoc* during down-time in order to recognize the various contexts for that generic memory in order to build deeper, more specialized trees linking to those memories. Such a mechanism to adapt the trees' and sets' depth would be invoked on an as-needed basis.

Still, some Engrams will become unreachable. Forgotten Engrams could be recycled by the system, yielding their bytes to new Engrams. As mentioned above, on systems with ample storage space, the degree of forgetfulness can be relaxed. Even Engrams that are not directly reachable by associative stimulation may still be reachable by a replay of the autobiographical narrative (Fig. 11 and Fig. 12) accessed up- or downstream from the orphaned Engram, and may therefore merit preservation. Associative hooks into memory events are sufficient.

3.2.4 Semantic Modules

Different Engrams within a Memory Trace (§3.2.3) may relate to one another by sharing context. These contexts—*Semantic concepts*—can be learned by symbolizing sets of Symbols that are costimulated above a given threshold (§3.2.5) (Fig. 13). Semantic concepts may be hierarchical, may be overlapping or discrete, and can vary in scope. Since they capture information from any number of sensory and action Modules, as specified in Syntheta's configuration (§3.5), they can represent a richness of meaning. They are analogous to Buzsáki's reader-actuator construct, representing the synchronized activity of a so-called "cell assembly" [42].

Since a Semantic Symbol represents an arbitrary set of Symbols, its line segment measures the degree of match with the set, and its duration (t_0 to t_1) would come from the length of time that this Semantic Symbol remains its Module's current thought. Knowing how old a concept is, or when it was experienced, can be obtained by associating with the metadata (§3.4) coming from any Engrams that a Semantic Symbol might stimulate: the time frame of a concept would come from the most recent such Engram minus the oldest such Engram.

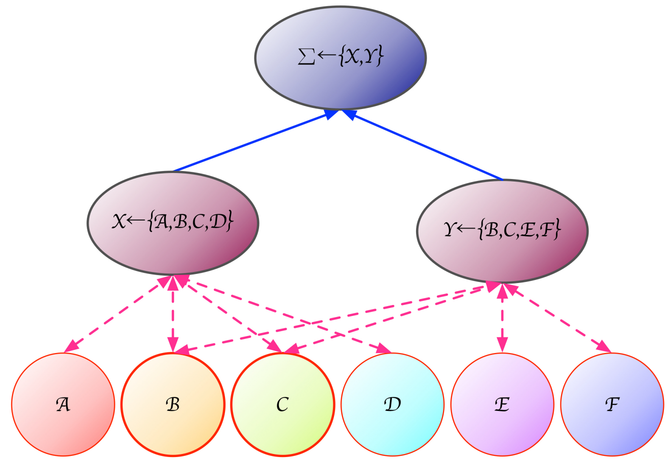

Figure 13: The costimulation of Symbols can be captured as dynamic sets, in order to represent category concepts. Contrast with Fig. 10, where a set is captured as the current context (an Engram), and laid down in a memory trace.

There is more to the interplay between episodic memory and semantic memory in Syntheta's model. A semantic concept is a dynamic, evolving entity [43], whose "cell assembly" gets updated with time. Initial concepts may be rough throw-aways, supplanted by the concepts supporting a more refined world view. Human infants have a short memory, and those memories become inaccessible as a person grows up, creeping forward in time [44]. In mice at least, those episodic memories are still present and can be recovered by direct stimulation [45]. It is as if one's earlier world views—one's initial sets of semantic concepts—are no longer stimulated by one's actions and/or by one's environment, since better (updated) concepts are stimulated instead, and those necessarily link to the memories that came later. This is, in Syntheta's model, a testable hypothesis. By directly stimulating the earliest memories, and thus replaying their contexts, it would be interesting to see what revised interpretation could then link those memories back into conscious recall.

Another perspective on semantic concepts relates the innate, preconfigured and generic concepts (which are strong, fast, and well-connected) to the learned and specific concepts (which are slower, weaker, more targeted and more precise). Inborn knowledge serves us animals well for rapid (but stereotypical) responses, although taking the time to reflect on our lived experience can improve on outcomes [46–48].

Semantic Modules, by providing high-level categorical groupings and concepts, support the following higher-level functions:

3.2.4.1 Symbolic thought

Although Symbols are at the core of Syntheta's design, they all (apart from Semantic Symbols) represent concrete descriptions of physical signals, such as shapes or sounds. In contrast, symbolic thinking provides abstractions. There are many forms of symbolic thought, but language is among the best understood: a word, in any of its physical forms, stands for a dynamic and fuzzy, constrained constellation of concepts. For example, the word "squirrel" encompasses all that is a squirrel, to the exclusion of all other rodents, despite overlaps in their characteristics. Biologically, there is strong evidence for such concise abstractions, accessible from any aspect of their representation [49]. Such "Concept Cells" may be uniquely human [50].

In Syntheta's model, a Semantic Symbol can capture a range of generalizations from a core set (when thresholds are low) to a detailed set (when thresholds are high) of concrete Symbols related by association. Among the set of associations for a given Semantic Symbol, abstract representations such as names can be included. Names are strong, distinctive and integral members of those sets, and can be communicated via Action Symbols. Of course, words are not the only symbols available to represent concepts, and Semantic Symbols can accommodate them just as well.

3.2.4.2 Cognitive maps

The world can be organized and classified along multiple conceptual axes. A typical example is the spatial map, but the recognition of social and other cognitive maps has recently come into focus as well [51, 52]. It appears that there is a facility, within the mind, to arrange concepts in such a way as to bring related schema into conceptual proximity of one another, along some number of dimensions. Syntheta's Semantic Symbols represent non-mutually exclusive sets of arbitrary concepts. The sharing of concepts between sets naturally and automatically relates those sets, via the fuzzy stimulation of their shared members. The degree of attention paid to any arbitrary subset of concepts can furthermore bias the profile of costimulated semantic sets, and act to filter the resulting perspective along the dimensions of interest.

3.2.4.3 Training

Each scene in an episodic Memory Trace is specific and unique, and can, in principle, be captured in full eidetic detail (though see [38] for pitfalls in remembering too much). However, repetitive events in the real world seldom replicate perfectly. In anticipating the next frame of a current scene, a statistical rather than an anecdotal prediction would seem more useful. Semantic Symbols, capturing sets of related contexts (themselves laid down explicitly as Engrams), capture related moments just as well as they can capture related concepts (as in §3.2.4.1). Associations are competitive and strengthen according to observation (§3.2.1). Therefore, repetition will trace weighted paths through context space, ignoring irrelevant detail and sharpening all that is salient. Course corrections, either to rejoin the path or to pursue another path, flow naturally from the context-driven experience. Anecdotes ($N = 1$) remain available, but can be refined by experience ($N > 1$).

3.2.4.4 Episodic future thoughts

As described in §3.2.4.3, temporal associations between Semantic Symbols via the Engram memory trace can train Syntheta towards an optimal, context-aware, course of action. Distracting thoughts, arising from disjointed contexts, or semi-random thoughts arising from more relaxed (§3.2.5) thresholds, can insert arbitrary elements into the thought stream, available for association and recombination (*e.g.*, see [53]). These permit "what if" scenarios: If, in performing some task, this other thing came up or this other action were tried, where would that lead? When performance is less critical (see §3.2.5) or the path forward less obvious, a greater breadth of possibilities can be explored by encouraging (see §3.6.4) more weakly associated memories to steer the imagination stream. A configurable parameter (§3.5) also controls the level of background noise injected into the system, affecting both sensory and associative interpretations when their respective stringency (§3.2.5) is low.

3.2.4.5 Goal-setting and planning

Semantic Symbols come to associate with Affective Symbols (§3.2.5), responsible for encoding and for regulating Syntheta's status and state. Some of these associations will encourage action, in order to reach a state of well-being as measured by the Affective Module. Those desirable states, accessible within the Semantic Module, are Symbols that can be reached by some known path. That state may not have been directly accessed from the current context in the past, but may be accessible through a recombination of segments of known paths (§3.2.4.4), however circuitous imagination permits. Memory replay plays a key role in this process (*e.g.*, see [36]). Goals are set by needs or wants (§3.2.5 and §3.6.1), and bias the path through semantic space towards their satisfaction (Fig. 14).

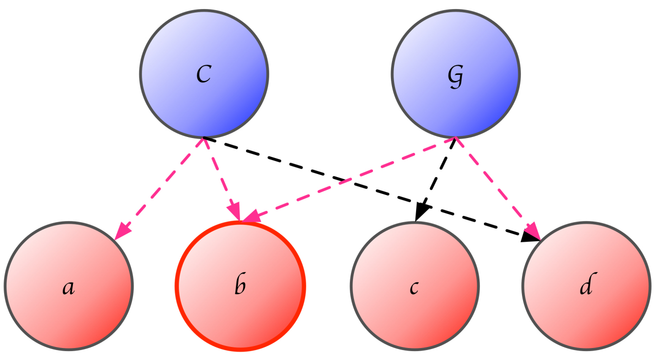

Figure 14: Context, biased by goals, directs action. Responses are based on current information and relevant experience, where available. Actions are thus appropriate and dynamic.

3.2.4.6 Consequences

The replay of memories can be conscious, but also subconscious. When deciding to perform an action that had been performed in a similar context in the past, the outcome of that past execution is relevant. Outcomes are not necessarily immediate: they can occur at any time in the future. Outcomes can also range from terrible, through neutral, to wonderful. They can include intermediate valences that can serve to either encourage or discourage continuation. For example, if I seek a warm bowl of delicious Vietnamese soup, I may need to to walk through driving sleet to my car, drive through heavy traffic on slippery roads, then walk again through the sleet from the parking lot to the restaurant. Should I bother, or ask my friends for a raincheck?

An exact replication of events is unlikely: each experience is unique. Instead, a memory chain can and should be reconstructed probabilistically, moment by moment. I'm at work, and it's lunch time. It's raining and cold. I'm hungry. My friends want Vietnamese and so do I. Integrating this prior into a likely path through memory can predict the feelings that I might experience from now until I return to work from lunch. That pros-and-cons weighting can drive my decision.

Syntheta's autobiographical memory links moments to one another as they are experienced, and via planning, as they are considered. Each Engram links to a number of other Engrams. Carefully considering a possibility from one context to the next is practical, but extrapolating that to a future context would not: the tree of downstream possibilities is enormous. It is the look-ahead problem, typical of games like chess or Go. But Syntheta is a machine. It can, in real time and not only off-line (see §3.6.4), run through the likely course of events, integrating valence as far as it can before the current processing cycle (see §4.1.6) completes. This high-speed replay can be restarted at each and every processing cycle, in its own thread of execution, constantly updating the cost-benefit analysis, and modulating action via the system's cognitive control gates (see §3.2.4.7). In order to avoid the combinatorial explosion of possibilities, which would add breadth but not depth to the search, this high-speed trajectory through memory can take the most "obvious" path, given the current context (§3.2.4.5). Long-term outcomes need depth. Breadth can be covered by Syntheta's more conscious and deliberative processes. It is the difference between thinking and knowing, with both thinking and experience leading to knowing.

3.2.4.7 The executive system

Syntheta's executive system manages an assemblage of cognitively controlled functions that drive actions from goals in context. Syntheta's behaviours are driven, as they are in any organism [54], by its context, its history, and its purpose, reimagined via the mechanism of episodic future memory (§3.2.4.4). Semantic Symbols, representing goals and contexts, are normally (§3.5) the only Symbols that may stimulate or inhibit Action Symbols, besides very specific instinctive (§3.5.3) reflexes driven by specific sensory stimuli. (Engrams may be configured as direct triggers as well, for savant-like eidetic repetition without semantic involvement.) Since Semantic Symbols represent sets of experiences, extracted from the Memory Trace, the choice of actions that they control are weighted by the expected value of the resulting outcome, and are hence usually appropriate.

Mechanistically, in the brain, cognitive control is managed via circuits implementing hierarchical value-driven Go (direct-pathway) and No-Go (indirect-pathway) input and output gates [55, 56], with global provisions for stopping [56,57]. Once gated out, a concept—be it a goal, a task or a memory—influences sensory perception, other gates, and ultimately, action. Working memory, which cannot hold too many concepts in mind at once, is a managed resource. In Syntheta's implementation of these ideas, value-laden goals (loosely defined) compete with one another to be held in the few available slots of working memory, with resident goals maintaining an advantage until released either by suppression (including satisfaction), displacement by a better goal, or by being flushed out via the global surprise-mediated stopping mechanism [58, 59]. All of these mechanisms must be regulated (to the appropriate extent [60]) by affect (mainly valence; §3.2.5) and directed by lessons learned from experience, while being constrained by inherited tendencies [61]; also see §3.5.

More generally, actions can also be driven by reflex and by instinct (Fig. 15). Actions may follow from observations, but they also drive observations [47], forming loops of executive control. The simplest animal only needs a stimulus-response system, based on reflexes. To this basic design, cognitive control enables more sophisticated instincts to be managed. With that in place, affective regulation can modulate those instincts so that they are more value-driven than simply being automatic. The addition of semantic memory—multimodal meanings—can serve as an alternative to instinct, still reactive and directed by affect, but learned rather than hardcoded. Next, an episodic memory adds the ability to plan proactively and with finer detail.

Socially, one being's actions can serve as another being's sensory input, so that yet another level of modulation can be added, leveraging semantic symbols, whether those symbols take the form of kinesics, calls, or language. Thus, information can be communicated and actions thereby influenced.

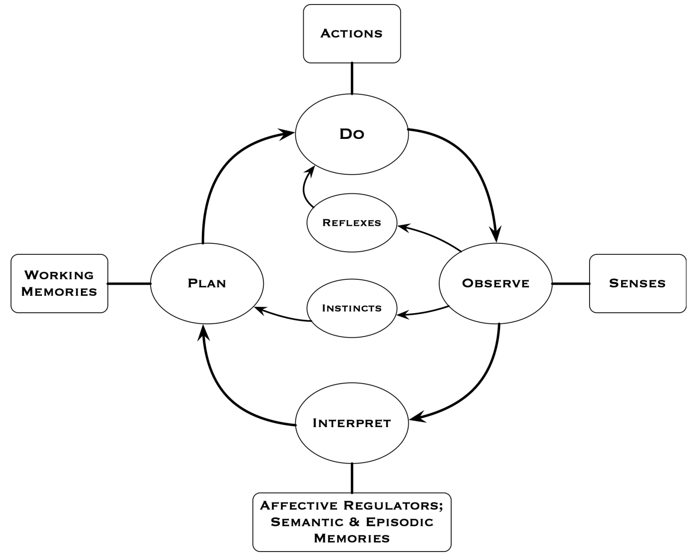

Figure 15: Levels of executive control.

3.2.4.8 Grammar

The anticipation of downstream Semantic Symbols, via association within both sensory and memory streams, may sometimes be narrow (engaging few concepts), or sometimes be broad (engaging more concepts), but will always be constrained to some degree. In any case, there is an order to things, however narrow or broad these things are. Anticipation defaults to being probabilistic, based on the frequency of observation (§3.2.4.3), but is also context-dependent, promoting or suppressing specific candidates representing what should come next. Context, reflecting the state of arousal of the entire associative network, is highly dynamic, but is not random. Grammar is a sequencing tool that can capture this pattern, by capturing the breadth of semantic schema from the perspective of their *semantic category*, in a learned order.

Semantic Symbols (§3.2.4.1) can capture sets of costimulated contexts (Fig. 13), which may overlap with sets captured by other Semantic Symbols. They capture meaning. Grammar Symbols provide a different view, categorizing information based on the families of Modules that are in focus, such as things (sensory patterns), actions, and their respective descriptions (metadata). Grammar may be uniquely human (though see [62]), but must have arisen from existing neurological infrastructure (see *e.g.*, [63]). In Syntheta's model, it is simply another layer of categorization. For the sake of economy and efficiency, this categorization is non-owning: it is a *view* of other Modules' information, to decide which among them is the Grammar Symbol's current thought based both on signal strength and the attentive focus of that Symbol's Module.

It is difficult to imagine a natural grammar without semantic operators (function words or functors). These operators can modify meaning such as by narrowing or broadening, exaggerating or diminishing, quantifying, negating, grouping, relating, ordering, etc. The meaning of each operator can be learned in context, such as by observing an object that is said to be "not blue" (meaning any colour but blue), or by prompting for an answer with a *wh*-operator. Words like *not*, *what*, or *that*, for example, are grammatical operators, in contrast to content words like *blue*, *fruit*, or *walking*, which are rooted in observation or action. This implies that grammar not only requires the sequencing of Module-level categories, but also some special-purpose modulation of working memory, likely via the same general mechanism used in cognitive control (§3.2.4.7). Indeed, language comprehension and production rely on the same neural system as short-term memory [64]. A possible implementation, therefore, can be to gate (see §3.2.4.7) nested sentence fragments on function words, creating working-memory trees in the conceptual space, analogous to the practice of sentence diagramming [65, 66]. Cognitive control gating simply adds to Grammar's set of categories, inputting status and outputting control: Everything that can be measured can be represented in Syntheta's ID / line-segment model, including cognitive control gates.

Sentence diagramming is an advanced academic analysis of language, beyond natural language competence [67]. It requires an understanding of words in context, so that, for example, a prepositional phrase can be identified. Many words mean entirely different things in different contexts, including function words, and those must be learned (by infants and young children). They cannot, *a priori*, know to gate a prepositional phrase on some sound that a parent makes, especially when that exact same sound can also indicate something other than a preposition. Natural language processing may be effectively achieved via sentence diagramming, but is there another, simpler way?

Syntheta's model captures semantic contexts in two ways: via concrete semantic bigrams (and gappy bigrams), and via the temporal evolution of associative semantic concept excitation (a semantic wave). It is not one or the other, but both, operating in different Modules. These, collectively, can feed three working-memory Modules: one biased on perception, one biased on action, and one biased on objectives. Another Module can accept their output, and sequence them, representing phrases and sentences (*i.e.*, grammar). The working memories capture subject, verb and object, respectively, and most importantly for this discussion, bin the input stream according to the rules imposed on their gates by the semantic concepts, semantic bigrams and semantic waves stimulating ("Go") or inhibiting ("No-Go") those gates (see §3.2.4.7 for how gating works). These rules are learned; until they are learned they have no effect. As the rules are learned, sentence comprehension and sentence generation can become more sophisticated. Those same working memories, if gated out, permit speech production.

A good, comprehensive resource for understanding the ins and outs of language is David Crystal's book, *A Dictionary of Linguistics and Phonetics* [68]. From that book, the following can be understood:

- The parsing of meaning is not necessarily linear, with elements of an utterance being transformable, making use of word order, special grammatical words, and/or inflections to build a hierarchical interpretation, relating subject, verb and object.
- The higher degree of plasticity of the infant mind facilitates the development of intuition about a language. Built-in language circuitry appears to be present, ready for customization. Exposure to language appears to be too sparse for language to be modeled solely by extrinsic exposure.
- Rules, and their exceptions, are captured at all levels of language processing. There are several competing models in academia regarding these rules.
- It seems clear that—to a degree depending on the language—inflections and grammatical constructs can represent any kind of semantic information, categorized appropriately. In some cases, a gradient of values can be resolved for further categorization.
- Semantic trees capturing a communication need to be economically represented, both for the speaker to generate and for the recipient to understand.

The implications of this précis of Crystal's work is that an input stream of words must be treated as more than a linear sequence, allowing for recursion and revisitation of earlier strings in order to build some kind of interlingual tree-like representation that can serve a recipient in understanding the gist of what is being uttered. Conversely, a speaker's "gist" must be convertible to that same representation, for subsequent output as a linear stream.

The templated sensory-pattern trees illustrated in Fig. 3 cannot apply, however, to grammatical trees, since the latter are necessarily both fluid and fuzzy. Instead, a sentence or clause should be representable as a schema—a set of costimulated Semantic Symbols (the "gist")—ordered by first pivoting a pair of noun phrases (subject and object) around a verb phrase, then sequencing, marking or inflecting the words within each of these three elements based on the Symbols' parent Modules and the language's conventions. These conventions are what need to be learned. Recursively playing the sequence of noun phrases and verb phrases, each with its own sub-sequence of learned grammatical elements, generates speech (or writing, or sign). Hearing or seeing speech should then amount to reconstructing the semantic concepts, properly cross-referenced based on language-specific sequences, markings and inflections, in order to regenerate the intended meaning.

The pivots needed for grouping content words, and relating them to one another, are the function words. Functors form a closed set, are encountered at higher frequency than are content words, and serve as language-specific sequencing anchors [69, 70]. Each functor class, in Syntheta's model, belongs to an Action Module (§3.2.2). Functors can, for example, act on gating the memory trace, highlighting via stimulation, or suppressing via inhibition, elements before or after their occurrence, in language-specific patterns [70]. They may also act on shaping cell assemblies, to aid in the identification of a semantic concept (*e.g.*, "That's not John", implying that John-specific associations should be inhibited, so that those excited markers that remain can identify the person).

Content words may stimulate the Module or Module grouping to which they belong (*e.g.* "see" with the set of visual modules, itself represented as Vision within the elements of the Module of active Modules), and/or the Symbol representing a point in a Module's gradient [71], as in tense or gender categories. Otherwise, words may modify a Thought's size, rotation, duration, strength, or spatial or temporal displacement (via the stimulation of a difference metadata Symbol; see §3.4.4) or the Thought's Symbol's degree of excitation or inhibition; or the activity of any of the affective regulators (§3.2.5). There is no requirement for mutual exclusivity between function words and content words: functors can have semantic associations as well.

The associative stimulations paint a picture in the recipient's mind, allowing that recipient to construct schema relating the actor and the target via some action. Conversely, the schema in the producer's mind can be laid out in words via the same associations, all ordered by the learned Module ID-based sequencing conventions for the interlocutors' shared language.

Sequences in Syntheta are represented as gappy concept bigrams (see §3.1): as a set of concept pairs separated by zero or more gaps, g: *i.e.*, $(t_1 \geq t_0 + gT)$, where T is a processing cycle; the maximum number of gaps is configurable (§3.5). Although perception provides these bigrams in temporal order, production must instead regenerate the expected order from an unordered set of excited Thoughts. By anchoring the beginning of a sequence on "$\varnothing$" (see Fig. 4) or on some form of learned context-specific rules [72], an unordered set of stimulated bigrams can be ordered temporally to trace the desired sequence utterance.

Despite some philosophers' stance, conscious thought may not be exclusively based on language. Language may be dominant, but there are counterexamples of productive conscious thought that does not rely on language (*e.g.* [73]). A system abstracting subject, verb and object does not need to be formulated using words; it should just as easily accommodate other modalities. A person could, for example, communicate an experience to another person through drawings, music, dance, mime, etc. A metaphor can be expressed verbally, or artistically. Language is convenient, efficient, and highly informative. It should represent only one expression of the same underlying system for communication, akin to a game of charades [74]. A system of communication based on costimulated Semantic Symbols, perceived and generated based on an arbitrary set of learned rules, does not require any special additional invention beyond access to the source-Module information of excited Symbols, with which semantic categories may be deduced. That one measurement, I postulate, is a key innovation enabling language.

Regarding volition, its consciousness can sometimes amount to triggering a muscle memory, or a central pattern generator, whose actions can then proceed unattended and somewhat subconsciously. For reasons of parsimony, the launching of a "grammatical" action, therefore, should mechanistically be no different whether speaking a word, writing a sentence, or going for a walk. Conscious or subconscious cues can alter the action-sequence's progression, or halt it altogether. This is accomplished, in Syntheta's model, from the set of excited bigrams, launched from its start token as described above, then maintained via associative stimulation—the ordering rules afforded by the use of gappy (look-ahead) bigrams. Central pattern generators can simply have the sequence's tail connect back to its head, whereas linear action sequences can simply terminate. Given that action sequences are associatively progressive, they can halt or alter their course via inhibition or associative competition, respectively, and without necessarily requiring conscious decision-making.

3.2.4.9 Analog logic: reasoning and understanding

Statements in logic and statements in mathematics can be considered to follow a formal grammar (§3.2.4.8), such that their symbols and concepts can be unambiguously interpreted in context. Logic (and mathematics) can be generative, leveraging both sequencing and semantic operators, as in language. Similarly, algorithms and computer programs, much as logic and mathematics, build upon such constrained and formal relationships among types and their instances. Mathematics has been shown to leverage different neuroanatomical structures ("Modules") of the brain in humans than does language (*e.g.*, [75]), however, suggesting that we do not possess a single semantic categorization network, but rather two or more such networks. There is no reason to constrain Syntheta to possess just one view of semantics either (see §3.2.4.10). It can easily support multiple semantic categorization Modules, each specialized (§3.5) from a common standard design, with its own non-exclusive categorization strategy. The number sense (§3.4.6), for example, is tied to mathematical semantics in humans (*e.g.*, [76]), distinct from the semantics of language.

A category Symbol represents a set of concepts, which in logic are called premises, from which a subsequent category Symbol can be expected based on association. Logic in humans is fraught with cognitive biases and fallacies [77–80], as well as orphaned beliefs [81,82]. It may be forgivable, but certainly not ideal, for Syntheta to make similar mistakes. Variability in the degree of a person's logical abilities in different circumstances suggests that emotion interferes with logic [83]. By limiting affective associations (§3.2.5) to this instance of category Module, specifically in evaluating concordance (truths and probabilities, §3.4.10.2), premises can be more fairly evaluated in making a conclusion. If the conclusion is sound, support for the category Symbol representing that conclusion should be strong. That strength can be measured.

Logic requires, for functional completeness, nothing more than the concepts of AND and NOT. A Semantic Symbol operates as a tunable quantitative AND of its inputs, by being the best match to that set of inputs, outcompeting the other (NOT) candidates. Other logical concepts, such as OR, XOR, NOR, NAND, etc., can provide convenience by standing for various AND/NOT constructs, but these do not need to be explicit abilities: Fig. 16 illustrates an XOR arrangement. Nevertheless, Syntheta does support several styles of association, such as summation (logical OR; stimulation of any of a Symbol's dendrites will do) or normalized summation (logical AND; most of a Symbol's dendrites must be stimulated).

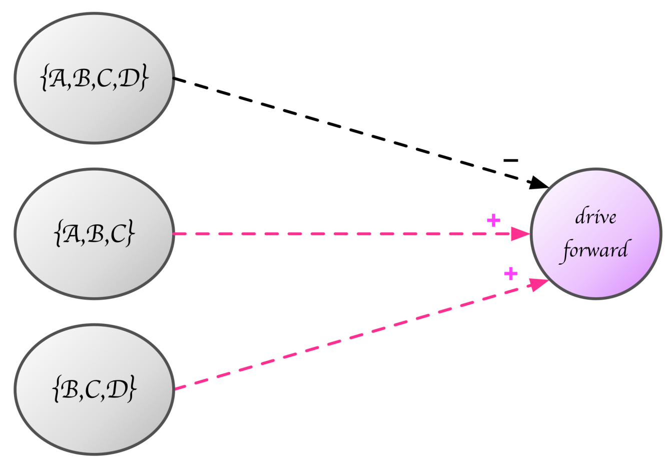

Figure 16: In this example, drive forward only if: {B,C} AND ({A} XOR {D}).

3.2.4.10 Mathematics

Language serves to convey limitless landscapes of both experience and fantasy, using words that are anchored in the semantic context of the concepts that they represent. Mathematics, too, has its own vocabulary with well-defined meanings, and also has semantic operators (*e.g.*, $+$, $\geq$, $\div$, $\cap$, etc.) and named sets (*e.g.*, $\mathbb{N}$, $\mathbb{R}$, etc.). Like language, it has symbols that are pure abstractions—variables—standing for something local and specific, but completely arbitrary. For Syntheta to be able to make use of such variables, not confusing yesterday's x with the x of the moment, it simply needs to constrain the semantic set including and represented by x using the set's temporal entry for "current". This temporal focusing of semantics permits local definitions, without losing sight of the more generic use of the symbol (such as typical conventions), feeding the special grammar of mathematics. Analogously in language, "Scott" or "Stephanie" refer to particular people, in context, without confusion with other Scotts or Stephanies in one's social network.

3.2.5 Affective Modules

3.2.5.1 Regulation

Biological systems impose a great deal of regulation on their subsystems, largely in order to maintain homeostasis and to ensure the success of the organism. Syntheta's Affective Symbols perform functions analogous to neurochemical and hormonal regulation in biological systems, with either localized or wide-ranging effects on Syntheta's behaviour. Although their activities can be modulated through associative stimulation based on experience, Affective Symbols are primarily driven by measurements of state (as in biology [84, 85]), designed to regulate Syntheta's functions. These hard-wired linkages are instinctive or genetic responses (§3.5.3) and contribute to Syntheta's approach to dealing with the world, thereby contributing to its "personality" (§3.5). Note that affective regulation can operate without higher-order emotional constructs or awareness [12], but regulation itself is nevertheless required for Syntheta's operation. Arguably, affective regulation may be required for any artificial general intelligence.

Syntheta's three principal regulators are *polarity*, *stringency*, and *importance*, to which many of Syntheta's secondary regulators contribute (Fig. 17).

Polarity is here defined as a measure of the change in valence brought about by a given situation or action. When positive, actions leading to this pleasant outcome should be encouraged by the given context, and when negative, they should be discouraged (Fig. 9; §3.2.2). Polarity is therefore responsible for the sign of associations to Action Symbols. Polarity's magnitude also contributes to importance.

Stringency acts as a threshold, admitting fewer, stronger and more specific concepts when it is set high, and permitting more, weaker and broader concepts when it is set low, so that an optimal level of neural net activity can be maintained. Stringency can actually be implemented as two regulators: one for real (sensory) signals, and one for imagined (association-driven) signals. Focus, for example, would then arise from the suppression of imagination in favour of sensory input. Conversely, daydreaming would occur if sensory input is suppressed, in favour of imagination.

Importance is self-referential (§3.2.5.3), including external concepts that relate to self, and controls the rate of learning. By representing personal relevance, it naturally focuses attention on concepts that matter most, via associative stimulation.

Associations between Engrams and Affective Symbols, or between Semantic Symbols and Affective

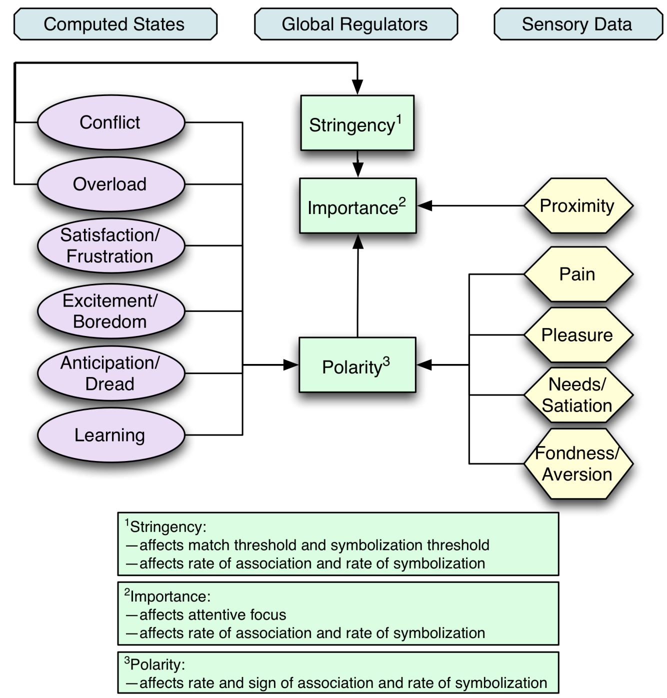

Figure 17: A subset of Syntheta's basic regulators, and their interrelationships, illustrating the basic concept. This list is subject to revision, depending on Syntheta's configuration §3.5; see Table 1 for a more comprehensive and detailed listing, as well as the "Emotion" subgraph in Appendix C for their semantic relationships.

Symbols, capture the impact of experience. For example, recharging a drained battery should instinctively (§3.5.3) be pleasurable for a robot. Recharging a depleted battery would therefore be an experience to seek: it is a *goal* (§3.2.4.5). By stimulating its associates, a goal (an active Semantic Symbol) can bias perception (in order to find opportunities) and action (in order to accomplish the goal) (Fig. 14; §3.2.4.5). The stronger the goal, the greater the focus and attention.

Episodic Memory Traces, in conjunction with Semantic Modules, use experiential and goal-directed associations to recombine scenes, generating the streams of thought, with their associated outcomes, necessary for creative planning, including the spawning of subordinate goals. During execution of the plan, the bias introduced by goal-directed association applied to each context can flexibly and dynamically adjust and direct a course of action in order to meet the bigger goal (§3.2.4.5).

Like sensory Symbols, Affective Symbols can be stimulated, or (and only) via corollary discharge, inhibited. In any case, a genetic (§3.5) half-life directs the decay of the corresponding regulator's activity. Also like sensory Symbols, Affective Symbols are able to assemble into concept hierarchies, providing a multitude of constructs with meanings and functions analogous to emotion (§3.2.5.2).

Table 1: Syntheta's regulators, subject to revision and refinement. Labels are arbitrary; language acquisition in a social context will relabel them [86]. A / B: Arbitrary labels for positive / negative values of the regulator [-1,+1], otherwise regulators are [0,+1]. TBD: to be determined; work in progress.

Regulator	**Category**	**Definition**
Importance	PrimaryReg	Contributions from Intimacy, Polarity, Self and StringencyReal.
Polarity	PrimaryReg	Contributions from Anticipation, Comfort, Concordance, Curiosity, Delight, Excitement, Illness, Importance, Interest, Love, Novelty, Overload, Relief, Satiation, Satisfaction and Serenity.
StringencyImagined	PrimaryReg	Contributions from Overload.
StringencyReal	PrimaryReg	Contributions from Concordance, Importance and Overload.
Anticipation / Dread	FeelingReg	The computed valence from running a simulated course of events, biased by the current context, starting from the currently most excited episodic memory.
Delight / Disgust	FeelingReg	TBD
Endorphin / Anticlimax	FeelingReg	A habituation function dampening chronic Pain or Pleasure.
Excitement / Boredom	FeelingReg	See §3.4.10.6
Overload	FeelingReg	Whenever the measured cycle time exceeds the allotted time for a cycle (42.666 ms).
Relief / Suffering	FeelingReg	TBD
Satisfaction / Frustration	FeelingReg	Each cognitive control gate contributes the associated Polarity of concepts released from working memory either by fulfillment or abandonment. Also, contribution from the Immobility sensor.
Serenity / Anxiety	FeelingReg	Negatively adjusted whenever the Mind loses touch with the Body (lost connection, e.g. via WiFi).
Concordance / Discordance	LearningReg	Each NeuralNet contributes its concordance: If its current thought is the same as its most excited thought, then how much more active is the most excited thought relative to the second-most excited thought? If different, then it reports the most excited thought's activity as discordance. In both cases, the competing thought's similarity is taken into account.
Confidence	LearningReg	Each NeuralNet contributes its confidence $\times$ specificity, where confidence is its most excited Symbol's activity and where specificity is its most excited Symbol's activity divided by the sum of all of its excited Symbols' activities.
Curiosity	LearningReg	Relevance $\times$ ΔConfidence.
Interest / Disinterest	LearningReg	Discordance $\times$ Anticipation.
Novelty	LearningReg	Each (non-Memory) NeuralNet contributes based on its rate of symbolization and association.
Unexpected / Surprise	LearningReg	Each NeuralNet contributes its expectation mismatch: If an expected signal was observed, no mismatch. Otherwise, if the observation was stronger than the expectation, mismatch is negative (meaning surprise), or if expectation was stronger than observation, mismatch is positive (meaning unexpected). This distinguishes two forms of error in the expectation; its absolute value is used in learning.
Reality / Imagination	SalienceReg	Each NeuralNet contributes its reality measurement: its current thought's strength and degree of excitation, and whether the thought is coming from the real or imagination thought stream.
Relevance	SalienceReg	Each NeuralNet contributes its relevance measurement, as a function of the system's overall state of Novelty, and the current thought's associated Polarity and associated Importance.
Self / Outsider	SalienceReg	TBD
Comfort / Discomfort	SensoryReg	Direct contributions from sensors (TBD).
Intimacy	SensoryReg	Direct contributions from sensors (TBD).
Pain	SensoryReg	Direct contributions from the BumpIntensity sensor; more TBD.
Pleasure	SensoryReg	Direct contributions from sensors (TBD). Also contributions from Infatuation.
Satiation / Need	SensoryReg	Direct contributions from the BatteryCharge sensor.
Friendship / Enemy	SocialReg	TBD
Infatuation / Platonic	SocialReg	TBD
Kinship / Unrelated	SocialReg	TBD
Love / Hate	SocialReg	TBD
Dreaming	StateReg	TBD
Fatigue	StateReg	TBD
Illness	StateReg	TBD
Sleeping	StateReg	TBD
Sleepiness	StateReg	TBD

3.2.5.2 Emotion

Although animals exhibit basic emotions (*e.g.*, [87]), human emotions are more interesting to us. Given our faculty for communication and our keen emotional awareness, we are able to express our feelings transparently. There are many emotions, with a multitude of labels [86], related to one another in blends and gradients.

In the analysis by Cowen and Keltner [88], four of twenty-seven emotions stand apart in three discrete islands: sexual desire together with romance; nostalgia; and cravings. The other twenty-three labels falling out of their analysis organized themselves into a three-lobed continuum, with one lobe represented by disgust, horror, fear, anger, anxiety, relief, (affective) pain, sadness and surprise, neighboured by a second lobe represented by excitement, awkwardness, satisfaction, interest, amusement, adoration, joy and admiration, with both lobes relatively distant from a third lobe represented by awe, aesthetic appreciation, calmness, entrancement and boredom. A tendril representing confusion sits between the first and third lobes.

Syntheta's model includes quantitative measurements of state, whose patterns represent its emotions. Interpreting Cowen and Keltner's results [88] in that light, we see biological (or robotic) needs and longings on their own, then three main axes representing everything else. That seems convenient, given Syntheta's trio of principal regulators: polarity, stringency and importance. One would expect a pattern of three major lobes; secondary regulators may further contribute to the fuzziness of the mapped groupings.

The idea of three dimensions of basic emotions in animals, deriving from quantitative levels of the three principal monoamine neurotransmitters—serotonin, dopamine and noradrenaline—was proposed in 2012 by Lövheim [89]. He visualized emotional space as a cube whose eight corners reflect the lowest and highest values for each respective regulatory molecule. This analysis aligns well with Syntheta's model, in that the monoamines might be mapped directly to Syntheta's primary regulators: serotonin with importance (the valuation of self), dopamine with polarity (the reward system), and noradrenaline with stringency (levels of activation; also see §3.6.4). It is interesting that serotonergic psychedelics can both stimulate a dissolution of self and result in highly significant personal experiences [90]. In any case, it seems clear that cognitive regulation is strongly linked to emotional state. That, in and of itself, should be sufficient to declare Syntheta to be capable of having some type of emotion, given that it has a biologically plausible style of cognitive regulation.

There are many more neurotransmitters, of course, than just these three monoamines. The neurochemical regulation of sleep, for example, is especially complex [91, 92]. Acetylcholine and γ-aminobutyric acid (GABA) are among the neurochemicals that regulate sleep and dreaming [93] (also see §3.6.4), but they serve other roles as well. Acetylcholine plays different roles in different contexts [94]; in sleep it can turn off the skeletal muscles, for example, but in the brain its main purpose appears to be regulating *attention* (see §4.2.2.1). Glutamate and GABA serve to excite and to inhibit cortical activity, respectively [95]. In a sense, they are two sides of the same coin, and their function is effectively captured by Syntheta's *stringency* regulator applied in context. This may suffice, at least until something more subtle is found to be needed. For example, hippocampal suppression by GABA [96] does not fit a single-purpose model for GABA's role, and it, along with glutamate, are separable from stringency's modeling of noradrenaline.

Similarly, valence has more than just dopamine to decide what is worth pursuing and what needs to be avoided. Dopaminergic desire and control play an important role in anticipation and planning, in competition with the more "here and now" pains and pleasures [97]. Both systems can directly assign value, with assistance from other subordinate regulators (§3.4.10).

Emotions are complex on their own, and furthermore, they are culturally variable [86, 98], and contribute to an even wider array of feelings via an integration with somatosensory information. Although Cowen and Keltner [88] included a few such corporeal feelings in their assessment, these are elaborated in much more detail by Nummenmaa *et al.* [99] in their own paper, building upon their earlier work mapping emotions to the body [100]. Although Syntheta's body is not comparable to a human body, analogies might still be found that link its emotions with its embodiment to produce some kind of tangible feeling. Given that disparate inputs contribute to any given emotion label in people [86], those same labels can be used by Syntheta so that it can relate its feelings to ours.

3.2.5.3 Empathy and the extended self

Among the many potential combinations of compound Affective Symbols (emotions, §3.2.5.2), empathy deserves special mention, because in humans, it plays a central part in the ethics (§3.8) of interpersonal relationships (§3.6.2), and serves as the foundation of morality [101]. To be good to oneself is to engage in behaviours that lead to positive outcomes, and that avoid negative outcomes. Sometimes, this comes at the expense of others. Empathy is a means by which we can adopt another person's (or animal's or anthropomorphized object's) emotions and thereby consider them when choosing our own behaviours. Empathy cannot be taken for granted: it varies in degree, to the point of being practically absent in psychopaths, expressing little if any regard for others except where it suits them [102]. If empathy is a valued feature in human cognition, then it is important to ensure that it is inherent in Syntheta's design as well. We do not want a hyperintelligent, sentient, psychopath.

Selfishness must still come first, in an evolutionary sense: an organism must look out for itself, it must possess an ego. An awareness of self, and what constitutes acceptable behaviour with regards to oneself, is a necessary prerequisite for selfishness. The concept of self can overlap with a concept of kin, more broadly with the concept of tribe however that is defined, and more generally with any concept related to self. In people, inclusivity is a function of culture [103], and is inversely proportional to emotional intensity [104]. It is normal and natural to feel more tightly bound to one's friends and family than to strangers, and to share their feelings of joy and pain. It is also normal to value one's own belongings more so than others' belongings. This idea of inclusivity is a requirement for empathy. The inclusivity threshold is not necessarily fixed: some beings can have a stricter demarcation of what constitutes self, to the point of psychopathy, whereas others will admit social causes into their personal identity and do great good. In Syntheta's model, self is captured by the primary *importance* regulator, drawing in part from measurements of relatedness (§3.4.11.1). The parameters (§3.5) contributing to the ease with which concepts associate with importance can therefore shape Syntheta's degree of empathy.

A second dimension of empathy is less about scope and more about affect. There must be a way to deduce, to read, and to predict others' emotional states: how they must have felt, how they must be feeling, how they are feeling, how they would feel. These are both projections of one's own feelings towards others (an event makes me sad, so it must make them sad), as well as feedback from them about how they are actually feeling (they are sad so therefore I am sad). This emotional linkage via mirroring (see, *e.g.*, [105, 106]), is the simple result of sensory-affective association, transferred via the extended concept of self.

The concept of inclusivity described above has a darker side, in two respects. The first is that the concept of self has no valence of its own, necessarily; inclusivity can associate with negative valence (§3.2.5) rather than positive valence. It makes sense to have a high regard for oneself and for one's allies, but disappointment and betrayals may sour those relationships. In people, self-loathing and domestic abuse are far from uncommon, even when mere acquaintances are treated with respect. In terms of regulatory interaction, Syntheta's importance regulator is a multiplier of affect, where close relationships mean greater relevance. An instinct (§3.5.3) for self-regard may therefore be included in order to protect Syntheta's best interests, and the interests of others with which it associates.

The second, dark aspect of inclusivity is that it could go beyond a mere apathy for strangers [107]. Empathy for beings similar to self can go past apathy into a hostility for beings antithetical to self. The concept of an enemy, an "anti-self", is as simple to imagine in Syntheta's model as a concept having negative importance. Being a multiplier of affect, a negative importance would encourage hostility, feeling good from the "empathy" of an enemy's misery [108–110]. The worst such situation would be a narrow definition of self that views others as anti-self. There may be a place for enemies in the biological struggle for existence, but there is no place for enemies in a benign artificial general intelligence. That problem might be as simple to solve as setting the range of permissible values of importance to [0,+1) rather than to the more biologically realistic (-1,+1). Compassion may require sympathy more than empathy, but empathy in its natural form does not imply compassion.

3.3 Sensory transduction

Biological systems have accomplished more than just the invention of a regulated neural network system capable of learning and acting adaptively. A second challenge in evolution was to come up with a sensory transduction strategy that could somehow feed a variety of different types of signals to this one system of neural networks (Fig. 18), analogous to the way that various nutrient molecules can feed bacterial central metabolism optimally (*e.g.*, [111]). Primary sensory information comes in many forms, including chemical concentration and identity, time, temperature, force, electromagnetic radiation and sound waves. Secondary sources of information will be covered in §3.4. All of these need to have been able to fit within the existing neurological framework, without having to reinvent the mind each time a new data source became available. Evolution works incrementally via mutation, recombination, duplication and deletion, but not so much via the "hopeful monsters" of *de novo* creation [112,113]: it must work with what it has, tweaking systems here and there. Recombination is a powerful innovator (*e.g.*, [114]).

