The Predictive Basis of Situated and Embodied Artificial Intelligence

Keith L. Downing
Complex Adaptive Organically-Inspired Systems Group (CAOS)
The Norwegian University of Science and Technology
Trondheim, Norway
keithd@idi.ntnu.no

ABSTRACT

While classic AI systems still struggle to properly incorporate common-sense knowledge, Situated and Embodied Artificial Intelligence (SEAI) aims to build animats that acquire a common-sense understanding of the world via interactions between simulated brains, bodies and environments. Neuroscientists believe that much of this common sense involves predictive models for physical activities, but the transfer of sensorimotor skill knowledge to cognition is non-trivial, indicating that SEAI may meet a daunting challenge of its own. This paper considers the neurological basis for procedural common sense and the possibilities for its transfer to conscious reasoning. This helps assess the prospects for SEAI to eventually surpass classic AI in the quest for generally intelligence systems.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: General

General Terms

Artificial Intelligence theory

Keywords

situatedness, embodiment, neural networks

1. INTRODUCTION

Classic AI systems, often called GOFAI (Good Old-Fashioned AI) systems, generally rely on the manipulation of ungrounded symbols under the strict constraints of mathematical abstractions such as logic and probability theory. These systems normally assume away all environmental and bodily factors to focus on cognition in a vacuum. This works well for chess but builds awkward robots. In fact, the lack of basic intuitions about body and world (i.e., common sense) was the downfall of many purely-cognitive GOFAI systems as well. Attempts to force-feed this general knowledge in a top-down manner into GOFAI systems failed miserably, since much of natural common sense exists not in platform-independent software, but in behavioral repertoires finely tuned to the structure and dynamics of body and environment.

Situated and Embodied AI (SEAI) researchers believe that GOFAI's most obvious and frustrating deficiency, common sense, comes only via the learned experiences of a body in a world. Whereas "I think, therefore I am" might have been an appropriate slogan for GOFAI, it's converse more aptly summarizes SEAI. That is, by living, we acquire common sense, which then supports more complex reasoning.

Although the SEAI philosophy is attractive, the cruel realities of robotics raise major obstacles. Whereas GOFAI began with a divergent radiation of impressive applications displaying many forms of (shallow) intelligence, SEAI seems to have converged on a menagerie of wall-following robots, all of which have very deep, functional (albeit implicit) understandings of their own body and domain: a barren floor surrounded by walls. There are some interesting exceptions, such as Robocup teams, floor-sweeping and lawn-mowing robots, etc., but to date, sensing and acting have not produced common-sense scaffolding for cognitive activities [19], such as *planning* [11] or mathematical reasoning [13].

After 20 years of SEAI, one expects more. GOFAI adherents can arguably write-off SEAI as overly-optimistic biological envy, but SEAI supporters counter that any other starting point for intelligence is ad-hoc and doomed to run aground. This research looks to neuroscience in assessing the long-term potential of SEAI, with special focus on predictive knowledge and the barriers to (and possible avenues for) its transfer to cognition.

2. BODIES, BRAINS AND PREDICTION

Neuroscientists generally agree that brains evolved to support complex motion; stationary organisms do not need them. Llinas [14] elaborates on this tie between motion and cerebral development and evolution, giving the example of a sea squirt, which is mobile during its early life stages but later becomes sessile, whereupon it digests its own brain!

In addition, brains appear to grow, during development, to fit their bodies. According to Edelman's theory of Neural Darwinism [8], neurons compete for synaptic connections during both development and learning, leading to *survival of the best networkers*. Deacon [6] uses Neural Darwinism as the basis for his Displacement Theory (DT), wherein

GECCO'05, June 25–29, 2005, Washington, DC, USA.

the networking competition during development produces brains scaled to fit the body's sensory and motor apparatus.

Weaving together the work of these 3 neuroscientists provides an interesting account of the origins of intelligence. Briefly, evolution has discovered the brain as a solution to the movement-control problem, since the emergent oscillatory dynamics of coupled muscles has severely limited complexity. It may suffice to pump blood through a multi-chambered heart, but it cannot control arm movements during tree climbing. Higher and higher layers of control evolved to realize more advanced sensing and acting. Throughout this ascent in complexity, brains redimensioned to fit evolving body types via an internal competition for cranial space and synaptic connections. In the transition from large-bodied apes to humans, the massive reduction of sensory and motor targets combined with a relatively constant cranial size reduced the demand for primary sensory and motor neurons, leaving extra space for higher-level structures, such as the prefrontal cortex (PFC), which appears to be a key prerequisite to symbolic reasoning and cognition. Of key relevance to SEAI, the PFC is also the highest level of motor control.

However, the symbol-processing brain is merely a re-dimensioned sensorimotor brain, partially exapted for cognitive endeavors. The complexities of sensing and acting are certainly no less problematic for us than for our distant ancestors, so the brain is still a sensorimotor controller, but with impressive reuse possibilities. The key questions for SEAI concern *what* is reused and *how*. One intriguing possible answer is *predictive knowledge*.