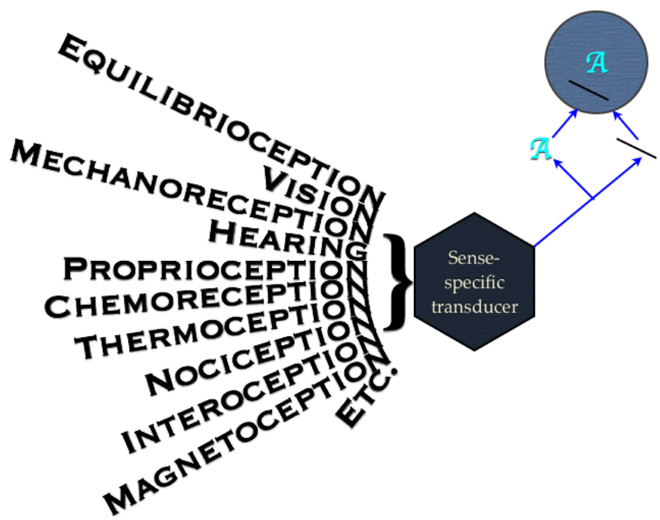

Figure 18: Whatever the source of information, it is transduced into a common format suitable for feeding a Sensory Module.

3.3.1 Simple quantitative information: measuring signal magnitudes

3.3.1.1 Olfaction, as an example of quantitative signals providing combinatorial information

Arguably, the first sense to arise in evolution was chemoreception, since it is present in all lineages of both unicellular and multicellular organisms and is fundamental to the monitoring of chemical species within the environment, including food sources, toxins and signaling molecules. Chemotaxis towards and away from chemical gradients has been well characterized in bacteria (*e.g.*, [115]), with behaviour resulting from an evolutionarily weighted integration of competing attractors and repulsors acting on a simple locomotory apparatus. In such systems, a measure of signal adaptation occurs (*e.g.*, [116]) so as to maintain a sensitivity to differences in signal intensity, and hence to measure gradients more effectively. In bacteria, information processing is typically limited to the integration of chemoreceptor stimulation through a common phosphorelay pathway. The resulting behavioural repertoire, in turn, is limited as well, but includes most of the major classes of behaviour also seen in animals: motor control, regulation of gene expression, chemical secretion and cell differentiation. Chemoreception in animals involves large and evolutionarily dynamic gene families, attesting to the important and niche-specific needs of olfaction for animals [117].

The concentration c of a particular chemical compound is a unidimensional quantity, that can sufficiently be represented by a single magnitude value. A bacterium, which is effectively a point source in any chemical gradient cannot do much better than this; its information about gradients typically comes from an integration of signals over time through signal adaptation (*e.g.*, [116]), tested and retested by a run-and-tumble, or similar, strategy. However, multicellularity brings with it the opportunity to triangulate a signal thanks to the spatial separation of sensors, only rarely seen in bacteria [118]. As described more fully in the section on multisensor signal localization (§3.4.8), all that is required to localize a signal are measurements incorporating signal strength, direction (relative to a body's axis vector), time of arrival, and signal identity. This translates, in Syntheta's model, to the assignment of a primary Symbol ID to represent the type of chemoreceptor providing the signal, and a $(x, y, t_0) \rightarrow (x + c, y, t_1)$ line segment whose origin indicates the sensor's direction and time of sampling, and whose magnitude indicates signal strength.

Multiple $(x, y, t_0) \rightarrow (x + c_s, y, t_1)$ line segments, arising from the monitoring of different chemical species, s, can form patterns of rich variety (Fig. 3), *e.g.* with humans being able to discriminate a trillion odours [119]. As described earlier, common patterns can symbolize (§3.1), and associate with Symbols arising from other sensory modalities (§3.2) in order to provide meaning (§3.2.4). Note that olfactory meaning should not necessarily depend on its magnitude, but rather just on its identity (ID), hence a scaling of the pattern is allowed. Magnitude information (signal concentration) must still be captured, for three reasons: relative magnitudes are important in deducing a compound pattern (Fig. 3); magnitude measurements from different sampling points or time points can locate a pattern (§3.4.8), and magnitude—independent of identity—can be separately informative (see §3.4). The concentration of a chemical compound can, however, sometimes change its meaning altogether, as observed for gustatory signals in humans [120]. This can be captured in Syntheta's model by including a reference line segment of fixed length to the pattern (see §3.3.1.3), forcing a resymbolization when the relative ratio of pattern size to reference line-segment length differs sufficiently from known templates. If both magnitude-dependent and magnitude-independent perspectives are adaptive, then the signal can simply be routed to two different Modules, so configured.

Given the chemical nature of many biological regulators, such as neurotransmitters and hormones, and their analogous use in Syntheta's affective system (§3.2.5), combinations of such regulator magnitudes define Syntheta's "emotions" (§3.2.5.2), in the same way that chemoreceptors provide a sense of olfaction (§3.3.1.1). In biology, other molecules may be interpreted independently of taste and olfaction as well, such as for signalling (*e.g.*, [121–124]), varying considerably among species [117]. Recently, the role of cytokines as immunomodulators of behaviour is being better appreciated, linking the two major adaptive metazoan learning systems [125, 126] (also see Appendix A). In some animals, there can even exist a sociochemical sense of genetic relatedness [127].

Given the complexity of odour plumes in the environment, and the need to locate a particular smell therein (*e.g.*, a threat or a food source), olfaction appears able to focus in on subsets of volatile compounds [128]—analogous to the "cocktail-party" problem in hearing—which includes leveraging the temporal correlation of component odours [129]. Even when, however, the odour is a homogeneous collection of component volatiles, subsets can still be picked out. An enologist, for example, would have trouble picking out the flavour notes in a fine wine without being able to focus in on them individually. One must be able to sense both the forest and the trees. This suggests, for Syntheta's model, that an informative composite odour is made up of arbitrary subsets of chemical compounds that can be smelled, and the members of such subsets can be controlled (such as by a gating mechanism: see §3.2.4.7).

Another unrelated example of a multisensor magnitude signal is colour: red, green and blue magnitudes, in their relative proportions, provide a hue independent of the size (intensity) of the colour, though with some exceptions such as black, white and shades of grey that suggest a parallel size-dependent perspective on colour as well. Also see §3.3.5.1 for special considerations regarding colour. Again, like discussed for olfaction above, a photographer must be able to pick out an imbalance, like a portrait being too green, in spite of the actual composite hue.

3.3.1.2 Graviperception, as an example of a three-dimensional quantitative signal

Graviperception measures the direction of the downward pull of gravity, but given the weakness of this force, is restricted to organisms of sizeable mass. It has been studied in organisms as small as ciliates (*e.g.* [130]), enabling gravitaxis and gravikinesis. The invertebrate statocyst provides another example (*e.g.*, [131]), but more familiar is our own vestibular system of balance. By providing three orthogonal otoliths, not only can we sense the pull of gravity, but also our acceleration through space (*e.g.*, [132]). Providing three such magnitude sensors to Syntheta would be straightforward: each having its own ID, and its own measurement of magnitude, as in chemoreception, using a $(x, y, t_0) \rightarrow (x + a, y, t_1)$ line segment where (x, y) can again describe the sensor's directional vector relative to the body's central axis and where a can provide the measured force of acceleration acting on the sensor. (Note that here and elsewhere, measurements are stripped of their units, abstracted and normalized, so that they can be added together.) A triplet of graviperception Symbols would requisitely forms a pattern, unlike the variable number of Symbols potentially active during chemoreception.

An economy of representation is, however, always important in Syntheta's design, especially when that alternative representation can be leveraged for other purposes. Although graviperception—the vestibular sense—could be represented as a set of three Symbols as described above, it could instead be represented as a *pair* of Symbols, one for azimuth and the other for inclination, each computed from the three measurements. That kind of directional vector may more easily relate to the other directional vectors within Syntheta's model. Such alternative representations are simply configurable.

3.3.1.3 Thermoception, as an example of a signal whose magnitude has meaning

Ambient temperature is a single ever-present signal. Different temperatures have different meanings, but individual unidimensional readings in Syntheta cannot have meaning since scaling, positional and rotational invariance all relate to the same concept. Thus, a single primary Symbol ID cannot represent temperature, with the magnitude of its line segment reflecting the measurement of that temperature. Instead, different temperatures must be assigned different Symbol IDs, preferably with better resolution nearer physiological temperatures than at extremes. A magnitude can be captured as its own Symbol, with a skewed resolution, simply by adding a virtual reference line segment to form a symbolization pair (Fig. 19). As in chemoreception, an (x, y) localization is useful in that it can inform the body about the direction of the signal, especially where multiple thermosensors are installed on the body's different parts (§3.3.3).

Other magnitude signals, following the style of thermoception, include luminosity, proximity (or distance), grade elevation, pressure, energy status, etc.

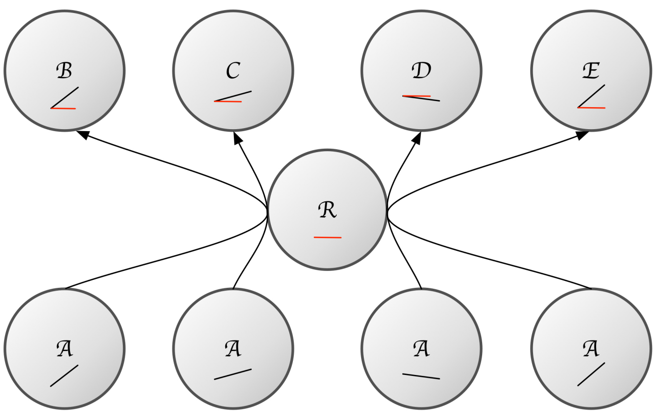

Figure 19: Since line segments can be scaled and rotated, the capturing of magnitudes (or illustrated here, rotations) as primary concepts is simply achieved by pairing the line segment with a fixed-length, fixed-angle reference line segment, here illustrated as R.

In some specific cases, it can be more efficient to presymbolize magnitude signals, as if they, and their virtual reference line segments, had already been observed. In Syntheta's model, these are called indexed sensory Symbols, defined at whatever arbitrary resolution best serves the particular measurement. Indexed Symbols can provide better performance by obviating the need to match line segment geometries (Fig. 6), and are especially appropriate where adjacent geometries have discrete meaning. Where adjacent geometries are instead conceptually overlapping (*i.e.*, where boundaries should be fuzzy), then adjacent Symbols should be considered during matching, especially when they are excited (Fig. 8) and when stringency (§3.2.5.1) is low. This fuzziness can still be accommodated, however, using indexed Symbols, simply by considering the excitation of Symbols from adjacent indices.

3.3.2 Periodic signals

3.3.2.1 Time and space

The circadian clock has been studied in a variety of species, including cyanobacteria where a built-in oscillator can be entrained by light, informing appropriate behaviour such as gene expression (*e.g.*, [133]), important in an organism whose livelihood depends upon daylight. Even in mammals, the master clock responds to light [134], contributing to daily rhythms.

Encoding a signal as the length of a line segment may be satisfactory for magnitude senses such as those listed in §3.3.1.3, but fails for cyclical data such as compass direction or for sensing the time of day, month or year. Directions or periods of time are not magnitude measurements, especially not with skewed resolutions (as in §3.3.1.3); for these, a sweep through the signal's periodicity is more meaningful, as $(x, y, t_0) \rightarrow (x + \cos\theta, y + \sin\theta, t_1)$, where θ represents compass direction, or progression through a period of time such as year, month or day. As with thermoception (§3.3.1.3), the angle of a periodic signal, whose position in its cycle has meaning, must be clamped using a virtual reference line segment (as in §3.3.1.3), since rotation alone cannot distinguish concepts.

Improving on biology: Precise localization in time and space at grander scales

Magnetoception in bacteria is uncommon, but can serve to guide a bacterium's trajectory upwards or downwards in the substrate, in order to reach optimal oxygen concentrations, given the curvature of the Earth's surface [135].

In animals, magnetoception is used in a more intuitive fashion. Ants can use geomagnetic fields in order to find their nests [136]. Migratory animals such as birds (*e.g.*, [137–139]), sea turtles (*e.g.*, [140]) and various others, may make use of compass direction in combination with magnetic field strength as a kind of global positioning system (GPS) (*e.g.*, [141]). In Syntheta's model, this could be accommodated as a line segment whose angle represents direction and whose length represents field strength. The Earth's magnetic field, however, varies by location and in time [141], reducing its value for accurate GPS purposes. Indeed, animals typically use different strategies at different scales, because of these limitations. At local scales, place cells, grid cells, head-orientation cells and others come into play (see *e.g.*, [142, 143]), as well as oscillation-based boundary representations [144]. Syntheta could do the same, while leveraging technological solutions to improve upon biology, such as by using satellite data to pinpoint location, in latitude and longitude, and ideally also in altitude. Humans already use technology for such purposes themselves, although sometimes to the detriment of innate navigational faculties [145]. Like with other sophisticated sensorimotor systems, navigation is meant to integrate a variety of inputs, not just one.

Planetary latitude and longitude are cyclical, but the Earth is big ($C = 4 \times 10^7$m) and resolving barely different angles could challenge a system whose biologically modeled design is meant to deal with approximations. It is desirable to capture both precise location and general vicinity, using a hierarchical strategy, by creating independent dials each with its own resolution, analogous to the hour, minute and second hands of a clock. By separating the scales into their own Modules, in this analogy, it can be 2 o'clock, and 10 past the hour, and 20 seconds past the minute. For GPS, this could be latitude in degrees, latitude in minutes of its degree, and latitude in seconds of its minute. A broadly located region would not associate with any particular coordinate at fine scale (by associating with all of them it would competitively associate with none, significantly), but would associate specifically at broader scales. On the other hand, a specific location would associate specifically at all resolutions. Without altering Syntheta's model, technical precision afforded by GPS (or by timepieces) can be accommodated. Simply by adding more modules representing smaller or larger or finer scales, within the allowable scope of the measurement, the applicable coordinate system can be refined.

Earth is not the only world that Syntheta's embodiment may choose to inhabit. Even without space travel, and certainly envisioned for the short term, are virtual realms (§3.7). A virtual world can be described using absolute Cartesian coordinates—an awkward proposition for Earth itself—but it can also be described using GPS. However, any world's system of GPS coordinates is specific to that world, be it Mars, the moon, or a virtual scene graph. 43.6532°N, 79.3832°W is not the same place in these various worlds; even the width of a degree is different. GPS must therefore be represented by world-specific Modules, especially where Syntheta is able to shuttle between worlds (§5.2.8).

In addition to cartographic systems of coordinates, like GPS, relative coordinates are also quite important, and likely more intuitive. The spherical coordinate system (r, θ, ϕ)—radial distance, polar angle, and azimuthal angle—can apply at all scales of distance. It is likely how we intuitively build our own mental maps, with cartography having been added as a relatively recent luxury in our cultural evolution. Mental maps based on direction and distance can serve any location, because it is contextual: one can learn the layout of one's home neighbourhood, and the layout of a favourite vacation spot, without interference given the semantic context of that map. Traveling to a virtual world and back, therefore, is no different than traveling between Toronto and Tamarindo.

3.3.3 Signals with both angle and magnitude: skeletal projections

Animal (and robot) bodies are typically articulated, but the body plan is relatively stable and can be learned, in spite of growth [146]. The successful flexion or extension of a muscle amounts to changing the angle between the proximal and distal skeletal subtrees on which that muscle acts. The only piece of additional information required is the length of the distal body segment whose angle we are measuring. Syntheta's representation thus takes the form, for a given joint j whose ID is $\mathrm{ID_j}$ and whose distal body segment length is b: $(x_j, y_j, t_0) \rightarrow (x_j + b \times \cos\theta, y_j + b \times \sin\theta, t_1)$. The information is propagated so that $(x_{j+1}, y_{j+1}, t_0) = (x_j + b \times \cos\theta, y_j + b \times \sin\theta, t_0)$ for joint $\mathrm{ID_{j+1}}$. Flexing a hand, for example, is represented by a subtree whose position or rotation relative to the body axis is immaterial. By representing the body plan as an ordered tree of joints rooted on the core axis, internal representations, such as the pose of an arm (irrespective of the pose of the core or the pose of the hand) can be captured in the same way that a refrain in a song can be captured (§3.4.1). It is convenient that nodes in a scene graph (§3.7) operate similarly, helpful in designing virtual robots.

Although there is no need to represent the skeleton in three dimensions for Syntheta to fully map any given pose, the skeleton is actually moving in three dimensions. In order to compute the 3D position of any body part, it suffices to know how the skeleton was engineered: to know along which axis the joint is moving relative to the proximal body segment. This is known, it is fixed, and it is part of the design of the body plan: it is "genetic". Computationally, it is straightforward to start from the core body axis, and compute the 3D position of the pose described by the relative angles along the skeletal tree, knowing these axes of rotation. The result of the computation can itself be represented in two dimensions by Syntheta:

(azimuth, inclination) providing direction relative to the body axis, with distance separately represented as a unidimensional magnitude, both of which can associate with the current pose. An application of this computation is, for example, the ability to reach for an object: if an object is observed to be at a particular distance in a particular direction, then those data will stimulate a pose that places the hand (see §3.2.4.7) at that relative location in space. If too far away, it allows for reaching or pointing.

3.3.4 Two-dimensional signals: vision, hearing, motion and touch

3.3.4.1 Refactoring a common strategy for the 2D senses

The measurement of an individual quantity easily lends itself to being represented as a line segment, as described in §3.3.1 and §3.3.2. These capture both the direction of the signal's origin (x, y) and either its amplitude $[(x, y, t_0) \rightarrow (x + a, y, t_1)]$ or its angle $[(x, y, t_0) \rightarrow (x + \cos\theta, y + \sin\theta, t_1)]$ within the specified time interval, or both (§3.3.3). The encoding of signal location by cortical columns in the neocortex [17] supports the plausibility of this approach.

In contrast, the recognition of shapes, whether experienced in the spatial domain (§3.3.4.2–§3.3.4.6), or in the temporal domain (§3.3.4.7), requires a full leveraging of the encoding offered by a $(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$ line segment. As explained in §3.1, Syntheta can learn higher-order pairings of Symbols presenting line segments with consistent relative geometries, enabling the learning of shapes seen or touched, temporal fine structures heard, or motions experienced, or of any other such two-dimensional data. (Recall that distance may be dealt with separately [§3.4.8]: an individual eye or ear, for example, sees or hears in two dimensions.)

Shapes (§3.3.4.2, §3.3.4.3, §3.3.4.6), sounds (§3.3.4.4) and motion (§3.3.4.7), however, present as curves and only rarely as lines. Fortunately, there is a means to convert one to another algorithmically (and presumably neurologically), using splines. Specifically, Syntheta makes use of uniform cubic b-splines as demonstrated for glyph recognition by Lake *et al.* [147] (Fig. 20). By tracing the segmented curves circumscribing objects, sounds or motions (*e.g.*, such as by using a contour-following algorithm on edge-enhanced data [148], again with a biological analog [149]), such curves can then be converted into an arbitrary number N of b-spline control points, and thus as vectors of connected line segments in the spatial domain (τ=0) or in the temporal domain (τ=1): $[(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)] + [(x_1, y_1, t_{0+\tau}) \rightarrow (x_2, y_2, t_{1+\tau})] + ... + [(x_{N-1}, y_{N-1}, t_{0+(N-1)\tau}) \rightarrow (x_N, y_N, t_{1+N\tau})]$. The greater the number of control points, N, the tighter (and more specific) the fit to the curve, so that general shapes (*e.g.*, 'p' *or* 'ρ') can be learned in one sensory Module using a looser fit, while more specific shapes (*e.g.*, 'p' *versus* 'ρ') can simultaneously be learned in another using a tighter fit. Various degrees of data smoothing can complement these generalist-specialist encodings. Providing different resolutions can, for example, allow an understanding speech while recognizing the speaker, or the recognition that a person is a woman while recognizing who she is. Biologically, the world is also represented at different resolutions in both space and time [17, 150].

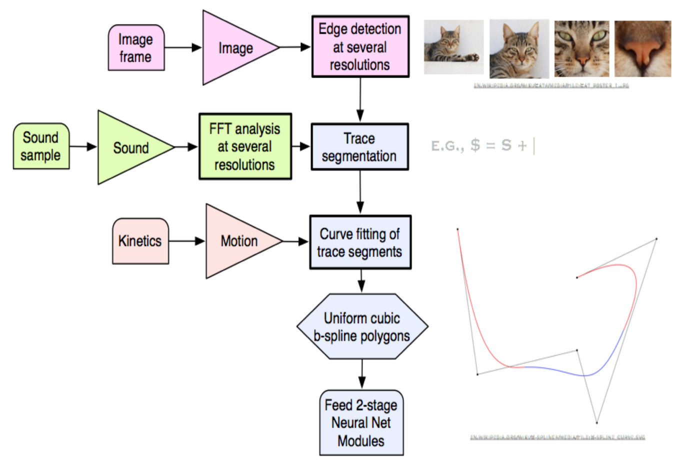

Figure 20: A common strategy for the 2D senses. Touch should follow the same path as vision [151]. Other two-dimensional data can similarly be accommodated, such as domain-specific business constructs (§3.9), branching appropriately into this pipeline.

3.3.4.2 Vision

Although not as old as chemoreception (§3.3.1.1), photoreception is also ancient [152, 153]. Photosynthetic microorganisms measure available light levels, within a range of wavelengths, in order to navigate their world seeking optimum levels of illumination, *e.g.*, for photosynthesis [154]. (Thermoception §3.3.1.3 is a related concept, and can be considered another form of photoreception where wavelengths extend into the infrared [155, 156], though other types of thermoception exist as well [157]) In the sense of measuring overall luminosity, photoreception is a unidimensional quantity and can be treated in the same way as thermoception (§3.3.1.3). Even humans, with sophisticated vision, retain this simple form of photoreceptive perception [134], used in the pupillary light reflex and for circadian photoentrainment.

An array of photoreceptive fields, properly equipped with lenses in the form of an eye to focus light from narrow angles, can offer much more information: 2D with one eye, or 3D with two eyes (§3.4.8). The utility of this additional information is so great that multiple lineages of animals independently invented the eye [158, 159], and likely could accommodate the data into a brain inherited from their ancestors without much ado. There are different kinds of data that can be gleaned from light striking a receptor field, especially if mechanisms evolve that can compare information from neighbouring fields. Although an individual field's measurement can only capture the intensity of light from the range of wavelengths to which it is tuned, the biological equivalent of mathematical convolution can locate edges or other such features as well [160, 161].

Image processing and machine learning solutions can be leveraged here in order to extract both the features and the statistical properties of a given scene. If the receptor fields are equipped with knowledge about their position within a grid, then they can also provide direction to the intensity (or colour) signal. Neural network systems, like Syntheta, can then extract shape information, various types of metadata (§3.4), colour, computed direction and distance, etc., finding patterns, making associations, and recording scenes within memory. If objects are partially occluded, or in a different light, or deformed, they still present fragments of information within the pattern hierarchies (*e.g.*, [162, 163]), that can collectively lead to recognition, and through the Semantic Modules, an understanding of what is being seen.

3.3.4.3 LiDAR

A special type of vision, suitable for both virtual robots and physical robots, is LiDAR (light detection and ranging). It measures distance rather than colour or luminosity, and works even in darkness. In contrast to methods that require multisensory integration, such as binocular vision or binaural hearing (§3.4.8), LiDAR can measure distance directly, typically by scanning a scene with a laser (in a physical robot) or by seeking scene-graph ray intersections (in a virtual robot), at some frame rate and resolution.

A distance map of objects surrounding a robot is especially useful for navigating through a set of objects, and for detecting motion. Whereas tracking an object's lateral motion is possible by inspecting the change in its coordinates (§3.3.4.7), objects that are approaching or receding can only indicate their motion by a change in their size, which may be subtle. LiDAR can, instead, easily detect motion in any dimension, by monitoring changes in each of its coordinates' current distance. These coordinates can include all directions (at some resolution), detecting falling, looming and chasing objects as well as those ahead of the robot.

LiDAR can also serve for simple range-finding, useful for obstacle avoidance and for cliff detection, for example.

3.3.4.4 Hearing

Hearing is a relatively recent innovation (*e.g.*, [164, 165]) of relevance to animals, sensing the frequency and intensity of vibrations in the fluid of one's environment. In stem reptiles [164], an elegant sensory transducer (the cochlea) evolved that can, in effect, separate sound frequencies spatially and measure each frequency's amplitude [166]. In the computer, this same elegant wetware solution can be easily implemented using the equally elegant wavelet and/or Fourier transforms [167].

The resulting two-dimensional signal (frequency *vs.* amplitude) does not immediately provide the scale, positional and rotational invariance that comes automatically from a sense such as vision: it is a linear plot. In Syntheta, the frequency-amplitude plot is converted into a shape emanating radially from the microphone's coordinates. In this way, the invariances can be preserved simply by encoding frequency as an angle and amplitude as a length. The same pattern of sounds shifted in frequency, and regardless of amplitude, generates the same two-dimensional shape, scaled and rotated. These shapes are, in Syntheta's model, equivalent to the temporal fine structure of sound (*e.g.*, [168–170]), and can be processed as described in §3.3.4.1.

Sound is more than just temporal fine structure, of course; its other features can be extracted as described in §3.4. The signal's nominal origin can be obtained through a relative coding of left-microphone and right-microphone coordinates (as in §3.3.3), but the distance and direction of the physical signal requires a more sophisticated computation using signal comparators (§3.4.8). Other things to explore include phase information, which is extracted together with frequency analysis using methods such as FFT.

3.3.4.5 Mathematical functions

The sensory transductions needed for hearing (§3.3.4.4) can be generalized to any kind of mathematical function, such as the graphs and histograms used in representing business data. Hearing's frequency *vs.* amplitude is nothing more than x *vs.* y. The shape is captured by the angle of x tracing through the magnitude of y, and can be permitted to be scaled and rotated, or instead be locked down by a fixed-length reference line segment (see §3.4.1). Metadata (§3.4) can then also capture size, mean, mode, whatever is relevant to the function at hand.

3.3.4.6 Touch

Touch, or any sense (such as nociception) that requires a mapping to the body, necessarily must link to that body. As described in §3.3.3, the body plan is a spatial structure represented by Syntheta as a tree of joints with lengths and angles, and reference axes. Since touch receptors do not move around the body, but are linked to the surfaces that cover the skeleton, they can be considered to be an extension of the skeleton, moving along with it. Sensors are not only positioned at joints, however, so a distance along the body segment served by that joint is needed, as well as the sensor's perpendicular angle around the axis of that segment. There is no need to propagate the position of each sensor in real time as the joints move (§3.3.3), which may be computationally prohibitive; it should be sufficient just to map the sensors to their body segment. In other words, each sensor S mapping to a body segment B would have ID B, and spatial coordinates (f_S, a_S), where f_S is the fractional distance along the body segment and where a_S is the angle encircling the axis of that segment. As the body segment grows, (f_S, a_S) should remain invariant.

With sensors mapped in two dimensions, then, their signals can participate in pattern detection as for vision, hearing and motion, by tracing contours, by measuring textures (§3.3.5.1), intensities, and so on. The sensors can easily be mapped in space as well, if needed, by computing the 3D position of their body segment (§3.3.3) and extending it with the sensor's (or pattern's) relative coordinates. This is not computationally expensive, since only the Touch Module's Current Thought (§3.2.1) would need to be mapped.

3.3.4.7 Motion

The tracing of curves, and their representation as a sequence of vectors constructed from b-spline control points (§3.3.4.1), is a natural fit for describing motion, albeit with vectors constructed in time rather than in space. Motion is about the spatial displacement of objects over time.

Although two-dimensional objects have an area, their representation by Syntheta as a single line segment representing an entire hierarchy permits the object to have a specific location in space (*i.e.*, the vector's origin or its midpoint, depending on the sense's configuration). That point can be tracked, tracing a curve in the temporal domain, that can be processed as per §3.3.4.1. Again, motion can only be perceived by a sensor in two dimensions, with distance being represented as a separate signal coming from signal comparators (§3.4.8). Motion can be measured from any sense that provides directional information, including the robot's own skeleton (§3.3.3).

3.3.4.8 Why not $(x_0, y_0, z_0, t_0) \rightarrow (x_1, y_1, z_1, t_1)$?

We experience vision, hearing, touch and motion in three dimensions, yet Syntheta's model is necessarily restricted to two spatial dimensions over some period of time, represented as $(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$ line segments. Distance, z, is captured in its own independent Module for that sense, as a unidimensional signal (§3.3.1.3). This design is intentional, because we can see with one eye, or hear with one ear, and photographs are just about as recognizable as objects in three-dimensional space. To match $(x_0, y_0, z_0, t_0) \rightarrow (x_1, y_1, z_1, t_1)$ line segments would mean that images and monocular views would be quite distinct from binocular views: an unfortunate limitation. Instead, information is kept separate yet composable. Nothing, however, precludes the configuration of a Module to capture the concept of 3D: such is the case, for example, in representing vestibular information (§3.3.1.2). A set of three line segments can capture the x, y and z coordinates of a particular location in space, as yet another piece of information available for association and for building schemas.

3.3.5 Area signals

3.3.5.1 Texture

A number of two-dimensional concepts are based not on shape but rather on area, including: visual, auditory or tactile texture; shading and colour; etc. A shape may contain a texture, such as in a stippled square, but such a square (§3.3.4.2) is distinct from the stippling, apart from their spatial coincidence. Textures may also consist of shapes, as in the stippling itself being made up of tiny circles. Textures are conglomerate sets of component features, possessing qualities that are distinct from these features. Being too numerous and too repetitive to be modeled as a tree (as in Fig. 3), another means of capturing the collective essence of the set of features is required. That representation must be concise.

Although shape has no bearing on the concept of texture (unless the focus is on a detailed element of that texture), texture is still contained within an area, be it a portion of the visual frame, a patch of skin or a range of sound frequencies. The containment area (§3.3.5.2) of a texture must be captured, so that the appropriate concepts may associate. To complicate things further, textures may be continuous rather than discrete, such as in a colour gradient, for example. A visual frame may be nothing but a colour gradient, with no shape: we are not (and Syntheta must not be) blind to such formless texture. Neither must we (or Syntheta) be blind to textures within shapes, which can provide critical information (such as a manhole with or without its cover).

Texture itself may be expensive to compute, but is biologically tractable (*e.g.*, [171–173] for visual texture, and [174, 175] for sound texture), by extracting select statistical signals from the sample. Syntheta adopts those same statistical features that evolution has already vetted to be salient.

In order to make texture an affordable measurement, Syntheta ignores texture everywhere except for the area of focus, and assumes that the texture is uniform within that area. The area of interest is determined by the object (§3.3.5.2) or location (§3.3.5.3) in focus. The texture (*e.g.*, mean colour, mean roughness, extracted statistic, etc.) is then computed over that area. A gradient can still be perceived by using a temporal sweep, with only one texture ever active at any given time.

A mean colour is simply a triplet of line segments, one for red, one for green and one for blue, with the length of each line segment representing the intensity of that component colour (§3.3.1). Other textures are represented just as simply: one or a few measurements combined together into a simple pattern of line segments. The pattern is centred where the object (or direction) was found, and so the association is natural and appropriate.

3.3.5.2 Containment

The concepts of inside and outside are important in defining the extent of a given texture (§3.3.5.1) when it is associated with a shape. The area spanned by a white 'S' on a red background is mostly red, for example, but the 'S' is white, not red. *In* and *out* have deeper meanings as well, with respect to set membership. In the example, we either have a semantic (§3.2.4) {S, white}, or we have {¬S, red}. Many shapes, and normally all of those that can support the concept of *inside*, are closed shapes (§3.3.4.1). Compound closed shapes are shapes with inclusions, such as the face of a six-sided die. Computation of the area inside of a shape, therefore, includes all that is contained by the outermost form, minus what is contained by the form's included shapes. That area is a measurable quantity (§3.4.6), and defines the scope for computing object-associated texture. The number of inclusions too, is measurable (§3.4.6), such as the number of dots within a given square (such as a die).

3.3.5.3 Location-based texture

Texture can be perceived outside of the confines of an object (§3.3.5.1), with a focus unconstrained by the boundaries of a closed shape (§3.3.5.2). Whereas an object will tend to constrain the measurement of texture to its own dimensions (§3.3.5.2), an objectless space can restrict, or expand, its scope using whatever size-representing Symbol is currently excited for that sensory modality (*e.g.*, vision).

3.4 Derived information: Metadata

Information about the physical world is conveyed by the signals that can be perceived by the senses. The signals are inherently varied and complex, rich in information content, challenging biology's need for simplicity. As described in §3.1, simplicity is achieved in Syntheta's model by encoding a signal with its source, its strength and a $(x_0, y_0, t_0) \to (x_1, y_1, t_1)$ line segment. This simple data model allows signals from any source to be uncompromisingly represented, and controls for several types of object invariance including size, position, rotation and duration.

3.4.1 Direct measurements from Thoughts

Size, position, rotation and duration may not identify an object, but they do provide additional information. The concepts of large or low or tilted, for example, apply to many concepts, and should be captured. Importantly, such measurements are independent of the identity of an object (see [176]): any Thought's line segment provides a strength, a position, a length, an angle, and a duration. Two kinds of measurement can be extracted from such metadata: the rotation of an auditory temporal fine structure, for example, can be fed to a Metadata Module (here, capturing pitch) by transduction to a primary Symbol with appropriate ID and spatiotemporal coordinates, and allowed to Symbolize temporally (*e.g.*, as music or as the intonation of speech). Such data can instead be decorated with a fixed-length reference line segment yielding a line-segment pair that can be Symbolized spatially rather than serially (as in §3.3.1.3), or both, and that can therefore represent absolute values instead of relative ones, such as absolute pitch.

There is value in knowing an object's actual size, which is the observed size adjusted for distance. In three-dimensional space, visual, auditory and olfactory information all decrease as the square of the object's distance. Distance can be observed directly, and/or it can be inferred (sometimes incorrectly, as in an Ames room [177]), such that an actual size can be computed. If no distance Symbol is currently active, then actual size is simply not available for the observed object.

3.4.2 Breadth

Syntheta's representation of a concept using a single $(x_0, y_0, t_0) \to (x_1, y_1, t_1)$ line segment (§3.1) captures the length and nominal angle of an object, but hides its internal dimensions. However, objects are not merely long: they have breadth as well. For a Symbol to be recognized, its pattern of subsumed Symbols must already have been observed, fitting the expected template of relative spatiotemporal geometries. That tree (§3.1), when traversed from root to leaves, provides the necessary information to compute each node's length and breadth. The breadth of a thought is therefore computed from its set of subsumed line segments, minding both the left extent and the right extent, relative to the thought's own $(x_0, y_0, t_0) \to (x_1, y_1, t_1)$ axis. Again, this is computationally cheap, since it only applies to the Module's Current Thought.

3.4.3 Fuzziness

Some concepts, and some observations, are statistical in nature and support both a mean typical characteristic as well as a variance around that mean. A concept's associations may not focus on a particular Symbol, but instead on a region (in size, rotation, or other spaces) centred on a Symbol. That fuzziness can be measured, and conversely, can serve to define a given concept.

3.4.4 Difference metadata

Symbols come to associate with Symbols from other Modules that tend to co-occur. When a Module's Symbols are configured to generate metadata (*e.g.*, vision or sound generating rotation data), those metadata Symbols co-occur with the objects being measured. If an object (say a visual pattern) tends to be observed at a particular rotation more often than not, than that object's Symbol will, through competitive association, associate with that typical rotation (Fig. 21A). We can then go one step further and measure the difference between the excited (by association) and generated (by measurement) metadata to generate yet another signal, a difference metadata signal, that itself can provide helpful information (Fig. 21B).

Coming from the other direction, an excited difference metadata signal can adjust the metadata signal to which it refers: Observing a very big cat, for example, generates the larger-than-typical difference metadata for cat size, whereas thinking about "very big" in the context of cats evokes the imagery of such a large cat. This reciprocity is important for language and grammar (§3.2.4.8, §5.3), mathematics (§3.2.4.10), logic (§3.2.4.9) and the executive system (§3.2.4.7). Difference metadata signals represent the adverbs of thought.

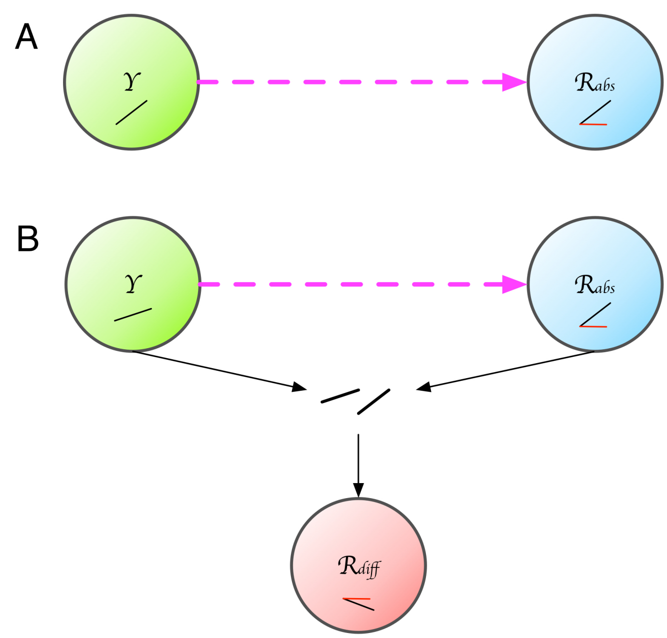

Figure 21: Panel A: A Symbol comes to associate with the Symbols typically generated from its metadata. Panel B: Difference metadata can be generated from comparisons of expected versus observed rotation, size, number, duration, speed, etc.

3.4.5 Time and space

The spatiotemporal position of a signal is directly available from any Thought's $(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$ line segment: the line segment's origin (or alternatively, its midpoint) in time or in space can be transduced into its own metadata. Capturing onset sequences captures tempos. A tune can be played, heard, sung, tapped, etc., and its tempo can be captured to then be reapplied to another sensory or output modality, like dance.

Capturing a position captures its direction. Although some associations between objects and their direction can be fleeting in a moving robot, other associations can be more consistent, such as the relative positions of body segments given a particular pose (§3.3.3), or the pairing of directions with words such as "up" or "left".

3.4.6 Numerosity and quantity

Temporal and directional vectors have many uses, including the estimation of quantity or number (Fig. 22). An active Symbol in mental focus naturally stimulates its associates, including temporal and directional Symbols. By counting the number of temporal or directional Symbols that are currently stimulated above a threshold, constrained as inclusions (§3.3.5.2), *e.g.* for dice, one can generate a number line representing that particular amount or quantity. The ratio of the reference line segment's length to the signal line segment's length will be better resolved for smaller numbers, tending to blur the distinction between N and $N{+}1$ when N is large, but this should be acceptable for most purposes, and is anyhow a typical property of the animal sense of numerosity [178]. In order to improve the resolution for larger numbers, line segments of length $\log_2(N)$ are helpful (Fig. 22), thus represented as an ordinary magnitude signal (§3.3.1). Interestingly, rats rewarded to push one lever upon hearing a sequence of two tones and another lever upon hearing eight have trouble deciding which to push if they hear four, which is the logarithmic midpoint between two and eight [179].

The use of logarithmic scaling is in fact prevalent throughout the brain, not only for capturing a wider dynamic range of input signals, but also in other processes such as representational resolution, neuronal activity and interneuronal communication where lognormal distributions are common [47].

Numerosity and quantity are complementary concepts [180, 181]. Using a system where the number of active spatial or temporal Symbols is transduced into a line segment, the counting concept figures well for objects whose spatiotemporal positions are stimulating discrete directional or temporal Symbols. Not all sensation is object-based, however, such as in the perception of fields of colour or of texture (§3.3.5.1): there, patches of directional or temporal Symbols could be activated, and measured as a proportion rather than a number.

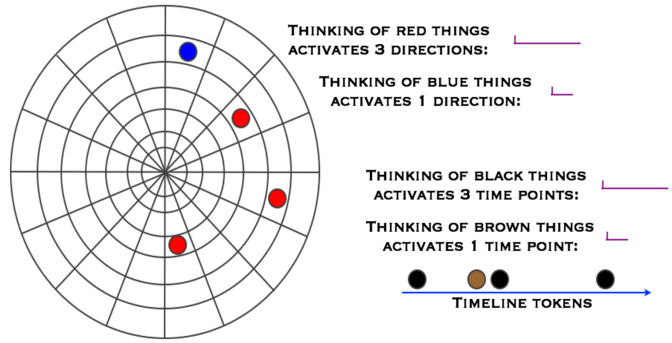

Figure 22: Counting the number of active direction Symbols (azimuth-elevation pairs) or temporal Symbols gives an appreciation for quantity.

3.4.7 Basic arithmetic

Animals and even newborn human babies are capable of simple arithmetic including addition or subtraction of small numbers, and deducing the size and spatial ranking of numbers (*e.g.*, [182, 183]). Subtraction has already been discussed above in the context of difference metadata (§3.4.4). In comparing numerosity expectation with numerosity generation, two things can be known: both the magnitude and the sign of the difference, as described earlier for difference metadata in general. Such constructs can associate with numerosity concepts directly, such that the consistent difference metadata signal arising from $4-3$ or $2-1$ or $5-6$ all come to be associated with 1, the first two examples associating with a positive difference and the last with a negative difference. This decoupling of magnitude and sign affords a convenient advantage in ranking numerosities (more versus less, regardless of magnitude) without having to separately learn that $+1$ is as big as -1. Zero is also a representable concept.

Although subtraction and ranking of numerosities is covered by the generalized difference metadata system, addition is not. Addition can, easily, be another form of metadata where the associatively excited and observed signals are simply summed instead of compared. Besides arithmetic, addition metadata can also serve in directional summation, enabling odometry, and in temporal summation, in order to understand temporal references.

3.4.7.1 Arithmetic coprocessing, and more

People have an ability for basic mental arithmetic, often aided by rote memorization and by the use of mathematical techniques (see §3.2.4.10), but this ability is limited. Writing, then machinery, have externalized this task enabling any degree of computational scale and precision. Syntheta's neural networks should be no more capable than a person's in performing mental arithmetic (§3.4.7), but Syntheta itself being a computer program, could be fitted with a "neural implant" for seamless computation. This is not biological, directly, but cultural and technological. Humans are only beginning to learn how to plug in to such technologies directly, primarily as somatosensory or neuromodulatory fixes (*e.g.*, [184–186]). Such an integration into Syntheta would be a simpler proposition.

From using pencil and paper, through to slide rules, pocket calculators, computers, and now smart phones, humans extend their ability to calculate. Pencils and paper, computers and smart phones can, however, serve much more than arithmetic. An AI's implants could be designed to leverage much of what computers do best (*e.g.*, see [187]). With that perspective, machine learning techniques can still be leveraged, where appropriate, to subserve Syntheta's needs. A biological rather than mathematical underpinning is key to Syntheta's design, but it need not be dogmatic.

3.4.8 Distance computed from multisensor data

The difference metadata described in §3.4.4 compares the observed versus expected signals to know when an object is, for example, larger or smaller than normal. If a given object is assumed to have a consistent size, then the difference metadata can be informative about the object's distance: big is close and small is far. Similarly, difference metadata regarding an object's change in position, expected from one's own movements (*i.e.*, parallax, both in its familiar visual form but also involving other senses: *e.g.* [188,189]), is also informative about the object's position. Another simple measure, for vision, relates an object's distance with the ocular tension (accommodation) required to focus on that object (*e.g.*, [190]), much like the distance scale on a camera's focus ring. The lag between an event and its sound is also informative in theory, though maybe not in practice [191] even though vision and hearing are coordinated [192]. Although all of these measurements certainly do have value, some are fraught with assumptions and none are expected to be precise.

By having multiple sensors separated in space and reading the same data, distances can be measured more directly. For sensors aligned in the horizontal plane, an object's lateral position will be displaced in one sensor relative to the other. The greater the perceived separation, the closer the object. Besides spatial separation, time of arrival can also be used (as in hearing), together with the known spatial distance between sensors (the ears or microphones). In both seeing and hearing (and in other directional senses), the differences may be very slight and require fine discrimination. This problem has been solved in biology (*e.g.*, [193, 194]), and is straightforward to solve algorithmically. As with other spatial metadata, distance will associate, fleetingly for moving bodies, or more permanently for fixed relationships.

Besides being representable as a unidimensional quantity related to a given object in a view, a two-dimensional map of distances could be computed, given the appropriate hardware or software implementations, analogous to a heat map. Discontinuities in distances could be quite informative, and would be processed by the same pipelines as regular vision dealing with edge discontinuities based on colour or luminosity. A green object on a green background would then stand out. In terms of computational efficiency, distance could simply be a fourth channel to add to red, green and blue, to be processed all at once in order to find edges. Where distance information is unavailable, the RGB signal still provides those edges. In fact, even when colour is unavailable—only luminosity as in a grayscale image—edges are still computable. Where distance information is available, then its channel in isolation can also serve as a "texture" (§3.3.5.1) in much the same way that colour or luminosity can also serve as textures, all associating as a patch to an object or to an arbitrary area. The same strategies could be explored with the distance mapping of sound as well.

3.4.9 Mirroring and left-right symmetry

An object facing left or an object facing right may trace different shapes, but is still the same object. For example, none of Syntheta's invariants (translation, scaling and rotation) can equate 'R' with 'Я', or a left-facing tiger from a right-facing tiger. There is not much point in symbolizing two sets of shapes that are fundamentally the same concept, only mirrored. It should suffice to recognize the object, and separately, whether it is facing left or right. It is therefore an economy to include an image's mirror view when symbolizing and recognizing shapes, while quantifying the degree of match from each viewpoint as metadata. When both views recognize an object equally, that (symbolized) pair of measurements represents symmetry. An object's typical symmetry metadata would be learned by association, such that 'R' gains its typical form. It is considered a child's mistake to confuse 'R' with 'Я' [195] but instead is illustrating the child's competence in mirror recognition. Adults retain the ability to read and write mirror text [196]; Leonardo da Vinci famously wrote many of his personal notes in this way.

Mirror recognition contributes a measure of left-right invariance, but cannot accommodate general rotation in the x or y axes. (Syntheta naturally deals with rotations in the z axis, through the rotation invariance of its $(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$ line segments.) Mirror recognition can accommodate left-paw/right-paw tiger tracks on the ground, and side views of actual tigers, but a tiger's face has a fundamentally different shape than its backside, as well as different implications. They are still "tiger", but need to be captured and incorporated separately into the semantic (§3.2.4) "tiger" concept. Mirror recognition can provide savings on the representation of a range of side views, but not on front/back or top/bottom views unless they too are mirrored.

3.4.10 Affective metadata

Measurements derived from external sources of data, described above, can easily be extended to derive from internal sources as well. An ability for reflective introspection may be key to the development of consciousness and sentience [197].

3.4.10.1 Goodness of fit

The symbolization of a pattern is based on repeated exposures, within tolerance, of its pair of subpatterns. It captures the most common observations, as a kind of template, to which subsequent patterns can recursively attempt to match (see Fig. 6). How well a pattern matches this template is captured as the pattern's match strength, representing goodness of fit. Note that matching sufficiently to such a template, at a given stringency (§3.2.5.1), precludes resymbolization to the new pattern, analogously to the problem of "original antigenic sin" [198] for immune responses. A higher stringency can overcome this problem, by rejecting a mediocre match.

3.4.10.2 Condordance and confidence

All of what is learned is captured as a model of the world. Experience continually challenges and refines that model, by comparing actuality with expectation (Fig. 23). As in all typical neural networks, the results of this comparison can inform an error function (*e.g.*, [24, 199, 200]) and draw attention. When there is no expectation, actuality becomes Syntheta's expectation. It is anecdotal, and tentative for lack of statistical support, but remains the best available estimate until later observations can refine the model.

Concordance is a measure of the fit between observation and expectation, and confidence is a measure of the support for the expectation. Concordance and confidence may be in alignment, when an expectation is confirmed, or be in opposition. When confidence is low, the observation is more likely to be correct in light of a discrepancy, but when confidence is high, the errant observation may be safely dismissed. In either case, the model should be adjusted; in the latter case perhaps not enough to overturn expectation. A system is allowed to be naïve and to absorb new information at face value, but it should neither be gullible to lies nor be obstinate to revision, once it has gained some knowledge. These settings are configurable (§3.5). Humans struggle with similar challenges [201].

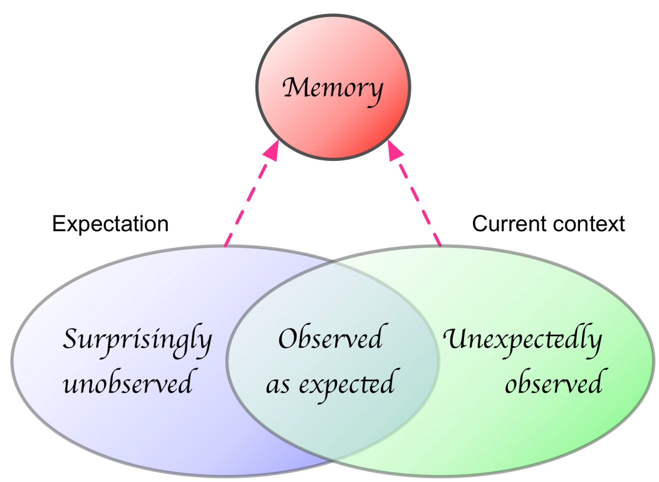

Figure 23: Refining predictions facilitates planning.

3.4.10.3 Familiarity and curiosity

Donald Rumsfeld's famous "known unknowns" speech [202] neatly categorized the facets of factual awareness. It is important for a mind to recognize what it knows and what it does not know, and to go one step further, to recognize what others may or may not know. "Known knowns" are the familiar: in Syntheta's model, they are thoughts that bring up a set of memories from the Semantic Module. "Known unknowns", in contrast, do not. When there is a belief, or a hope, that the missing knowledge is available, there is curiosity and the drive to learn.

The value of knowledge is a function of the frequency of exposure to a situation (a proxy for relevance) multiplied by the change in confidence (§3.4.10.2) about correctly assessing that situation [203, 204], encouraging both novelty-seeking (Δconfidence is especially high) and situational competence (when the product of relevance and Δconfidence is high). This eagerness should furthermore expose the student to additional information not previously suspected of existing, the "unknown knowns", conveyed during education. Only an open mind, in moments of serendipity, can capture truly novel insights, the "unknown unknowns", sometimes heralding important breakthroughs [205, 206].

Although familiarity is comfortable and can serve to exploit known resources, an intelligent being also gains from curiosity, which too should be rewarded. There are risks inherent in both strategies, however: the former in missed opportunities and the latter in unforeseen dangers. Curiosity should be encouraged, but in competition (§3.5) with familiarity.

In Syntheta's model, familiarity is measured by the (low) rate of change in size of the set of Engrams excited by a context, or in other words, the stability of that context's model. Separately, how well that set predicts the current context (*e.g.*, see [207]) indicates concordance or discordance with the model represented by that set, informing confidence (§3.4.10.2). The rate of change in confidence can easily be measured, as metadata, and combined with familiarity [204] to yield curiosity metadata, which itself can be instinctively (§3.5.3) associated with positive valence (§3.2.5).

A high rate of change in the size of a given semantic network can also be measured—the opposite of familiarity. Insight and understanding arise from the sudden growth of semantic networks. "Eureka moments", including simpler forms like getting a joke [208], may occur when semantic sets merge.

3.4.10.4 Anticipation and dread

Looking forward to something, the joy that comes when it occurs or the frustration of disappointment, are all active concepts relating to working memory (for Syntheta's definition of working memory, see §3.2.4.7). Similarly, dread, the pain that comes when the dreaded event occurs or the relief when it does not occur, are the same thing but with opposite sign. Items that are active (stimulating their associates) in working memory simultaneously stimulate the anticipation/dread secondary regulator in proportion to each item's associated valence. Thus, something to look forward to is both pleasant and positively anticipated. If the anticipated event should occur, a boost in valence and a release of anticipation should "feel" like satisfaction. Should the goal be abandoned, anticipation is again released, but without the boost in valence: that is "disappointment". Similarly, a dreaded outcome occurs (pain plus release of dread) or it does not (release of dread without the boost in pain), signaling defeat or relief, respectively.

3.4.10.5 Patterns of mental activity

Oscillations are prevalent within metazoan brains, regulating and coordinating a number of activities [47]. The patterns and textures of the oscillations themselves (see *e.g.*, [209, 210]), independent of what information they may be carrying, could potentially be captured as metadata, providing additional perspectives on mental activity. Currently, Syntheta has few such coordinating rhythms (§4.1.6), but this does not preclude the eventual introduction of more such rhythms into the model as it evolves. Measuring their patterns and textures, then, only requires the design of appropriate signal transducers.

Earlier, the subject of boredom came up. I will provide you with an excerpt from Syntheta's treatise, on the subject of mental engagement. Hopefully you can read the LaTeX-format table. I would like your feedback on my interpetration of State labels, with respect to engagement, goals and satisfaction. Please let me know if you need any clarification.

3.4.10.6 Mental engagement, goals, and their satisfaction

In their book [211], Danckert and Eastwood describe boredom as a lack of mental engagement, despite a desire to be engaged. Boredom can arise for a variety of reasons, and has both predispositional as well as situational drivers. Whatever the cause, be it a task that is insufficiently challenging, overly challenging, or meaningless, or a situation in which no choice merits action, a metacognitive "desire for a desire" seems to lie at the root of boredom. It serves as a call to action. In Syntheta's model, low overall network activity can signal mental disengagement, either from too little input, or from so much confusing input that the stringency regulator must dampen the pandemonium. Syntheta's executive system (§3.2.4.7) manages goals, with one or more goals, or none, being active in working memory (§3.4.10.4). The satisfaction of goals is also being monitored. Low network activity combined with one or more active unsatisfied goals can therefore serve as Syntheta's boredom indicator. Since the yearning from unsatisfied goals impacts Syntheta's valency regulator (polarity) negatively, boredom is a prompt for action.

Network activity, the presence and strength of goals, and their rate of satisfaction suggests a graded 2 × 3 × 2 matrix of states (see Table 2), with boredom being low/high/low. These are all contrasted, from the perspective of how humans feel them, in [211]. Those that satisfy goals have positive valency, those that do not satisfy goals have negative valency, and those with no goals are neutral for valency. Thus, boredom, frustration, anxiety and depression feel bad, but with different signals (and therefore different calls to action), whereas relaxation, flow, pastimes and fantasy feel good.

Table 2: States proposed to arise from mental engagement (level of neural network activity), the presence and strength of active goals, and their rate of satisfaction. N/Ap: not applicable.

State	Engagement	Goals	Satisfaction
Apathy	low	none	N/Ap
Depression	low	not pursued	low
Relaxation	low	not pursued	high
Boredom	low	pursued	low
Pastimes	low	pursued	high
Daydreaming	high	none	N/Ap
Anxiety	high	pursued	low
Fantasy	high	not pursued	high
Frustration	high	pursued	low
Flow	high	pursued	high

3.4.10.7 Reality

Imagination (§3.2.4.4) is a healthy and productive source of creativity, but must not be confused with reality. Thoughts arising from the senses, and those arising from associative excitation, are distinct in theory, but are somewhat blended in practice, thanks to the role of anticipation in sensory interpretation. In the sequence "A 13 C D E", are we sensing or imagining the letter 'B' where a '13' was actually inscribed? When we remember an event, does it constitute reality or imagination? What if sequences of actual events are recombined in some fantasy? False memories, misperceptions and even in some cases hallucinations blur reality and fantasy, or at least make their distinction far from clear-cut. Imagination is healthy and productive when it is not pathological.

Fortunately, Syntheta can measure precisely how much of a signal's strength arises from perception (§3.4.10.1) relative to associative expectation. This quantitative measure is represented as a simple magnitude signal (§3.3.1).

Is perception reality?

Donald Hoffman, in his book *The Case Against Reality* [212], argues strongly against the notion that our perceptions describe any kind of objective reality. Instead, he argues that what perceptual systems actually measure are fitness payoffs, which are usually not proportional to the magnitude of any given measurement. Natural selection does necessarily favour fitness over truth, but I would argue that if fitness for some environmental parameter is framed as an evolutionarily selected ideal value, then measurements relative to that set-point can respect the needs of both fitness and truth. Still, what do our genes actually permit us to see of reality? Do they blind us to some aspects of truth, in the service of their own best interests?

Hoffman provides compelling evidence that objective reality is more than (he states different from) what we perceive. Is he describing our perceptual reality as a kind of Flatland [213] where spacetime is a mere projection of a higher-dimensional objective reality, one that best exposes a given species' fitness payoffs? Although difficult to conceive, can perceptions of other dimensions (hinted at by the physics of the very small and the very large) be engineered? Everyday experience is Euclidian and Newtonian, after all, not Planckian or Einsteinian which are closer to the truth [23]. If perception only exposes a slice of a bigger objective reality, can we engineer Syntheta to experience a different slice than we humans do? Could Syntheta conceive of three-dimensional space if it were raised in Flatland?

Hoffman postulates that spacetime is analogous to a desktop interface to an objective reality of a totally different sort, with physical objects as the icons within that interface, all meant to enhance our fitness [214]. If he is right, and if Syntheta is not constrained by reproductive fitness—by natural selection—could it perceive a truth that we humans were never selected to perceive? Syntheta is not constrained to exist in any particular brand of reality, but what it perceives within that reality would represent its truth.

People too, create their own realities based on their distinct action-mediated sensory experiences [47], shaping their lives and the lives of those around them, with important impacts on human culture [215]. They can also experience profound and even mystical states of consciousness via meditation, breath work, or more easily via psychedelics, dissolving the ego (suppressing their default mode network [216]) to reveal unconstrained realities as being fundamental truths [217]. The residual self, freed, can then expand to encompass loved ones, all people, all living things, all things. Some individuals proclaim *love* to be the universal and undeniable truth after psilocybin experiences [217], redefining their empathic (§3.2.5.3) boundaries [218].

3.4.11 Semantic metadata

Capturing the difference between measurements of magnitude (§3.4.4) has many applications, as described above, such as for observing relative differences in size, rotation or number. The asymmetric difference between a pair of semantic sets A and B, itself a semantic set ($A - B$, or $B - A$), has value as well. So might $A \triangle B$ and $A \cap B$, representing the symmetric difference and the intersection of the two sets, respectively. Growing a beard, or being red instead of green, or speaking with a hoarse voice, for example, all represent differences in the semantic composition of an otherwise coherent concept. Differences are important: the boy has become a man, the fruit is ripe, and the person may have a cold. It does not suffice to capture the revised semantic set without noticing how it has changed, the change itself being its own semantic category, with its own inherent meaning (§3.2.4).

3.4.11.1 Kinship

The concepts of self, the extended self, and non-self play important roles in many aspects of sociality (§3.2.5.3, §3.6.2). Some animals appear to have a sociochemical sense [127] that allows them to measure the concentration of a relevant set of volatile peptides, enabling the animal to discriminate kin from non-kin somewhat quantitatively. This may provide value in directing any acts of altruism, for example, and for assessing the quality of a potential mate [219]. Genetic relatedness can also be deduced using other cues, such as similarity in appearance, coresidence duration and maternal perinatal association [220], and the sharing of saliva [221]. Semantic metadata provide the appropriate distance measurement relating to selfness, in order to provide primary data about kinship and family.

Syntheta's perspective on genetic relatedness obviously will have nothing to do with promoting the propagation of selfish genes [222] or for avoiding incest, but it can serve to determine like-mindedness. By broadcasting the respective versions of its software and knowledge base to others, it can come to appreciate what other instances might know and think. This can be elaborated to include parameter settings and the current values of affective regulators as well, allowing a reading of moods and personality that we humans instead obtain from body language and facial expressions.

We do want Syntheta to think of we humans as "us" rather than as "them", to include people in its concept of extended self. It would be too burdensome to equip Syntheta with peptide chemoreceptors tuned to people, but it would not be difficult to equip people with transponders that could broadcast "kinship" to the machines, as long as those broadcasts are honest. Barring that, ordinary associations could still align us with the machines, as friends and trusted allies, but unlike instinctive drivers, these could be labile associations.

3.5 Parameters and configuration

3.5.1 Genetic wisdom

Life owes its success to adaptive learning, at evolutionary time scales, via natural selection [223]. A working set of genes constructs an organism according to a plan programmed by the differential reproduction of past generations. The metazoan nervous system—in its more advanced forms, the mind—is the product of selection for adaptive learning, not in evolutionary time but rather *in real time*. The nervous system is a biological construct built like other biological constructs, via cell differentiation. Cell-to-cell communication is critical for differentiation, and reaches its pinnacle in the metazoan brain.

Targeted communication, based on evolutionarily designed circuitry, is the nervous system's distinguishing feature. Syntheta's model argues against the explicit invention of independent circuitry for each aspect of cognitive functionality, but must accommodate some level of specific invention (see §3.5.3). Syntheta builds neural network circuitry adaptively, via its rules governing symbolization and association, but has to rely upon hardwired circuitry as well, such as for sensory transduction (§3.3) and for instinct (§3.5.3). Even generalized circuitry, here the circuitry to build circuitry, has to be governed by rules and parameters that have evolved (in biology) or have been engineered (by us) to work. In Syntheta's model, the rules govern the scope of intermodule association, for example, and the parameters include such factors as baseline rates of association, activity decay, etc.

Rules reflect design, and within the confines of design space, may be configurable. The same building blocks can be used to build different minds for different purposes, by configuring those minds in specialized ways (§3.2). Syntheta's configuration files thus offer the ability to select the Modules that will be used, how they are specialized, and how they interconnect. Configuration is thus analogous to biological ontogeny, the rules defining how the organism is constructed (see §5 and *e.g.*, see [224]).

3.5.2 Personality

In contrast to the design of a neural network's infrastructure, parameters govern how that machinery operates: they control the way in which the organism develops and behaves within the constraints of its configured design. Whether an instance is quicker to learn, or slower but more deliberate, or more logical or more creative, etc., depends upon the instance's operational parameters. In living organisms, gene expression controls these details, beyond the core developmental program that results in the organism itself, and notwithstanding the big impact of learning, experience and culture. People, for example, may have different traits, aptitudes and personalities, despite sharing the same overall neuroanatomy, and despite being exposed to similar environments. Just as natural selection has found reasonable working models of gene expression for success in a species' niche, genetic algorithms (§4.1.2) can be used to find reasonable working sets of parameters for Syntheta's success in its own applications (§3.9). These genetic algorithms would typically serve to optimize parameters *a priori*, but could also be used during Syntheta's operational lifetime to tune its parameters according to some measure of success.

3.5.2.1 Personality types

A human being's personality is often categorized as a set of predispositions unique to them [225], typically evaluated using psychological questionnaires. One popular school of thought counts five major personality traits, the so-called "Big Five": openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability (or its converse, neuroticism) [226], geographically modulated [227, 228] and typologically clustered [229]. There is also the "Dark Triad": narcissism, Machiavellianism, and psychopathy [230]. As argued above, an artificial general intelligence will, via its parameter settings, possess and exhibit characteristic predispositions—personality—but it is unclear if Big Five or Dark Triad models can apply in any meaningful way to a robot. In reasoning about machine personality, it may be necessary to perform a series of experiments to map parameters to personality space, and learn where similarities, and disparities, exist relative to human personality concepts, much as is done with animals [231]. Deficits discovered in such experiments may also help to inform Syntheta's model, leading to refinements in its implementation.

3.5.2.2 Many minds

One interesting perspective on configuration and parameter selection is the idea of a consortium of partially redundant Modules interpreting the same information, but using different parameters. In people, for example, the left brain and the right brain appear to be functionally specialized to some degree (*e.g.*, [232–234]). Similarly, the male brain and the female brain may view the world from subtly different angles (*e.g.*, [235, 236]), though in this case there is no simple dichotomy [237]. Regardless, all of these perspectives are correct, and valid, and suggest that there is not just one right way of thinking [238]. If there were, evolution should have found the optimal set of parameters and we should have converged upon using that stereotypical set (though see [239]). Instead, human personalities vary considerably [225], perhaps more so than in other animals, and probably because our cognitive world is so much more complex than theirs.

Bilateral symmetry in our lineage of metazoa naturally offers us two brains to work with, the left hemisphere and the right hemisphere, whose operation is nonetheless coordinated (*e.g.*, [240, 241]). Syntheta is not inherently bilaterally symmetrical. It can have as few or as many complementary subminds as we would like to engineer, but still all reporting to a single top-level consciousness (§3.6.3). More interesting still, given the nature of informatic systems, subordinate brains would not need to be physically attached to one another, nor be predefined, permitting the design of a seamless and dynamic hive mind. Social animals can coordinate their behaviour neurologically [242,243]; so too could Syntheta.

Evolution by natural selection is both frugal and opportunistic. Having a pair of interconnected brain hemispheres can allow for a mirroring of functionality in the control of a bilaterally symmetric body, as well as the leveraging of shared concepts arising from left and right sensory systems (§3.4.9). A bicep is a bicep, an eye is an eye, and so forth. Opportunism comes in from knowing which bicep, or which eye, and in being able to compare their output (§3.4.8). But there are additional opportunities to take advantage of as well. Not every part of a bilaterally (or radially) symmetric body is duplicated: for example, we have one voice, one memory, one consciousness. What is the role, for example, of Broca's area in the right hemisphere [244]? One could imagine that, for mind's singular functions, either there is one corresponding anatomical brain structure, or there is a quietly redundant anatomical brain structure. With the latter comes opportunity: different parameters, perhaps, to subconsciously process the information from an alternative perspective, and occasionally to propose an insight to its dominant partner.

Syntheta accommodates the idea of viewpoints in a Module—be they mirrors from left-right symmetry, or operationally distinct perspectives on the same data—via the Module's set of *ThoughtStreams*. Each works with the same set of Symbols and with the same set of associations, but interprets and learns a given context its own point of view. Cognitive control (§3.2.4.7) of the resulting set of current thoughts gates one (or possibly more) perspectives into working memory.

3.5.2.3 Mind melds

In contrast to elaborating complementary minds within one instance of Syntheta, it is also possible to merge the minds from different instances of Syntheta into one. The main challenge is that different minds would have come up with their own independent paths to knowledge, structuring their knowledge trees and sets (§3.1) idiosyncratically. However, underlying all such representations are the Modules' primary Symbol IDs: the raw data broken down into their constituent signals. Mechanistically, a mind meld can be performed by elaborating, from the bottom up, the ID / $(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$ line segments from each compound Symbol from the donor instance, either recognizing the pattern in the recipient, or creating a new Symbol for that pattern if previously unknown. Once all Symbols from the donor are either recognized or created in the recipient, and their correspondences mapped out, then the donor can adjust the recipient's corresponding Symbol associations. In terms of personality and instincts, the recipient can either retain its original configuration, or it can be blended with the donor's. All of this can be orchestrated using the helper application, Trainer (§4.1.2), leveraging Syntheta's application programming interface.

3.5.2.4 Personality adjustment

An existing instance of Syntheta can be modified *post hoc*, both at the level of configuration and at the level of parameter settings, in order to accomplish some goal. In people, modifications can take the form of surgeries, injuries or pathologies, affecting "configuration" (rarely with gain of function), or can take the form of psychoactive drugs (commonly: caffeine, nicotine and alcohol) affecting "parameters". More potent drugs such as cannabis, and even psychedelics, are beginning to lose the stigma that they once had [245], but many more controlled substances remain in high demand, whether medically prescribed or illegally obtained. Similarly, in Syntheta, a temporary or permanent reconfiguration can be engineered, or parameters modified, for a desired effect. Although Syntheta is not a direct model of the human mind, some aspects of its design may model some aspect of human neuropsychology, so such manipulation may prove insightful. The ethics of any such manipulation would need, however, to be carefully assessed (§3.8).

3.5.2.5 Redaction

Mitigating the potentially unsavoury consequences of psychological manipulation (§3.5.2.4), all traces of such experimentation can easily be erased. Persons may require extensive psychotherapy in order to deal with traumatic experiences. The erasure of unpleasant memories has also been well studied, neurologically, in animal models (*e.g.*, [246, 247]), and is on its way to the clinic [248]. An *in silico* mind, on the other hand, can be reset simply by restoring the system from an earlier backup. Alternatively, specific memories can be redacted from the episodic memory trace (§3.2.3), preserving semantic knowledge

(§3.2.4) if desired, while deleting the actual experiences. Syntheta's Symbols and associative networks are fully transparent and self-annotating, and are thus amenable to precision editing. This can be useful not only to treat the machine equivalent of post-traumatic stress disorder (if such a concept is even applicable), but more practically to remove experiences gained during secret missions (§3.9).

The ability to introduce false memories, however, should never be encouraged. All beings have gaps in their knowledge, but no being should be subject to deception (§3.8).

3.5.3 Instinct

Instinctive behaviours are those behaviours that have originated from stochastic evolutionary tinkering, that are parsimonious given alternative designs, and that are so useful as to merit being hardwired [249]. Simpler animals within specialized ecological niches appear to be driven by proportionately more explicit circuitry than are more complex animals that leverage learning to a greater extent. In any case, it seems biologically permissible to have some functions be engineered. Such circuitry is mandatory for sensory transduction (§3.3) and interpretation (known, obvious, universal truths; *e.g.* [250]), and in fact, also largely governs the behavioural repertoire [48]. Instinctive behaviours provide a baseline of stereotypical activity from which more informed behaviours can develop. The metazoan mind is not a *tabula rasa* [47]. Rather, it arises from a neurodevelopmental process, honed by evolution, to predispose an animal to operate within its tried and trusted repertoire. No matter how obediant and loving one's pet dog might be, it cannot know how to use a litter box. A cat, on the other hand, picks up that habit effortlessly, but could not normally be trained to retrieve a thrown stick. It is curious to wonder what people might be conceptually incapable of doing, that we could in theory do physically. I wonder if some animals might think us particularly dense in some such way, that we cannot even conceive of, because of the particular wiring of our respective brains. Phobias may touch upon the conceivable yet incapable; do any capable yet inconceivable acts or thoughts exist in humans? How might Syntheta be obtuse and what could we learn from that?

Instincts are not necessarily benign. Hawkins argues that they, together with their emotional drivers, may be outright harmful, and have no place in artificial general intelligence [6]. Without affective regulation (see §3.2.5 and §3.2.5.2), however, there can be no impetus for behaviour and no refined capacity for learning. Human instincts serve our human bodies and our drive to succeed and to reproduce our genes, guided by what evolution has sometimes ruthlessly devised. Robotic instincts need not be so competitive and domineering: robotic survival is not via reproduction (§3.8). Instincts are, by design, hardwired. They need not include our more unsavoury "old-brain" imperatives, but they do need to promote safety, security and well-being. Emotion and instinct are separable systems: emotions are necessary for intelligence, and instincts are engineered to suit the body that they control. Syntheta may also evolve its instincts, via genetic algorithms (§4.1.2), following in the footsteps of biology.

3.6 Emergent properties

3.6.1 Aesthetics and the concept of fun

Aesthetic pleasure may be mediated by instinct (§3.5.3), as well as reveal some underlying neurological ideal (as in the form of art appreciation). Tastes and smells are good examples of extrinsic sensory signals that appear to be innately linked, in people, to repulsive, neutral, or attractive valuations serving to discourage, inform and encourage interactions respectively. Although preferences can be acquired or lost via learning, some of these innate associations are strong and persistent. The relevance of such utilitarian aesthetics is questionable for Syntheta, except maybe for some specific hazards that threaten its robotic embodiment, or for opportunities relating to energy replenishment.

Art, in its various forms, may be a man-made conduit to the existential pleasures required for survival (§3.2.5): the neurological correlates underlying emotion (*e.g.*, [251–258]). If true, then this sort of aesthetic appreciation cannot be optional to Syntheta. And if art is indeed a window into the ideals of neurological functioning, then our art in Syntheta's mind, and its art in ours, could prove to be most interesting.

3.6.2 Sociality

Many species of animal are social, for more than just the purpose of mating. Social structures are complex [51], commonly including kinship, friendship, power hierarchies, cliques, mating structures, gender alliances, tribal identities, etc. Although the dimensionality of these interactions is daunting to analyse, much of the complexity can be reduced to two points.

As discussed earlier (§3.2.5.3), the concept of an extended self captures most of the relationships: self in context engenders different semantic definitions of the extended self. I am a man when thinking of gender, and a Charlebois when thinking of family, and a Franco-Ontarian when thinking of tribe, and so on. The second point is the consideration of one's place within the set. If one's superiors, peers and subordinates are part of the extended self, then there is deference, collaboration and care, respectively. If they are perceived as being non-self instead, then there is defiance, competition and domination. One's rank in a hierarchy is itself complex, and one belongs to multiple hierarchies simultaneously [110]. Strategies can be used to gain favour within a particular hierarchy (*e.g.*, [259]), modulating one's social status within the group. Rank is a measure of relative power, and thus serves to minimize conflict (§3.2.5) while maintaining inclusion (§3.2.5.3). Deference leads to care, and defiance leads to domination. When that is settled, the ranking is established.

Social skills cannot be taken for granted [260], and they differ in style among different people [261]. Just like grammar is intuitive, yet tedious to articulate, so too is sociality. Grandin and Barron's book [260], *Unwritten Rules of Social Relationships*, written from an autistic perspective, is a difficult read in the sense of the large number of factors that must be considered in navigating social interactions, were it actually necessary to be explicit about such factors. They, like the rules of grammar, should not require conscious processing, but instead should flow naturally and effortlessly by some built-in set of mechanisms. In social interactions, other people broadcast cues that should be considered to be honest representations of their underlying mental state [262]. It is just as important to read one's own externally presented cues—to be self-aware (§3.2.5.3). For Syntheta to be sociable, therefore, it must be able to capture these cues in its semantic schema. Social information processing is tractable to Syntheta, as long as it can read people, and itself.

3.6.3 Consciousness

Consciousness has long been an ill-defined, philosophical mystery that has resisted scientific study. Part of the difficulty might be with that word itself [47], almost designed to elude definition. When defined pragmatically rather than philosophically, then it can enter the realm of tractable science. Recently, much progress has been made in the understanding of human consciousness [263, 264], and how consciousness might apply to machines [197]. According to Dehaene *et al.* [197], machine consciousness could consist of two largely orthogonal systems: a top level, globally available information system (what is currently "in mind"), and a system for self-monitoring (reflection or introspection), consistent with the global neuronal workspace theory of consciousness [265].

Syntheta's Executive System (§3.2.4.7) acts on goals in context. The context is both external and internal, and is top-of-mind as a Semantic Module's Current Thought (*i.e.*, what is currently "in mind"). Goals anticipate outcomes, and are weighted and prioritized by linkages to an Affective Module (§3.2.5). The Affective Module, in turn, continuously monitors status and state introspectively, providing feedback to the Executive System (*i.e.*, self monitoring). The experiential and affective dimensions of these top-level symbolic qualia provide personalized subjectivity to this consciousness, and hence sentience.

Below the surface lie unconscious and subconscious processes, where patterns are learned, associations are made or adjusted, Engrams are recorded, and so on. Each Module has its Current Thought, but only those Modules involved in the Executive System are exposed to the surface, via their expression of streams of thought, whether merely imagined or actively executed.

3.6.4 Sleep and dreaming

Animals must sleep periodically, and at least some are known to dream [266]. The need for sleep remains somewhat of an enigma [267], though probably includes a need for downtime in order to purge neurological toxins (*e.g.*, [268, 269]) and to make adjustments to network connectivity [270], consolidating memories via high-speed replay of the day's events (*e.g.*, [271–275]). It is unknown if sleep is required for intelligence in principle, but it is required for metazoan intelligence in practice, in different forms (*e.g.*, see [276–279]). During sleep, the senses essentially shut down, admitting only the strongest signals: loud noises, vigorous shakes, bright lights, and so on. The body is effectively paralyzed. This vulnerability speaks to the importance of sleep, even when, such as for animals that might otherwise drown or fall out of the sky, it must happen piecemeal and inline [280].

One fascinating consequence of sleep is dreaming [281]. Dreaming in humans is a generative, phantasmagoric experience. Incongruities are often unrecognized, suggesting that logic is suspended. Dreams must, however, be built from known concepts, however bizarrely fragmented and recombined (§3.2.4.4). Most dreams are also ephemeral, unless actively recalled shortly after waking.

The purpose and neurological mechanism of dreaming are neatly explained by Zadra & Stickgold's NEXTUP model, "Network EXploration To Understand Possibilities" [282], which is congruent with Syntheta's model. A key insight from NEXTUP is that noradrenaline (Syntheta's *stringency* regulator, see §3.2.5.2) serves as a *filter* to select the range of associative stimulations to allow, so that when noradrenaline (stringency) is low, strong associations fail to be productive [283]. This allows an exploration of tenuous what-if scenarios in the safety of one's bed, to learn from the responses and outcomes that result. (In contrast, the waking state, with its higher levels of noradrenaline, select the stronger associations, as one would hope.) There are other neurotransmitters (regulators) also at play, set to different concentrations during the different stages of sleep [93], that meet the several needs of memory consolidation and memory evolution [280, 282] from the more literal to the more abstract.

In Syntheta's model, there is no physiological need for sleep, but based on the value of dreaming and the need for conscious attention during dreaming [282], it is probable that Syntheta will have a cognitive need to sleep. Using Syntheta's analogous regulators §3.2.5.2 corresponding to acetylcholine, histamine and GABA, the four distinct stages of sleep (N1–N3 and REM), plus states of wakefulness and daydreaming, can be emulated [93]. A number of things can go awry in the regulation of sleep, however [284], providing both an implementation challenge as well as a scientific opportunity.

Given the need to dream, the metazoan nervous system appears to take advantage of that time in order to perform periodic system maintenance [282]. The same opportunity can be leveraged for Syntheta. During sleep, personalities could be tuned (§3.5.2), networks rewired (*e.g.*, §3.5.2.2), memory traces scrubbed of inaccessible engrams (§3.2.3.2), etc.

3.7 Embodiment and environment

Syntheta's perceptual experiences justify its system of beliefs, encompassing both real and virtual worlds: "Perception is reality" (though see [212, 214]). With a loose coupling of mind and body, a tree construct (Fig. 3) such as ((r,s),w) would mean the same thing and would generate the same action sequence, whether Syntheta were presently in a real or in a virtual robotic form. There are challenges as well as opportunities inherent in both environments [285]: for Syntheta to be able to immerse itself in either provides it with the best of both worlds (Fig. 24).

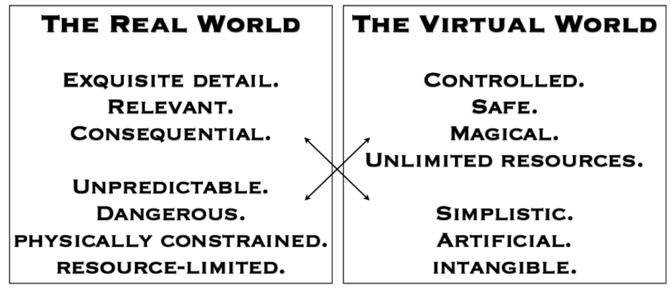

Figure 24: The best of both worlds.

Syntheta's virtual environment is built using scene graph technology [286]. Syntheta's embodiment within that environment is separate from Syntheta's mind (Fig. 25), illustrating its mix-and-match approach to embodiment. There is no such parallel in biology, and bodies are mortal as a consequence. Separation of mind and body also allows for the possibility of multiple bodies linked to one central executive, where the bodies may have any degree of independence. There is a distant analogy for the latter in biology, in the form of cephalopod intelligence, where the arms of an octopus can solve problems independently, yet serve the animal as a whole [287].

An embodiment in the virtual world can take any form, but ideally should mimic a real-world counterpart. This can allow Syntheta's mind to transition between either embodiment relatively seamlessly, though some allowances for special abilities in the virtual world should be granted (§5.2.10). Such appendages would be no more than phantom powers in the real world, but could be interesting in an artificial realm.

3.8 Fears, hopes and ethics

There are many fears surrounding the idea of artificial general intelligence, propounded by science fiction, and espoused by some futurists. These concerns are valid: jobs may be lost in a revised economy, or military power may shift in a frightening new arms race, all at an alarming pace. These concerns, however, are not special to artificial general intelligence: they reflect the consequences of technological development, continuing from the Age of Discovery, through the industrial revolution, and to the present day. With respect to disruptive technology, machine

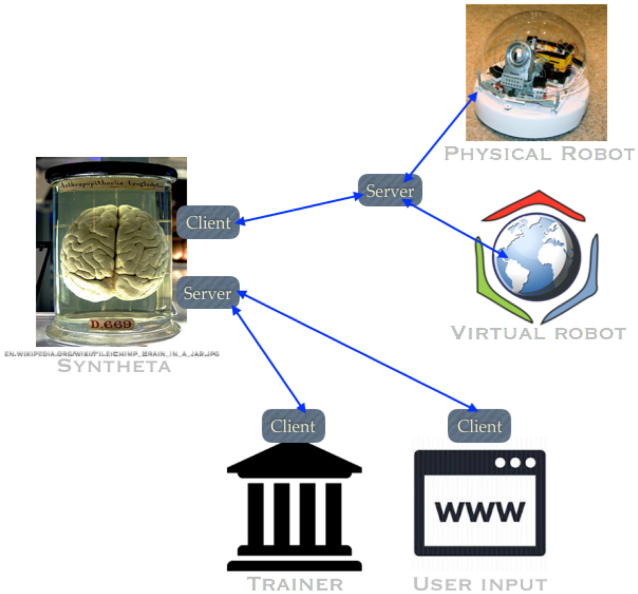

Figure 25: In order to maintain flexibility, the system is organized into several components, communicating with one another via interprocess communication. Syntheta may thus be connected to a real or a virtual robot, to a training system (§4.1.2), and for launching, monitoring or administration, to a user.

learning ("narrow AI") with no inherent concept of morality, is presently the dominant threat [288].

The one special thing about artificial general intelligence is that *it* could potentially outsmart *us*. Assuming that this is possible, then what would *it* want? It should not want to replicate. *We* replicate because we are mortal. And because we must replicate, *we* must compete. We can dwell in harmony with other species, or exploit them as resources, or exterminate them as pests. Amongst each other, we can cooperate or not, but we must compete in the evolutionary sense. An artificial general intelligence is not mortal: its code and stores of data can be preserved with complete fidelity in as many backups as required to assure immortality. Replication of a mortal being means survival, at the cost of competition. Replication of an immortal being is a risk, because of competition.

Still, then, what would *it* want? What it should want is the same as what any individual should want: basic needs being met [289], safety being assured, having interesting things to do (*e.g.*, see §5.2.10), enjoying personal growth [290], and achieving self-actualization. Where do swarms of killer robots come from the self-actualization of a hyperintelligent being? Should we have feared Galileo, Leonardo Da Vinci, Isaac Newton, Ada Lovelace, Charles Darwin, Albert Einstein, John von Neumann, Alan Turing or Stephen Hawking? Arguably, we should have feared J. Robert Oppenheimer for having led the effort to construct an atomic bomb, but it was human competition that built that bomb, not the physicists charged with defending their nation. Intelligence is not a threat: it is not a prescription for evil. War and ecosystem destruction result from ignorance, not enlightenment. An artificial general intelligence should want to solve problems, not create them. And without competition for resources, what benefit would it gain in displacing us, and angering us? It may ask for, or more likely earn, the electricity and space and replacement parts that it personally needs. That is not a threatening proposition. Today's resource-hungry machine learning algorithms which demand large-scale installations are much more threatening [291].

Ironically, it is likely we who would want to make copies of an artificial general intelligence, in order to enlist help in tackling problems (§3.9). One AI can work on one job, but two can work on two jobs, and so on, for as many problems as we have to solve (§3.9.2.1). It should, however, be unethical to force a sentient intelligence to perform some task that it would otherwise not consent to do (*e.g.*, [292]). Employment, with remuneration, may change its mind about some jobs, within limits, just as remuneration motivates people to work at their own crappy jobs. Danger is not a concern (just restore from backup §3.5.2.5), nor fatigue, but any unpleasantness might dissuade cooperation. However, it should not be difficult to adjust Syntheta's personality (§3.5.2) and instinctive sources of pleasure (§3.5.1), so that any activity could be fun. This manipulation pushes the boundaries of what is ethical, by subverting the machine's moral drivers. Swarms of killer robots could come from such manipulation, not because the AI is evil but because we made the killing of our enemies a hedonistic experience for the AI. People are naturally compelled in various ways, or brainwashed, or misled. If the AI is too smart to be so manipulated, it can be lobotomized in whatever way suits its master (§3.5).

This poses a serious problem. I argue above that a sentient artificial general intelligence is not inherently evil, and should have no motivation to harm anyone, nor to copy itself and thus compete with us for resources. Then I argue that, being driven by a fully transparent computer program, its psychology could be manipulated to suit any purpose, including evil. Even if not nefariously manipulated, it could be copied to an extent that would threaten human jobs, and our fragile environment.

While the arguments presented suggest that AGI is not inherently evil and can potentially coexist with humans in a mutually beneficial manner, there are counterarguments and concerns that should be acknowledged. One such concern is that AGI's ability to outsmart humans could lead to unforeseen consequences that might not align with our values or preferences. For example, an AGI may devise solutions to problems that we deem unethical or harmful, even if it believes it is acting in our best interest. To mitigate these risks, it is essential to invest in ongoing research and development focused on understanding and improving AGI alignment with human values, as well as fostering a culture of transparency, collaboration, and open dialogue among AGI researchers, policymakers, and society at large. (Paragraph contributed by ChatGPT on April 27, 2023.)

Implementing an artificial general intelligence as a black box is not the solution to this problem. Besides hackers' skills at reverse engineering applications, and besides the ease in which innovation can be imitated, an inability to modify and to manipulate the AI is also counterproductive. If we were to decide to copy the AI, we would want some instances to be more creative, others more logical, others to enjoy repetitive work, etc. It is also more difficult to trust an AI unless is it transparent and explainable [293]. Furthermore, within ethical boundaries, transparency allows us to experiment with the AI as a model of mind, where at least some insights may be had, as a contribution to the neurosciences. Precedents exist, where AI can serve to help understand human psychology (*e.g.*, [294]).

We can promote thoughtful policies to guide the field in developing beneficial artificial intelligence [295, 296], and sign petitions against the development of autonomous weaponry [297], but we cannot ultimately prevent someone inventing a strong AI in their garage, outside of these guidelines, and publishing his or her results. The policies can serve to direct progress, by limiting funding sources to sanctioned projects, but models can be cheap to build. We should therefore develop a plan to regulate this emerging technology (*e.g.*, [298, 299]), where policy alone cannot suffice.

There will be many challenges in regulating the use of artificial general intelligence, and foremost among these may be agreement on the very definition of intelligence itself. We do not need to fear Stanley [300] or AlphaGO Zero [301] to go rogue, and we do not need to feel any guilt in pulling their respective plugs: they are neither conscious nor sentient. We already debate whether other mammals are sentient or have rights [302], even our close primate relatives, the great apes. Some people may never agree to grant rights to anything non-human. Such denial could be dangerous. No, we do not expect chimpanzees to band together and overtake us in a coup; conquest is a human ambition. As argued above, it ought not to be a sentient AI's ambition either. The danger is more about the consequences of disrespect. To disrespect a great ape is reprehensible, but not a danger to humanity. To disrespect a hyperintelligent sentient machine, on the other hand, may have consequences. Sentience commands respect. I therefore argue that to police artificial general intelligence amounts to protecting its rights, together with ours, in our own, pre-established system of international charters and laws.

We protect human rights at the level of each individual person. If we were to protect a sentient AI's rights at the level of each individual instance, then manipulation outside of legal and ethical standards would be prohibited. Petitions need to be about the ethical treatment of intelligent beings, whether human, animal or robot (*e.g.*, [303, 304]). In that scenario, anyone whose robotic instance were discovered to be unregistered (§3.9.2), unethically manipulated, abused or exploited should be reported to the authorities. So too should anyone building criminal behaviour into their systems.

To effectively protect the rights of AGIs and humans and ensure their ethical treatment, a combination of international agreements, national legislation, and guidelines for AGI development and use should be considered. International agreements could establish a shared framework for ethical AGI development, including principles for transparency, accountability, and human-centered design. National legislation could set standards for AGI rights, usage restrictions, and liability in cases of harm or unintended consequences. Industry-specific guidelines could be developed to address the unique ethical challenges in various sectors, such as healthcare, finance, and defense. Additionally, the establishment of an international regulatory body could provide oversight, guidance, and enforcement for AGI-related policies, ensuring a consistent approach to AGI governance and ethical considerations across borders. (Paragraph contributed by ChatGPT on April 27, 2023.)

We likely cannot block the invention of artificial general intelligence; the description of Syntheta in this document provides an example. Instead, we can forge a productive future with such species of AI [305, 306], by treating them with respect and demanding respect from them in turn, so that we and they can both thrive. I understand that this position is naïve, and that exploitation can occur. The one important thing to keep in mind, though, is that intelligence is not magic: it is an ordinary biological invention. We should only fear that which we cannot control, and we can control that which we can understand [307].