During motion control, delays in sensory feedback can cause regulatory instability, and living organisms have sensory processing machinery that is too slow to complement their motor abilities. To combat this mismatch, neuroscientists and control theorists agree that the brain needs predictive models [20, 14]. Given a current sensorimotor context, these predict the next context and use it to calculate the feedback error signal. Since the predictor runs internally, it produces an estimated future state long before the sensory system can provide the actual state. Areas such as the cerebellum, basal ganglia and hippocampus are often cited as subconscious centers for the acquisition and use of these models, which neuroscience often posits as the basis of common sense [17, 14]. However, the neural links between our explicit, verbal understanding of causal concepts such as *drop* and *shatter* and our subconscious *feel* for them are unclear. Yet, it seems that the ultimate success of SEAI's vertical scaling to cognition critically depends upon the existence of these ties, or, at least, the possibility of realizing them in artificial intelligences.

Prediction, which Llinas calls the ultimate function of the brain [14], is therefore the linchpin in a very tempting, motivating argument for continued SEAI research:

1. Complex movement requires an ability to predict the immediate future.

2. Prediction involves various functional mappings between and among states and actions.

3. These mappings constitute basic common sense.

4. Thus, the demands of movement provide the basis for cognition.

Although superficially straightforward, the neurological foundations for this line of reasoning are rather unstable, as discussed below.

3. MODELS AND MEMORIES

Though SEAI began with Rodney Brooks' [4] direct assault on GOFAI, with the battle cries *intelligence without representation*, and *the world is its own best representation*, two decades of experience with robotic embodiments of pure behaviorism reveal critical limitations of representation-free minds. In mammals, neuroscience has discovered strong correlations between neural states and rich sensorimotor contexts [5, 2], thus indicating models of some form. But where are they and how are they acquired?

As a starting point, neuroscientists differentiate between two types of memories: declarative and nondeclarative (or procedural) [17]. A good deal (possibly all) of explicit, conscious human knowledge resides in the neocortex, particularly the higher-level association regions of the parietal, temporal and frontal lobes. These declarative memories are of either a) specific objects or situations (i.e., episodic) or b) general concepts (i.e., semantic). Declarative memories are easily formed from single-exposure incidents, often those of emotional significance. However, the consolidation process is far from a snapshot-and-cache scenario. Rather, the hippocampus (HC) appears to store a reasonably rich version of the snapshot and then gradually (over the course of days,weeks or even years) off-loads the memory back to those neocortical areas that stimulated the HC during the original experience.

Procedural memories are directly tied to sensorimotor activity. This knowledge is implicit in the sense-and-act machinery and appears inaccessible to consciousness, yet, it is believed to be the basis of our intuitive (i.e. common sense) understanding of the world [17]. Whereas declarative memories are processed in the HC and later off-loaded to the cortex, procedural memories are learned *in place*, in areas such as the basal ganglia, cerebellum, and amygdala, along with sensory and motor cortices. Typically, procedural memories require multi-trial learning.

From the standpoint of acquisition, the two memory types are intuitive. A declarative (particularly an episodic) situation may only arise once, yet survival can be greatly enhanced by storing (at least a very abstract) representation of it. For example, the memory of a single glimpse of a tenacious predator can govern a lifetime of avoidance behavior. Conversely, procedural skills lend themselves to frequent rehearsal; i.e., the situation arises many times and need not be cached in an HC-like organ. Here, the world really can be its own best representation and the neural sensorimotor areas can gradually adapt to the recurring world context on their own. Procedural skill learning includes operant and classical conditioning, and sequence learning. Interestingly, it also includes many forms of categorization: we can form many sensory classes without using the HC. Thus, many concepts in the brain are neither hard-wired, nor explicit, but learned directly by cognitively inaccessible areas of the sensory cortex [17].

At the cellular level, the learning of declarative and nondeclarative memories is quite similar [12, 17], but from a systems neurological perspective, they differ dramatically. Specifically, two key areas for procedural learning, the cerebellum and basal ganglia, perform supervised and reinforce-

ment learning, respectively, while the hippocampus and cortex realize unsupervised associative learning of declarative information [7].

4. THE SENSORIMOTOR HIERARCHY

When viewed as a sensorimotor controller, the brain exhibits a clear hierarchy of tightly interconnected modules, as shown in Figure 1. For simple reflex actions, signals travel from sensors to the spinal cord and then immediately back to the muscles. Activities involving proper timing or finesse often call on the cerebellum, which recommends actions based on learned associations between complex sensory contexts and motor responses that often enlist many muscles and body parts. Further up the hierarchy, the basal ganglia select and sequence high-level contexts that represent enduring (for seconds or minutes) sensorimotor or cognitive states. Finally, the neocortex, comprised of sensory, motor and frontal areas, provides long-term storage for the contexts that trigger basal gangliar and cerebellar loops.