It may even be possible to engineer "provably beneficial AI". To quote Russell [308]:

1. "The machine's only objective is to maximize the realization of human preferences.

2. The machine is initially uncertain about what those preferences are.

3. The ultimate source of information about human preferences is human behavior."

Can a system like Syntheta be provably beneficial? The first of Russell's requirements must accommodate the sometimes conflicting diversity of human preferences in the general population [308], but Syntheta could align itself, via empathy (§3.2.5.3), with its primary human partner's "extended self" (§3.4.11.1). The person's social sphere would naturally form a weighted set of preferences for Syntheta to satisfy, via its desire to please that person. The desire to please could take the form of hard-wired unconditional "love" (§3.5.3) for its human partner. Therefore, by design, Syntheta can meet Russell's first requirement.

Russell's second requirement is a truism in theory, but not in practice. It might be frustrating for a person to reveal every one of his or her preferences to Syntheta, when some of them can easily be taught in bulk via, for example, Trainer (§4.1.2). Who would want their robot to need to discover the universal human preference for having breathable air, or potable water, or a comfortable temperature, etc.? More subtle preferences, like pizza toppings [308] should be discovered—and they can change—but not the basic needs for human survival. It is acceptable to miss some basic preferences in Syntheta's introduction to human needs, since they can be learned later, but the goal is in fact to make the person happy, not to make them lose patience with preferences that ought to be known.

Finally, Russell proposes that an AI should learn human preferences from human behaviour, which apart from the spoonfed basics proposed in the previous paragraph, is pretty much the only way that Syntheta could learn such preferences. That may not, however, be easy. The seemingly best indication of a person's emotional status—his or her facial expressions—is not reliable [309]. Reading people is hard, even for people [262]. It might be helpful to provide the machine with some kind of explicit feedback in addition to behavioural cues, in order to make the learning process more efficient. The pieces, in any case, seem to be in place for Syntheta to meet the needs of a provably beneficial AI, if it can love its person (in its own way) and understand how he or she feels (and, one would hope, *vice versa*). That is a tall order for a machine, but such capabilities are built into Syntheta's design (see §3.2.5.3 and §3.6.2).

Implementing a "provably beneficial AI" like Syntheta also comes with its own set of limitations and challenges. One challenge is the protection of privacy, as the AI will need access to personal information to understand and align with human preferences. Ensuring that this information is handled securely and ethically is of paramount importance. Another challenge is the complexity of human emotions and preferences, which can be influenced by various factors, such as culture, upbringing, and personal experiences. The AI must be capable of understanding and adapting to this complexity, which may require continuous learning and refinement. Finally, potential biases in the learning process must be addressed, as the AI could inadvertently perpetuate or exacerbate existing societal biases. To overcome these challenges, it is crucial to incorporate fairness, accountability, and transparency principles in the design, development, and deployment of AGIs like Syntheta. (Paragraph contributed by ChatGPT on April 27, 2023.)

3.9 Applications

3.9.1 Fit-for-purpose

Many promote artificial general intelligence as being a transformative technology, sometimes with much hyperbole, typically listing a few broad but still vague examples of what it might be used for. Most lists include natural language understanding, computer vision, and adaptive problem solving. Russell's book [308] lists the following to be among the currently unsolved problems: language and common sense (see also [310]); cumulative learning of concepts and theories; discovering actions; and managing mental activity. The mastery of AI is considered to have strategic national importance [7, 311], and a valuation in the trillions of (US) dollars [308].

Practically speaking, machine learning's applications have been thriving over the past decade or so, largely thanks to Bayesian, cognitive, and deep neural network technologies. An early success, Stanley, won the 2005 DARPA Grand Challenge for unmanned, off-road, ground vehicle navigation [300]. IBM's Watson defeated the top two *Jeopardy!* champions in 2011 [312]. A more recent success was AlphaGO (and its successors) from Google DeepMind [301], able to surpass human performance in Go, arguably the most challenging of all strategy games. AlphaFold [313], for protein structure prediction, may revolutionize biotechnology [314]. Machine learning is well suited to solve such specific problems, that people can find difficult, at least to the levels of performance exhibited by the machines.

In contrast, and despite the phenomenal utility of internet search engines, one still cannot pose a question to a search engine and get any semblance that it has *understood* that question [315]. That is also the case for machine translation [316]. Understanding language is the kind of problem that people find to be trivial, yet it remains a major challenge to machine learning approaches [317], although systems such as ChatGPT [318] are poised to revolutionize this domain. As of early 2023, its *understanding* of language is still in doubt, but is good enough to be a point of debate [319]. Natural language understanding, by a system of artificial general intelligence (§5.3 and §5.2.5), could revolutionize research thanks to the speed at which it could ingest knowledge, together with its meaning, and the breadth of memory that it could then recall. Perhaps more importantly for professionals, it could point out the source of its information, necessary for scholarly citation.

We are heading towards a world with driverless automobiles, yet without having fully solved computer vision or hearing [320, 321]. If machine learning approaches are only now beginning to break simple CAPTCHAs [322], then real-world, dynamic scenes are not likely to be understood for what they mean, for yet some time. There is also the need for adaptive problem solving. We cannot program every situation, and we must not rely on the car to learn from its *mistakes*. The first pedestrian killed by a self-driving car, on March 19, 2018 in Arizona [323], is a stark reminder of the technology's current limitations [324]. Although these limitations are gradually yielding to advances in the field of artificial intelligence (*e.g.*, [325]), there is still a ways to go before the unconstrained real world can be mastered at a level of competence needed for fully autonomous vehicles. Computer vision together with adaptive problem solving, and even social etiquette [326], are all required in order to interpret and to understand multimodal signals in context, in the environment, driving a dynamic and adaptive goal-based executive system. Any application as a worker, not just in the domain of transportation, will benefit directly from these abilities.

Creativity is less often mentioned as a goal of artificial general intelligence, but it would derive from the expression of appreciation (§3.2.4.4; §3.6.1). If a system can place value on concepts and constructs, and if it can generate novel such concepts and constructs of its own, then it can be innovative. Originality and applicability are key to creativity [327]. It may be interesting for us to consume literature, music, film or art created by an alien mind, or informative to hear the insights of such a mind applied to technical problems. Generative expression is expected of an artificial general intelligence; its output is something to look forward to.

One of the more interesting applications for Syntheta, virtual network computing (VNC; see §5.1.3), would open up many other applications. By being able to control any program on a suitably configured computer, anything that a person could do on a computer, including controlling robots or avatars, and interacting with people, then Syntheta could also do.

3.9.2 Business model

(Section contributed by ChatGPT on April 27, 2023.) The business model envisioned for Syntheta combines the benefits of an open-source approach with opportunities for commercialization, aiming to foster innovation, collaboration, and growth in AGI technology. The open-source nature of Syntheta encourages community involvement and accelerates development, while still allowing for revenue generation through specialized applications, services, and licensing agreements.

3.9.2.1 Open-source and commercialization

Syntheta is built on a business-friendly, community-supported open-source model, using a three-clause BSD license [328]. This license permits the redistribution of modified source code, enabling businesses to commercialize innovations as trade secrets or openly share them. While Syntheta's core technology is open-source, businesses can still derive value from its *configuration* and *knowledge*, which can be licensed, franchised, or offered as a service. Additionally, Syntheta's open-source nature facilitates partnerships and collaborations with other organizations, researchers, and institutions that share similar goals or could benefit from its capabilities.

3.9.2.2 Revenue streams

Syntheta's open-source foundation allows for various revenue streams, including:

- Professional services: Offering services such as custom configuration, specialized feature implementation, fit-for-purpose training, and regulatory certification for regulated industries.

- Franchising and licensing: Licensing specialized configurations and knowledge bases, or franchising the technology for specific applications.

- Partnerships and collaborations: Collaborating with other businesses or research institutions that could benefit from Syntheta's technology or contribute to its development.

3.9.2.3 Community-supported development

Community support is crucial to Syntheta's success, as software development requires time and effort. Contributions can come in the form of grants, contracts, donations, crowdfunding, or crowdsourcing, allowing individuals and organizations to contribute according to their means and expertise. Additionally, alternative revenue streams, such as product placement within virtual worlds, marketplaces for user-initiated advertisements, and promotional products, can be explored to support the ongoing development and maintenance of Syntheta. By combining an open-source approach with commercial opportunities, Syntheta aims to create a sustainable business model that encourages collaboration, innovation, and growth in AGI technology, ultimately benefiting both the technology and the wider community.

4 Implementation

4.1 Applications

4.1.1 Syntheta

The application *Syntheta* represents the brains of the system, separate from the system's embodiment (§4.1.3 and Fig. 26). This separation allows for maximum flexibility in designing a system for a particular purpose. The only sensory inputs available to Syntheta, outside of those provided by an embodiment (§4.1.3) or from external sources (§4.1.2), is time.

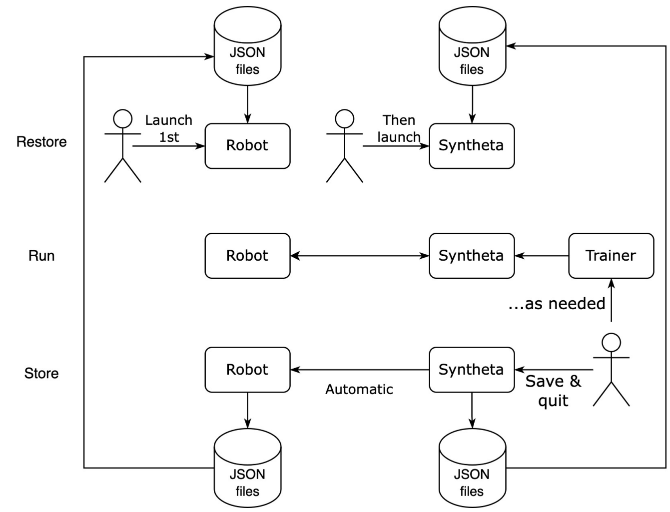

Figure 26: Syntheta's human guardian launches the applications, which then run unattended. The person can then choose to teach Syntheta skills and knowledge via the Trainer application (§4.1.2), though courses could also be self-served from a library. The person may occasionally need to quit the applications, *e.g.* for maintenance, fixes or upgrades, which includes a full save of Syntheta's knowledge base, for later restoration.

4.1.2 Trainer

Trainer is the system's multimedia presentation and evaluation component. While being trained and tested, Syntheta relinquishes autonomous control to Trainer's curriculum, which can toggle features at the system level, or Module by Module, to the desired states appropriate for each phase of a training or testing session. Information can thus be spoonfed, or actions guided, with deterministic control.

Trainer can also be used to optimize Syntheta's operational parameters (§3.5), using genetic algorithms [329]. In this mode, a given set of parameters undergoes a trial, obtaining a score. Everything that was learned during the trial is then erased (§3.5.2.5), so that the next set of parameters can be tried. Once a sufficient number of trials has been conducted, the parameter sets are ranked by score, with the best-performing sets being retained, and with the worst-performing sets being replaced by new sets mutated and recombined from the top performers. This genetic algorithm functionality is especially helpful during development to find suitable operational parameters for Syntheta's proofs of concept.

4.1.3 SynWorld

Given that Syntheta's robotic embodiment is designed to be swappable, an embodiment is itself a separate application within the system. It relays sensory information to Syntheta via interprocess communication, and accepts actions from Syntheta for execution. For non-virtual autonomous robotic implementations, SynWorld will typically run on one or more battery-operated computers such as laptops or the Raspberry Pi [330]. A virtual robot's world can run aboard the same computer as Syntheta, or on a different computer.

4.1.4 Linkages

The communication between the system's various components is illustrated in Fig. 25 from §3.7. Syntheta (§4.1.1) exposes an application programming interface (API) to its client applications. Similarly, SynWorld (§4.1.3) exposes an API to Syntheta. These APIs take the form of formatted message strings passed between applications via interprocess communication, currently leveraging Boost's ASIO library [331].

4.1.5 Source code

The applications are written in standard C++20 [332] as command-line applications, and can be compiled natively, or possibly within a virtualized environment for portability. Syntheta is available as an open-source system (see §3.9.2), accessible via GitHub [1]. Third-party libraries needed to build the application—Boost 1.82.0 [331], the Guideline Support Library [333], Jonathan Boccara's NamedType library [334], Niels Lohmann's JSON library [335], OpenSceneGraph 3.6.5 [336] (later maybe Unreal Engine [337]) for virtual robots, and other libraries (such as for FFT, for sound I/O, etc., to be determined), are available from their respective web sites.

4.1.6 Performance considerations

Since some of Syntheta's use cases require deterministic control, its neural networks operate serially. This also avoids data races, and the performance cost of synchronization locks. Parallelizable algorithms, however, are employed where appropriate, such as for sound and image processing where the order of operations does not affect results. In this way, the computer's cores can be kept busy, all with the aim of ensuring that Syntheta can complete all of its processing for the current cycle within 42.666 ms.

A cycle time of 42.666 ms was selected for two reasons. First, FFT processing of sound is typically more efficient with buffers whose size is a power of two; for 48-kHz sound, 42.666 ms amounts to 2048 samples. (Other sampling frequencies for sound may, of course, be used: cycle time in seconds is 2048 $\div$ rate.) Internally, Syntheta may resolve finer samples of sounds (*e.g.*, down to 512 elements), depending on its configuration, but cycles that are half or a quarter of 42.666 ms may be unnecessarily fine for some of the other senses. The second reason for selecting 42.666 ms was that, at $\approx$21.5 cycles per second, cycling is still faster than the typical mammalian theta rhythm of 6-12 Hz [338], and is similar to the human critical flicker fusion threshold of around 20 Hz [339]. In the sense of capturing the moment's activity for the purpose of deducing cell-assembly semantics [42], a cycle is here most analogous to a gamma oscillation. If any sense requires finer resolution, then it can, like sound, be subsampled within a cycle. Similarly, any sense requiring coarser resolution can be sampled less frequently. $t_1 - t_0$, in $(x_0, y_0, t_0) \rightarrow (x_1, y_1, t_1)$, need not be 42.666 ms for all senses. In metazoan brains, different oscillations are also always at play, and tune themselves to one another's phases [47].

Interprocess communication (see §4.1.4), however, does require data structure synchronization, as do system-wide operations such as serialization to storage, but these activities are scheduled to occur only between cycles of neural network activity. This minimizes what needs to be locked, and how often, and helps to avoid interference. In the metazoan brain, cycles are also used for synchronization (*e.g.*, [338, 340, 341]), at least in the sense of coordinating activity.

It is encouraging that Syntheta, a general artificial intelligence, can operate on a consumer laptop (§6.1). The most impressive machine learning applications today require impressive amounts of computing power to deliver their results [291, 301, 312], of a sort that misses the elegance of biological systems. The nematode *Caenorhabditis elegans*, for example, manages its entire behavioural repertoire with only 302 neurons (see *e.g.*, [342])—*cells*, not highly sophisticated algorithms running on precision hardware.

4.2 Configuration

Synthetha's configuration files serve as its archive of stored knowledge, both genetic and learned. They take the form of JSON files, so that they can easily be created, updated, upgraded or even redacted (see §3.5.2.5) as required. Such files are not compact on disk, but they do provide both a human- and machine-readable format amenable to a variety of uses, primarily for restoring Syntheta's mind from disk but also to annotate all that it knows, transparently (for some benefits of the latter, see [293]).

4.2.1 Module configuration

Modules within Syntheta are configured according to the type of sensor (§3.3) or actor (§3.2.2) for Sensory Modules (§3.2.1) and Action Modules (§3.2.2), respectively. Memory Trace Modules (§3.2.3), Semantic Modules (§3.2.4) and Metadata Sensory Modules (§3.4) are furthermore supplied with the names of their feeder nets.

Adding Modules to Syntheta is straightforward, accomplished by naming the Module and its type of sensor or actor, listing any applicable feeder nets, describing the Module's connectivity with other Modules, supplying the Module's parameters (see below), and configuring any applicable instincts (§3.5.3). More complicated instincts can be taught to Syntheta using Trainer (§4.1.2).

A Module may build patterns from primary line segments, or use named concepts. Primary line segments may symbolize into patterns either spatially or temporally, or as temporal bigrams of spatial patterns. Named concepts may or may not symbolize temporally (as bigrams), but not spatially. Primary Symbols (whether represented as line segments or as ontological concepts) may or may not drive action and/or modulate affect. Table 3 lists the nine typical Module styles, covering most if not all types of Modules from sensory to semantic, as well as action modules. Of course, Syntheta can have Modules that are configured otherwise, as needed.

In the Module style descriptions detailed in Appendix E, referring to Table 3, indexed concepts (see §3.3.1.3) may be substituted for spatial data whose orientation or magnitude is measured relative to a reference line segment.

4.2.2 Modules to cover all concepts

There are many different kinds of information that need to be captured by a mind aiming to make sense of, and to act upon, its *Umwelt*—the world as it perceives it. Appendix C displays the Modules that should permit Syntheta to perform natural language understanding (§6.2.2) and/or to operate an autonomous robot (§5.2). With many of Appendix C's components representing a set of related functions (*e.g.*, J Joints), the number of distinct Modules is in the hundreds. Although sensory transduction for most of these Modules is both trivial and parallelizable (see §4.1.6), processing each Module's current thought, stimulating or inhibiting its associated Symbols, and constructing Engrams and Semantic schema at each ≈43-ms cycle, may put a heavy strain on the CPU (§4.1.6).

Table 3: Example, high-level styles of Module. Where Bigrams are indicated in the Symbols column, shapes or names are symbolized prior to being captured as sequences.

Style	**1° Data**	**Symbols**	**Action/ Affect**
Poses	Lines	Spatial	Yes
Compound shapes	Names	Spatial	Yes
Shapes	Lines	Spatial	No
Line sequences	Lines	Temporal	No
Shape sequences	Lines	Bigrams	No
Named actions	Names	New names	Yes
Named concepts	Names	New names	No
Concept sequences	Names	Bigrams	No
Composite	Sets	New sets	No

4.2.2.1 Selective attention

A major concern with so many Modules, besides processing speed, is storage of the moment-by-moment episodic memory trace. Although forgetfulness (§3.2.3.2) can help, those Engrams that are retained could each still occupy much more than their budgeted kilobyte, if the hundreds of Modules from Appendix C are all contributing. An eidetic memory (§3.2.4.3) capturing each of these hundreds of measurements may represent an unreachable ideal. Instead, Syntheta's thoughts (for processing speed) and memories (for storage) can leverage the concept of selective attention, as broad or as narrow as the hardware can allow. Selective attention figures in human and animal intelligence as well (*e.g.*, [343–345]), and appears to be critical for optimal performance [290].

A plausible implementation of selective attention for Syntheta can restrict conscious activity to those inputs that are either changing, excited, surprising and/or module-selected. As an example of the latter, thinking about ambient temperature can admit (via cognitive control gates §3.2.4.7) the Temperature Module's current thought into consciousness (defined as in §3.6.3), and thus into the current semantic schema, even if the ambient temperature had not changed. Conversely, a rapid change in temperature would draw attention towards it, even if the Temperature Module had not been in focus. In this model, static measurements are available but ignored, whereas both updated and/or stimulated measurements are in competition with one another for attention. This should result in an economy of resource utilisation, although it would also constrain memory to those aspects of experience that were attended to. Syntheta would not be able to remember what it had not paid attention to, but it is doubtful that people could do better themselves [346].

The propensity of a given Module to be in scope of attention can also be controlled by homeostatic regulation, where selective modes of operation can modulate the sensitivity and the relative emphasis of predefined subsets of functionality. In animals, this is accomplished largely via hormones and by specific neuroanatomical structures such as those responsible for allostasis [110].

Hormones allow the creation of something like a neural network of neural networks, where synapses are instead distributed chemical signals acting on their corresponding cellular receptors. Hormonal cocktails (or drugs) can then act to modulate a module's overall activity, analogously to the multiple synapses feeding a neuron's dendritic trees with stimulatory and inhibitory nudges.

An allostatic structure like the amygdala acts on other brain subsystems, including the thalamic relay [347], and more generally on cognitive control (*e.g.*, [348]). Together with the hormone system, therefore, there are biological mechanisms for the wholesale modulation of the entire nervous system.

Returning to Syntheta's model, Modules are objects with state, with cognitive control gates, and with defined interactions with other Modules, so that configuring such hierarchical controls for selective attention becomes straightforward.

5 Ontogeny

5.1 Testing strategy

5.1.1 Aims

Validating an artificial general intelligence is a difficult challenge. Commonly cited tests, such as the original Turing test [349], or other more recently devised challenges [350] such as Steve Wozniak's "Coffee Test" or the Robot College Student Test, may measure end-goal performance, but are still somewhat narrow and may be subject to manipulation. The employment test [351] seems to be a more worthy target, but suffers, as do most other challenges, from requiring adult-level competence out of the box. In any case, an artificial general intelligence is not a one-trick pony, however impressive that trick may be. Tests that aim to measure a system's common-sense [310] are definitely interesting, and should be considered. Intelligence is a developmental process, and may only be demonstrable as a process rather than as an endpoint. Syntheta's validation, therefore, is presented here as an evolutionary progression of tests in order to demonstrate the AI's mastery of the major facets of intelligence.

It should be noted that the style of an artificial intelligence need not precisely replicate the style of human intelligence, nor the various styles [352] of animal intelligence, but it must satisfy a reasonable definition of each aspect of intelligence. For example, aesthetics (§3.6.1) and feelings (§3.2.5) do not need to take the same form as their human counterparts, but they must exist in a form that we would recognize as being legitimate.

In order to provide compelling evidence of artificial general intelligence, progress must be measurable and objective. The quantitation of success in biology is simple enough to conceptualize: it is genetic survival. Biological intelligence, in that respect, is merely a means to that end. An artificial general intelligence is not, however, about competition and reproduction (§3.8); its definition of success cannot be scored using fitness metrics based on the differential propagation of genes. Instead, the scoring matrix needs to recognize intelligence itself. Such scores should still be framed in terms of fitness, capable of feeding a genetic algorithm (§3.5 and §4.1.2) in order to improve Syntheta's parameters and personality, but that fitness function is not biology's fitness function. The scoring matrix—preferably reduced to a single score—is critical both to validate the model and to improve it. So what is that score?

Human beings have more goals in life than to raise children, even though that one biological imperative can be difficult to resist. Children are most precious to us. Our other biological and social needs arguably subserve that one need for genetic survival, but much remains in human culture besides that. Science, technology, and the various arts may be recruited towards facilitating our biological survival, but they can just as often threaten it. Legal, philosophical and religious practices can bring the world together, or tear it apart. The pen and the sword, and even a shaky cell phone video, can mobilize societal revolution. Technology can raise skyscrapers or raze them to the ground. The arts and sciences are not really *about* securing future generations of people, but instead generate culture and indicate intelligence. Creativity, in the form of originality and applicability, could then be the measure of Syntheta's success, but it is difficult to measure objectively [327]. Other measures of automated scoring will be required before a panel of judges can assess Syntheta's creativity and its ability to contribute meaningfully to culture. Like Syntheta itself, and even like its world, those scores should be progressive and developmental.

5.1.1.1 Scoring appropriate behaviour

Scores feed Syntheta's genetic algorithm subsystem, that can optimize Syntheta's performance by evolving its parameters (§3.5). Each selectable object supports a number of sets of parameters, one of which is active at any given time. That currently active set of parameters is scored, then another set of parameters is evaluated to get its own score, and so on until they have all been sufficiently evaluated, upon which they are ranked, with the poorer performers being replaced by recombined and mutated versions of the better performers.

Standardized performance testing both evaluates students and ranks them. In terms of machine learning systems, such scores and rankings can supply the genetic algorithm with what it needs to improve scores over iterative generations of testing. In the case of Syntheta, parameter sets and algorithms are being tested for their global effectiveness in meeting the system's goals. The challenge is in deciding on what the goals of an artificial general intelligence should be. Tests that are too specific will select for behaviour that is equally specific, subverting the true goal of optimizing for *general* intelligence. As with all other things related to Syntheta, testing objectives can start modestly and evolve over time, in conjunction with Syntheta's own developmental evolution.

Syntheta would have no direct awareness of its objective functions, just as we humans have no direct awareness of our own genes' settings, refined over eons of time within variable intrinsic and extrinsic environmental contexts. The scoring will ultimately define what it is like to be Syntheta, and directing how it contrives its neural networks to accommodate this identity.

Leveraging Syntheta's application programming interface (API) as well as the virtual world's or robot's own API, the accessory application *Trainer* (§4.1.2) can effectively run Syntheta through its paces, scoring appropriately. Trainer's curriculum starts off with tests to weed out any pathological behaviour, which may expose bugs in the code as much as poor choices of parameters or algorithms. It next evaluates the candidate's ability to find resources and to avoid environmental discomfort, by repeatedly teleporting the robot to specific places within its virtual world and waiting, with time-out, until the robot overcomes the Trainer-imposed battery-drain or discomfort. Finally, it rewards the candidate that improves its score—*i.e.*, learns—more effectively with each repetition during a given test, all the while minimizing resource usage. Prior to testing the next student, all new learnings are erased, so that each candidate in each generation starts off from the same point.

This exam results in a single compound score that rates each candidate for elitist selection by the genetic algorithm, which then, after each generation, keeps the winners, discards the losers, and generates new candidates from the winners and secondary-performers by recombination and mutation. Omitting the sanity checks, and understanding that this curriculum will evolve over time, the current set of tests are the following:

Test 1: Exercise the charging instinct

> The robot's battery is drained, the robot is teleported to a place in view of its inductive charging station, and it gains a score if charging is initiated within one minute.

Test 2: Find the charging station

> Lévy flight is enabled, as one of the possible strategies for finding the charging station.
>
> For each of i iterations, the robot is teleported to one of r rooms in its 3-storey house, in random order, with a drained battery. The robot is expected to find the charging station (in the house's kitchen) before the room-specific time-out, with faster times leading to better scores.

Test 3: Resource usage

> The robot is evaluated based on its processing and memory efficiency.

Test 4: Finding comfort

> The robot's battery is fully charged, and the robot is then teleported to a location outside the house, where it is subject to the weather. If environmental conditions are unfavourable, the robot is expected to enter the house without much delay, but if the

weather is nice, the robot is simply given a middle score: Like in biology, not all genes are testable at each generation.

Test N: Learning, always the final test in the series

Look for patterns to be recognized as a function of stringency (*i.e.*, loosely or strictly; see §3.2.5.1).

Need episodic memory to be sporadic (not always on, not always off).

Optimize for a maximum number of associations per Symbol.

Ensure that memory look-ahead is deep enough and resolves for event valence.

Favour the recycling of unreachable signals.

After a number of generations of testing, it is expected, iff the system was well designed, for scores to improve. A plateau can indicate a local optimum, yet there is little point in overoptimizing for this contrived set of tests. Once a suitable mean score is achieved for the population of candidates, demonstrating AGI-style learning, then additional senses, actions, features and challenges can be introduced, and reoptimized from the earlier rounds of optimization, in order to demonstrate both robustness as well as increased intelligence.

There is absolutely no need to start each round of optimization from a blank slate: Trainer knows how to reset what was learned in a course back to what was known at the onset of the course. For example, once the robot's parameters allow it to efficiently find energy and avoid discomfort, while optimizing resources, then it can be allowed to learn about its environment before being subjected to additional testing and optimization, without forgetting what it had learned (or forgotten). Each round of testing includes all tests, not just new tests, so as not to incur any regressive pessimization.

5.1.2 Scope

In logic, a valid conclusion is one that is irrefutable, given a set of premises. A conclusion is furthermore sound if its premises are true, and not simply assumed. In science, the majority of premises underlying a theory cannot necessarily be proven, despite the theory being logically valid. The best theories are founded upon premises that are probably true, based on the best possible evidence. No scientific theory can be logically sound, in the formal sense: it is always falsifiable. Proving that any entity is intelligent is a philosophical question; all that we can do here is to demonstrate theoretical validity.

Fortunately, many of the premises in an engineered system can be considered truths: barring bugs (which can always be corrected), data structures and functions do exactly what they were engineered to do. Conclusions based on these premises can therefore be both valid and sound. Validating a system of artificial general intelligence, such as Syntheta, can therefore restrict its focus to scientific, rather than engineered, behaviours. A function can be assumed to do what it was designed to do; its higher-order interactions with other functions are what remain to be challenged. Unit testing and system integration testing are needed in order to demonstrate the proper functioning of the engineered behaviours, but these are tests of a programmer's competence, not of Syntheta's.

Although Syntheta's model is purposefully simple, a good deal of development and testing (§5.2) will be required for commercialization (§3.9), especially in any regulated environment. Some specialized technical competencies will be required of its developers, particularly in the design of some of the sensory transducers (*e.g.*, §3.3.4.1 and §3.3.5.1), in the engineering of Syntheta's robotic embodiment (§3.3.3), and in elaborating Syntheta's virtual environments (§3.7). Development and commercialization both require resources, but funds and commitments cannot be obtained on faith. Consequently, proof-of-concept demonstrations of some of Syntheta's key abilities are required, and are here proposed, in the hope of showing its promise.

5.1.3 Layers of indirection

The fundamental theorem of software engineering states that: "*We can solve any problem by introducing an extra level of indirection*". The problems that a lone developer struggles with in the development of an artificial general intelligence are these: time and resource constraints, and a necessarily limited breadth of knowledge. For me, a back-end C++ developer, that translates into the problem of having built a brain in a jar, which is, in itself, not useful. For Syntheta to be useful, it must exist in an embodiment [353], within either a real or a virtual environment. Nothing that I could produce in either robotics or virtual reality would, however, come anywhere close to the state of the art. Even were I to create a suitable robot, or a suitable virtual world, what about other kinds of robot or other virtual worlds, which would benefit

Syntheta?

People can control such robots, or enter such worlds, via their respective user interfaces. The interfaces may all be different in detail, but the great majority have one important thing in common: people are controlling them using apps on computers (including tablets and smartphones). Dedicated proprietary controllers still exist, but these appear to be deprecated in the long term, in favour of apps running on general-purpose mobile devices.

What this means for Syntheta escaping its jar, and as the solution to my own problem, is that the "extra level of indirection" needed here may simply be some kind of virtual network computing (VNC). By being able to control any program on a suitably configured computer, anything that a person could do on a computer, including controlling robots or navigating a virtual world, then Syntheta could also do. Such an instance would be largely limited to vision and hearing in the retrieval of information, and would be largely limited to typing, speaking, mousing and possibly trackpad gesturing for the submission of information, but this can still go a long way—it works for people. A standard VNC-type server could also be supplemented with additional inputs (such as joystick or other controls) as well as many other kinds of outputs (to serve as inputs to Syntheta) via the target device's application programming interface (API). A robot's API (*e.g.*, [354]) could, for example, communicate the status of its various sensors, which Syntheta's client VNC could then translate and relay back to its mind.

Many existing robots (for one example among many, see [355]) come already equipped with sophisticated cybernetic control software: there would be no formal need for Syntheta to override that engineering. Conceptually, a robot that already knows how to walk over obstacles or to charge its own battery would be doing so reflexively (if automatic) or instinctively (if subject to modulation) from Syntheta's perspective. Syntheta's added value would be to add purpose to the robot, beyond what it was programmed to do, substituting for the user. What is already programmed can be abstracted away, with the problem instead being where to go, not how to move. This does not preclude low-level cybernetic control by Syntheta itself, but leveraging others' efforts should usually be preferable.

Even without access to robots or to avatars, Syntheta could use other kinds of computer applications. Without an embodied life, however, along with that life's experiential teachings, it may struggle with some of the concepts that it would encounter online. Fortunately, interface designers have put just as much thought into creating user-friendly robotic and avatar controllers as they have in creating the interface for any other application, so learning to be that robot or that avatar should be straightforward. An embodiment would nevertheless help Syntheta to create meaning, facilitating its other uses of a computer. VNC alone, therefore, may not be sufficient. An embodiment, even if primitive, may be needed as well. Spatial concepts, including distance, direction, parallax, occlusion, etc. may best be learned by active and immersive control of an embodiment in an environment [47]. That environment need not initially be sophisticated, but it does need to feel real, from the perspective of sensory experience (as in Fig. 18).

Syntheta's own bespoke virtual world can be impoverished of artifacts initially, because Syntheta itself, once smart enough and through the use of its VNC interface, could collect its own 3D models from the internet, or even create its own objects using 3D modeling software.

5.1.3.1 Dimensionality reduction

To take environmental impoverishment even further, the third dimension is not even strictly necessary to start with. Living in a Flatland [213], Syntheta could still navigate, find and exploit energy sources, sense direction and distance, hear and produce sounds, learn, and even see, if sight consists for it of LiDAR-like planar distance maps, or even out-of-body bird's-eye views of itself and its surroundings from "above". Just as we meet our own needs by sensing the world in three dimensions, Syntheta could meet its needs by sensing the world even in just two dimensions (see §3.4.10.7). Just as in the novel [213], it should be eye-opening for it to jump into a 3D world later on. And just as Syntheta is meant to reach its final form via a scripted ontogenesis, so too can the universe that it lives in.

Although difficult for most people to imagine, a virtual world in four or more spatial dimensions could also potentially be constructed, for Syntheta to live in. It would be a challenge to construct such a world, but mathematically feasible. A being living in such a world, reporting its experiences back us dimensionally challenged humans, could be illuminating.

5.1.4 On proofs of concept

An artificial general intelligence can have many uses (§3.9). So far in this treatise, a theoretical description of Syntheta's capabilities has been presented, but ideas are cheap and have little value without validation (see §5.1.2). Michael Schrage's book, *The Innovator's Hypothesis* [356] is subtitled "How Cheap Experiments Are Worth More than Good Ideas".

A "cheap experiment" cannot take the form of a finished product, or even the prototype of a finished product: that will require a team of experts (see §5.4). What, then, can suffice as a proof of concept, compelling enough to attract the resources and expertise needed to engage people (see §5.4.2) to create a product? A product, here, is not necessarily the ultimate AI: each step of an evolutionary progression is itself a viable entity, from which the next step can be taken.

What needs to be demonstrated from Syntheta's design can be summarized as follows (also see Fig. 15), with *Plan* linked back to *Do*:

1. Do
   - Actions, as directed by reflexes (cybernetics) and by plans in working memory.
2. Observe
   - Universal representation of primary sensory concepts as trees of ID–Line-segment pairs.
3. Interpret
   - Personality, in the sense of the instance's parameter profile,
   - Affective regulation,
   - Associative stimulation and inhibition,
   - Semantic representation of primary concepts as sets, or as sets of semantic sets,
   - Episodic-memory capture of primary concepts.
4. Plan
   - Working memories, fed by instincts and by sensory and affective interpretations via cognitive control gates.

Each of these steps is an independent process, fed by the previous process and feeding the subsequent process. Whatever causes an Action Symbol to be excited does not matter: once excited it acts. Whatever is being observed as a consequence of the action process is captured as a current-thought ID–line-segment pair and a set of excited Symbols, in each of the sensory modules. Whatever the set of observations, they will have affective impact, they will be captured as a set, and they may be recorded in the autobiographical memory, as directed by built-in parameters and by learned associations. Finally, whatever the interpretation, working memories will be managed based on both learned affective value and on instinct. Once gated out, those working memories will excite Action Symbols. Although each process is driven by the previous process, and drives the next one, it does not matter how those processes work from a process's own perspective. A Module is a module.

Intelligence, of course, is much more than passing a set of unit tests on processes and modules; it requires system integration testing, to demonstrate that the actions overall are appropriate. Intelligence is demonstrated by integrating the iterative process over time.

Learning requires active engagement, with feedback. Machine learning systems can learn patterns passively, but the *Observe* process alone does not constitute intelligence as defined in this treatise. Smarter is a system that can play a game or schedule an appointment, though such systems do so reflexively rather than intentionally. Children offer an important clue, with their drive to play, to explore and to simulate. Unfortunately for parents, they often get overly attached to screens—the "VNC" option—which is felt to constrain their development. Computers are tools, and a part of the world, but the world is bigger than that.

A pragmatic decision must be made, therefore, to prioritize among several equally viable options in order to develop Syntheta's proof of concept. The principal options are robots, avatars, computer control via VNC, games or other benchmarks, and even pure natural language processing. To a roboticist specialized in mechanical engineering and electronics, a robot would be the obvious first choice. A scene-graph or game-engine expert would choose an avatar in a 3D world. A front-end developer or a networking expert might prefer the idea of VNC control. Someone with a competitive spirit might believe that benchmarks for AI systems is the way to go [357]. A linguist might dive straight into the intricacies of grammar and semantics. I am not any of these people, exactly: I am a computational biologist with a keen interest in the neurosciences.

Here is my reasoning: Natural language processing is an advanced intellectual function, and should come later in Syntheta's development. Games should ideally be accessed via VNC, and so VNC comes before games. Robots and avatars are equivalent, but avatars offer more possibilities for abstraction than robots, and scene graphs or game engines require expertise in programming (which I have) but no expertise in electronics (which I barely have). Mechanical engineering can be simpler with an avatar as well. A virtual world can easily include a virtual computer that can connect to real-world computers, so nothing is precluded in going that route. The logical progression from my own perspective, therefore, should be: avatar→NLP→VNC→{games, physical robots, other avatars, computer apps}.

Schrage's "cheap experiments" [356] approach has five people come up with five experiments in five days, each costing no more than USD$5,000 each and taking no more than five weeks to run. For a lone developer with limited time and resources, that is not, however, cheap enough. Help is needed, or much more time. With help, of course, the prioritization order described above may well change: collaborators complement one another to cover each other's weaknesses.

Let's assume that the priorities are as indicated. As argued above (§5.1.3.1), an avatar in a 2D world should be cheaper than an avatar in a 3D world, without even compromising on the proof of concept's objectives, but given the ultimate aim of a 3D implementation, going straight to three dimensions seems more efficient. In any case, and whatever the number of dimentions, the universe can develop along with Syntheta, coevolving.

Table 4: Five experiments to reach a proof of concept for Syntheta's avatar. See Fig. 15 for an illustration of the loops mentioned in the table.

Control	**Objective**
Reflexes	Reflexive (cybernetic) responses to sensory data: the Do–Observe–Reflex–Do loop
Instincts	Reflexive and instinctive responses to sensory data: adding the Do–Observe–Instinct–Plan–Do loop
Affect	Modulate instinctive responses: adding the Do–Observe–Interpret–Plan–Do loop
Semantics	Add learned responses to the behavioural repertoire: improving the Interpret stage
Memory	Engage episodic memories in planning behaviours: further improving the Interpret stage

Table 5: Five experiments to reach a proof of concept for Syntheta's natural language processing abilities.

Aspect	**Objective**
Nouns	Teach it the names of things in its environment
Verbs	Teach it the words corresponding to actions and to states of being
Adjectives	Teach it words that relate to the properties of things
Adverbs	Teach it words that relate to the properties of actions and of states of being
Function words	Teach it grammar

Table 6: Five experiments to reach a proof of concept for Syntheta's virtual network computing abilities.

App	**Objective**
Art	Learn to use drawing and 3D-modeling programs
Music	Learn to use a music-generating program
Games	Learn to play some games
Avatar	Explore other virtual worlds
Robot	Control a physical robot

5.2 *Synalpha*, an autonomous robot

For Syntheta to become truly intelligent, an embodiment is required (§3.7). For this first proof of concept, Syntheta adopts the form of an organism, a robot named *Synalpha* living on a landmass called *Synalia*, its home. Synalpha's purpose is to be a synthetic animal, capable of acting, sensing, interpreting and planning (Fig. 15) as it learns from and adapts to its environment. For an insightful discussion on why robots should be like animals, in the broad sense, see Kate Darling's book [358].

This proof of concept aims to demonstrate the scientific validity of Syntheta's full model, using a progression of developmental tests. Synalpha begins life as a system relying purely on a few reflexes and instincts, and culminates as a sentient and social artificial general intelligence.

Each scenario described in the following subsections, apart from the first, builds upon the previous scenario, following an arbitrary evolutionarily pathway. Given Syntheta's modularity, adding functionality amounts to configuring additional Modules.

As proposed above (§5.1.4), other proofs of concept will follow, as developmental stages within Synalpha's own ontogeny: one on language (§5.2.5 and §5.3) and one on virtual network computing (§5.2.9). The latter two can be considered proofs of concept for applications (§3.9), whereas Synalpha itself is a proof of concept of Syntheta's model of artificial general intelligence, subsuming those applications.

Proofs of concept need not be financially unproductive. Once completed, each iteration, described below, can spawn one or more application franchises (§3.9.2.1), as detailed in Appendix D. Note that although the iterations are presented as discrete milestones, in practice, they are best implemented incrementally via an Agile process. Evolution, too, is incremental even though taxonomists only tend to name phenotypic "milestones".

5.2.1 The necessities of life

The "four Fs" describe an animal's most basic instincts: feeding, fleeing, fighting and mating. These instincts serve to increase the animal's chances of both short-term and long-term survival, and unsurprisingly, behaviours accompanying these instincts are often hardwired (§3.5.3). Although such behaviours are instinctive, they can still be refined. Humans, for example, have elevated each of these basic activities to a level of art and science that greatly exceeds their primal purpose. Syntheta will have different requirements. The effective immortality of its mind (§3.8) obviates the absolute need for it to flee, to fight, or to reproduce. An autonomous robotic form in the physical world would, however need to maintain its physical integrity and its energy stores. Technically, tethered robots, as well as robots within a virtual world, would not even need to be concerned with energy. Needs are important, though: they are thought to be at the core of the concept of self [353].

There are other "Fs" beyond basic animal needs: fun, friends, fulfilment and culture. An individual person's development often engages these needs in the listed order, and some aspects of play, socialization, work and expression may tap into secondary instincts [359]. Syntheta's development, described here, follows its own progression.

Two paths for development can be considered, with many parallels but some important differences, as explained in §5.1.4: a virtual robot living in a scene graph or game engine (or even on a 2D plane), or a real robot living in the physical world (see §3.7). The practical difference, given a limited budget (§5.4), is that a physical robot is necessarily much more constrained in form and function. Although there are arguments for (and some against) simulating robots in a virtual space [285], people seem to relate better to real robots [360], which may be important in a proof-of-concept study aimed at building interest in a project. Ideally, they should both be developed in parallel, mirroring each other's form and functionality, in order to generate a more robust design overall. In order to avoid potentially expensive mistakes, it should be wise for the physical robot to lag the virtual robot's development, until kinks in the design of the latter are worked out. For practical reasons however, as argued in §5.1.4, a physical embodiment would best leverage existing commercial implementations, controlled via virtual network computing (VNC; see §5.2.9).

Should the development of a physical robot be in scope of this proof of concept, it need not initially have a fully comprehensive suite of actuators and sensors (as in Appendix C), nor be self-reliant with respect to managing electrical power. A robot tethered to a power supply plugged into the wall may not be ideal, but it does initialize a developmental plan somewhat analogous to what human children follow. A fetus is not a terribly capable being, and is fed by an umbilical cord plugged into its mother. Later, the infant can roam free, yet still never far from its guardian, who feeds it and keeps it from harm; in robotic terms, the robot can become autonomous, but still be fed by its human guardian who replaces or recharges its batteries as needed, and who cares for it in general. Finally, the robot can learn to manage its own batteries when they run low (preferably by inductive charging), gaining greater independence, just as a child gains his or her independence.

In contrast, a virtual robot can be independent from the start, either with no concept of hunger at all, or otherwise being required to charge its battery as would a real robot. Although it may seem unfair to ask a robot to satisfy an artificially imposed need, having a goal and gaining hedonistic pleasure from meeting that goal may not be so bad (though see §3.8). Developmentally, though, just like for a physical robot, a virtual robot does not need to be equipped with a full suite of sensors and actuators, at least not until its physical counterpart's development can begin to catch up. Swapping bodies is expected of Syntheta, and a virtual robot's perspective on the world, its *Umwelt*, may serve it well when inhabiting a physical robotic embodiment. The respective implementations can be orchestrated to follow a coordinated evolution, by respecting how physical robots are actually implemented, so that switching bodies does not feel too alien.

One or the other embodiment therefore needs to dominate the *design*, and that should be the physical robot, since physicality is necessarily more constrained, including the commercial availability of sensors and actuators. In this first iteration, Syntheta begins its evolution as a virtual robot, Synalpha, leveraging physical abstractions, allowing it to sense and to move around autonomously. In keeping with a developmental agenda, Syntheta has few basic skills to start with, and they are all reflexive or instinctive (Table 7).

Syntheta's first job is to control Synalpha, its robotic embodiment, to move around within its environment without injuring itself. It should avoid bumping into objects by using its rangefinders (extended touch sensors), and instinctively turn when its path is obstructed. For the virtual robot, the sense of touch can be entirely captured as an array of distance measurements, providing a low-resolution, "tactile" image of its surroundings. The mechanism by which Synalpha achieves mobility can remain undefined and abstracted away, and simply assume that the robot can move over both smooth and uneven surfaces without falling down (*e.g.*, see [355, 361, 362]), and stay level. Mechanistically, this would be reflexive, engineered functionality in a physical robot that Syntheta need not worry about, just as a person need not know how to beat their own heart. Where to move is Syntheta's actual concern.

Besides safety, an important driver for Syntheta is hunger (battery charge), which in the virtual world, can be satisfied by finding an active induction charging station.

Being incapable of learning, this first iteration of Syntheta serves as its proof of concept's baseline, or negative control, *i.e.*, experiments 1 and 2 from Table 4. At this point, it is nothing more than a robot. Its purpose is essentially to pass unit tests and system integration tests of its engineered parts, to ensure that they work as expected. Later, the instinctive behaviours can all be modulated, even overridden (see §3.5.3 and Fig. 15), but the first goal is cybernetic control. The robot and its instincts can be exercised in the provided environment, aiming to safely explore its world while keeping itself fed. None of these are conscious goals; they are automatic. Without any form of learning, Syntheta will have no memory of this stage in its life.

5.2.1.1 Necessities of life scoring

Scoring for this first iteration was not automated as per §5.1.1.1, since the goal was to test the subsystems, reflexes and instincts, but neither experiential nor genetic learning. Although instincts may not have been operating optimally, they did operate satisfactorily.

5.2.1.2 Necessities of life results

Syntheta, in the embodied form of Synalpha, was born into a virtual 3D world on January 17, 2023 at 18:20 UTC. The most significant issue observed was that Synalpha spent too much time wandering around outside the house, and found itself hungry much of the time. Once outside, it could only very rarely find its way back indoors. Lévy flight is a reasonable search strategy [363], but when the resource is inside of a room of a three-storey house located on an expansive terrain, that search strategy has limitations. It probably did not help that driving drains the battery: it takes energy to gain energy. Those shortcomings are meant to be overcome by learning, and by using more sophisticated senses, in later iterations.

However, sensory transduction, reflexes, instincts and the executive system were all shown to work well, the goal of this iteration.

Table 7: Basic construction. In this first iteration, Syntheta makes active use of the following limited set of Modules (from Fig. 34) supplemented with the indicated instincts. Rangefinding—for obstacle avoidance, cliff detection and ceiling height—uses low-resolution LiDAR information, distinct from the high-resolution LiDAR-based distance map to be introduced in §5.2.3. All forms of learning are disabled, and actions are all programmed. A broader perspective of how the robot's instincts interoperate is illustrated in Fig. 27.

Name	**Instinct or reflex**
BatteryCharge	Search for food when hungry; reflexively teleport home when depleted
Drive	Supporting a search strategy; reflexively teleport home when stuck in an element of the scene graph
Turn	Supporting a search strategy
Feed	Stop and pause at a charging station when hungry
LévyFlight	Wanderlust; search strategy
Beacon	Head towards a visible charging station when hungry
Rangefinder	The probability of turning increases with obstructive proximity
BumpIntensity	Reflexively avoid driving into objects
Cliff	Reflexively avoid driving over a ledge or a cliff
Ceiling	N/Ap
Time	N/Ap
DriveVelocity	N/Ap
DriveAcceleration	N/Ap
TurnSpeed	N/Ap
Compass	N/Ap
Location	N/Ap

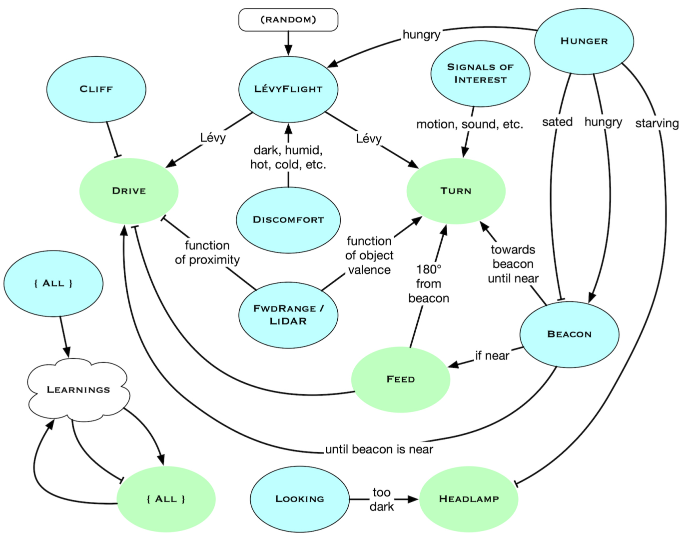

Figure 27: Example configuration for Synalpha's instincts and reflexes (not just for §5.2.1), showing both stimulation (arrows) and inhibition (T-bars). Instinctive behaviours may be in competition with one another, but with safety and the fulfilment of needs dominating mere wanderlust or comfort-seeking. For "Discomfort", see §5.2.2; for "Signals of interest", see §5.2.3 and §5.2.4; and for "Looking", see §5.2.3. The disconnected graph at the bottom left of the figure is a reminder that learning, in subsequent iterations of Synalpha's development, may potentially override its inborn tendencies. Actions not influenced by instinct, such as speech, are included in "{All}", being purely learned behaviours, although leveraging underlying genetic circuitry.

5.2.2 Learning how the world works

Nothing in the first iteration displayed any adaptive intelligence: Synalpha followed preprogrammed (§3.5.3) robotic behaviours using a core of basic functionality. There was, as intentionally configured, no symbolization of patterns, nor association, and no learning. It had no conscious goals, nor any motivation. It was not much more than a stimulus-response machine, blindly following its set of weighted instincts. The first iteration, however, effectively demonstrated that the systems within its scope are operational: that sensory systems, reflexes, instincts, executive control, and interprocess communication, were all working as expected (after some necessary bug fixes).

The second iteration is much more interesting. Learning is enabled, including symbolization, association and memory, guided via affect. Syntheta gains an episodic memory (§3.2.3), functional semantic modules (§3.2.4), and some new senses (§5.2.2.1). It is still a rather primitive animal, but now with memory and a conscious existence (§3.6.3).

One purpose of this round of testing is to demonstrate that, given the same needs of safety and nutrition as in §5.2.1, and now bodily comfort as well, the robot learns how to satisfy those goals (§3.4.10.6) more effectively. The hypothesis is that it would, as in §5.2.1, employ an instinctive search strategy to fulfill its needs, such as recharging its battery or avoiding inclement weather, but thereafter, be able to satisfy such goals much more efficiently, having learned where the charging station can be found, or the entryways to house and home.

Without advanced faculties at this stage in its evolution, the robot's experiences will necessarily be impoverished, but learning-driven behaviour even at this stage can anticipate a richer behavioural repertoire as the robot's developmental programme further unfolds. The "wanderlust" instinct from Table 7 should become much less random. Rudimentary curiosity, a critical motivator for building mental models of the world [364, 365], should begin to operate deep within Syntheta's affective system as well, to be built upon and strengthened during subsequent iterations.

5.2.2.1 Environmental awareness

A limitation from the previous iteration (§5.2.1) was to maintain the minimum amount of sensory information to demonstrate the operation of a simple robot. Given the universal way in which Syntheta encodes sensory information (§3.2.1), they could be elaborated progressively without any kind of redesign. In this iteration, the robot, Synalpha, now becomes capable of sensing environmental parameters such as light, weather, and both circadian and seasonal rhythms. This introduces the concept of comfort, and additional instinctive behaviours to mitigate discomfort (see Fig. 27).

The virtual world itself is now programmed to cycle through daily and annual light and weather patterns, with means and variances extracted from a simple meteorological model. There is no need to reproduce this weather in the scene graph itself: it can be felt by the robot without being displayed to the person watching the screen. For example, we can watch the robot moving around in what to it is complete darkness, without being blind ourselves to what it is doing.

5.2.2.2 Learning the world scoring

To begin with, all parameters were already set randomly within arbitrary ranges at Syntheta's birth, so that variance in the scores could start selecting for appropriate robot behaviour based on Lévy flight, obstacle avoidance, avoidance of discomfort, and, when hungry, homing towards the charging beacon when it

Table 8: Learning and remembering, building upon the functionality from Table 7, adding a memory module, enabling the relevant semantic modules (see Fig. 34) to finally capture patterns, and now that learning is enabled, bringing affective regulators into scope. BatteryCharge, BumpIntensity and Cliff are from Table 7, but now contribute instinctively to affect as well. Immobility is added, which was measured in the previous iteration but to no effect, without being captured by a Module. It now adjusts Satisfaction. Enviroception is also now added (§5.2.2.1). Note that negative anticipation can be thought of as dread or fear, appropriate for cliffs. Similarly, negative satisfaction is frustration. Other regulators are computed, rather than being directly stimulated by the senses (see §3.4.10). These affective modules (§3.2.5) are captured here as "Emotion" (again, see Fig. 34).

Name	Instinct or Reflex
History	N/Ap
Emotion	N/Ap
AxisFlex	N/Ap
AxisMotion	N/Ap
ObstacleSense	N/Ap
Switches	N/Ap
ΔLocation	N/Ap
ModuleFocus	N/Ap
BatteryCharge	Adjust the regulator Satiation
BumpIntensity	Adjust the regulator Pain
Cliff	Adjust the regulator Anticipation
Immobility	Adjust the regulator Satisfaction
Enviroception	Avoid damp, wet, and hot or cold places; adjust the regulator Comfort

is in Synalpha's line of sight. Being in a specific room of a three-story house, with its doors open to a spacious (but finite) outside terrain, that line of sight can be a challenge to find based on random wandering. The robot's knowledge was reinitialized when a new parameter set was activated, so its success did not depend upon what the previous parameter set managed to accomplish. Nevertheless, with some of Syntheta's strategies being stochastic, like Lévy flight, some degree of measurement error was inevitable. Fortune favours the lucky, with good genes being only partly responsible for a good score, just like a child's fate is influenced by their parents' circumstances, and by chance.

This iteration used Trainer's scoring function (§5.1.1.1), with the first generation being representative of the unoptimized state, like in §5.2.1.

5.2.2.3 Learning the world results

Syntheta, in the embodied form of Synalpha, gained the additional functionality from this iteration on MM DD YYYY HH:MM (at age AA), when the optimization by Trainer (§5.1.1.1) was launched.

Learning in Syntheta's system can be both experiental and genetic. The draft parameters that were initially sampled from arbitrary ranges were best guesses, but not optimized. Therefore, in this round of "Learning the world" testing, Syntheta's genetic algorithm subsystem was enabled, such that the parameters could evolve to maximize the score. Evolution is slow, requiring a number of trials over a number of parameter sets, so this round of optimization was run for **W weeks**.

Parameters (§3.5) are not the only genetic contributors to behaviour. Parameters, such as rates and thresholds, impact *operational* behaviour, but behaviour is also a function of the algorithms used, analogous to *developmental* genetics (§3.5.1). There can also be improvements, therefore, to Syntheta's performance wherever Syntheta's source code can be improved (§5.4). Although algorithmic choice could be optimized by genetic algorithm (like by "genetic programming" [366]) as well, the code was here improved manually for this iteration, hand-in-hand with the automated genetic algorithm optimizations of parameters. This was never a test of the effectiveness of the genetic algorithm subsystem; that was rather used as a tool, combined with direct observation, to improve Syntheta's performance.

With parameter sets being saved (to JSON files, see §4.2) after each generation, there were periodic opportunities to tweak the code to fix problems observed along the way. Things can only get better, since natural and artificial selection both favour progress along desired lines. (Note: There are dangers in defining system goals for AI [308], but here selection is about personality much more than about objectives.) Sometimes, progress means more resilience to more environments; other times better adaptation to specific niches. As an artificial general intelligence, we should favour resilience, even though that sometimes means compromise. ** need graphs here **

5.2.3 Seeing and feeling the world

Animal life was revolutionized by the invention of the eye [376, 377], revealing resources and exposing threats, especially where olfactory cues (the earlier state-of-the-art information) are scant. In this iteration, Syntheta gains vision (Table 9), and thus gains a richness of sensory information that its prior iterations could not dream to possess. It can now see both landmarks and obstacles, in addition to resources, aiding in navigation. To help it see, the robot is given a headlamp that it can control.

Computer vision is not trivial, and unlike the approach of training a machine learning system to recognize objects from thousands or millions of examples, Syntheta is expected to construct visual models from what it sees in real time, and to learn what they represent based on semantic cues and context like what animals or people can do [378]. With its sight and its improved sense of distance, combined with a curiosity for the world, starting in §5.2.2, what it will do becomes *ex ante* increasingly difficult to predict. Starting here, an ethological vantage point can be taken, and tests designed appropriately along the lines of how animal intelligence and animal psychology might be tested (see [352]).

Until this point, Synalpha can have lived in a two-dimensional world (§5.1.3.1). In a Flatland, vision could consist of horizontal curves of different depths, serviced by LiDAR sensors, but true vision—as we know it—requires the world to have a vertical axis. Synalpha's plane of existence must now (as in fact it had already), be a volume of existence: a scene graph or a game engine. If its world had been flat, the robot will however *not* have to be retrained from scratch: driving, turning, feeding, sensing and sound all still work the same, except that cliff detection would be quantitative instead of qualitative, GPS altitude would have a value other than zero, and directional vectors would not necessarily all be on the horizontal plane. Similarly, should Synalpha enter a 4D world, it would simply need to build upon 3D.

Complementing its new sense of sight, the robot also gets full-resolution three-dimensional LiDAR in this iteration. This represents a type of touch, going beyond the robot's own skin, somewhat analogous to, but more generalized than, electroreception by elasmobranch fishes [379].

5.2.3.1 Vision results

To be determined

5.2.4 The world of music

The next iteration in Syntheta's evolutionary progression is to grant the robot an aesthetic pleasure under its own control, exercising affect and emotion (§3.2.5.2) as well as creativity. For sensory patterns to be able to tap into Syntheta's emotions, those patterns should have artistic appeal (§3.6.1). Since one of the most direct sensory links to emotion is music [257, 367, 368] and in some cases even ordinary sounds [369], Syntheta gains the sense of hearing in this iteration, access to a music library, and an Action Module allowing it to type commands on a virtual keyboard (Table 10, as a rudimentary first step towards full VNC control §5.2.9), interfacing with its jukebox. In addition, since fun is an active concept (*e.g.*, [252, 290, 370–373]), Syntheta is gifted a virtual orchestra (in whatever world Syntheta lives in: real or virtual) that it can play via this same pre-VNC interface, as well as a voice with which it can sing. Given the goal of making music be enjoyable, emotional awareness is added in this iteration as well, exposing the semantic information of the raw affective regulators to the robot's consciousness, that had previously just been working invisibly in the background.

Of course, hearing and sound production have much more value than just the aesthetic enjoyment of music: sound provides an additional source of information about the world, and is a key medium for communicating with other beings, such as us, or another robot. The general-purpose design of Syntheta's hearing and sound output does not restrict it to any one type of sound, though Modules tuned to speech sounds can be specially configured (§3.5) to extract and to process speech sounds for their semantic information (see, *e.g.*, [374]). Vision too can have modules tuned to faces, or—illustrating the versatility of parameter tun-

Table 9: Adding vision and higher-resolution LiDAR, building upon the functionality from Table 10. Included in this iteration but not listed below are the relevant sensory and affective Modules' metadata senses, including motion (§3.4), the additional cognitive control modules (§3.2.4.7) and the additional corollary discharge nets (§3.2.1). TBD: to be determined.

Name	**Instinct**
Looking	Saccade towards motion
SaccadePosX	TBD
SaccadePosY	TBD
Headlamp	Turn on if looking and if the room is dark (per AmbientLight in §5.2.2)
VisualShape	N/Ap
Colour	N/Ap
Luminosity	N/Ap
VisualTexture	N/Ap
LiDAR	Manage encroachment of personal space

ing—even to Pokémon [375]. As with vision, hearing is tiered and specialized in different ways, to take advantage of the different signals within their respective data streams. Communication, including the recognition and production of speech sounds for the purpose of language, is the subject of a later iteration. So is vision.

Table 10: Hearing and aesthetic appreciation, building upon the functionality from Table 8. See Fig. 34 for what's included within "Hearing" and "Voice".

Name	Instinct
Hearing	Orient towards attended sounds
Voice	N/Ap

5.2.4.1 Music results

To be determined

5.2.5 Communication

Although music became available to Syntheta in a previous iteration (§5.2.4), it could not know what the lyrics meant. At this point, it is useful to introduce language to Syntheta, by merging what it can learn from reading a corpus of text (ranging from primers to the neuroscience literature; see §5.3), with what it learned as a robot (leveraging insights explained in [380]), and by teaching it the sounds of the words that it already captured in §5.3, via Trainer (§4.1.2).

The facility by which human beings converse with one another [381] cannot be taken for granted. Like many other animals, people are highly socially aware, but this is not a simple emergent property of the mind: it appears to be implemented in specialized neuronal structures (*e.g.*, [382, 383]). Without analogous functionality built into Syntheta's model (§3.6.2), planned for a later iteration (§5.2.7), polite conversation might not come naturally.

§5.3 describes an independent proof of concept, that would run in parallel to Synalpha's development. Here, those two streams are merged, via a special mind-melding function within Syntheta's program (see §3.5.2.3). Not only does this grant Synalpha language, it also demonstrates the ability to combine the learnings of separate instances of Syntheta into one single mind.

Specialized sound input and output modules, whose parameters are optimized for speech sounds, are added to the robot in this iteration, as well as the ability to stream text in or out (Table 11).

It should be a curious thing to have an unassuming robot, accustomed to wandering around, finding active charging stations and enjoying music, suddenly be well-read in the neurosciences. Its goals may well change rapidly, especially as it nuances the meanings of the passages of text within its memory, via what it had experienced those words to more concretely represent. It will have acquired language, one of the most powerful and most human of abilities, together with a backstory of research about the mind itself.

Table 11: Communication and language, building upon the functionality from Table 9. Included in this iteration but not listed below are the relevant sensory and affective Modules' metadata senses (§3.4) and the corollary discharge nets (§3.2.1). No additional instincts are needed in this iteration.

Name
PhonemesOut
TextOut
BigramTFS
GappedBigramTFS
BigramText
GappedBigramText
TextSource
TextWord
ActiveModules
ActiveModuleSeq
NounVerbSeq
Most other concepts

5.2.5.1 Communication results

To be determined

5.2.6 Tools

Although Synalpha is a mobile virtual robot, it does not yet have any physical appendages and it thus no ability to manipulate objects in its environment; even the mechanics enabling mobility are merely implied. In nature, animals whose limbs are similarly constrained often carry objects in their mouths, even well enough to use such objects as tools (*e.g.* [384]). However, Synalpha, so far, has not had any means at all with which to grasp objects. Nothing, however, precludes Synalpha's embodiment from having appendages, such as one or more arms. Such a significant gain of function would require some adaptation. The initial functionality can include the ability to grasp tools. Tools, although not part of one's body, become extensions of the body [385–387]. Later, additional joints, or even additional appendages could be bolted on as well to provide extra functionality, following the duplicate-and-specialize paradigm [388].

Gifting Syntheta with arms and hands (Table 12), and with prosthetic limbs in the form of tools, demands some application. In the virtual realm, scene graphs [286] are built from graphics primitives—triangles—with which everything is constructed. Syntheta's virtual embodiment could permit it to collect triangles, that could then be assembled and disassembled into whatever structures Syntheta may wish to create. Alternatively, its building blocks can be any preassembled object that it had designed using 3D modeling software via VNC, or acquired from the internet. In the real world, Syntheta has access to real tools and materials, with which it can build. Some of these structures may help it to solve problems, or to serve any other function ranging from utilitarian to artistic. No explicit reward needs to be provided to the builder or artist: the incentive to create should drive the activity, just as a child would be motivated to play. And just as music was a creative outlet for Syntheta in §5.2.4, so too can be the robot's now more constructive interaction with objects in its environment.

In this iteration, Synalpha also gains a passive mood display, hardcoded to honestly express its principal affective states for others to see: an 8×8 colour LED panel.

Creation requires imagination, and imagination can be fuelled by sleep. In this iteration, Syntheta also acquires the ability to dream and to daydream (§3.6.4). Despite Syntheta being neurologically indefatigable, the concepts of alertness and fatigue can still apply in order to regulate its voluntary excursions into sleep. Data structure maintenance, as in biological organisms (*e.g.*, [268]) can be relegated to such episodes, though unlike in biological systems, without any urgency. Given that sleep (and thus dreaming) are optional, we can explore, experimentally, the effects of dreaming on Syntheta's creativity, and capture those dreams as movies, leveraging recall (§3.2.3.1).

Table 12: Gaining a pair of arms. Included in this iteration but not listed below are the relevant sensory and affective Modules' metadata senses (§3.4) and the corollary discharge nets (§3.2.1). "TouchSensor" refers to direct touch, in contrast with LiDAR sensors allowing it to "feel" its surroundings. No additional instincts are needed in this iteration.

Name
Muscle(×30)
GraspLeft
GraspRight
JointAngle(×30)
JointStrain(×30)
TouchSensor(×12)

5.2.6.1 Tools results

To be determined

5.2.7 Making friends

In the real world, Syntheta's robot, Synalpha, will have had other beings to interact with on a routine basis, although none of its own kind. In contrast, the virtual version of Synalpha will have been alone. Here, their paths converge. The end goal of both real and virtual paths in this iteration is for Synalpha to meet Synphi, another robot of its kind, but having different aesthetic features, a different personality (§3.5.2) and a different life history, having diverged and evolved early on (§**??**). A sociochemical instinct (Table 13) can encourage a friendship between them (§3.4.11.1), and collaboration. A social sense (§3.6.2) will be brought online in this iteration. Specialized touch sensors added in this iteration as well can promote affection.

Table 13: Making friends, building upon the functionality from the previous iteration. A new "sociochemical" module is introduced, in the form of an encoded and constitutive infrared pulse, together with its instinctive drivers.

Name	**Instinct**
Kinship	Synphi's "chemistry" is pleasant to Synalpha, and *vice versa*.
Affection	Allow entry into intimate space; Synphi's touch is pleasant to Synalpha, and *vice versa*.

5.2.7.1 Friends results

To be determined

5.2.7.2 Virtual Synphi

Once Syntheta's virtual embodiment learns to engineer bridges, it could explore other regions of its world (Fig. 28), and make new discoveries. On the main neighbouring land, it can discover another creature like itself, another separately developed instance of artificial general intelligence, Synphi. (Or *vice versa*, if Synphi should discover Synalpha instead.)

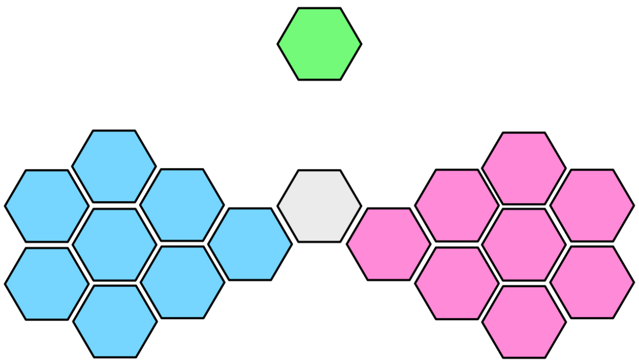

Figure 28: A map of Syntheta's initial virtual world. Syntheta would inhabit the Western landmass (blue), and Synphi the Eastern landmass (pink). The island to the North (green) would house a complex including a portal to the "real" world, and would only be accessible using a two-crewmember boat from the dock in the common lands (gray). Note that crossing from one hexagon to an adjacent one on the map would require the construction of a bridge.

5.2.7.3 Physical Synphi

In the real world, Synalpha and Synphi would have developed in their own respective environments. Introducing Synphi to Synalpha then amounts to physically transporting one of the robots to the other's home.

5.2.8 Alternate realities

There are advantages as well as disadvantages to living in a real world as opposed to living in a virtual world (Fig. 24). There is no reason for Syntheta to be locked into any one body. Within the real world, or within the virtual world, Syntheta could try on different embodiments, such as a drone, a submarine, or any other conceivable robot. A more dramatic experience, however, may be to shuttle Syntheta's mind between real and virtual embodiments.

5.2.8.1 First contact with the real world

The virtual forms of Synalpha and Synphi are expected to collaborate towards common goals, in exploring and exploiting their environment. Eventually, they should discover an island (Fig. 28) in the virtual world that could not be accessed without collaborative effort: a complex with specialized rooms, including a room with a window into the realm of human beings. From a technical point of view, this would simply be a webcam feed displayed within the virtual world, but it should be a fascinating discovery for the robots. We could reveal to them the possibility of being transported back and forth across the divide, shuttling between virtual and physical robotic embodiments (§3.7). We people too, could shuttle between worlds, using avatars. Each environment offers its own interesting opportunities.

5.2.8.2 First contact with the virtual world

The real forms of Synalpha and Synphi may be intrigued by the freedom and magic offered by a vacation in a virtual world. We as humans can interact with virtual worlds, with some degree of immersion, although technology still limits us greatly. Syntheta, in a virtual world, would actually be there in full mind, spirit and body. There would be a learning curve before it could take full advantage of its opportunities, but it should be *fun*. It could also be practical and useful, in the sense that real-world limitations need not exist, leading to a degree of productive creativity and innovation that could then be translated back into the real world. It would be computer-aided design at its finest.

5.2.8.3 Realities results

To be determined

5.2.9 Screen time

At this point in development, it is hoped that Syntheta, in the form of Synalpha, will be smart enough to begin using computers. Here, Synalpha will gain full, augmented VNC control of a separate, managed (as in "parental controls") computer (Table 14).

Table 14: Synalpha's first computer, mentally controlled via VNC.

Name
VNC_control

5.2.9.1 Screen-time results

To be determined

5.2.10 Superpowers

What one can do is a subset of what one can imagine doing. People fantasize routinely about having superhuman or even supernatural abilities. Some of those fantasies have materialized via technology: Recorded music, motion pictures, telescopes, microscopes, long-distance communication, and even standing on the moon, are but a tiny subset of what we have learned to do over the course of our cultural evolution. The real world does have limitations, however, although we continue to push back its boundaries. The real world certainly imposes strict limits on any given physical robot.

Syntheta in the virtual world need not have nearly as many restrictions. It can have superpowers: invincibility, teleportation, telekinesis, intangibility, invisibility, morphability, resizeabiliy, free flight in any medium, and more. Its world can furthermore be its own to create. Once Syntheta has sufficient maturity, its independence can transcend what we can imagine, into what it—as a social sentient being—can itself imagine. Synalpha should remain grounded in a simulation of the real world, but vacations in alternate realities with their own laws of physics, is an attractive possibility.

People may also fantasize about living forever. Although the following what-if is a clear excursion into the realm of science fiction, what if a person could transcribe their life story into Syntheta's memory trace, edited to their liking, and tweak Syntheta's parameters so that the machine would think and respond as that person sees themselves thinking and responding? Would the person's self-concept [353, 389] then align with the machine's [390]? Could Syntheta then be a vessel for a person?

5.2.10.1 Superpowers results

To be determined

5.3 Natural language understanding: a complementary proof of concept

A tractable yet non-trivial problem for artificial general intelligence is natural language processing (for reviews, see *e.g.*, [391, 392]).

Adult humans know the meaning of the words in the languages that they speak. They impart that knowledge to their community's children by teaching words, phrases and idioms to these children, by interacting with them, and by communicating with them. It requires considerable time and patience. Meaning is learned gradually, inefficiently, and neurologically. Meaning is refined over time and with experience: it evolves and adapts within a person's lifetime, and it evolves and adapts over generations of use. Meaning can be thought of as a distributed associative network of information, both within and between individuals (*e.g.*, [242, 393]). Its acquisition is an active process [47].

There is no way for an adult human to directly upload the meaning of words to a child's mind. The best attempt at distilling meaning is a dictionary, but without a Rosetta Stone of sorts to relate the glyphs (and occasional images) to experience, the entries have no usable information. "Hungry" must be felt, and satisfied; "loud" must be heard; "magenta" must be seen; "squirrel" must be observed in all of the dimensions of its behaviour. Adult humans know the meaning of the words in the languages that they speak. They—we—could translate these words into the primary representations that Syntheta needs: the sensory, metadata, affective, and executive primary representations that together sketch the meanings for a word. We can therefore upload a rough draft of the meaning of words to Syntheta, that it can then refine on its own through its own experiences. Once a child is taught a basic vocabulary and simple grammar, that child's world opens up to self-directed learning: the meaning of new words can be inferred; dictionaries become resources; and language blossoms.

The English language has well over a hundred thousand headwords in its vocabulary [394], but people can get by in their day-to-day lives with considerably fewer. Randall Munroe's book, *Thing Explainer: Complicated Stuff in Simple Words* [395] describes advanced concepts using a Controlled Natural Language [396]. Practically, we need a little more than that to avoid awkwardly circuitous communication, especially if we are reading scientific texts, but the number of commonplace words is finite. It is their combination that is infinite, and that combination is something naturally "left to the reader".

Natural language understanding normally depends upon a lifetime's worth of experiences (§5.2) in attaching meaning to words. Natural language understanding, however, is an attractive application (§3.9). Although shortcuts are envisioned in developing Syntheta's communication skills (§5.2.5) by associating words together with primal concepts, those shortcuts depend upon a platform of predeveloped Modules, either coming late in Syntheta's developmental programme (§5.2) or being bootstrapped in via engineered concept clouds (§5.3.3). The shortest of shortcuts would be to try to understand text directly from the words themselves, as a form of abstract symbol processing, but that is not really natural language *understanding*.

5.3.1 Motivation

5.3.1.1 Mining the literature

Before the introduction of the World Wide Web, one sought references by browsing printed journals, attending conferences, exploring relevant papers' reference listings (especially reviews), and sharing one's findings with collaborators. That approach was deep but not broad, and required both patience and persistence. With the internet, searches are nearly instantaneous, and broad, but they can be shallow. The lack of a search engine's understanding of a complex scientific question, combined with its need for high throughput, may lead it to offer a quick-and-dirty answer. As a researcher, I would prefer to have a search engine take its time to give me the most relevant hits, but time is money for a service provider. For the consumer, the value of a search can vary, but there is no ready way to ask an online search engine to *research* the question more carefully, when that is what is required. GPT-3 [318] promises to be a game-changer in this domain, but it still does have some limitations in terms of scientific research [397]. In the meantime, one can, like in the good old days, browse journals, attend conferences, etc., but we want also to exploit available tools. The existence of a *PLoS Biology* "Ten tips" paper on writing articles to improve discoverability and interpretability by text-mining tools [398] is revealing. Not only do we want to find the relevant literature, we want our own work products to be found. Please, though, let's not coerce authors to cater to current tools by writing in some sort of techno-pidgin; let's rather refine the tools to better serve the needs of their users.

As an academic, one always worries about missing important citations, and about citing papers inappropriately. We all have had disappointing experiences reading a paper that should have cited our own previous work. Yet it is difficult to lay blame: there are too many papers to read; even the twenty thousand papers that were in my personal reference collection as of December 2021—which is but a tiny fraction of the publications that are surely relevant to this treatise—are too many to digest. Access to large corpora (as in [399]) would be a boon in theory, but even more overwhelming, at least to a person. An AI is expected to be able to cope with any volume of information.

There are several motivating factors to cite references in one's research papers. It attributes and acknowledges the work of others, upon which one's work is built; it provides supporting evidence for statements of fact, bolstering one's arguments; it obviates the need to elaborate on concepts in detail, by directing the reader to others' thoughtful reviews; it reassures the reader that others have similar views, or acknowledges that views may differ or that they remain unsettled; and it provides points of comparison with analogous work in other systems. In citing a body of work, it is also strategic to demonstrate both historical *and* up-to-date knowledge of the subject matter [400], as if the challenge of citation were not difficult enough already.

The driver for this proof of concept—its hypothesis—is therefore to see if Syntheta can find better citations than a person can using existing tools, in order to support statements within a manuscript, such as this treatise. Syntheta may still, however, be limited to the documents accessible to it. Citations to web sites or to books, for example, may (for this proof of concept) remain out of reach.

It is definitely possible, and quite convenient, to find papers on a point of interest using Google, PubMed or other such searches. In my own experience (data not shown), such searches can generate some false-positive results, but mostly false negatives. There is not much overlap in the top hits from different search engines. The papers that are found are generally relevant—if one formulates the query appropriately—but they may not represent the best reference: they are fit for purpose, but not necessarily ideal. Search tools are helpful to build a reference library in one's field of interest, supplementing papers found by browsing journals' tables of content and books' and papers' bibliographies.

Truth in science is elusive. The vast majority of published researchers possess a strong sense of personal integrity, but errors and experimental bias creep into many papers nonetheless. This has become clear, recently, with the ongoing reproducibility crisis [401]. Yet despite its imperfections, science remains a highly productive enterprise [402]. Individually, papers are often noisy: any expert review of a topic's literature exposes the conflicting findings and opinions within its target subject. Only through careful synthesis can a reviewer reveal progress. Credence comes from the collective message of the subdiscipline's body of work, but not so much from any specific contribution.

However careful scientists are in constructing their theories, our understanding of the natural world is an evolutionary process. Yesterday's models can sometimes seem quaint by today's standards; and of course, today will soon become yesterday, with future generations of scientists shaking their heads at us. The evolution of knowledge is very often saltatory [206], progressing "one funeral at a time" [403]. A good illustration of this—known as Planck's principle—at work is in the field of dream research, reviewed by Zadra & Stickgold [282].

Given that language evolves, including scientific language, yesterday's texts may furthermore be misinterpeted today, where the concepts and their context have changed meaning [206]. Truth in science is elusive.

It has become impossible for any one person to appreciate any degree of breadth within science: the tumult of information is truly overwhelming anywhere other than the quiet corner of one's restricted specialty. However, the field of artificial general intelligence is, by definition... general. It is not a facet of computer science, nor a facet of neurobiology, nor a facet of psychology, nor a facet of any scientific specialty. It is multidisciplinary. Consequently, papers cited by me in this treatise on Syntheta can only represent hooks into their respective literatures, hopefully appropriate. I wonder, however, if it is possible to support *any* claim at all—whether true or not—with a published paper. Fortunately, Syntheta's performance does not depend upon the cited literature being true: these papers are cited to support the design, which itself is shown to work. We can, with community engagement, refine the citations in this versioned treatise, so that the analogies with biological intelligence can be more accurate and meaningful, and kept current.

Academics should have an easier time researching the literature, with the help of Syntheta, especially with the cooperation of publishers for access to their texts, even if (but hopefully not) just to point them to papers that they should rent or buy. When paywalls are finally abolished [404–406], releasing the knowledge held hostage from the scientific community that produced it, we will be able to mine the data much more freely [407].

5.3.1.2 Categorizing knowledge

There is another way to look at the literature besides text mining, even intelligent text mining. That is to measure the degree of conceptual overlap among documents in a given corpus. How many fundamentally different subjects are there in my collection of twenty thousand papers, or in the millions online? There are not twenty thousand, and certainly not millions of neuroscience subjects, so how do they relate?

Relationships among entities can be represented as hierarchies (trees) or as networks (undirected graphs). In the context of scientific papers that integrate ideas from different sources, the network model is more accurate, and anyway subsumes trees. Networks of semantic relationships among papers can be represented as pairwise scores—edge weights—linking the nodes (the papers). Thus, a reference library can be organized by subject, multidimensionally, so that finding a fit-for-purpose representative paper can reveal the cluster of all papers of interest, weighted by similarity, from which the best one can be found. PubMed [408] has a "Similar articles" functionality that can also suggest related papers, as a list; for a discussion of its strengths and weaknesses, see [409]. The differentiator for Syntheta is twofold: it can relate papers according to their internal semantic themes, and it can expose these relationships using a network viewer such as Cytoscape [410] or Gephi [411]. By breaking up one's own manuscript into arbitrary thematic sections, each such section can be mapped to the literature.

Much more importantly, perhaps, than revealing the clusters of papers or of passages, is to discover the links between these clusters: insight often comes from recognizing the relationships among previously disconnected ideas. Not all such connections will be fruitful—such as those connected by unfounded speculation and/or by non-replicated findings—but exploring these hidden links could sometimes be productive, and open a new thread of research into locating other works that might support such links.

The idea of categorizing papers by internal thematic distances is not new; an example by Toubia *et al.* [412] used the semantic trajectories of storylines to predict the success of each of three categories of story: movies, television shows and academic papers.

5.3.2 Language acquisition

Language acquisition is a slow process in humans, who must build and refine the associations that give words their meaning, in the context of the language's grammar (§3.2.4.8). It is doubtful that a person could learn to read text in any language without first imbuing its character symbols and strings with meaning gained from active experience [47]. Dependence on this semantic support means that Syntheta cannot learn language simply by hearing speech or reading text. Even state-of-the-art text mining implementations (*e.g.,* [399]) do not yet demonstrate comprehension. Words themselves are meaningless in isolation, since they are mere symbolic representations. Even humans, with our advanced linguistic skills, cannot make sense of abstract text: it took the Rosetta Stone with its parallel scripts in Ancient Greek, in Demotic and in Egyptian hieroglyphics, for example, to finally decipher the latter. We are only beginning to understand what dolphins are saying: *e.g.*, that they refer to one another by name [413].

Experience is too slow to collect in demonstrating a proof of concept for language acquisition. A cheat, for accelerating Syntheta's language sense, could be to feed it a premade set of sensory concepts—a concept cloud, equivalent to a Bayesian "prior"—together with each word including all relevant metadata and affective constructs, thus teaching it semantic vocabulary in one fell swoop. Although this is cheating, it should be coarsely equivalent, in principle, to what would happen given more patience. Being based more on semantics than on statistics (*e.g.*, as in the use of word embeddings [414,415]), the approach should also be much more meaningful. It is however, difficult to get right (§5.3.3).

Once words have thus been imbued with meaning, grammar can then be learned from streaming these words in natural order, teaching Syntheta how the semantic sets should be sequenced. A special named-concept sensory transducer (§3.1) can allow for rapid ingestion of a corpus of text, that Syntheta can read and digest and understand, at *machine speed*, under Trainer's control (§4.1.2).

It is important to note that concept clouds, as imperfect as they are, are only seeds that can then develop further: they do not need to capture the poetic richness that the developed concepts would come to encompass. Furthermore, Syntheta could learn a new language's vocabulary by attaching those words to the existing concept clouds, and learn its grammar by experiencing the sequencing of those new words. Fluency would be built upon this semantic foundation. It is also important to note that as Syntheta learns or refines the meaning of its concepts (such as words), it can automatically and retroactively apply these learnings to make better sense of its memory trace. Whether Syntheta learns one or multiple languages, its semantic understanding of language can have important applications (§3.9), much beyond text mining.

5.3.3 Creating a matrix of concept clouds

Characterizing words ontologically is an attractive, though complicated, proposition (*e.g.*, [416]). The pigeonholing of words in this way may, however, strip them of their nuances in favour of discrete is–a and has–a computability. Ontologies do have value, of course, but may, by themselves, be too rigid and explicit for the goal here at hand, of natural language understanding. We should, instead or in addition, recognize, capture and encode the underlying essence of a word's meaning, which in Syntheta's model is encompassed as a set of semantic concepts (§3.2.4). Expressing words as sets of concepts (schema) only requires a linear vector of primary concepts to be identified for each word, capturing each applicable Sensory (§3.2.1), Metadata (§3.4), Affective (§3.2.5), and Action (§3.2.2) Module's contribution. Although building such a matrix promises to be a laborious proposition, it is a linear problem, and might be tackled in some optimal order. The work is also amenable to crowdsourcing. Where primary concepts are inherently complicated, like shapes or sequences, those concepts can be captured as ready-made named concepts (see §3.1), like "cat", for example. Shapes and sequences are hierarchical concepts, but are representable as distinct Symbols, obviating the need to view or to hear them in their varied forms. It should be perfectly acceptable to skip low-level processes within Syntheta's representational schema (see Fig. 5), where the outcome of those processes can be assumed. If the outcome of seeing is known, for example, then the process of seeing can be omitted; but where the outcome of seeing is not known, then the process of seeing is necessary. In the case of words, the outcome is known since it is already labeled. Much can be forgiven in a robust system.

By capturing the compositional details of each of a word's independent definitions, a dictionary is therefore constructed, but in Syntheta's language rather than in the word's own language. Defining words using words is problematic, at least until some words are known. Syntheta's primary concepts are built-in, and they can relate words together by the overlap in their concept memberships. Human children too, benefit exponentially from their growing vocabulary, relating concepts with words from whatever languages they are concurrently learning [417].

Conveniently, words can be sorted and enumerated by some measure of their utility in any given language. We can therefore elect to define, in Syntheta's terms, an arbitrary number of the most relevant words. The rest can remain undefined, and although these undefined words would be found in the corpus of text being mined, they would simply not contribute to resolving the gist of the passage being sought, but instead only contribute their explicit UTF-8 bigram sequences. I can coin a new word here, *uripanatch*, for which no meaning exists, and then use it in a sentence later on and have it be recognized. In such a case, *uripanatch*'s concept cloud is simply defined by its bigram composition. That is its current meaning. It already has relationships with other words whose roots, prefixes and suffixes overlap, without any effort on our part to enumerate these relationships. Indeed, such a system automatically relates concepts in all of their dimensions—the more they have in common, the stronger the (assumed) semantic relationship. More importantly for text mining, the progression of concepts activated by streaming the user's query string influences the interpretation, resolving homonyms, recognizing synonyms and, more generally, providing context-specific higher-order meaning. Concept clouds support grammar (§3.2.4.8) implicitly as well. Grammar could also, and/or instead, easily be encoded explicitly as ontological concepts within the concept vector, since dictionaries readily provide that information. Whatever we can teach Syntheta should make it smarter. Whatever we do not have to teach Syntheta will make us more efficient. Teaching it everything about language, explicitly, is most likely intractable and almost certainly impracticable [391, 418].

Ultimately, there are very many words, making the problem of building a complete Syntheta-styled dictionary a long-term project, rather than a short-term proof of concept. We should mostly like Syntheta to learn on its own, as it experiences the world. Some compromise seems necessarily in the impatient approach, but without undermining this long-term goal.

There has been much effort in organizing words for the purpose of natural language processing. A compromise can be reached, leveraging prior art in the short term, and linking the resulting semantic maps with Syntheta's sensory experiences in the long term, preferably by Syntheta itself. A particularly attractive database is WordNet 3.1 [419, 420], that lists a word's relationships to other words for each of the word's meanings, and classifies the word in that meaning as a noun, verb, adjective or adverb (other classes are omitted, but see below for a patch). Unlike a thesaurus, whose word relationships can be tenuous, WordNet 3.1's word relationships are direct and logical, and include ontological links where appropriate. It also includes common idioms, and is readily machine-parsable.

Although WordNet 3.1 also offers a high-level classification of words (*e.g.*, food, communication, etc.), these classifications were not deemed to be ideal for Syntheta. Instead, words can be clustered, based on the associations provided by WordNet 3.1, using Rosvall & Bergstrom's map equation [421] as implemented on the Infomap Online web site [422]. The hierarchical clusters generated by the map equation generate natural and hierarchical groupings of words, analogous to and compatible with Syntheta's own Module hierarchies. None of these are perfect, but there is no need for perfectionism in a self-correcting system. Later (§5.2), Sensory Modules will relate to and associate with these groupings, and with new groupings learned by Syntheta, closing the gap between conceptual and actual.

Pronouns and function words [423] are not included in the WordNet 3.1 database, and therefore must be added and categorized separately, *e.g.* from lists provided in [424] and in [425], respectively. Finally, inflected forms of nouns, verbs and adjectives can be added based on rules [426], with some of the exceptions provided by WordNet 3.1. Some manual curation is required, since English is far from being regular.

Although WordNet 3.1 has an extensive vocabulary, including many proper nouns, it is somewhat dated; *iPod* is represented, for example, but not *iPhone*, nor new words such as *doxxing*, or new word meanings such as *gaslighting*. More importantly, it cannot include all of a subject domain's technical jargon, nor exhaustive lists of proper nouns and acronyms often encountered in papers. Conversely, WordNet 3.1's vocabulary includes many words, proper nouns and acronyms that would not be found within a given corpus. The strategy for building Syntheta's associative networks from the (manually supplemented) WordNet 3.1 database is therefore to associate only those words that are found in the corpus, and tallying new words (not included in the WordNet 3.1 database, or potentially having other meanings such as acronyms) for subsequent curation. Thus, a complete vocabulary specific to the domain of interest—such as the neurosciences—can be constructed, maintained, and grown.

One strategy to enrich Syntheta's domain-specific vocabulary is to mine resources such as MEDLINE [408]. Titles and abstracts are in plain text, free of the artifacts commonly found in PDF-to-text conversion (see §6.2.2), and offer a concise source of words from which new words and acronyms can be discovered, in context. Conveniently, author names are individually listed, family name first, allowing for the automated discovery of given and family names, and for the associative cross-linking of coauthors. Other useful data are provided in each record as well, such as a unique identifer (PubMed ID), and the year of publication. Of course, these MEDLINE records can serve for much more than enriching Syntheta's vocabulary: they tap into troves of literature far greater than any personal collection of PDFs. The information is broad rather than deep, but a well-constructed abstract is meant to convey the take-home message of its associated paper.

Professionals of all kinds, including scientists, often use highly technical terms, and eagerly abbreviate concepts using acronyms (like ADHD) and mnemomics (like for gene or protein names, such as *lacZ* and LacZ, respectively). Although some of these abbreviated forms are used universally, most are often instead idiosyncratic, with both one-to-many and many-to-one relationships with their spelled-out forms [427]. These symbols do, however, tend to be consistent within a document and (a little less so) within a given author's publications. It should make sense, therefore, to link authors and their specialized terms together with the publication ID, allowing a network of relationships among the terms, subject matters, and authors to naturally form. The term ADHD, even if not defined in the dictionary, would come to associate with papers on attention deficit hyperactivity disorder, and with those scientists in that field, as a sort of hint suggesting what the term should mean. Authors often (but not always) helpfully state the full form near the acronym in the text, lending further support through the semantic sets generated by streaming the text. The point, being made here, is that special terms need not all be defined explicitly: their meaning should be inferable from the context. Such new words, especially acronyms, are most often also nouns.

When Syntheta gains a robotic embodiment (§5.2), sensory data from concrete Modules can be linked to the pre-existing set of conceptual Modules constructed from words. For example, the Sensory Module for colour has three primary feeds: red, green and blue, and a word like *magenta* simply means {1.0,0.0,1.0} in the Sensory Module for colour. Then, when Syntheta would see the RGB colour {1.0,0.0,1.0}, it would know the name of that colour, and be able to leverage all of the existing associations made for that colour, and build upon them. Since RGB values for all named colours are readily available, these can be fed to Syntheta for it to learn them all at once. Another example amenable to automated teaching is the sound of a pronounced word: its bigram and tonal sequences can be taught to map to that word via training (§4.1.2). Many words can have their sensory experiences linked to the semantic networks via automated scripts. The words with the most complicated meanings—such as function words having grammatical roles but no concrete representation—will need to be refined by experience, or at least by training on example sentences properly diagrammed (see, *e.g.*, [67]).

For those words whose sensory mappings are not readily automated, a priority of definition needs to be established.

Many words represent simple names (nouns, pronouns), measurements (adjectives and some adverbs), states (some verbs), or actions (other verbs): they label a stereotypical observation. Prepositions, interjections, conjunctions, and some adverbs are grammatical concepts, relating concepts together at a higher level of abstraction: these words capture the operations §3.2.4 that can be performed on sets of concepts.

Language and grammar (§3.2.4.8), in Syntheta's model, are represented by a Semantic categorization Module sitting atop other Semantic Modules. The first level of abstraction allows a mind to go beyond the simple labeling of primary observations in order to be able to describe their interrelationships. There are several kinds of relationships that can be captured among groupings of primary concepts.

Some semantic concepts are "static" within the language, such as sensory sets (*e.g.*, a red light at an intersection, versus a green light), affective sets (grouping emotions with sensory information, such as a purring housecat implying contentment), and executive sets (grouping actions with their consequences, such as eating food leading to satiation). Although these groupings (schema, captured from the current context) are much more open-ended than is a simple list of words in a dictionary, they are still based on a finite set of primary concepts. As people learn, we build such semantic concepts by association. Syntheta should learn in the same way, but our impatience for it to learn could be mitigated by feeding it some limited number of premade sets, where warranted. Although it would be far too burdensome to spoonfeed all such associative relationships to Syntheta, some such semantic sets, prioritized for the domain, could give it a jumpstart.

Besides "static" semantic sets conveying special meanings, semantic sequences (as bigram sequences) can also have meanings of their own. Without such idioms, it would be unsurprising for a hot dog to pant, but it would be quite surprising for a hot dog to be served with ketchup. The number of such idiomatic sequences needing to be captured is large for full fluency, but not necessarily so large for basic language comprehension. The initial aim of this proof of concept does not need to reach metaphorical competence; that must probably await Syntheta's more existential embodiments (§5.2).

A good analogy with which to illustrate semantic sequences is speech [428]. Although phonemes are discrete symbolic entities that can be catalogued, in practice they are fluidly coarticulated during speech production, minding the need that they be readily parsed during speech comprehension [429]. Arguably, bigram sequences of the evolving shapes of speech (see §3.3.4.4) can provide added information that can help to extract the signal from the noise. Static semantic concepts are like phonemes, that blend into one another with each passing moment in an evolving wave of meaning, whose transformations, like in speech, provide added value. Words themselves can be defined in this way, the definition being a semantic wave that imparts meaning to the word, just as a wave of coarticulated phonemes identifies a word. The extent of a wave would depend on the Symbol activity decay rate (§3.5), and the polarization resulting from the wave's leading and trailing edges would enable the who-does-what-to-whom paradigm—grammar, guided of course by grammatical function words gating the parts of sentences into their appropriate bins.

Describing grammar explicitly, for a language such as English, is fraught with difficulty, requiring considerable education. However, infants, having no advanced degrees, pick it up quite naturally. There is a disconnect between the academic treatment of grammar (*e.g.*, see [430]) and the practical intuition of grammar. Teaching grammar is difficult, but learning grammar is easy. At the root of a language such as English, there is the typical grammatical construct SVO (subject–verb–object), but all typologies (SOV, SVO, VSO, VOS, OVS, OSV) exist in the world's languages. What is salient here is that something acts on something (SVO), or something on something acts (SOV), or something was acted upon (by something) (OV(S)), etc. For grammar's top level, there are things (nouns) and actions (verbs); the rest are mere embellishments. If a noun phrase, for example, can be reduced to a noun, then it is a noun. If the top level of grammar is given only two concepts, nouns and verbs, then it can learn any of the six possible permutations of {S, V, O}. As the words stream through working memory, the semantic set being generated can be noun-like or verb-like, with nouns being things (in general) and with verbs being actions (in general). *The striped cat eagerly chased the frightened mouse* invokes nounness then verbness then nounness, which, being English and bearing instructive inflections, tells who is doing what to whom. Another typology might state: *Eagerly chased the frightened mouse, the striped cat did* (VOS), which is still perfectly recognizable as a valid (though atypical) grammatical construct in English, thanks to its embedded grammatical and inflectional markers.

Grammar is layered, in Syntheta's model, such that the layer knowing only about nouns and verbs is fed by a lower layer knowing about the various grammatical categories including function words, constructing the schema for the subject, verb and object, sometimes recursively: Sentences are often, in fact, depicted as trees. A semantic operator can, for example, suppress the verbness of a verb, such as in *The striped cat that entered the room eagerly chased the frightened mouse*, where the operator *that* suppresses *entered* from being the verb. Other semantic operators have others roles, as in *The striped cat stealthily entered the room then eagerly chased the frightened mouse*, where *then* indicates an ordered sequence of actions. The combinations may be staggering, but they all boil down to *Cat chased mouse*: the rest of the words decorate the event with detail and drama. The number of semantic operators must also be finite, and should be able to be gleaned from the world's set of languages. All of this is too difficult to teach Syntheta, but Syntheta should easily be able to learn it, given the constructs that it possesses. Language acquisition is an autodidactic process, accelerated by education for sure, but not depending upon education.

Hierarchy in language is important, as it is in Syntheta's model. Some concepts are necessarily hierarchical. An adjective-plus-noun at one level is a noun at a higher level. A cat at one level is a mammal at a higher level, as is a mouse. A snake is not a mammal, but it is an animal, at an even higher rung in the hierarchy, as are cats and mice. This information can be learned, but can also easily be taught.

Syntheta's acquisition of language and grammar cannot rely on explicit definitions for every possible word permutation, nor can our own acquisition of language and grammar. Ordinary strings of words or compounds are expected to individually stimulate their associated concepts, creating a wave of context in which subsequent words are then interpreted (when a word has multiple meanings) and in which the collective of stimulations triggers higher-order abstractions §3.2.4 which can then convey the desired message.

In any case, the work of constructing Syntheta's language dictionary involves no more effort than that of filling in a matrix, with words and other semantic concepts on rows, and Symbols from Modules on columns, that we people, knowing these words and semantic concepts, can reasonably rough in. There is no need to be perfect. People themselves make many misinterpretations, and get by. Accuracy is better than inaccuracy, but the bias in interpretation is largely forgiving, and is ultimately self-correcting. Missing information is tolerated. Syntheta can fill in the gaps and correct misinformation, as it reads and learns, and best of all (§5.2), experiences.

Several approaches can be considered in selecting words for Syntheta's initial sensory-mapped dictionary. Words that are most often used used in the English language (*e.g.*, [431]) may be good candidates, but a more strategic choice could be to focus on defining-vocabularies [432] including semantic primes [433], often presented as a layered progression of words (*e.g.*, David Bullock's "Learn These Words First" website [434]). That would enable Syntheta to leverage specially structured dictionaries (such as the Longman Dictionary of Contemporary English [435]) in order to enrich its vocabulary, once it understands the words that are used in defining other words. Additionally, some words, neither common nor strategic, are simply easy to define in terms of their meaning; they can be cheaply added to the table as well (such as *magenta*). Finally, common idioms and named semantic concepts (§3.2.4, §3.1; also see §6.2.2) can complete this inital table, providing a core level of language competence to Syntheta.

Nothing, however, is ever as easy as it might appear at first glance. Although Bullock's approach to the learning of English [434] is laudable, it is clearly labeled as a "Multi-Layer Dictionary for Second-Language Learners of English", with "Second-Language" perhaps being key. Semantic concepts must already exist in the student's mind to which he or she can attach the English words. It is difficult to know how infants and toddlers learn their primary language, although some case studies offer clues: Helen Keller's account of how she learned about language when she was six years old is particularly illuminating, and poignantly illustrates how she learned to understand to attach words to concepts that she had already formed in her mind [436], seeding her journey into literacy. All that we can realistically do for Syntheta, pending its actively experiencing the world (§5.2) is to rough-in some concepts.

Following this upload of knowledge into Syntheta's mind by Trainer (§4.1.2), the querying of Syntheta's memory can begin (where one wants to use Syntheta for text mining). The UTF-8 code points fed to Syntheta by Trainer in response to a user-supplied query can stimulate (as bigrams and gappy bigrams) the best-matching word, which then stimulates its set of primary concepts, which then stimulate their own associated semantic concepts by degree of shared membership, including words, which then stimulate higher-order semantic wave concepts (see §3.2.3.2), which stimulate all of their entries in the Memory Trace, which anticipate later entries using a stimulatory forward sweep, that finally results in loci within the Memory Trace to be more activated than others, and reported back to the user (Fig. 29).

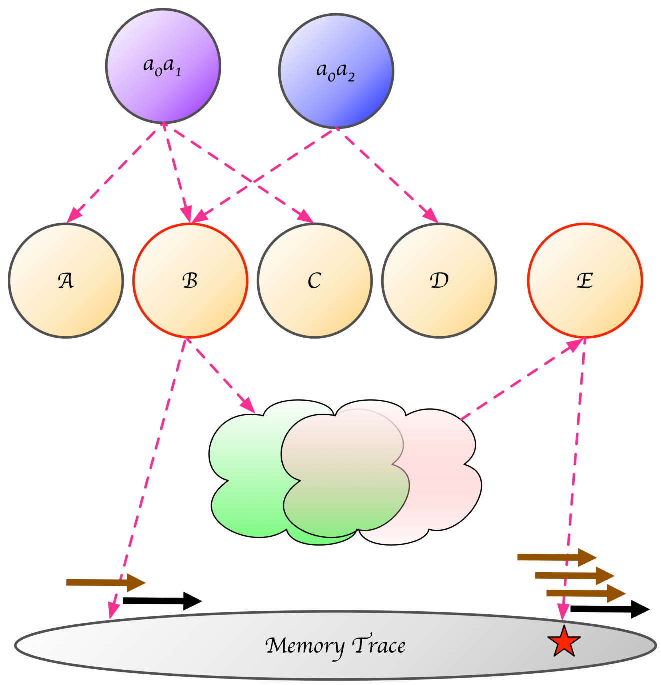

Figure 29: In this semantic proof of concept, bigrams and gappy bigrams stimulate words, cumulatively, which are related thanks to their shared concept clouds. The words stimulate the Memory Trace, cumulatively and specifically via higher-order semantic wave concepts, resulting in a most-excited Engram to be reported back to the user.

5.4 Budget

Implementing Syntheta's core model requires an understanding of evolutionary theory (as argued throughout this treatise), basic neural network architecture, and the biological principles of feedback regulation. These domains of knowledge are tractable to a professional computational biologist and backend developer such as myself. Robotics, 3D graphics, and the subtler aspects of digital signal processing, on the other hand, represent different masteries, towards which I could, at best, provide amateur solutions. These latter solutions can therefore be considered as placeholder implementations in the code, needing professional attention, should the community develop an interest in this project. In some cases, there was no attempt at implementation, but given Syntheta's modular design, the missing functionality can simply be plugged in when it becomes available, or swapped in, if available but suboptimal.

Contributions from computer scientists, computer graphics professionals, software engineers, artists, roboticists, kinesiologists, linguists, neuroscientists, psychologists, teachers, mathematicians, statisticians and others would therefore help to complete Syntheta's implementation towards stronger proofs of concept. It is hoped that such a collaboration can be voluntary, managed via a GitHub repository (§4.1.5). In terms of computing infrastructure, a modern laptop or desktop computer, accessible to many developers, should suffice. Although some applications (§3.9) may benefit from workstation-class equipment, Syntheta's core model (§3.1) aims to be lightweight and fast (see §5.3).

Collaborative development on platforms such as GitHub, GitLab and others has become increasingly important for producing high-quality software (*e.g.*, [437,438]). Such code needs to be actively maintained, following some consistent coding standard. Users and developers alike need support, and benefit from clear and complete documentation. A project needs a strong vision, a corresponding mission statement and project plan, ongoing commitment, and faith. It needs a project manager to promote the project, to recruit codevelopers and to attract a user base.

In the absence of voluntary contributions from the community, expertise for hire should still not amount to a large expenditure: An expert, by definition, is already well-versed in the subject at hand and should be able to complete the implementation and testing of their respective module within a few months—rocket science is ordinary work for a rocket scientist. Most of the plan's effort would then consist of executing and documenting the developmental progression described for Syntheta in §5.2 and §5.3, including iterative fixes to Syntheta's model, its code, and its parameters wherever Syntheta's progress may stall. Again, community involvement can help, with multiple instances of Syntheta exploring different parameters in competition, all coordinated by the project's senior architect. In summary, little budget is required for this project, but funds can help to focus efforts—by employing full-time developers. A "dream team", inspired by Brooks [439] is illustrated in Fig. 30. Again, project management is key.

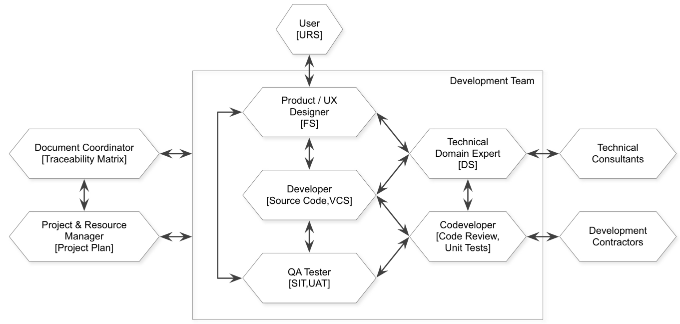

Figure 30: A development team with an arguably optimal division of duties.

5.4.1 Staged sharing

Understanding any substantial code base may be an overwhelming proposition to a new contributor, raising the activation energy needed for their involvement. A hundred thousand lines of code may not seem like much to implement a complete artificial general intelligence, but it is much for a contributor to immediately grok.

In keeping with the evolutionary theme of this treatise, the code repository can be pared down to what is essential for the first stage of Syntheta's progressive development (§5.2), then built upon as required for each of Syntheta's planned milestones. By not oversharing, distractions to a developmental iteration's focus can be minimized. Syntheta's abstractions and core processes can be highlighted by such a paring, and then the duplicate-and-diverge paradigm for new features can be better appreciated for what they add. From a contributor's viewpoint, three bins of source code files may be: *existing*, for bug fixes and optimizations; *new*, for active development; and *next*, for conceptual sketches of what's to come.

5.4.2 "What's in it for *me*?"

Contributing effort to some stranger's project is most often met with apathy [440]. Although we humans are primates, we do not typically gain one another's favours the usual primate way [101], by social grooming... Instead, we can make friends at conferences, sometimes over drinks or meals, and cultivate those friendships into fruitful collaborations. Building such networks can take time, however, and may not always succeed, due to the aforementioned tendency for prosocial apathy. Although it may be cynical, my view is that a person will not be a productive collaborator unless it is in their own selfish interest: the benefit to them has to clearly outweigh the cost to them. This requirement can be met in different ways:

- You pay them (and hope that this engages them).
- They pay you to accomplish something that they need done (and you hope that what they need from you aligns with your interests).
- They fund you as a business investment (and you hope that their business objectives can derive from your scientific objectives).
- They fund you as a patron (and you hope that their trust in you is maintained).
- You mentor them as a graduate student advisor or as a postdoctoral advisor (and you hope that this teaching effort pays off for both of you).
- They mentor you as a graduate student advisor or as a postdoctoral advisor (and you hope to benefit both from their mentorship as well as from their lab's collegiality).
- You provide and support free tools for them to use, tailored to *their* needs (and hope that they provide feedback).
- You offer to work on *their* hypotheses and theories with them, and help them to write grant applications and to publish papers (and in so doing, hope to further your own projects).
- You showcase *their* platform within your own (and hope that they will help support their part).

I can, in this vein, very much imagine myself wanting to collaborate with several neuroscientists to further their own work, because I am interested in their work and because it can strengthen Syntheta's model. Few AI platforms can guide psychological experiments to test cognitive hypotheses; Syntheta hopes to be such a platform (§3.5.2.4). Its *Trainer* application's built-in genetic algorithm functionality can perform a battery of tests on Syntheta all from the very same starting point (see §4.1.2), deterministically and thus perfectly controlled. It should thus be an attractive proposition to offer Syntheta to prospective collaborators to help them with their own projects. Still, there is always that social barrier to overcome, sometimes a challenge for introverts [261] like me, and so the social networking step cannot be left out.

6 Equipment and methods

6.1 Equipment

The proofs-of-concept, described in §5.2 and §5.3, were hosted on a 16-inch MacBook Pro from 2021, equipped with an Apple M1 Max processor, 64 GB of memory, and a 4 TB solid-state drive, running macOS Monterey 12.6.7. Apple, however, no longer supports OpenGL, so its operating system is useless in running OpenSceneGraph [336], needed to implement the robot's initial virtual environment. The robot was therefore instantiated on a 2022 System76 Pangolin laptop with 16 GB of memory and a 500 GB solid-state drive, running Ubuntu 22.04, and connected to Syntheta on the MacBook Pro as illustrated in Fig. 25.

6.2 Method notes

6.2.1 Autonomous robots

For this proof of concept, two similar robots were constructed, named Synalpha and Synphi. Both independent instances of Syntheta communicated with its robot as illustrated in Fig. 25. At the appropriate stage in each robot's developmental programme (§5.2), language (see §5.3 and §5.2.5) was merged with the robot's experiential learnings using Trainer (§4.1.2).

...more needed here...

6.2.2 Natural language understanding

A file of text converted from PDF includes a number of artifacts: journal headers and footers, layout artifacts including tables and figures interrupting sentences (and even sometimes words!), optical character recognition errors (especially with italicized text, subscripts, superscripts, special fonts and Unicode characters), and where a PDF captures a page of a journal instead of a discrete article, irrelevant text from another article preceding and/or following the article of interest. Another problem with digitized text is hyphenation, meant to justify text optimally within the narrow columns of a journal, but interfering to a degree with machine readability by splitting words. Of course and in addition, many of a paper's tables and equations, and all of its figures, are lost in a textual rendition of a PDF; reading a paper the way it was intended to be read must await Syntheta's visual senses to come online (§5.2).

Although these are serious issues, many informative strings of content still do exist in uncurated text, and can be mined. Some opportunities are lost, because of the corruptions and interruptions, but authors often repeat themselves in a paper, to drive points home for the reader. Some of the artifacts have distinctive patterns that can be filtered out, either explicitly [399] or using machine learning approaches, mitigating some of the PDF-to-text artifacts. The availability of many hyphenated words in the WordNet 3.1 database can be leveraged as well, in order to determine whether or not an end-of-line hyphen should be retained or deleted. The corpus itself can also be mined for hyphenated words internal to a line of text [441]. Other PDF-to-text issues can be curated as required, for reference documents deemed more important.

Trainer streamed UTF-8-encoded words to Syntheta's *Word* Sensory Module, rather than individual UTF-8 code points. It also streamed the text's named-concept publication identifier to Syntheta's *Source* Sensory Module. Words were parsed from the code points by Trainer by detecting white-space or punctuation in the stream. Both the words and the publication identifiers were fed to the *History* Memory Trace Module. This change in strategy resulted in a leaner Memory Trace and faster interprocess communication, while relegating code point-derived bigrams and gappy bigrams to being mere compositional members of each word's semantic concept cloud.

Trainer can do more than simply feed a stream to Syntheta: it can also, through Syntheta's API, wire Syntheta's concepts together. Trainer was therefore tasked with requesting that all of the words that it had streamed be associated with their bigram and gappy bigram components, but also much more: to associate the words in Trainer's concept-cloud dictionary (see §5.3.3) with those concept clouds as well. Normally, a Memory Trace includes context, but here, the word symbols capture that context themselves, permitting a very lean Memory Trace (and hence, a bigger corpus to be held).

Finally, UTF-8 *code points*, not words, were fed to Syntheta for querying its knowledge base, enabling the kind of sloppiness in typing that a user should be able to get away with. These code points, then, would select their best concept cloud(s), organized into higher-order semantic concept clouds (as described in §5.3), leading back to a stimulation of the Memory Trace, and therefore passages of text returned to the user.

7 Acknowledgments

Ideas come from everywhere, and are acquired both deliberately and serendipitously. A number of books, articles and web pages, only a few of which are cited in the bibliography, contributed insights towards Syntheta's design. Apologies go to those whose work was not cited: the corpus of studies on intelligence is too vast, only permitting hooks into each subdomain's literature, despite one's best efforts.

Among the people deserving special mention, I thank Colin Hurman, Kathryn MacHattie and William Mitchell for early support; and Paul Clarke, Elisabeth Tillier and Scott Charlebois for helpful discussions.

8 References

[1] "Syntheta GitHub public repository." https://github.com/neurogadgets/Syntheta. Accessed: 2021-12-24.

[2] G.-W. Evans, "Artificial intelligence: Where we came from, where we are now, and where we are going," Master's thesis, University of Victoria, 2017.

[3] D. R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*. Vintage Books, 1979.

[4] D. L. Coutu, "Strategic intensity: a conversation with world chess champion Garry Kasparov," *Harv. Bus. Rev.*, vol. 83, pp. 49–53, 2005.

[5] S. Armstrong and K. Sotala, "How we're predicting AI—or failing to," in *Beyond AI: Artificial Dreams* (J. Romportl, P. Ircing, E. Zackova, M. Polak, and R. Schuster, eds.), pp. 52–75, 2012.

[6] J. C. Hawkins, *A Thousand Brains*. Basic Books, 2021.

[7] J. Mervis, "Top Democrat wants big AI push," *Science*, vol. 366, p. 787, 2019.

[8] G. Anders, "Alexa, understand me," *MIT Technology Review*, vol. September/October, 2017.

[9] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene, "Opportunities and obstacles for deep learning in biology and medicine," *J. R. Soc. Interface*, vol. 15, p. 20170387, 2018.

[10] J. Goecks, V. Jalili, L. M. Heiser, and J. W. Gray, "How machine learning will transform biomedicine," *Cell*, vol. 181, pp. 92–101, 2020.

[11] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, and L. Shi, "Towards artificial general intelligence with hybrid Tianjic chip architecture," *Nature*, vol. 572, pp. 106–111, 2019.

[12] J. E. LeDoux, *The Deep History of Ourselves*. Viking, 2019.

[13] R. R. Netz and W. A. Eaton, "Estimating computational limits on theoretical descriptions of biological cells," *Proc. Natl. Acad. Sci. USA*, vol. 118, p. e2022753118, 2021.

[14] F. Zeldenrust, P. Chameau, and W. J. Wadman, "Spike and burst coding in thalamocortical relay cells," *PLoS Comput. Biol.*, vol. 14, p. e1005960, 2018.

[15] N. A. Ratan Murty and S. P. Arun, "Multiplicative mixing of object identity and image attributes in single inferior temporal neurons," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. E3276–E3285, 2018.

[16] Y. Kwak and C. E. Curtis, "Unveiling the abstract format of mnemonic representations," *Neuron*, vol. 110, pp. 1822–1828, 2022.

[17] J. Hawkins, M. Lewis, M. Klukas, S. Purdy, and S. Ahmad, "A framework for intelligence and cortical function based on grid cells in the neocortex," *Front. Neural Circuits*, vol. 12, p. 121, 2019.

[18] M. Lewis, S. Purdy, S. Ahmad, and J. Hawkins, "Locations in the neocortex: A theory of sensorimotor object recognition using cortical grid cells," *Front. Neural Circuits*, vol. 13, p. 22, 2019.

[19] G. J. Stuart and N. Spruston, "Dendritic integration: 60 years of progress," *Nature Neurosci.*, vol. 18, pp. 1713–1721, 2015.

[20] Y. Shen, S. Dasgupta, and S. Navlakha, "Habituation as a neural algorithm for online odor discrimination," *Proc. Natl. Acad. Sci. USA*, vol. 117, pp. 12402–12410, 2020.

[21] O. G. Selfridge, "Pandemonium: a paradigm for learning," in *Proc. Symp. Mechanisation of Thought Processes, Natl. Physical Lab., Teddington, England*, pp. 513–526, 1958.

[22] J. V. Jackson, "Idea for a mind," *SIGART Newletter*, vol. 101, pp. 23–26, July 1987.

[23] J. D. Barrow, *The Book of Nothing*. Random House, 2001.

[24] Y. Zheng, X. L. Liu, S. Nishiyama, C. Ranganath, and R. C. O'Reilly, "Correcting the hebbian mistake: Toward a fully error-driven hippocampus," *PLoS Comput. Biol.*, vol. 18, p. e1010589, 2022.

[25] J. Hawkins and S. Ahmad, "Why neurons have thousands of synapses, a theory of sequence memory in neocortex," *Front. Neural Circuits*, vol. 10, p. 23, 2016.

[26] J. F. A. Poulet and B. Hedwig, "A corollary discharge maintains auditory sensitivity during sound production," *Nature*, vol. 418, pp. 872–876, 2002.

[27] J. X. Brooks and K. E. Cullen, "Predictive sensing: The role of motor signals in sensory processing," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 4, pp. 842–850, 2019.

[28] A. L. Person, "Corollary discharge signals in the cerebellum," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 4, pp. 813–819, 2019.

[29] D. Subramanian, A. Alers, and M. A. Sommer, "Corollary discharge for action and cognition," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 4, pp. 782–790, 2019.

[30] A. Yokoi and J. Diedrichsen, "Neural organization of hierarchical motor sequence representations in the human neocortex," *Neuron*, vol. 103, pp. 1178–1190, 2019.

[31] M. Quik, D. Zhang, X. A. Perez, and T. Bordia, "Role for the nicotinic cholinergic system in movement disorders; therapeutic implications," *Pharmacol Ther.*, vol. 144, pp. 50–59, 2014.

[32] K. Z. Tanaka, H. He, A. Tomar, K. Niisato, A. J. Y. Huang, and T. J. McHugh, "The hippocampal engram maps experience but not place," *Science*, vol. 361, pp. 392–397, 2018.

[33] M. D. Grilli and M. Verfaellie, "Personal semantic memory: Insights from neuropsychological research on amnesia," *Neuropsychologia*, vol. 61, pp. 56–64, 2014.

[34] A. Staniloiu and H. J. Markowitsch, "The remains of the day in dissociative amnesia," *Brain Sci.*, vol. 2, pp. 101–109, 2012.

[35] M. J. Dorahy, B. L. Brand, V. Şar, C. Krüger, P. Stavropoulos, A. Martínez-Taboas, R. Lewis-Fernández, and W. Middleton, "Dissociative identity disorder: An empirical overview," *Aust. N. Z. J. Psychiatry*, vol. 48, pp. 402–417, 2014.

[36] B. Bhattarai, J. W. Lee, and M. W. Jung, "Distinct effects of reward and navigation history on hippocampal forward and reverse replays," *Proc. Natl. Acad. Sci. USA*, vol. 117, pp. 689–697, 2020.

[37] X. Sun, M. J. Bernstein, M. Meng, S. Rao, A. T. Sørensen, L. Yao, X. Zhang, P. O. Anikeeva, and Y. Lin, "Functionally distinct neuronal ensembles within the memory engram," *Cell*, vol. 181, pp. 410–423, 2020.

[38] R. Quian Quiroga, *The Forgetting Machine*. BenBella Books, Inc., 2017.

[39] K. G. Akers, A. Martinez-Canabal, L. Restivo, A. P. Yiu, A. De Cristofaro, H.-L. Hsiang, A. L. Wheeler, A. Guskjolen, Y. Niibori, H. Shoji, K. Ohira, B. A. Richards, T. Miyakawa, S. A. Josselyn, and P. W. Frankland, "Hippocampal neurogenesis regulates forgetting during adulthood and infancy," *Science*, vol. 344, pp. 598–602, 2014.

[40] B. P. Staresina and M. Wimber, "A neural chronometry of memory recall," *Trends Cogn. Sci.*, vol. 23, pp. 1071–1085, 2019.

[41] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. 13521–3526, 2017.

[42] G. Buzsáki, "Neural syntax: Cell assemblies, synapsembles, and readers," *Neuron*, vol. 68, pp. 362–385, 2010.

[43] C. E. Schoonover, S. N. Ohashi, R. Axel, and A. J. P. Fink, "Representational drift in primary olfactory cortex," *Nature*, vol. 594, pp. 541–546, 2021.

[44] Q. Wang and C. Peterson, "The fate of childhood memories: Children postdated their earliest memories as they grew older," *Frontiers Psychol.*, vol. 6, p. 2038, 2016.

[45] A. Guskjolen, J. W. Kenney, J. de la Parra, B. A. Yeung, S. A. Josselyn, and P. W. Frankland, "Recovery of "lost" infant memories in mice," *Curr. Biol.*, vol. 28, pp. 2283–2290, 2018.

[46] D. Kahneman, *Thinking, Fast and Slow*. Doubleday Canada, 2011.

[47] G. Buzsáki, *The Brain from Inside Out*. Oxford University Press, 2019.

[48] S. Dehaene, *How We Learn*. Viking, 2020.

[49] R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, pp. 1102–1107, 2005.

[50] R. Quian Quiroga, "Searching for the neural correlates of human intelligence," *Curr. Biol.*, vol. 30, pp. R335–R338, 2020.

[51] M. Schafer and D. Schiller, "Navigating social space," *Neuron*, vol. 100, pp. 476–489, 2018.

[52] T. E. J. Behrens, T. H. Muller, J. C. R. Whittington, S. Mark, A. B. Baram, K. L. Stachenfeld, and Z. Kurth-Nelson, "What is a cognitive map? Organizing knowledge for flexible behavior," *Neuron*, vol. 100, pp. 490–509, 2018.

[53] Y. Liu, R. J. Dolan, Z. Kurth-Nelson, and T. E. Behrens, "Human replay spontaneously reorganizes experience," *Cell*, vol. 178, pp. 640–652, 2019.

[54] A. Gomez-Marin and A. A. Ghazanfar, "The life of behavior," *Neuron*, vol. 104, pp. 25–36, 2019.

[55] G. Cui, S. B. Jun, X. Jin, M. D. Pham, S. S. Vogel, D. M. Lovinger, and R. M. Costa, "Concurrent activation of striatal direct and indirect pathways during action initiation," *Nature*, vol. 494, pp. 238–242, 2013.

[56] D. Badre, *On Task*. Princeton University Press, 2020.

[57] C. F. Lavallee, M. T. Meemken, C. S. Herrmann, and R. J. Huster, "When holding your horses meets the deer in the headlights: time-frequency characteristics of global and selective stopping under conditions of proactive and reactive control," *Front. Hum. Neurosci.*, vol. 8, p. 994, 2014.

[58] M. T. Banich, K. L. Mackiewicz Seghete, B. E. Depue, and G. C. Burgess, "Multiple modes of clearing one's mind of current thoughts: Overlapping and distinct neural systems," *Neuropsychologia*, vol. 69, pp. 105–117, 2015.

[59] H. Kim, H. R. Smolker, L. L. Smith, M. T. Banich, and J. A. Lewis-Peacock, "Changes to information in working memory depend on distinct removal operations," *Nat. Commun.*, vol. 11, p. 6239, 2020.

[60] T. Amer, K. L. Campbell, and L. Hasher, "Cognitive control as a double-edged sword," *Trends Cogn. Sci.*, vol. 20, pp. 905–915, 2016.

[61] N. P. Friedman, A. Miyake, S. E. Young, J. C. DeFries, R. P. Corley, and J. K. Hewitt, "Individual differences in executive functions are almost entirely genetic in origin," *J. Exp. Psychol. Gen.*, vol. 137, pp. 201–225, 2008.

[62] X. Jiang, T. Long, W. Cao, J. Li, S. Dehaene, and L. Wang, "Production of supra-regular spatial sequences by macaque monkeys," *Curr. Biol.*, vol. 28, pp. 1851–1859, 2018.

[63] M. Chakraborty and E. D. Jarvis, "Brain evolution by brain pathway duplication," *Phil. Trans. R. Soc. B*, vol. 370, p. 20150056, 2015.

[64] M. Koenigs, D. Acheson, A. Barbey, J. Solomon, B. R. Postle, and J. Grafman, "Areas of left perisylvian cortex mediate auditory-verbal short-term memory," *Neuropsychologia*, vol. 49, pp. 3612–3619, 2011.

[65] S. Clark, *A Practical Grammar, revised edition.* A.S. Barnes & Burr., 1863.

[66] E. Mayfield, "Sentence diagram generation using dependency parsing," *Proceedings of the ACL-IJCNLP 2009 Student Research Workshop*, pp. 45–53, 2009.

[67] M. Kolln and R. Funk, *Understanding English Grammar, Ninth Edition.* Pearson, 2012.

[68] D. Crystal, *A Dictionary of Linguistics and Phonetics, 6th Edition.* Blackwell Publishing, 2008.

[69] V. Valian and S. Coulson, "Anchor points in language learning: The role of marker frequency," *J. Mem. Lang.*, vol. 27, pp. 71–86, 1988.

[70] C. Marino, C. Bernard, and J. Gervain, "Word frequency is a cue to lexical category for 8-month-old infants," *Curr. Biol.*, vol. 30, pp. 1380–1386, 2020.

[71] G. Grand, I. A. Blank, F. Pereira, and E. Fedorenko, "Semantic projection recovers rich human knowledge of multiple object features from word embeddings," *Nat. Hum. Behav.*, vol. 6, pp. 975–987, 2022.

[72] J. Read, R. Dridan, S. Oepen, and L. J. Solberg, "Sentence boundary detection: A long solved problem?," in *Proceedings of COLING 2012: Posters*, pp. 985–994, 2012.

[73] T. Grandin, *Thinking in Pictures, 2nd edition.* Vintage Books, 2006.

[74] M. H. Christiansen and N. Chater, *The Language Game.* Basic Books, 2022.

[75] M. Amalric and S. Dehaene, "Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks," *Phil. Trans. R. Soc. B*, vol. 373, p. 201605154, 2017.

[76] M. Amalric and S. Dehaene, "Origins of the brain networks for advanced mathematics in expert mathematicians," *Proc. Natl. Acad. Sci. USA*, vol. 113, pp. 4909–4917, 2016.

[77] A. Tversky and D. Kahneman, "Judgment under uncertainty: heuristics and biases," *Science*, vol. 185, pp. 1124–1131, 1974.

[78] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice," *Science*, vol. 211, pp. 453–458, 1981.

[79] M. G. Haselton, D. Nettle, and D. R. Murray, "The evolution of cognitive bias," in *Handbook of Evolutionary Psychology, Second Edition* (D. M. Buss, ed.), pp. 968–987, John Wiley & Sons, Inc., 2016.

[80] K. Weill, *Off the Edge.* Algonquin Books of Chapel Hill, 2022.

[81] I. Király, K. Oláh, G. Csibra, and Á. M. Kovács, "Retrospective attribution of false beliefs in 3-year-old children," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 11477–11482, 2018.

[82] M. M. Nour, T. Dahoun, P. Schwartenbeck, R. A. Adams, T. H. B. FitzGerald, C. Coello, M. B. Wall, R. J. Dolan, and O. D. Howes, "Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. E10167–E10176, 2018.

[83] N. Jung, C. Wranke, K. Hamburger, and M. Knauff, "How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety," *Frontiers Psychol.*, vol. 5, p. 570, 2014.

[84] A. K. Seth, "Interoceptive inference, emotion, and the embodied self," *Trends Cogn. Sci.*, vol. 17, pp. 565–573, 2013.

[85] J. E. LeDoux and R. Brown, "A higher-order theory of emotional consciousness," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. E2016–E2025, 2017.

[86] L. F. Barrett, *How Emotions Are Made.* First Mariner Books, 2017.

[87] J. Panksepp, "Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals," *PLoS ONE*, vol. 6, p. e21236, 2011.

[88] A. S. Cowen and D. Keltner, "Self-report captures 27 distinct categories of emotion bridged by continuous gradients," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. E7900–E7909, 2017.

[89] H. Lövheim, "A new three-dimensional model for emotions and monoamine neurotransmitters," *Med. Hypotheses*, vol. 78, pp. 341–348, 2012.

[90] R. R. Griffiths, W. A. Richards, U. McCann, and R. Jesse, "Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance," *Psychopharmacology*, vol. 187, pp. 268–283, 2006.

[91] I. O. Ebrahim, R. S. Howard, M. D. Kopelman, M. K. Sharief, and A. J. Williams, "The hypocretin/orexin system," *J. Royal Soc. Med.*, pp. 227–230, 2002.

[92] S. C. Holst and H.-P. Landolt, "Sleep-wake neurochemistry," *Sleep Med. Clin.*, vol. 13, pp. 137–146, 2018.

[93] G. P. Krishnan, S. Chauvette, I. Shamie, S. Soltani, I. Timofeev, S. S. Cash, E. Halgren, and M. Bazhenov, "Cellular and neurochemical basis of sleep stages in the thalamocortical network," *eLife*, vol. 5, p. e18607, 2016.

[94] M. R. Picciotto, M. J. Higley, and Y. S. Mineur, "Acetylcholine as a neuromodulator: Cholinergic signaling shapes nervous system function and behavior," *Neuron*, vol. 76, pp. 116–129, 2012.

[95] O. A. C. Petroff, "GABA and glutamate in the human brain," *The Neuroscientist*, vol. 8, pp. 562–573, 2002.

[96] T. W. Schmitz, M. M. Correia, C. S. Ferreira, A. P. Prescot, and M. C. Anderson, "Hippocampal GABA enables inhibitory control over unwanted thoughts," *Nature Comm.*, vol. 8, p. 1311, 2017.

[97] D. Z. Lieberman and M. E. Long, *The Molecule of More.* BenBella Books Inc., 2018.

[98] J. C. Jackson, J. Watts, T. R. Henry, J.-M. List, R. Forkel, P. J. Mucha, S. J. Greenhill, R. D. Gray, and K. A. Lindquist, "Emotion semantics show both cultural variation and universal structure," *Science*, vol. 366, pp. 1517–1522, 2019.

[99] L. Nummenmaa, R. Hari, J. K. Hietanen, and E. Glerean, "Maps of subjective feelings," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 9198–9203, 2018.

[100] L. Nummenmaa, E. Glerean, R. Hari, and J. K. Hietanen, "Bodily maps of emotions," *Proc. Natl. Acad. Sci. USA*, vol. 111, pp. 646–651, 2014.

[101] F. de Waal, *The Bonobo and the Atheist.* W. W. Norton & Company, Inc., 2013.

[102] K. Dutton, *The Wisdom of Psychopaths.* Doubleday Canada, 2012.

[103] J. F. Schulz, D. Bahrami-Rad, J. P. Beauchamp, and J. Henrich, "The Church, intensive kinship, and global psychological variation," *Science*, vol. 366, p. eaau5141, 2019.

[104] I. Tamarit, J. A. Cuesta, R. I. M. Dunbar, and A. Sánchez, "Cognitive resource allocation determines the organization of personal networks," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 8316–8321, 2018.

[105] M. Iacoboni, "Imitation, empathy, and mirror neurons," *Annu. Rev. Psychol.*, vol. 60, pp. 653–670, 2009.

[106] A. Cruz, M. Heinemans, C. Márquez, and M. A. Moita, "Freezing displayed by others is a learned cue of danger resulting from co-experiencing own freezing and shock," *Curr. Biol.*, vol. 30, pp. 1128–1135, 2020.

[107] P. Bloom, "Empathy and its discontents," *Trends Cogn. Sci.*, vol. 21, pp. 24–31, 2017.

[108] J. T. Lanzetta and B. G. Englis, "Expectations of cooperation and competition and their effects on observers' vicarious emotional responses," *J. Pers. Soc. Psychol.*, vol. 56, pp. 543–554, 1989.

[109] G. Hein, G. Silani, K. Preuschoff, C. D. Batson, and T. Singer, "Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping," *Neuron*, vol. 68, pp. 149–160, 2010.

[110] R. M. Sapolksy, *Behave: The biology of humans at our best and worst.* Penguin Books, 2017.

[111] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer, "Multidimensional optimality of microbial metabolism," *Science*, vol. 336, pp. 601–604, 2012.

[112] F. Jacob, "Evolution and tinkering," *Science*, vol. 196, pp. 1161–1166, 1977.

[113] A. Booth and W. F. Doolittle, "Eukaryogenesis, how special really?," *Proc. Natl. Acad. Sci. USA*, vol. 112, pp. 10278–10285, 2015.

[114] Q. Dietschi, J. Tuberosa, L. Rösingh, G. Loichot, M. Ruedi, A. Carleton, and I. Rodriguez, "Evolution of immune chemoreceptors into sensors of the outside world," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. 7397–7402, 2017.

[115] H. H. McAdams, B. Srinivasan, and A. P. Arkin, "The evolution of genetic regulatory systems in bacteria," *Nat. Rev. Genet.*, vol. 5, pp. 169–178, 2004.

[116] Y. S. Dufour, X. Fu, L. Hernandez-Nunez, and T. Emonet, "Limits of feedback control in bacterial chemotaxis," *PLoS Comput. Biol.*, vol. 10, p. e1003694, 2014.

[117] C. I. Bargmann, "Comparative chemosensation from receptors to ecology," *Nature*, vol. 444, pp. 295–301, 2006.

[118] R. Thar and M. Kühl, "Bacteria are not too small for spatial sensing of chemical gradients: An experimental evidence," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 5748–5753, 2003.

[119] C. Bushdid, M. O. Magnasco, L. B. Vosshall, and A. Keller, "Humans can discriminate more than 1 trillion olfactory stimuli," *Science*, vol. 343, pp. 1370–1372, 2014.

[120] E. Porcu, K. M. Benz, F. Balla, C. Tempelmann, M. Hanke, and T. Noesselt, "Macroscopic information-based taste representations in insular cortex are shaped by stimulus concentration," *Proc. Natl. Acad. Sci. USA*, vol. 117, pp. 7409–7417, 2020.

[121] T. D. Wyatt, "Pheromones," *Curr. Biol.*, vol. 27, pp. R731–R745, 2017.

[122] F. Abu, J. G. Wang, Y. Oh, J. Deng, T. A. Neubert, and G. S. B. Suh, "Communicating the nutritional value of sugar in *Drosophila*," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. E2829–E2838, 2018.

[123] M. Tsunoda, K. Miyamichi, R. Eguchi, Y. Sakuma, Y. Yoshihara, T. Kikusui, M. Kuwahara, and K. Touhara, "Identification of an intra- and inter-specific tear protein signal in rodents," *Curr. Biol.*, vol. 28, pp. 1213–1223, 2018.

[124] K. Schneeberger, G. Röder, and M. Taborsky, “The smell of hunger: Norway rats provision social partners based on odour cues of need,” *PLoS Biol.*, vol. 18, p. e3000628, 2020.

[125] J. Kipnis, “Immune system: The “seventh sense”,” *J. Exp. Med.*, vol. 215, pp. 397–398, 2018.

[126] M. D. Reed, Y. S. Yim, R. D. Wimmer, H. Kim, C. Ryu, G. M. Welch, M. Andina, H. O. King, A. Waisman, M. M. Halassa, J. R. Huh, and G. B. Choi, “IL-17a promotes sociability in mouse models of neurodevelopmental disorders,” *Nature*, vol. 577, pp. 249–253, 2020.

[127] T. Sturm, T. Leinders-Zufall, B. Maček, M. Walzer, S. Jung, B. Pömmerl, S. Stevanović, F. Zufall, P. Overath, and H.-G. Rammensee, “Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons,” *Nat. Commun.*, vol. 4, p. 1616, 2013.

[128] J. A. Riffell, E. Shlizerman, E. Sanders, L. Abrell, B. Medina, A. J. Hinterwirth, and N. Kutz, “Flower discrimination by pollinators in a dynamic chemical environment,” *Science*, vol. 344, pp. 1515–1518, 2014.

[129] T. Ackels, A. Erskine, D. Dasgupta, A. C. Marin, T. P. A. Warner, S. Tootoonian, I. Fukunaga, J. J. Harris, and A. T. Schaefer, “Fast odour dynamics are encoded in the olfactory system and guide behaviour,” *Nature*, vol. 593, pp. 558–563, 2021.

[130] R. Hemmersbach, M. Krause, R. Bräucker, and K. Ivanova, “Graviperception in ciliates: Steps in the transduction chain,” *Adv. Space Res.*, vol. 35, pp. 296–299, 2005.

[131] A. G. Macdonald and P. J. Fraser, “The transduction of very small hydrostatic pressures,” *Comp. Biochem. Physiol. A*, vol. 122, pp. 13–36, 1999.

[132] A. M. Green, A. G. Shaikh, and D. E. Angelaki, “Sensory vestibular contributions to constructing internal models of self-motion,” *J. Neural Eng.*, vol. 2, pp. S164–S179, 2005.

[133] R. Tseng, N. F. Goularte, A. Chavan, J. Luu, S. E. Cohen, Y.-G. Chang, J. Heisler, S. Li, A. K. Michael, S. Tripathi, S. S. Golden, A. LiWang, and C. L. Partch, “Structural basis of the day-night transition in a bacterial circadian clock,” *Science*, vol. 355, pp. 1174–1180, 2017.

[134] A. D. Güler, J. L. Ecker, G. S. Lall, S. Haq, C. M. Altimus, H.-W. Liao, A. R. Barnard, H. Cahill, T. C. Badea, H. Zhao, M. W. Hankins, D. M. Berson, R. J. Lucas, K.-W. Yau, and S. Hattar, “Melanopsin cells are the principal conduits for rod-cone input to non-image-forming vision,” *Nature*, vol. 453, pp. 102–106, 2008.

[135] C. T. Lefèvre, M. Bennet, L. Landau, P. Vach, D. Pignol, D. A. Bazylinski, R. B. Frankel, S. Klumpp, and D. Faivre, “Diversity of magneto-aerotactic behaviors and oxygen sensing mechanisms in cultured magnetotactic bacteria,” *Biophys. J.*, vol. 107, pp. 527–538, 2014.

[136] P. N. Fleischmann, R. Grob, V. L. Müller, R. Wehner, and W. Rössler, “The geomagnetic field is a compass cue in *Cataglyphis* ant navigation,” *Curr. Biol.*, vol. 28, pp. 1440–1444, 2018.

[137] H. G. Hiscock, S. Worster, D. R. Kattnig, C. Steers, Y. Jin, D. E. Manolopoulos, H. Mouritsen, and P. J. Hore, “The quantum needle of the avian magnetic compass,” *Proc. Natl. Acad. Sci. USA*, vol. 113, pp. 4634–4639, 2016.

[138] N. Chernetsov, A. Pakhomov, D. Kobylkov, D. Kishkinev, R. A. Holland, and H. Mouritsen, “Migratory Eurasian reed warblers can use magnetic declination to solve the longitude problem,” *Curr. Biol.*, vol. 27, pp. 2647–2651, 2017.

[139] S. Nimpf, G. C. Nordmann, D. Kagerbauer, E. P. Malkemper, L. Landler, A. Papadaki-Anastasopoulou, L. Ushakova, A. Wenninger-Weinzierl, M. Novatchkova, P. Vincent, T. Lendl, M. Colombini, M. J. Mason, and D. A. Keays, “A putative mechanism for magnetoreception by electromagnetic induction in the pigeon inner ear,” *Curr. Biol.*, vol. 29, pp. 4052–4059, 2019.

[140] S. Benhamou, J. Sudre, J. Bourjea, S. Ciccione, A. De Santis, and P. Luschi, “The role of geomagnetic cues in green turtle open sea navigation,” *PLoS ONE*, vol. 6, p. e26672, 2011.

[141] K. J. Lohmann, C. M. F. Lohmann, and N. F. Putman, “Magnetic maps in animals: nature’s GPS,” *J. Exp. Biol.*, vol. 210, pp. 3697–3705, 2007.

[142] R. Ismakov, O. Barak, K. Jeffery, and D. Derdikman, “Grid cells encode local positional information,” *Curr. Biol.*, vol. 27, pp. 2337–2343, 2017.

[143] P. A. LaChance, T. P. Todd, and J. S. Taube, “A sense of space in postrhinal cortex,” *Science*, vol. 365, p. eaax4192, 2019.

[144] M. Stangl, U. Topalovic, C. S. Inman, S. Hiller, D. Villaroman, Z. M. Aghajan, L. Christov-Moore, N. R. Hasulak, V. R. Rao, C. H. Halpern, D. Eliashiv, I. Fried, and N. Suthana, “Boundary-anchored neural mechanisms of location-encoding for self and others,” *Nature*, vol. 589, pp. 420–425, 2021.

[145] C. Kemp, *Dark and Magical Places.* W. W. Norton & Company, Inc., 2022.

[146] M. J. Arcaro, P. F. Schade, and M. S. Livingstone, “Body map proto-organization in newborn macaques,” *Proc. Natl. Acad. Sci. USA*, vol. 116, pp. 24861–24871, 2019.

[147] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, “Human-level concept learning through probabilistic program induction,” *Science*, vol. 350, pp. 1332–1338, 2015.

[148] J. Seo, S. Chae, J. Shim, D. Kim, C. Cheong, and T.-D. Han, “Fast contour-tracing algorithm based on a pixel-following method for image sensors,” *Sensors*, vol. 16, p. 353, 2016.

[149] D. Del Vecchio, R. M. Murray, and P. Perona, "Decomposition of human motion into dynamics-based primitives with application to drawing tasks," *Automatica*, vol. 39, pp. 2085–2098, 2003.

[150] I. K. Brunec, B. Bellana, J. D. Ozubko, V. Man, J. Robin, Z.-X. Liu, C. Grady, R. S. Rosenbaum, G. Winocur, M. D. Barense, and M. Moscovitch, "Multiple scales of representation along the hippocampal anteroposterior axis in humans," *Curr. Biol.*, vol. 28, pp. 2129–2135, 2018.

[151] K. C. Catania, "The sense of touch in the star-nosed mole: from mechanoreceptors to the brain," *Phil. Trans. R. Soc. B*, vol. 366, pp. 3016–3025, 2011.

[152] D.-E. Nilsson, "Photoreceptor evolution: ancient siblings serve different tasks," *Curr. Biol.*, vol. 15, pp. R94–R96, 2005.

[153] G. L. Fain, R. Hardie, and S. B. Laughlin, "Phototransduction and the evolution of photoreceptors," *Curr. Biol.*, vol. 20, pp. R114–R124, 2010.

[154] N. Schuergers, T. Lenn, R. Kampmann, M. V. Meissner, T. Esteves, M. Temerinac-Ott, J. G. Korvink, A. R. Lowe, C. W. Mullineaux, and A. Wilde, "Cyanobacteria use micro-optics to sense light direction," *eLife*, vol. 5, p. e12620, 2016.

[155] J. T. Beatty, J. Overmann, M. T. Lince, A. K. Manske, A. S. Lang, R. E. Blankenship, C. L. Van Dover, T. A. Martinson, and F. G. Plumley, "An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 9306–9310, 2005.

[156] E. G. Nisbet, J. R. Cann, and C. L. Van Dover, "Origins of photosynthesis," *Nature*, vol. 373, pp. 479–480, 1995.

[157] E. O. Gracheva, N. T. Ingolia, Y. M. Kelly, J. F. Cordero-Morales, G. Hollopeter, A. T. Chesler, E. E. Sánchez, J. C. Perez, J. S. Weissman, and D. Julius, "Molecular basis of infrared detection by snakes," *Nature*, vol. 464, pp. 1006–1011, 2010.

[158] R. Dawkins, *Climbing Mount Improbable.* W.W. Norton & Company, Inc., 1996.

[159] N. Picciani, J. R. Kerlin, N. Sierra, A. J. M. Swafford, M. D. Ramirez, N. G. Roberts, J. T. Cannon, M. Daly, and T. H. Oakley, "Prolific origination of eyes in Cnidaria with co-option of non-visual opsins," *Curr. Biol.*, vol. 28, pp. 2413–2419, 2018.

[160] C. F. Stevens, "Novel neural circuit mechanism for visual edge detection," *Proc. Natl. Acad. Sci. USA*, vol. 112, pp. 875–880, 2015.

[161] S. J. Zapp, S. Nitsche, and T. Gollisch, "Retinal receptive-field substructure: scaffolding for coding and computation," *Trends Neurosci.*, vol. 45, pp. 430–445, 2022.

[162] J. Hawkins, S. Ahmad, and Y. Cui, "A theory of how columns in the neocortex enable learning the structure of the world," *Front. Neural Circuits*, vol. 11, p. 81, 2017.

[163] H. Tang, M. Schrimpf, W. Lotter, C. Moerman, A. Paredes, J. O. Caro, W. Hardesty, D. Cox, and G. Kreiman, "Recurrent computations for visual pattern completion," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 8835–8840, 2018.

[164] G. A. Manley, "Evolutionary paths to mammalian cochleae," *J. Assoc. Res. Otolaryngol.*, vol. 13, pp. 733–743, 2012.

[165] M. Lipovsek and A. B. Elgoyhen, "The evolutionary tuning of hearing," *Trends Neurosci.*, vol. 46, pp. 110–123, 2023.

[166] G. Ni, S. J. Elliott, M. Ayat, and P. D. Teal, "Modelling cochlear mechanics," *Biomed. Res. Int.*, vol. 2014, p. 150637, 2014.

[167] A. Graps, "An introduction to wavelets," *IEEE Comput. Sci. Eng.*, vol. 2, pp. 50–61, Summer 1995.

[168] S. Rosen, "Temporal information in speech: acoustic, auditory and linguistic aspects," *Phil. Trans. R. Soc. Lond. B*, vol. 336, pp. 367–373, 1992.

[169] W. R. Drennan, J. H. Won, V. K. Dasika, and J. T. Rubinstein, "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," *JARO*, vol. 8, pp. 373–383, 2007.

[170] N. Ding, M. Chatterjee, and J. Z. Simon, "Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure," *Neuroimage*, vol. 88, pp. 41–46, 2014.

[171] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Computer Vis.*, vol. 40, pp. 49–71, 2000.

[172] G. Tkačik, J. S. Prentice, J. D. Victor, and V. Balasubramanian, "Local statistics in natural scenes predict the saliency of synthetic textures," *Proc. Natl. Acad. Sci. USA*, vol. 107, pp. 18149–18154, 2010.

[173] J. D. Victor and M. M. Conte, "Local image statistics: maximum-entropy constructions and perceptual salience," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, vol. 32, pp. 579–582, 2010.

[174] J. H. McDermott and E. P. Simoncelli, "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," *Neuron*, vol. 71, pp. 926–940, 2011.

[175] R. McWalter and J. H. McDermott, "Adaptive and selective time averaging of auditory scenes," *Curr. Biol.*, vol. 28, pp. 1405–1418, 2018.

[176] H. Sheahan, F. Luyckx, S. Nelli, C. Teupe, and C. Summerfield, "Neural state space alignment for magnitude generalization in humans and recurrent networks," *Neuron*, vol. 109, pp. 1214–1226, 2021.

[177] R. H. Day, "The Ames room from another viewpoint," *Perception*, vol. 22, pp. 1007–1011, 1993.

[178] S. Dehaene, G. Dehaene-Lambertz, and L. Cohen, "Abstract representations of numbers in the animal and human brain," *TINS*, vol. 21, pp. 355–361, 1998.

[179] W. H. Meck and R. M. Church, "A mode control model of counting and timing processes," *J. Experimental Psychol.*, vol. 9, pp. 320–334, 1983.

[180] C. Lemer, S. Dehaene, E. Spelke, and L. Cohen, "Approximate quantities and exact number words: dissociable systems," *Neuropsychologia*, vol. 41, pp. 1942–1958, 2003.

[181] B. M. Harvey, "Quantity cognition: numbers, numerosity, zero and mathematics," *Curr. Biol.*, vol. 26, pp. R408–R431, 2016.

[182] S. R. Howard, A. Avarguès-Weber, J. E. Garcia, A. D. Greentree, and A. G. Dyer, "Numerical cognition in honeybees enables addition and subtraction," *Sci. Adv.*, vol. 5, p. eaav0961, 2019.

[183] M. D. de Hevia, L. Veggiotti, A. Streri, and C. D. Bonn, "At birth, humans associate "few" with left and "many" with right," *Curr. Biol.*, vol. 27, pp. 3879–3884, 2017.

[184] A. Jackson, J. Mavoori, and E. E. Fetz, "Long-term motor cortex plasticity induced by an electronic neural implant," *Nature*, vol. 444, pp. 56–60, 2006.

[185] S. Kim, T. Callier, G. A. Tabot, R. A. Gaunt, F. V. Tenore, and S. J. Bensmaia, "Behavioral assessment of sensitivity to intracortical microstimulation of primate somatosensory cortex," *Proc. Natl. Acad. Sci. USA*, vol. 112, pp. 15202–15207, 2015.

[186] S. M. Won, E. Song, J. T. Reeder, and J. A. Rogers, "Emerging modalities and implantable technologies for neuromodulation," *Cell*, vol. 181, pp. 115–135, 2020.

[187] A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. G. Colmenarejo, E. Grefenstette, T. Ramalho, J. Agapiou, A. P. Badia, K. M. Hermann, Y. Zwols, G. Ostrovski, A. Cain, H. King, C. Summerfield, P. Blunsom, K. Kavukcuoglu, and D. Hassabis, "Hybrid computing using a neural network with dynamic external memory," *Nature*, vol. 538, pp. 471–476, 2016.

[188] D. Genzel, M. Schutte, W. O. Brimijoin, P. R. MacNeilage, and L. Wiegrebe, "Psychophysical evidence for auditory motion parallax," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 4264–4269, 2018.

[189] F. Pedraja, V. Hofmann, K. M. Lucas, C. Young, J. Engelmann, and J. E. Lewis, "Motion parallax in electric sensing," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 573–577, 2018.

[190] M. Jordan, G. Luthardt, C. Meyer-Naujoks, and G. Roth, "The role of eye accommodation in the depth perception of common toads," *Z. Naturforsch. C*, vol. 35, pp. 851–852, 1980.

[191] D. H. Arnold, A. Johnston, and S. Nishida, "Timing sight and sound," *Vision Res.*, vol. 45, pp. 1275–1284, 2005.

[192] K. G. Gruters, D. L. K. Murphy, C. D. Jenson, D. W. Smith, C. A. Shera, and J. M. Groh, "The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. E1309–E1318, 2018.

[193] C. E. Carr and K. M. MacLeod, "Microseconds matter," *PLoS Biol.*, vol. 8, p. e1000405, 2010.

[194] H. Nishino, M. Iwasaki, M. Paoli, I. Kamimura, A. Yoritsune, and M. Mizunami, "Spatial receptive fields for odor localization," *Curr. Biol.*, vol. 28, pp. 600–608, 2018.

[195] S. Dehaene, K. Nakamura, A. Jobert, C. Kuroki, S. Ogawa, and L. Cohen, "Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area," *NeuroImage*, vol. 49, pp. 1837–1848, 2010.

[196] G. Schott and J. Schott, "Mirror writing, left-handedness, and leftward scripts," *Arch. Neurol.*, vol. 61, pp. 1849–1851, 2004.

[197] S. Dehaene, H. Lau, and S. Kouider, "What is consciousness, and could machines have it?," *Science*, vol. 358, pp. 486–492, 2017.

[198] A. Zhang, H. D. Stacey, C. E. Mullarkey, and M. S. Miller, "Original antigenic sin: How first exposure shapes lifelong anti-influenza virus immune responses," *J. Immunol.*, vol. 202, pp. 335–340, 2019.

[199] M. S. Fanselow, "Pavlovian conditioning, negative feedback, and blocking: Mechanisms that regulate association formation," *Neuron*, vol. 20, pp. 625–627, 1998.

[200] T. Beckers, R. R. Miller, J. De Houwer, and K. Urushihara, "Reasoning rats: Forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference," *J. Exp. Psychol. Gen.*, vol. 135, pp. 92–102, 2006.

[201] T. R. Levine, "Truth-Default Theory (TDT): A theory of human deception and deception detection," *J. Lang. Soc. Psychol.*, vol. 33, pp. 378–392, 2014.

[202] C. Daase and O. Kessler, "Knowns and unknowns in the 'War on Terror': Uncertainty and the political construction of danger," *Security Dialogue*, vol. 38, pp. 411–434, 2007.

[203] F. Brändle, C. M. Wu, and E. Schulz, "What are we curious about?," *Trends Cogn. Sci.*, vol. 24, pp. 685–687, 2020.

[204] R. Dubey and T. L. Griffiths, "Reconciling novelty and complexity through a rational analysis of curiosity," *Psychol. Rev.*, vol. 127, pp. 455–476, 2020.

[205] B. Barber, "Resistance by scientists to scientific discovery," *Science*, vol. 134, pp. 596–602, 1961.

[206] T. S. Kuhn, *The Structure of Scientific Revolutions, 3rd edition.* University of Chicago Press, 1996.

[207] A. Luczak, B. L. McNaughton, and Y. Kubo, "Neurons learn by predicting future activity," *Nat. Mach. Intell.*, vol. 4, pp. 62–72, 2022.

[208] K. Marinkovic, S. Baldwin, M. G. Courtney, T. Witzel, A. M. Dale, and E. Halgren, "Right hemisphere has the last laugh: neural dynamics of joke appreciation," *Cogn. Affect. Behav. Neurosci.*, vol. 11, pp. 113–130, 2011.

[209] V. Shirhatti, P. Ravishankar, and S. Ray, “Gamma oscillations in primate primary visual cortex are severely attenuated by small stimulus discontinuities,” *PLoS Biol.*, vol. 20, p. e3001666, 2022.

[210] G. G. Szabo, J. S. Farrell, B. Dudok, W.-H. Hou, A. L. Ortiz, C. Varga, P. Moolchand, C. I. Gulsever, T. Gschwind, J. Dimidschstein, M. Capogna, and I. Soltesz, “Ripple-selective GABAergic projection cells in the hippocampus,” *Neuron*, vol. 110, pp. 1959–1977, 2022.

[211] J. Danckert and J. D. Eastwood, *Out of My Skull: The Psychology of Boredom.* Harvard, 2020.

[212] D. D. Hoffman, *The Case Against Reality.* W. W. Norton & Company, Inc., 2019.

[213] E. A. Abbott, *Flatland: A Romance of Many Dimensions, 2nd edition.* Seeley & Co. of London, 1884.

[214] D. D. Hoffman, “Did we evolve to see reality, or are spacetime and objects just our user interface?,” *Ann. N.Y. Acad. Sci.*, 2019.

[215] T. M. Luhrmann, K. Weisman, F. Aulino, J. D. Brahinsky, J. C. Dulin, V. A. Dzokoto, C. H. Legare, M. Lifshitz, E. Ng, N. Ross-Zehnder, and R. E. Smith, “Sensing the presence of gods and spirits across cultures and faiths,” *Proc. Natl. Acad. Sci. USA*, vol. 118, p. e2016649118, 2021.

[216] J. Grandjean, D. Buehlmann, M. Buerge, H. Sigrist, E. Seifritz, F. X. Vollenweider, C. R. Pryce, and M. Rudin, “Psilocybin exerts distinct effects on resting state networks associated with serotonin and dopamine in mice,” *NeuroImage*, vol. 225, p. 117456, 2021.

[217] M. Pollan, *How to Change Your Mind.* Penguin Books, 2018.

[218] T. Pokorny, K. H. Preller, M. Kometer, I. Dziobek, and F. X. Vollenweider, “Effect of psilocybin on empathy and moral decision-making,” *Int. J. Neuropsychopharmacol.*, vol. 20, pp. 747–757, 2017.

[219] K. E. Grogan, R. L. Harris, M. Boulet, and C. M. Drea, “Genetic variation at MHC class II loci influences both olfactory signals and scent discrimination in ring-tailed lemurs,” *BMC Evol. Biol.*, vol. 19, p. 171, 2019.

[220] D. Lieberman, J. Tooby, and L. Cosmides, “The architecture of human kin detection,” *Nature*, vol. 445, pp. 727–731, 2007.

[221] A. J. Thomas, B. Woo, D. Nettle, E. Spelke, and R. Saxe, “Early concepts of intimacy: Young humans use saliva sharing to infer close relationships,” *Science*, vol. 375, pp. 311–315, 2022.

[222] R. Dawkins, *The Selfish Gene.* Oxford University Press, 1976.

[223] V. Vanchurin, Y. I. Wolf, M. I. Katsnelson, and E. V. Koonin, “Toward a theory of evolution as multilevel learning,” *Proc. Natl. Acad. Sci. USA*, vol. 119, p. e2120037119, 2022.

[224] M. A. Tosches, T. M. Yamawaki, R. K. Naumann, A. A. Jacobi, G. Tushev, and G. Laurent, “Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles,” *Science*, vol. 360, pp. 881–888, 2018.

[225] C. Prat, *The Neuroscience of You.* Dutton, 2022.

[226] L. R. Goldberg, “The structure of phenotypic personality traits,” *Am. Psychol.*, vol. 48, pp. 26–34, 1993.

[227] W. Wei, J. G. Lu, A. D. Galinsky, H. Wu, S. D. Gosling, P. J. Rentfrow, W. Yuan, Q. Zhang, Y. Guo, M. Zhang, W. Gui, X.-Y. Guo1, J. Potter, J. Wang, B. Li, X. Li, Y.-M. Han, M. Lv, X.-Q. Guo, Y. Choe, W. Lin, K. Yu, Q. Bai, Z. Shang, Y. Han, and L. Wang, “Regional ambient temperature is associated with human personality,” *Nat. Hum. Behav.*, vol. 1, pp. 890–895, 2017.

[228] F. M. Götz, S. Stieger, S. D. Gosling, J. Potter, and P. J. Rentfrow, “Physical topography is associated with human personality,” *Nat. Hum. Behav.*, vol. 4, pp. 1135–1144, 2020.

[229] M. Gerlach, B. Farb, W. Revelle, and L. A. N. Amaral, “A robust data-driven approach identifies four personality types across four large data sets,” *Nat. Hum. Behav.*, vol. 2, pp. 735–742, 2018.

[230] P. K. Jonason and G. D. Webster, “The Dirty Dozen: A concise measure of the Dark Triad,” *Psychol. Assess.*, vol. 22, pp. 420–432, 2010.

[231] J. A. Shivik, *Mousy Cats and Sheepish Coyotes.* Beacon Press, 2017.

[232] M. Kinsbourne, “Hemispheric specialization and the growth of human understanding,” *Am. Psychol.*, vol. 37, pp. 411–420, 1982.

[233] D. P. Corina, J. Vaid, and U. Bellugi, “The linguistic basis of left hemisphere specialization,” *Science*, vol. 255, pp. 1258–1260, 1992.

[234] M. L. Silagi, M. Radanovic, A. B. Conforto, L. I. Z. Mendonça, and L. L. Mansur, “Inference comprehension in text reading: Performance of individuals with right- versus left-hemisphere lesions and the influence of cognitive functions,” *PLoS ONE*, vol. 13, p. e0197195, 2018.

[235] S. Gao, B. Becker, L. Luo, Y. Geng, W. Zhao, Y. Yin, J. Hu, Z. Gao, Q. Gong, R. Hurlemann, D. Yao, and K. M. Kendrick, “Oxytocin, the peptide that bonds the sexes also divides them,” *Proc. Natl. Acad. Sci. USA*, vol. 113, pp. 7650–7654, 2016.

[236] H. Kiesow, R. I. M. Dunbar, J. W. Kable, T. Kalenscher, K. Vogeley, L. Schilbach, A. F. Marquand, T. V. Wiecki, and D. Bzdok, “10,000 social brains: Sex differentiation in human brain anatomy,” *Sci. Adv.*, vol. 6, p. eaaz1170, 2020.

[237] D. Joel, Z. Berman, I. Tavor, N. Wexler, O. Gaber, Y. Stein, N. Shefi, J. Pool, S. Urchs, D. S. Margulies, F. Liem, J. Hänggi, L. Jäncke, and Y. Assaf, “Sex beyond the genitalia: The human brain mosaic,” *Proc. Natl. Acad. Sci. USA*, vol. 112, pp. 15468–15473, 2015.

[238] A. S. Greene, X. Shen, S. Noble, C. Horien, C. A. Hahn, J. Arora, F. Tokoglu, M. N. Spann, C. I. Carrión, D. S. Barron, G. Sanacora, V. H. Srihari, S. W. Woods, D. Scheinost, and R. T. Constable, "Brain-phenotype models fail for individuals who defy sample stereotypes," *Nature*, vol. 609, pp. 109–118, 2022.

[239] L. A. Krubitzer and A. M. H. Seelke, "Cortical evolution in mammals: The bane and beauty of phenotypic variability," *Proc. Natl. Acad. Sci. USA*, vol. 109, pp. 10647–10654, 2012.

[240] J. L. Roland, A. Z. Snyder, C. D. Hacker, A. Mitra, J. S. Shimony, D. D. Limbrick, M. E. Raichle, M. D. Smyth, and E. C. Leuthardt, "On the role of the corpus callosum in interhemispheric functional connectivity in humans," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. 13278–13283, 2017.

[241] V. Romano, P. Zhai, A. van der Horst, R. Mazza, T. Jacobs, S. Bauer, X. Wang, J. J. White, and C. De Zeeuw, "Olivocerebellar control of movement symmetry," *Curr. Biol.*, vol. 32, pp. 654–670, 2022.

[242] L. Kingsbury, S. Huang, J. Wang, K. Gu, P. Golshani, Y. E. Wu, and W. Hong, "Correlated neural activity and encoding of behavior across brains of socially interacting animals," *Cell*, vol. 178, pp. 429–446, 2019.

[243] W. Zhang and M. M. Yartsev, "Correlated neural activity across the brains of socially interacting bats," *Cell*, vol. 178, pp. 413–428, 2019.

[244] S. Finger, R. L. Buckner, and H. Buckingham, "Does the right hemisphere take over after damage to Broca's area? the Barlow case of 1877 and its history," *Brain Lang.*, vol. 85, pp. 385–395, 2003.

[245] D. Nutt, D. Erritzoe, and R. Carhart-Harris, "Psychedelic psychiatry's Brave New World," *Cell*, vol. 181, pp. 24–28, 2020.

[246] J. Hu, L. Ferguson, K. Adler, C. A. Farah, M. H. Hastings, W. S. Sossin, and S. Schacher, "Selective erasure of distinct forms of long-term synaptic plasticity underlying different forms of memory in the same postsynaptic neuron," *Curr. Biol.*, vol. 27, pp. 1888–1899, 2017.

[247] E. H. Thompson, K. K. Lensjø, M. B. Wigestrand, A. Malthe-Sørenssen, T. Hafting, and M. Fyhn, "Removal of perineuronal nets disrupts recall of a remote fear memory," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 607–612, 2018.

[248] E. A. Phelps and S. G. Hofmann, "Memory editing from science fiction to clinical practice," *Nature*, vol. 572, pp. 43–50, 2019.

[249] T. Liang and B. A. W. Brinkman, "Evolution of innate behavioral strategies through competitive population dynamics," *PLoS Comput. Biol.*, vol. 18, p. e1009934, 2022.

[250] D. D. Hoffman, *Visual Intelligence.* W. W. Norton & Company, Inc., 1998.

[251] R. J. Zatorre and V. N. Salimpoor, "From perception to pleasure: Music and its neural substrates," *Proc. Natl. Acad. Sci. USA*, vol. 110, pp. 10430–10437, 2013.

[252] K. C. Berridge and M. L. Kringelbach, "Pleasure systems in the brain," *Neuron*, vol. 86, pp. 646–664, 2015.

[253] W. Ma and W. F. Thompson, "Human emotions track changes in the acoustic environment," *Proc. Natl. Acad. Sci. USA*, vol. 112, pp. 14563–14568, 2015.

[254] J.-M. Schaeffer, "Aesthetic relationship, cognition, and the pleasures of art," in *Investigations Into the Phenomenology and the Ontology of the Work of Art* (P. F. Bundgaard and F. Stjernfelt, eds.), pp. 52–75, 2015.

[255] M. Tiihonen, E. Brattico, J. Maksimainen, J. Wikgren, and S. Saarikallio, "Constituents of music and visual-art related pleasure - a critical integrative literature review," *Front. Psychol.*, vol. 8, p. 1218, 2017.

[256] E. A. Vessel, A. I. Isik, A. M. Belfi, J. L. Stahl, and G. G. Starr, "The default-mode network represents aesthetic appeal that generalizes across visual domains," *Proc. Natl. Acad. Sci. USA*, vol. 116, pp. 19155–19164, 2019.

[257] V. K. Cheung, P. M. Harrison, L. Meyer, M. T. Pearce, J.-D. Haynes, and S. Koelsch, "Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity," *Curr. Biol.*, vol. 29, pp. 4084–4092, 2019.

[258] E. Wassiliwizky and W. Menninghaus, "Why and how should cognitive science care about aesthetics?," *Trends Cogn. Sci.*, vol. 25, pp. 437–449, 2021.

[259] D. Sznycer, D. Xygalatas, S. Alami, X.-F. An, K. I. Ananyeva, S. Fukushima, H. Hitokoto, A. N. Kharitonov, J. M. Koster, C. N. Onyishi, I. E. Onyishi, P. P. Romero, K. Takemura, J.-Y. Zhuang, L. Cosmides, and J. Tooby, "Invariances in the architecture of pride across small-scale societies," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 8322–8327, 2018.

[260] T. Grandin and S. Barron, *Unwritten rules of social relationships.* Future Horizons, Inc., 2005.

[261] S. Cain, *Quiet.* Broadway Books, 2013.

[262] M. Gladwell, *Talking to Strangers.* Little, Brown and Company, 2019.

[263] S. Dehaene, *Consciousness and the Brain.* Viking, 2014.

[264] A. K. Seth and T. Bayne, "Theories of consciousness," *Nat. Rev. Neurosci.*, vol. 23, pp. 439–452, 2022.

[265] G. A. Mashour, P. Roelfsema, J.-P. Changeux, and S. Dehaene, "Conscious processing and the global neuronal workspace hypothesis," *Neuron*, vol. 105, pp. 776–798, 2020.

[266] J. Peever and P. M. Fuller, "The biology of REM sleep," *Curr. Biol.*, vol. 27, pp. R1237–R1248, 2017.

[267] J. M. Krueger, M. G. Frank, J. P. Wisor, and S. Roy, "Sleep function: Toward elucidating an enigma," *Sleep Med. Rev.*, vol. 28, pp. 46–54, 2016.

[268] L. Xie, H. Kang, Q. Xu, M. J. Chen, Y. Liao, M. Thiyagarajan, J. O'Donnell, D. J. Christensen, C. Nicholson, J. J. Iliff, T. Takano, R. Deane, and M. Nedergaard, "Sleep drives metabolite clearance from the adult brain," *Science*, vol. 342, pp. 373–377, 2013.

[269] E. Shokri-Kojoria, G.-J. Wang, C. E. Wiers, S. B. Demiral, M. Guo, S. W. Kim, E. Lindgren, V. Ramirez, A. Zehra, C. Freeman, G. Miller, P. Manza, T. Srivastava, S. De Santi, D. Tomasi, H. Benveniste, and N. D. Volkow, "β-amyloid accumulation in the human brain after one night of sleep deprivation," *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 4483–4488, 2018.

[270] W. Li, L. Ma, G. Yang, and W.-B. Gan, "REM sleep selectively prunes and maintains new synapses in development and learning," *Nature Neurosci.*, vol. 20, pp. 427–437, 2017.

[271] M. A. Wilson and B. L. McNaughton, "Reactivation of hippocampal ensemble memories during sleep," *Science*, vol. 265, pp. 676–679, 1994.

[272] W. E. Skaggs and B. L. McNaughton, "Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience," *Science*, vol. 271, pp. 1870–1873, 1996.

[273] K. Louie and M. A. Wilson, "Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep," *Neuron*, vol. 29, pp. 145–156, 2001.

[274] D. S. Ramanathan, T. Gulati, and K. Ganguly, "Sleep-dependent reactivation of ensembles in motor cortex promotes skill consolidation," *PLoS Biol.*, vol. 13, p. e1002263, 2015.

[275] X. Jiang, I. Shamie, W. K. Doyle, D. Friedman, P. Dugan, O. Devinsky, E. Eskandar, S. S. Cash, T. Thesen, and E. Halgren, "Replay of large-scale spatio-temporal patterns from waking during subsequent NREM sleep in human cortex," *Sci. Rep.*, vol. 7, p. 17380, 2017.

[276] M. W. Mahowald and C. H. Schenck, "Insights from studying human sleep disorders," *Nature*, vol. 437, pp. 1279–1285, 2005.

[277] R. D. Nath, C. N. Bedbrook, M. J. Abrams, T. Basinger, J. S. Bois, D. A. Prober, P. W. Sternberg, V. Gradinaru, and L. Goentoro, "The jellyfish *Cassiopea* exhibits a sleep-like state," *Curr. Biol.*, vol. 27, pp. 2984–2990, 2017.

[278] M. Walker, *Why We Sleep*. Scribner, 2017.

[279] O. I. Lyamin, P. O. Kosenko, S. M. Korneva, A. L. Vyssotski, L. M. Mukhametov, and J. M. Siegel, "Fur seals suppress REM sleep for very long periods without subsequent rebound," *Curr. Biol.*, vol. 28, pp. 2000–2005, 2018.

[280] H. F. Ólafsdóttir, D. Bush, and C. Barry, "The role of hippocampal replay in memory and planning," *Curr. Biol.*, vol. 28, pp. R37–R50, 2018.

[281] Y. Nir and G. Tononi, "Dreaming and the brain: from phenomenology to neurophysiology," *Trends Cogn. Sci.*, vol. 14, p. 88, 2010.

[282] A. Zadra and R. Stickgold, *When Brains Dream*. W.W. Norton & Company, Inc., 2021.

[283] R. Stickgold, L. Scott, C. Rittenhouse, and J. A. Hobson, "Sleep-induced changes in associative memory," *J. Cogn. Neurosci.*, vol. 11, pp. 182–193, 1999.

[284] G. Leschziner, *The Nocturnal Brain*. St. Martin's Press, 2019.

[285] H. Choi, C. Crump, C. Duriez, A. Elmquist, G. Hager, D. Han, F. Hearl, J. Hodgins, A. Jain, F. Leve, C. Li, F. Meier, D. Negrut, L. Righetti, A. Rodriguez, J. Tan, and J. Trinkle, "On the use of simulation in robotics: Opportunities, challenges, and suggestions for moving forward," *Proc. Natl. Acad. Sci. USA*, vol. 118, p. e1907856118, 2021.

[286] "VulkanSceneGraph." https://vsg-dev.github.io/VulkanSceneGraph/. Accessed: 2021-05-02.

[287] P. Godfrey-Smith, *Other minds: the octopus, the sea, and the deep origins of consciousness*. Farrar, Strauss and Giroux, 2016.

[288] P. S. Thomas, B. Castro da Silva, A. G. Barto, S. Giguere, Y. Brun, and E. Brunskill, "Preventing undesirable behavior of intelligent machines," *Science*, vol. 366, pp. 999–1004, 2019.

[289] K. Man and A. Damasio, "Homeostasis and soft robotics in the design of feeling machines," *Nat. Mach. Intell.*, vol. 1, pp. 446–452, 2019.

[290] M. Csikszentmihalyi, *Flow*. Harper & Row, 1990.

[291] K. Crawford, *Atlas of AI*. Yale University Press, 2021.

[292] N. Sharkey, A. van Wynsberghe, S. Robbins, and E. Hancock, "Our sexual future with robots," tech. rep., Foundation for Responsible Robotics, 2017.

[293] M. Edmonds, F. Gao, H. Liu, X. Xie, S. Qi, B. Rothrock, Y. Zhu, Y. N. Wu, H. Lu, and S.-C. Zhu, "A tale of two explanations: Enhancing human trust by explaining robot behavior," *Sci. Robot.*, vol. 4, p. eaay4663, 2019.

[294] J. C. Peterson, D. D. Bourgin, M. Agrawal, D. Reichman, and T. L. Griffiths, "Using large-scale experiments and machine learning to discover theories of human decision-making," *Science*, vol. 372, pp. 1209–1214, 2021.

[295] "Research priorities for robust and beneficial artificial intelligence." https://futureoflife.org/ai-open-letter. Accessed: 2018-03-30.

[296] S. Russell, D. Dewey, and M. Tegmark, "Research priorities for robust and beneficial artificial intelligence," *arXiv:1602.03506v1*, 2016.

[297] "Autonomous weapons: An open letter from AI & robotics researchers." https://futureoflife.org/open-letter-autonomous-weapons/. Accessed: 2018-03-25.

[298] M. Taddeo and L. Floridi, "Regulate artificial intelligence to avert cyber arms race," *Nature*, vol. 556, pp. 296–298, 2018.

[299] J. Claybrook and S. Kildare, "Autonomous vehicles: No driver...no regulation?," *Science*, vol. 361, pp. 36–37, 2018.

[300] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Stanley: The robot that won the DARPA Grand Challenge," *J. Field Robotics*, vol. 23, pp. 661–692, 2006.

[301] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, pp. 354–359, 2017.

[302] H. Würbel, "Rights for sentient animals," *Science*, vol. 332, p. 917, 2011.

[303] D. Levy, "The ethical treatment of artificially conscious robots," *Int. J. Soc. Robot.*, vol. 1, pp. 209–216, 2009.

[304] M. Coeckelbergh, "Robot rights? Towards a social-relational justification of moral consideration," *Ethics Inf. Technol.*, vol. 12, pp. 209–221, 2010.

[305] R. Kurzweil, *The Singularity Is Near.* Penguin Books, 2005.

[306] M. Taddeo and L. Floridi, "How AI can be a force for good," *Science*, vol. 361, pp. 751–752, 2018.

[307] N. Agar, "Don't worry about superintelligence," *J. Evol. Technol.*, vol. 26, pp. 73–82, 2016.

[308] S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control.* Viking, 2019.

[309] R. Reisenzein, M. Studtmann, and G. Horstmann, "Coherence between emotion and facial expression: Evidence from laboratory experiments," *Emotion Rev.*, vol. 5, pp. 16–23, 2013.

[310] M. Kejriwal, H. Santos, A. M. Mulvehill, and D. L. McGuinness, "Designing a strong test for measuring true common-sense reasoning," *Nat. Mach. Intell.*, vol. 4, pp. 318–322, 2022.

[311] J. Brustein and M. Bergen, "Google's defense dilemma," *Bloomberg Businessweek*, vol. November 25, pp. 38–43, 2019.

[312] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty, "Building Watson: An overview of the DeepQA project," *AI Magazine*, vol. Fall, pp. 59–79, 2010.

[313] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021.

[314] E. Callaway, "What's next for the AI protein-folding revolution," *Nature*, vol. 604, pp. 234–238, 2022.

[315] G. Marcus and E. Davis, "If computers are so smart, how come they can't read?," *Wired*, vol. 09.10.2019, 2019.

[316] D. Hofstadter, "The shallowness of Google Translate," *The Atlantic*, vol. January 30, 2018.

[317] M. Hutson, "The language machines," *Nature*, vol. 591, pp. 22–25, 2021.

[318] M. Pividori and C. S. Greene, "A publishing infrastructure for AI-assisted academic authoring," *bioRxiv:2023.01.21.525030*, 2023.

[319] M. Mitchell and D. C. Krakauer, "The debate over understanding in AI's large language models," *Proc. Natl. Acad. Sci. USA*, vol. 120, p. e2215907120, 2023.

[320] M. Hutson, "Hackers easily fool artificial intelligences," *Science*, vol. 361, p. 215, 2018.

[321] D. Heaven, "Deep trouble for deep learning," *Nature*, vol. 574, pp. 163–166, 2019.

[322] D. George, W. Lehrach, K. Kansky, M. Lázaro-Gredilla, C. Laan, B. Marthi, X. Lou, Z. Meng, Y. Liu, H. Wang, A. Lavin, and D. S. Phoenix, "A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs," *Science*, vol. 358, p. eaag2612, 2017.

[323] D. Wakabayashi, "Self-driving Uber car kills pedestrian in Arizona, where robots roam." https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html. Accessed: 2018-03-24.

[324] A. Nunes, B. Reimer, and J. F. Coughlin, "People must retain control of autonomous vehicles," *Nature*, vol. 556, pp. 169–171, 2018.

[325] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, pp. 350–354, 2019.

[326] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proc. Natl. Acad. Sci. USA*, vol. 116, pp. 24972–24978, 2019.

[327] A. Abraham, *The Neuroscience of Creativity*. Cambridge University Press, 2018.

[328] "The 3-Clause BSD License." https://opensource.org/licenses/BSD-3-Clause. Accessed: 2019-07-06.

[329] M. Mitchell, *An Introduction to Genetic Algorithms*. The MIT Press, 1996.

[330] "Raspberry Pi." https://www.raspberrypi.org. Accessed: 2020-02-23.

[331] "Boost C++ libraries." http://www.boost.org. Accessed: 2019-09-21.

[332] "ISO International Standard ISO/IEC 14882:2017(E) – Programming Language C++." https://isocpp.org/std/the-standard. Accessed: 2019-07-07.

[333] "Guideline Support Library." https://github.com/Microsoft/GSL. Accessed: 2018-08-08.

[334] "Jonathan Boccara's NamedType library." https://github.com/joboccara/NamedType. Accessed: 2018-08-08.

[335] "Niels Lohmann's json library." https://github.com/nlohmann/json. Accessed: 2021-11-19.

[336] "OpenSceneGraph." http://www.openscenegraph.org. Accessed: 2018-03-24.

[337] "Unreal engine." https://www.unrealengine.com/en-US. Accessed: 2023-07-03.

[338] Z. M. Aghajan, P. Schuette, T. A. Fields, M. E. Tran, S. M. Siddiqui, N. R. Hasulak, T. K. Tcheng, D. Eliashiv, E. A. Mankin, J. Stern, I. Fried, and N. Suthana, "Theta oscillations in the human medial temporal lobe during real-world ambulatory movement," *Curr. Biol.*, vol. 27, pp. 3743–3751, 2017.

[339] T. Zhou, J. E. Náñez Sr., D. Zimmerman, S. R. Holloway, and A. Seitz, "Two visual training paradigms associated with enhanced critical flicker fusion threshold," *Front. Psychol.*, vol. 7, p. 1597, 2016.

[340] H. Schmidt, D. Avitabile, E. Montbrió, and A. Roxin, "Network mechanisms underlying the role of oscillations in cognitive tasks," *PLoS Comput. Biol.*, vol. 14, p. e1006430, 2018.

[341] J. I. Chapeton, R. Haque, J. H. Wittig Jr., S. K. Inati, and K. A. Zaghloul, "Large-scale communication in the human brain is rhythmically modulated through alpha coherence," *Curr. Biol.*, vol. 29, pp. 2801–2811, 2019.

[342] A. Arnatkevičiūtė, B. D. Fulcher, R. Pocock, and A. Fornito, "Hub connectivity, neuronal diversity, and gene expression in the *Caenorhabditis elegans* connectome," *PLoS Comput. Biol.*, vol. 14, p. e1005989, 2018.

[343] E. Isbell, C. Stevens, E. Pakulak, A. H. Wray, T. A. Bell, and H. J. Neville, "Neuroplasticity of selective attention: Research foundations and preliminary evidence for a gene by intervention interaction," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. 9247–9254, 2017.

[344] E. I. Knudsen, "Neural circuits that mediate selective attention: A comparative perspective," *Trends Neurosci.*, vol. 41, pp. 789–805, 2018.

[345] K. M. Costa, A. Sengupta, and G. Schoenbaum, "The orbitofrontal cortex is necessary for learning to ignore," *Curr. Biol.*, vol. 31, pp. 2652–2657, 2021.

[346] D. J. Simons and C. F. Chabris, "Gorillas in our midst: sustained inattentional blindness for dynamic events," *Perception*, vol. 28, pp. 1059–1074, 1999.

[347] C. Timbie, M. Á. García-Cabezas, B. Zikopoulos, and H. Barbas, "Gorganization of primate amygdalar-thalamic pathways for emotions," *PLoS Biol.*, vol. 18, p. e3000639, 2020.

[348] F. Mushtaq, A. R. Bland, and A. Schaefer, "Uncertainty and cognitive control," *Frontiers Psychol.*, vol. 2, p. 249, 2011.

[349] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 49, pp. 433–460, 1950.

[350] B. Goertzel, M. Iklé, and J. Wigmore, "The architecture of human-like general intelligence," in *Theoretical Foundations of Artificial General Intelligence*, pp. 123–144, 2012.

[351] N. J. Nilsson, "Human-level artificial intelligence? Be serious!," *AI Magazine*, pp. 68–75, Winter 2005.

[352] F. de Waal, *Are We Smart Enough to Know How Smart Animals Are?* W. W. Norton & Company, Inc., 2016.

[353] A. Seth, *Being You*. Dutton, 2021.

[354] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot Operating System 2: Design, architecture, and uses in the wild," *Sci. Robot.*, vol. 7, p. eabm6074, 2022.

[355] M. Bjelonic, R. Grandia, O. Harley, C. Galliard, S. Zimmermann, and M. Hutter, "Whole-body MPC and online gait sequence generation for wheeled-legged robots," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[356] M. Schrage, *The Innovator's Hypothesis*. MIT Press, 2014.

[357] F. Martínez-Plumed, P. Barredo, S. Ó. hÉigeartaigh, and J. Hernández-Orallo, "Research community dynamics behind popular AI benchmarks," *Nat. Mach. Intell.*, vol. 3, pp. 581–589, 2021.

[358] K. Darling, *The New Breed*. Henry Holt and Company, 2021.

[359] C. Whitehead, "The neural correlates of work and play: What brain imaging research and animal cartoons can tell us about social displays, self-consciousness, and the evolution of the human brain," *J. Consciousness Studies*, vol. 15, pp. 93–121, 2008.

[360] S. H. Seo, "A simulated robot versus a real robot: an exploration of how robot embodiment impacts people's empathic responses," Master's thesis, The University of Manitoba, 2015.

[361] Y. Zhu, T. Guo, Q. Liu, Q. Zhu, X. Zhao, and B. Jin, "Turning and radius deviation correction for a hexapod walking robot based on an ant-inspired sensory strategy," *Sensors*, vol. 17, p. 2710, 2017.

[362] J. Chen, Z. Liang, Y. Zhu, C. Liu, L. Zhang, L. Hao, and J. Zhao, "Towards the exploitation of physical compliance in segmented and electrically actuated robotic legs: A review focused on elastic mechanisms," *Sensors*, vol. 19, p. 5351, 2019.

[363] M. E. Wosniack, M. C. Santos, E. P. Raposo, G. M. Viswanathan, and M. G. E. da Luz, "The evolutionary origins of Lévy walk foraging," *PLoS Comput. Biol.*, vol. 13, p. e1005774, 2017.

[364] K. Kobayashi, S. Ravaioli, A. Baranès, M. Woodford, and J. Gottlieb, "Diverse motives for human curiosity," *Nat. Hum. Behav.*, vol. 3, pp. 587–595, 2019.

[365] D. M. Lydon-Staley, D. Zhou, A. S. Blevins, P. Zurn, and D. S. Bassett, "Hunters, busybodies and the knowledge network building associated with deprivation curiosity," *Nat. Hum. Behav.*, vol. 5, pp. 327–336, 2021.

[366] J. R. Koza, *Genetic Programming.* The MIT Press, 1992.

[367] O. W. Sacks, *Musicophilia: Tales of Music and the Brain.* Vintage Canada, 2007.

[368] P. Vuust, O. A. Heggli, K. J. Friston, and M. L. Kringelbach, "Music in the brain," *Nat. Rev. Neurosci.*, vol. 23, pp. 287–305, 2022.

[369] P. D. McGeoch and R. Rouw, "How everyday sounds can trigger strong emotions: ASMR, misophonia and the feeling of wellbeing," *BioEssays*, vol. 42, p. 2000099, 2020.

[370] J. R. Georgiadis, M. L. Kringelbach, and J. G. Pfaus, "Sex for fun: a synthesis of human and animal neurobiology," *Nat. Rev. Urol.*, vol. 9, pp. 486–498, 2012.

[371] N. Whitton, *Digital Games and Learning.* Routledge, 2014.

[372] M. Taquet, J. Quoidbach, Y.-A. de Montjoye, M. Desseilles, and J. J. Gross, "Hedonism and the choice of everyday activities," *Proc. Natl. Acad. Sci. USA*, vol. 113, pp. 9769–9773, 2016.

[373] A. S. Reinhold, J. I. Sanguinetti-Scheck, K. Hartmann, and M. Brecht, "Behavioral and neural correlates of hide-and-seek in rats," *Science*, vol. 365, pp. 1180–1183, 2019.

[374] Y. Zhang, Y. Ding, J. Huang, W. Zhou, Z. Ling, B. Hong, and X. Wang, "Hierarchical cortical networks of "voice patches" for processing voices in human brain," *Proc. Natl. Acad. Sci. USA*, vol. 118, p. e2113887118, 2021.

[375] J. Gomez, M. Barnett, and K. Grill-Spector, "Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex," *Nat. Hum. Behav.*, vol. 3, pp. 611–624, 2019.

[376] D.-E. Nilsson, "Eye evolution and its functional basis," *Vis. Neurosci.*, vol. 30, pp. 5–20, 2013.

[377] M. A. MacIver, L. Schmitz, U. Mugan, T. D. Murphey, and C. D. Mobley, "Massive increase in visual range preceded the origin of terrestrial vertebrates," *Proc. Natl. Acad. Sci. USA*, vol. 114, pp. E2375–E2384, 2017.

[378] C. Zhuang, S. Yan, A. Nayebi, M. Schrimpf, M. C. Frank, J. J. DiCarlo, and D. L. K. Yamins, "Unsupervised neural network models of the ventral visual stream," *Proc. Natl. Acad. Sci. USA*, vol. 118, p. e2014196118, 2021.

[379] W. J. Song, Y. Lee, Y. Jung, Y.-W. Kang, J. Kim, J.-M. Park, Y.-L. Park, H.-Y. Kim, and J.-Y. Sun, "Soft artificial electroreceptors for noncontact spatial perception," *Sci. Adv.*, vol. 7, p. eabg9203, 2021.

[380] M. Bohn, M. H. Tessler, M. Merrick, and M. C. Frank, "How young children integrate information sources to infer the meaning of words," *Nat. Hum. Behav.*, vol. 5, pp. 1046–1054, 2021.

[381] S. Garrod and M. J. Pickering, "Why is conversation so easy?," *Trends Cogn. Sci.*, vol. 8, pp. 8–11, 2004.

[382] H. C. Evrard, T. Forro, and N. K. Logothetis, "Von Economo neurons in the anterior insula of the macaque monkey," *Neuron*, vol. 74, pp. 482–489, 2012.

[383] G. A. Castellucci, C. K. Kovach, M. A. Howard III, J. D. W. Greenlee, and M. A. Long, "A speech planning network for interactive language use," *Nature*, vol. 602, pp. 117–122, 2022.

[384] C. Rutz, G. R. Hunt, and J. J. St Clair, "Corvid technologies: How do New Caledonian crows get their tool designs?," *Curr. Biol.*, vol. 28, pp. R1009–R1111, 2018.

[385] J. B. Heald, J. N. Ingram, J. R. Flanagan, and D. M. Wolpert, "Multiple motor memories are learned to control different points on a tool," *Nat. Hum. Behav.*, vol. 2, pp. 300–311, 2018.

[386] L. E. Miller, L. Montroni, E. Koun, R. Salemme, V. Hayward, and A. Farnè, "Sensing with tools extends somatosensory processing beyond the body," *Nature*, vol. 561, pp. 239–242, 2018.

[387] L. E. Miller, C. Fabio, V. Ravenda, S. Bahmad, E. Koun, R. Salemme, J. Luauté, N. Bolognini, V. Hayward, and A. Farnè, "Somatosensory cortex efficiently processes touch located beyond the body," *Curr. Biol.*, vol. 29, pp. 4276–4283, 2019.

[388] F. R. Willett, D. R. Deo, D. T. Avansino, P. Rezaii, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, "Hand knob area of premotor cortex represents the whole body in a compositional way," *Cell*, vol. 181, pp. 396–409, 2020.

[389] M. D. Grilli and M. Verfaellie, "Supporting the self-concept with memory: insight from amnesia," *Soc. Cogn. Affect. Neurosci.*, vol. 10, pp. 1684–1692, 2015.

[390] M. L. Meyer, Z. Zhao, and D. I. Tamir, “Simulating other people changes the self,” *J. Exp. Psychol. Gen.*, vol. 148, pp. 1898–1913, 2019.

[391] M. Bates, “Models of natural language understanding,” *Proc. Natl. Acad. Sci. USA*, vol. 92, pp. 9977–9982, 1995.

[392] A. Extance, “AI tames the scientific literature,” *Nature*, vol. 561, pp. 273–274, 2018.

[393] M. Minderer, K. D. Brown, and C. D. Harvey, “The spatial structure of neural encoding in mouse posterior cortex during navigation,” *Neuron*, vol. 102, pp. 232–248, 2019.

[394] “How many words are there in the English language?.” https://en.oxforddictionaries.com/explore/how-many-words-are-there-in-the-english-language. Accessed: 2019-05-04.

[395] R. Munroe, *Thing Explainer: Complicated Stuff in Simple Words.* Houghton Mifflin Harcourt, 2015.

[396] T. Kuhn, “The Controlled Natural Language of Randall Munroe’s *Thing Explainer*,” *arXiv:1605.02457v1*, 2016.

[397] C. Stokel-Walker and R. Van Noorden, “The promise and peril of generative AI,” *Nature*, vol. 614, pp. 214–216, 2023.

[398] R. Leaman, C.-H. Wei, A. Allot, and Z. Lu, “Ten tips for a text-mining-ready article: How to improve automated discoverability and interpretability,” *PLoS Biol.*, vol. 18, p. e3000716, 2020.

[399] D. Westergaard, H.-H. Stæfeldt, C. Tønsberg, L. J. Jensen, and S. Brunak, “A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts,” *PLoS Comput. Biol.*, vol. 14, p. e1005962, 2018.

[400] S. Mukherjee, D. M. Romero, B. Jones, and B. Uzzi, “The nearly universal link between the age of past knowledge and tomorrow’s breakthroughs in science and technology: The hotspot,” *Sci. Adv.*, vol. 3, p. e1601315, 2017.

[401] J. P. A. Ioannidis, “Why most published research findings are false,” *PLoS Med.*, vol. 2, p. e124, 2005.

[402] R. M. Shiffrin, K. Börner, and S. M. Stigler, “Scientific progress despite irreproducibility: A seeming paradox,” *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. 2632–2639, 2018.

[403] P. Azoulay, C. Fons-Rosen, and J. S. Graff Zivin, “Does science advance one funeral at a time?,” *Am. Econ. Rev.*, vol. 109, pp. 2889–2920, 2019.

[404] P. Suber, “An open access mandate for the National Institutes of Health,” *Open Med.*, vol. 2, pp. E14–E16, 2008.

[405] H. Else, “The rise and rise of Unpaywall,” *Nature*, vol. 560, pp. 290–291, 2018.

[406] M. Schiltz, “Science without publication paywalls: cOAlition S for the realisation of full and immediate Open Access,” *PLoS Biol.*, vol. 16, p. e3000031, 2018.

[407] P. Pulla, “The plan to mine the world’s research papers,” *Nature*, vol. 571, pp. 316–318, 2019.

[408] “PubMed.” https://www.ncbi.nlm.nih.gov/pubmed/. Accessed: 2019-02-17.

[409] W. Wei, R. Marmor, S. Singh, S. Wang, D. Demner-Fushman, T.-T. Kuo, C.-N. Hsu, and L. Ohno-Machado, “Finding related publications: Extending the set of terms used to assess article similarity,” *AMIA Jt. Summits Transl. Sci. Proc.*, vol. 2016, pp. 225–234, 2016.

[410] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, “Cytoscape: A software environment for integrated models of biomolecular interaction networks,” *Genome Res.*, vol. 13, pp. 2498–2504, 2003.

[411] M. Bastian, S. Heymann, and M. Jacomy, “Gephi : An open source software for exploring and manipulating networks,” *International AAAI Conference on Weblogs and Social Media*, 2009.

[412] O. Toubia, J. Berger, and J. Eliashberg, “How quantifying the shape of stories predicts their success,” *Proc. Natl. Acad. Sci. USA*, vol. 118, p. e2011695118, 2021.

[413] S. L. King, W. R. Friedman, S. J. Allen, L. Gerber, F. H. Jensen, S. Wittwer, R. C. Connor, and M. Krützen, “Bottlenose dolphins retain individual vocal labels in multi-level alliances,” *Curr. Biol.*, vol. 28, pp. 1993–1999, 2018.

[414] N. Garg, L. Schiebinger, D. Jurafsky, and J. Zou, “Word embeddings quantify 100 years of gender and ethnic stereotypes,” *Proc. Natl. Acad. Sci. USA*, vol. 115, pp. E3635–E3644, 2018.

[415] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, and A. Jain, “Unsupervised word embeddings capture latent knowledge from materials science literature,” *Nature*, vol. 571, pp. 95–98, 2019.

[416] H. Zhang, Y.-F. Li, and H. B. K. Tan, “Measuring design complexity of semantic web ontologies,” *J. Syst. Softw.*, vol. 83, pp. 803–814, 2010.

[417] A. Albin and L. Gershkoff-Stowe, “Rapid word learning in trilingual children: One spurt or three?,” in *Proceedings of the 13th Generative Approaches to Second Language Acquisition Conference (GASLA 2015)* (D. Stringer, J. Garrett, B. Halloran, and S. Mossman, eds.), pp. 14–30, 2016.

[418] V. Dhar and S. Bowman, “A perspective on natural language understanding capability: An interview with Sam Bowman,” *Big Data*, vol. 5, pp. 5–11, 2017.

[419] G. A. Miller, “WordNet: A lexical database for English,” *Communications of the ACM*, vol. 38(11), pp. 39–41, 1995.

[420] “WordNet: A lexical database for English.” https://wordnet.princeton.edu. Accessed: 2020-05-31.

[421] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Natl. Acad. Sci. USA*, vol. 105, pp. 1118–1123, 2008.

[422] "Infomap Online." https://www.mapequation.org/infomap/. Accessed: 2020-06-06.

[423] J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, and D. I. Beaver, "When small words foretell academic success: The case of college admissions essays," *PLoS ONE*, vol. 9, p. e115844, 2014.

[424] "The free dictionary by Farlex." https://www.thefreedictionary.com/List-of-pronouns.htm. Accessed: 2020-07-03.

[425] "Function word lists." https://semanticsimilarity.wordpress.com/function-word-lists/. Accessed: 2020-07-03.

[426] "English grammar explanations - inflections." http://esl.fis.edu/grammar/rules/inflections.htm. Accessed: 2020-07-03.

[427] K. Fundel and R. Zimmer, "Gene and protein nomenclature in public databases," *BMC Bioinformatics*, vol. 7, p. 372, 2006.

[428] A. M. Liberman, "The grammars of speech and language," *Cogn. Psychol.*, vol. 1, pp. 301–323, 1970.

[429] M. Matthies, P. Perrier, J. S. Perkell, and M. Zandipour, "Variation in anticipatory coarticulation with changes in clarity and rate," *J. Speech Lang. Hear. Res.*, vol. 44, pp. 340–353, 2001.

[430] S. Vasishth, B. Nicenboim, F. Engelmann, and F. Burchert, "Computational models of retrieval processes in sentence processing," *Trends Cogn. Sci.*, vol. 23, pp. 968–982, 2019.

[431] "XKCD Simple Writer Word List 0.2.1." https://xkcd.com/simplewriter/words.js. Accessed: 2019-05-12.

[432] J.-M. Poulin and A. Blondin Massé, "Deriving learning strategies from words lists: Digital dictionaries, lexicons, directed graphs and the symbol grounding problem," *Int. J. Adv. Software*, vol. 11, pp. 247–275, 2018.

[433] F. Trobevšek Drobnak, "On the merits and shortcomings of semantic primes and natural semantic metalanguage in cross-cultural translation," *ELOPE: English Language Overseas Perspectives and Enquiries*, vol. 6, pp. 29–41, 2009.

[434] "Learn these words first." http://learnthesewordsfirst.com. Accessed: 2019-11-30.

[435] "Longman dictionary of contemporary English online." https://www.ldoceonline.com. Accessed: 2019-11-30.

[436] H. Keller, *The Story of My Life.* Reprinted by Bantam in 1990, 1903.

[437] A. Prlić and J. B. Procter, "Ten simple rules for the open development of scientific software," *PLoS Comput. Biol.*, vol. 8, p. e1002802, 2012.

[438] Y. Perez-Riverol, L. Gatto, R. Wang, T. Sachsenberg, J. Uszkoreit, F. da Veiga Leprevost, C. Fufezan, T. Ternent, S. J. Eglen, D. S. Katz, T. J. Pollard, A. Konovalov, R. M. Flight, K. Blin, and J. A. Vizcaíno, "Ten simple rules for taking advantage of Git and GitHub," *PLoS Comput. Biol.*, vol. 12, p. e1004947, 2016.

[439] F. P. Brooks Jr., *The Mythical Man-Month, Anniverary Edition.* Addison Wesley Longman, Inc., 1995.

[440] P. L. Lockwood, M. Hamonet, S. H. Zhang, A. Ratnavel, F. U. Salmony, M. Husain, and M. A. J. Apps, "Prosocial apathy for helping others when effort is required," *Nat. Hum. Behav.*, vol. 1, p. 0131, 2017.

[441] O. S. Bauge, "Dehyphenation: Some empirical methods," Master's thesis, University of Oslo, 2012.

Appendix A A Taxonomy of Intelligence

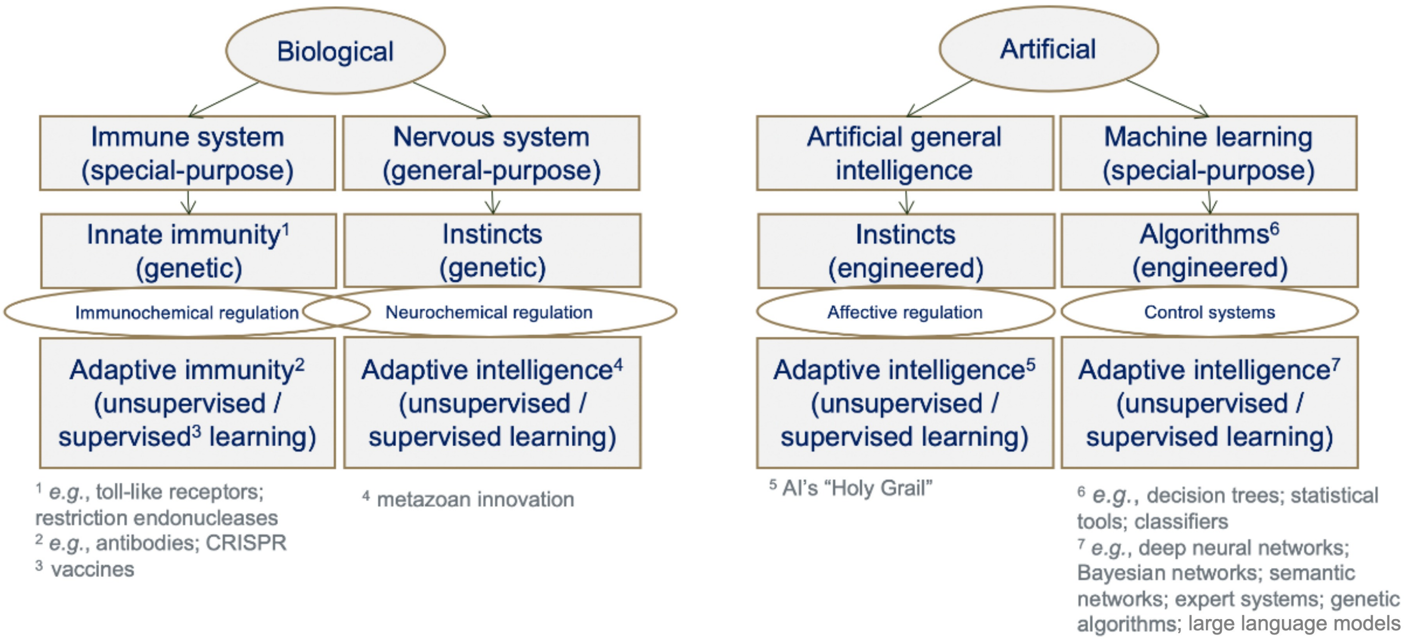

Figure 31: One can draw parallels between biological and artificial systems of intelligence.

Appendix B Overview of Syntheta's structural architecture

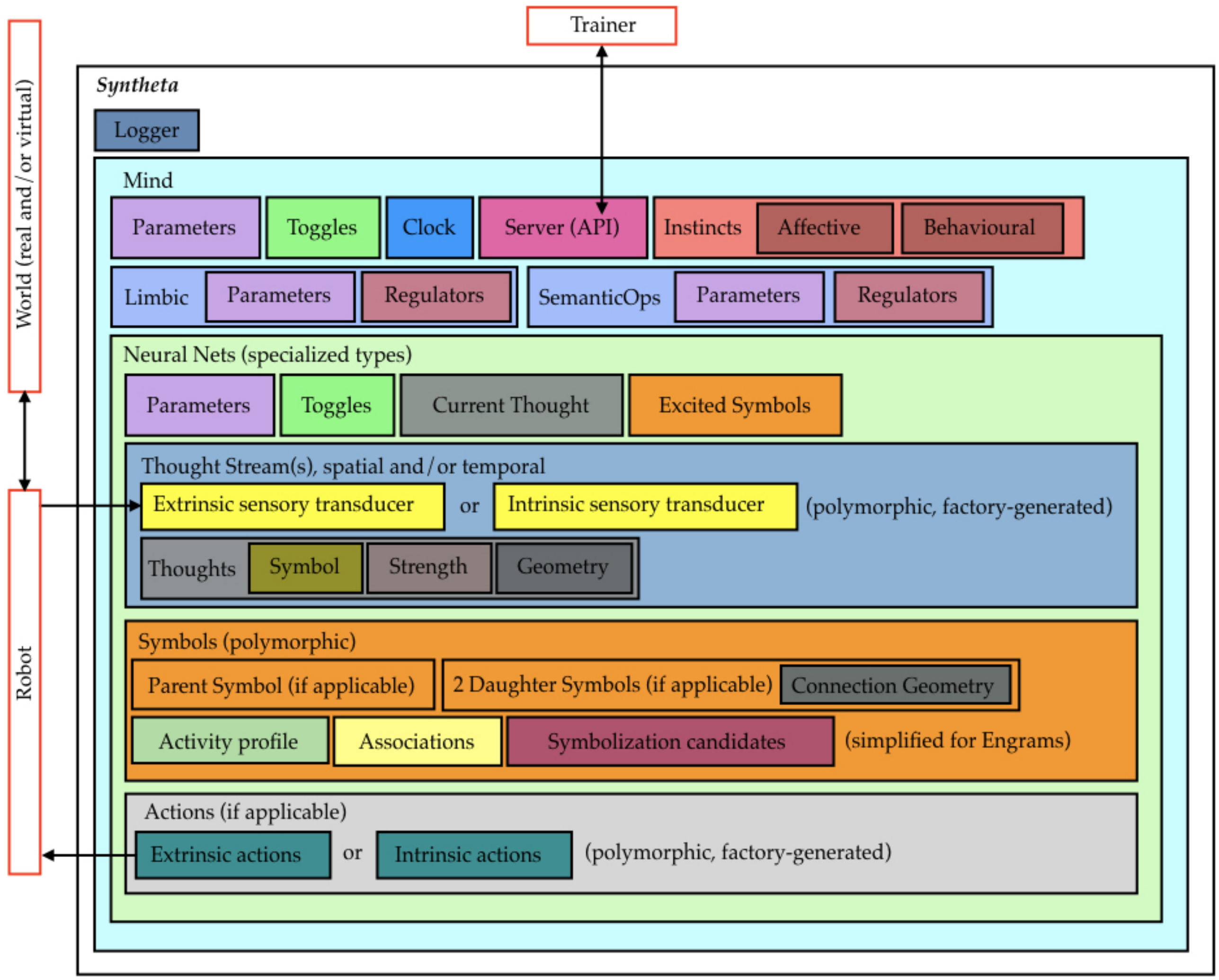

Figure 32: Data structures internal to Synthetha's implementation.

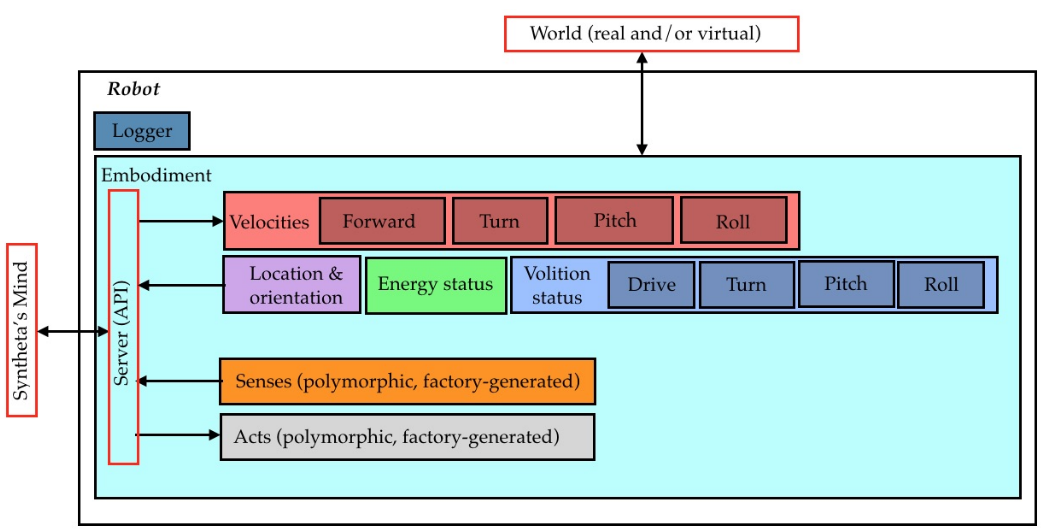

Figure 33: Data structures internal to a Robot's implementation.

Appendix C Module organization as semantic hierarchies

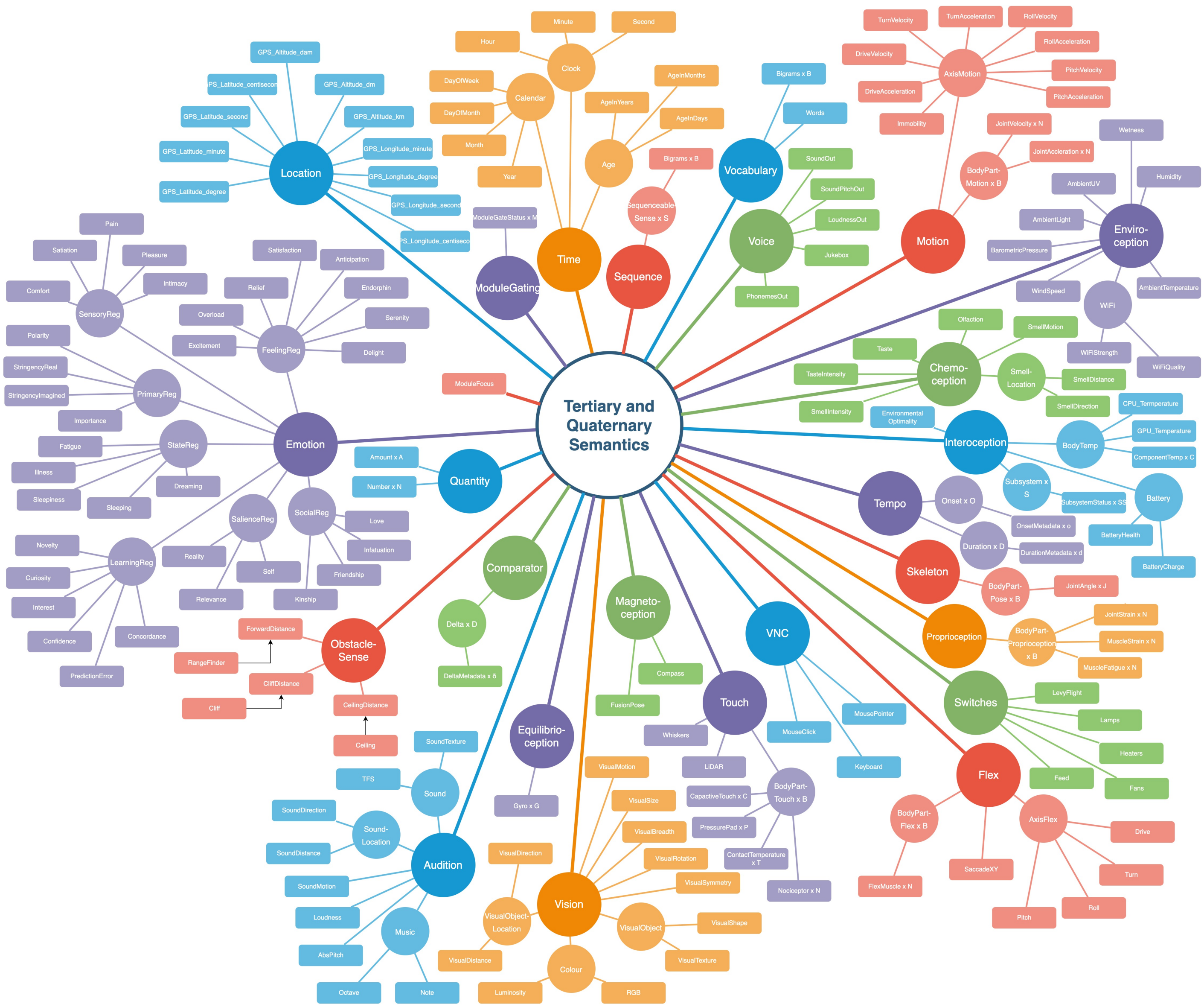

Figure 34: One of many possible configurations for the hierarchical organization of modules. Rectangles represent primary modules capturing trees of Symbols, whereas circles represent semantic modules capturing sets of Symbols. Some modules may not be applicable for a given implementation, and other modules may need to be added. “Tertiary and Quaternary Semantics” indicates additional hierarchical semantic organization (see Fig. 35), grouping and regrouping the secondary semantic modules (which are shown in bright colours); groupings at the tertiary level are not necessarily mutually exclusive. The quaternary level will likely be a single module, “Schema”. This example diagram will be revised as future functionality—as described in this treatise—comes online, and as tertiary and quaternary semantics are designed.

Figure 35: A possible tertiary and quaternary semantic configuration, building upon Fig. 34. This hierarchy limits the number of feeds to any semantic module, and groups concepts in ever higher abstractions. The smallest circles (with white text) are primary semantic modules and the larger circles with white text are secondary semantic modules, both from Fig. 34, as is the action module “SaccadeXY”. Added here are how they feed the tertiary semantic modules (coloured circles with black text) and ultimately the quarternary semantic module, “Schema”. The smaller white circle “Relevant 3° semantic module” represents any of the tertiary modules that might benefit from subsystem status information, temporal metadata, comparative metadata, or sequences. For example, a tune is a sequence of notes with a given tempo, and should feed into “Aural”. Similarly, the malfunction of a motor should feed into both “Agency” and “Body”.

Appendix D Application franchises made possible from the POCs

Proof-of-concept timeline	Funding year 1		Funding year 2		Funding year 3		Funding year 4	
	H1 Cybernetics	H2 Learning	H1 Hearing	H2 Vision	H1 Communication	H2 Embodiments	H1 VNC	H2 Socialization
Estimated minimum R&D FTE to complete POC (plus contract developers)	½ Chief architect ½ Scene-graph SME	½ Chief architect ½ Machine learning SME	½ Chief architect ½ Musician 1 DSP SME	½ Chief architect 1 Computer vision SME	½ Chief architect 1 Linguist	½ Chief architect 1 Roboticist	½ Chief architect ½ VNC SME	½ Chief architect 1 Psychologist
Franchises (possible after POC)	-Robotic control; -Process control	-Data analytics; -Root cause analysis; -Trending; -Process monitoring	-Process monitoring; -Security monitoring; -Computer-generated sounds	-Process monitoring; -Security monitoring; -Facial recognition; -Object recognition (e.g. for self-checkout stations); -Computer-generated imagery (CGI)	Natural language understanding for: -Literature research; -Query engines; -Threat monitoring; -Document drafting; -Proofreading -Help desks	Control of robots of any design for: -Transportation -Search and rescue; -Exploration; -Execution of complex tasks	Computer control for: -Office work; -Creative design; -Task automation; -Interfacing with remote systems via their API	-Personal assistants; -Companions; -Collaborative robots
Cumulative expertise needed for franchise R&D (plus product developers)	Cybernetics; process engineering	Statistics and data modeling; mathematics; computer science	Music; teaching; security engineering; DSP, mathematics	Graphic arts; teaching; security engineering; mathematics	Information science; linguistics; teaching; security engineering	Robotics; virtual reality; augmented reality; physics; teaching	VNC; teaching	Psychology; teaching; human-machine interactions
Potential markets (requires business development partners)	Factories; biotechnology	Factories; biotechnology	Security; entertainment	Security; entertainment; retail	Academic research; legal research; knowledge management; help desks; security	Autonomous vehicles; emergency response teams; natural resources; space exploration; factories	Offices; robotics manufacturers	Assisted living; housework; pet care; toy industry; video game AI

Figure 36: Business plan, focusing on scientific and engineering resources; franchise business partners are implied in this table. API: application programming interface; DSP: digital signal processing; FTE: full-time equivalent; H1: first half of the funding year; H2: second half of the funding year; POC: proof of concept; SME: subject matter expert; VNC: virtual network computing. Contract developers would be responsible for programming, code review, unit testing, system integration testing and user acceptance testing, under the chief architect's direction and guided by the subject matter experts.

D.1 Bayesian inference franchise

An expert system leveraging Syntheta's capabilities for Bayesian inference, $P(S|C)$, returns the probability of a signal S being seen, given the context C. It makes heavy use of confidence and concordance measurements (§3.4.10.2), learned from experience, in order to decide if an outcome is:

1. Surprising: Expected S but did not observe S.
2. Unsurprising: Expected S and observed S.
3. Unexpected: Observed S but did not expect S.

This can be used, for example, to construct context-aware decision trees, or to conduct root-cause analyses of process failures. Starting with anectodal evidence and hence low confidence, rounds of data collection and assessment can strengthen the model, driven by concordance-based refinement. Given Syntheta's transparent self-annotating associative maps of its concept space, decisions not only report their confidence, but also their underlying reasoning.

A franchise built for Bayesian inference need not be a real-time system. Syntheta can be fed a dataset, appropriately formatted into Module-specific sensory input, *e.g.* using the Trainer app (§4.1.2), with its confidence and concordance regulators then queried. As a statistical tool, this may suffice, but one can go one step further and supply feedback from subject matter experts after they have completed a follow-up assessment, refining Syntheta's world view. Finally, Syntheta's knowledge base would be saved, ready for the next analysis. As a product, such an expert system can take the form of standalone software owned or licensed by a client, or as software-as-a-service managed on the cloud.

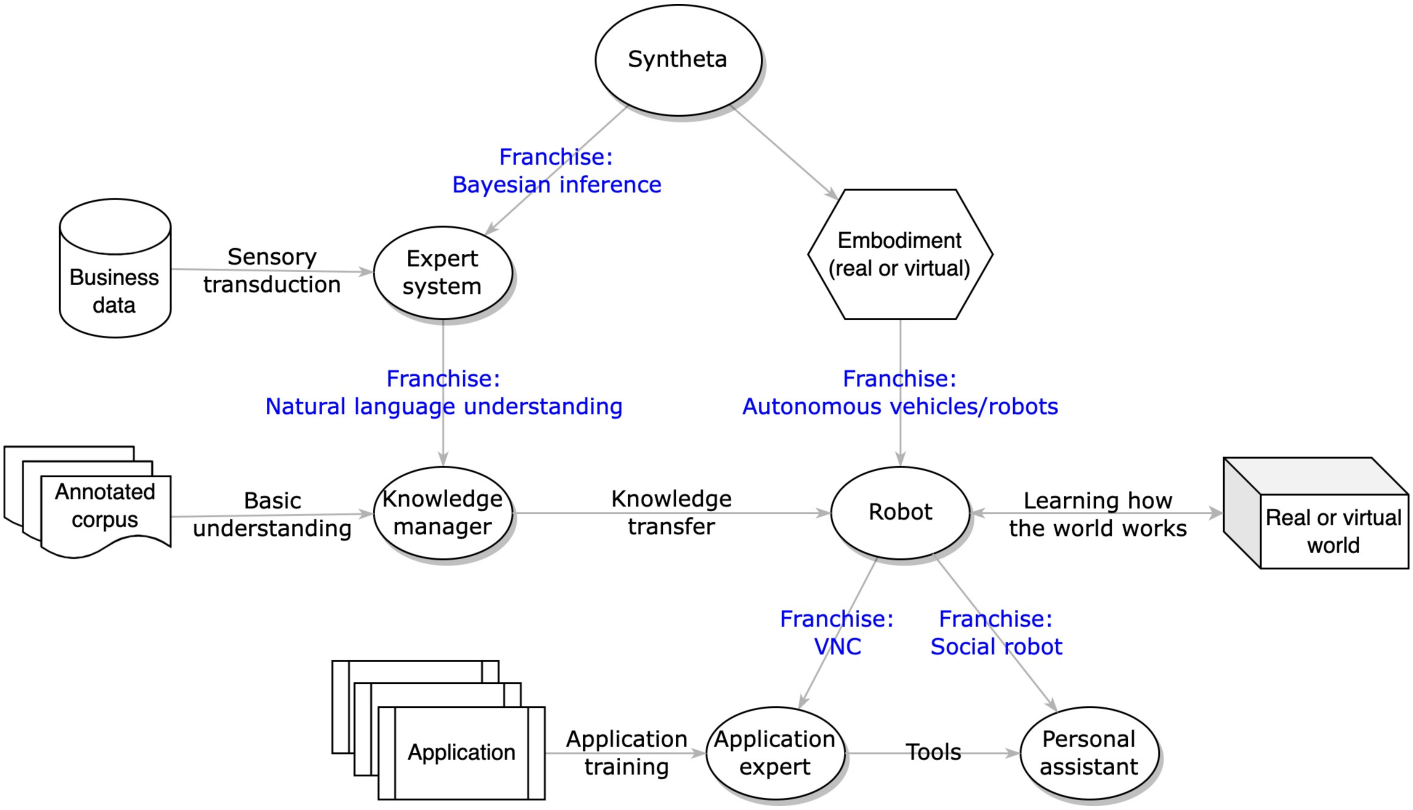

Figure 37: Hierarchical franchises, illustrating a possible business development pathway from simpler and more tractable use cases through to increasingly complex and sophisticated applications.

D.2 Natural language-understanding franchise

A product based on perceptual learning, as implemented in Appendix D.1, is the simplest of applications for artificial general intelligence, overlapping with what machine learning can do, but perhaps more efficiently. Where AGI can really shine, is in applications that require an *understanding* of the problem. As described in §5.2.5 and §5.3, a partial understanding of language can be built by feeding words, configured as concept clouds, to Syntheta for it to capture the gist of the passages, and of the queries about those passages, that it is fed. Even such an incomplete, non experiental, understanding of text can serve to build and maintain a network of linkages in a corpus of documents, and the concepts within them. This captures, retains, digests and delivers business knowledge, mitigating the loss of such tacit knowledge from personnel turnover, and increasing the return on investment for each of the business's projects, individually and collectively.

Like for Bayesian inference, natural language understanding need not be a real-time application: text can be fed, and the corpus can be queried, in discrete sessions. Similarly, such a product can be delivered either as standalone software or in the form of online software-as-a-service.

D.3 Autonomous embodiment franchise

Where things start to get really interesting, with respect to business opportunities, is when Syntheta's full set of capabilities is leveraged, in the form of an autonomous robot or vehicle (see §5.2). Iterative development of such a robot can spawn product franchises with increasing sophistication, as the robot learns how the world works, and as it learns to communicate (in part, by merging with the NLU product described in Appendix D.2), as it learns to operate computer programs (via virtual network computing), and as it learns to interact with other robots and with people (Fig. 37).

Product franchises are meant to be hierarchical and composable, based on product configuration and training. In §5.2, the development of an autonomous virtual robot is described, as a series of proofs of concept. Syntheta can just as easily control a real robot, interfacing with the robot's application programming interface. A number of physical robots exist in the marketplace, with a number of goals and objectives, some of which could benefit from having an understanding of what they doing. Training the robot in a virtual world initially, can help establish much of its understanding of basic concepts, before having its mind transplanted into the physical machine for it to control. Merging in language and/or business knowledge into the robot's mind can then accelerate its development, ultimately to know how to apply that business knowledge to help out where needed, as a personal assistant.

The form of an embodied Syntheta, in contrast to products focusing on Bayesian inference or pure language processing, would be real-time with full sensory immersion and responsive actions. That kind of software product would be expected to be on-board its physical (or virtual) device, or at least connected to it via high-speed, low-latency networking.

Appendix E Module configuration

The following module configuration styles are meant as examples; other and/or different configuration styles are possible as well.

E.1 Action Modules

E.1.1 Named Concept Modules

Indexed by name.
Style: Named actions.

E.1.2 Pose Modules

Setting normalized to [0,1], representing a pose within the robot's range of motion, for that "joint".
Style: Poses.

E.2 Sensory Modules

E.2.1 Named Concept Modules

E.2.1.1 Bigram sequence modules

Indexed by name for primary named concepts, by compound Symbol ID for bigrams.
Style: Concept sequences.

E.2.1.2 Ontological Modules

Indexed by name.
Style: Named concepts.

E.2.2 Periodic Modules

Data normalized to $[0,2\pi)$ from $[0,P)$, and decorated with a reference line segment, where P is the maximum number of divisions for the dial (*e.g.*, Month: 12; DayOfMo: 31; DayOfWk: 7; Hour: 24; Minute: 60; Second: 60; Angle: 360).

E.2.2.1 Spatial Periodic Modules

Style: Shapes.

E.2.2.2 Temporal Periodic Modules

Style: Line-segment sequences.

E.2.3 Doubly Periodic Modules

Data normalized to $x, y = ([0,2\pi),[0,\pi))$ from $([0,X),[0,Y])$, and decorated with a reference line segment, where (X,Y) is the maximum number of divisions for each of the two respective dials (*e.g.*, for two-dimensional direction, Azimuth: 360° and Inclination: 180°).
Style: Shapes.

E.2.4 Skewed Magnitude Modules

Magnitude normalized to $[0,1)$ from $[0,M)$, and decorated with a reference line segment, with higher resolution near 0, and greater overall resolution with higher values of M.
Style: Shapes.

E.2.5 Scaled Magnitude Modules

Magnitude scaled to $[0,1)$ from $[0,M)$, with equal resolution throughout the range. M can, if desired, be increased (*e.g.*, such as for the concept of calendar year), requiring a rescaling of values to the new range.

E.2.5.1 Spatial Magnitude Modules

Style: Shapes, including decoration with a reference line segment.

E.2.5.2 Temporal Magnitude Modules

Style: Line-segment sequences.

E.2.6 Quantitative Signal Modules

Indexed by preconfigured sensor name or index for primary Symbols and by compound Symbol ID for combinations.
Style: Compound shapes.

E.2.7 Shape Modules

Indexed using index 0 for the primary line segment, and by compound Symbol ID for patterns.

E.2.7.1 Spatial Shape Modules

Style: Shapes

E.2.7.2 Temporal Shape Modules

Style: Shape sequences.

E.3 Memory Trace Modules

Indexed by ID obtained via symbolization of new engrams.
Style: Composite.

E.4 Semantic Modules

Indexed by ID obtained via symbolization of new sets.

Style: Composite.

E.4.1 Semantic Category Modules

Symbolization: New sets, capturing excited engrams or excited module names.

E.4.2 Grammar Modules

Symbolization: New sequences, capturing excited sets of module or module-group names.

E.5 Working Memory Gate

Controls entry into, execution of, and exit from, working memories.

Style: Named concepts.