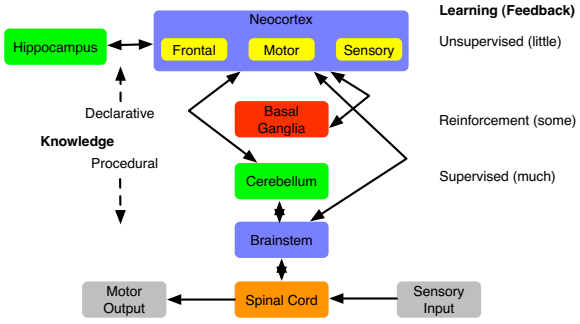

Figure 1: Mammalian sensorimotor control hierarchy

In general, higher points in the hierarchy are more consciously accessible, while lower modules perform unconscious acts. However, at least the 3 highest layers are strongly involved in both sensorimotor and cognitive activity. Hence, a lot of cognition a) utilizes the classic sensorimotor machinery, but b) is beyond conscious monitoring or control.

With respect to common-sense and predictive knowledge, their localization is highly ambiguous in the neuroscientific literature, due to both cross-experimental and cross-species differences. In general, we consider predictive knowledge as that involving associations between body-world states/contexts and other such states and/or actions. For example, knowledge that one context normally follows another is predictive, as are mappings from state-action pairs to consequent states.

4.1 The Cerebellum

Approximately half of the human brain's neurons reside in the cerebellum, which has long been known for its vital role in the learning and control of complex motions [12, 2]. As shown in Figure 2, the cerebellum receives convergent inputs from the sensory and high-level cortices. These are transferred from mossy fibers to granular cells, whose axons form parallel fibers (PFs) along the outer layer of the cerebellum. Purkinje cells (PCs) then read the parallel lines, with 10^5 - 10^6 synapsing on each PC dendritic tree. When PCs fire, they inhibit cells in the deep cerebellum, thus blocking or reducing particular muscular contractions.

Each PC receives signals from a single climbing fiber (CF); each CF contacts 1-10 nearby PCs. CFs relay signals from the inferior olive, which is stimulated by somatosensory inputs such as touch and temperature. The sensory afferents of an inferior olive cell are located near the muscles controlled by the PCs of that olivary cell's climbing fiber. Hence, the CF gives feedback directly related to the local action controlled by its PC. For example, if a muscular contraction causes a nearby joint to rotate excessively, the pain signal from the joint to the CF (via the inferior olive) would train the nearby PCs to reduce the strength of future contractions.

Plasticity at the CF-PC synapse relies on post-synaptic long-term depression (LTD) [2]. When a CF forces a PC to fire strongly, those PC dendrites that were recently activated by parallel fibers undergo chemical changes that reduce their sensitivity to glutamate (the neurotransmitter used by PFs). Hence, the influence of those PFs on the PC declines.

Although there is good topographic correspondence between CF-PC pairs and the body areas that they serve, an individual PC does not control a single muscle, but, via its indirect connections to the motor cortex, is one of many influences to several higher-level neurons, each of which affects several muscles. Thus, the learning induced by a single climbing fiber involves a graded behavioral change in several functionally-related muscles, and proper motion control emerges from a multitude of these microscopic adjustments.

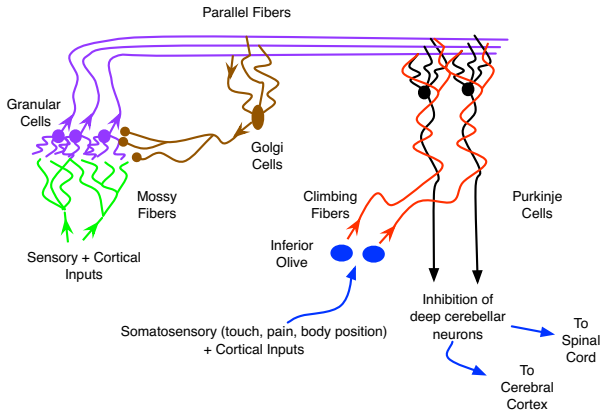

Figure 2: Basic organization of the cerebellum.

In general, cerebellar decisions are strongly contingent upon low-level sensorimotor feedback, and cerebellar plasticity is a) frequent, b) governed by simple sensory signals, and c) targeted toward a small set of Purkinje cells. Due to this high update frequency and locality, cerebellar learning is often considered supervised [7].

Several neuroscientists postulate causal models in the cerebellum, as summarized in [20]. In a nutshell, the cerebellum consists of many modular tracts or microzones, and each is believed to encode both a forward and inverse causal model related to a specific situation. The forward model computes expected future states when given the current state and action, while the inverse model computes an action when given a desired future state. As a feedforward controller, the cerebellum utilizes inverse models to provide motor-response recommendations. As a feedback controller, it uses predicted future states from the forward model to

compute predicted errors, which are then used to generate the next round of motor signals. In familiar situations, the forward model circumvents the need for actual sensory feedback, whose time-consuming processing causes delays that degrade regulation. Also, [20] sees a cooperative arrangement in a model pair whereby the inverse model whose corresponding forward model provides the most accurate prediction of the future will, in turn, have higher precedence among all the inverse models when recommending the next action.

4.2 The Basal Ganglia

Animals gain a selective advantage from learning associations between bodily and environmental indicators and value-laden results, e.g. rewards or punishments. By predicting desirable or dangerous situations from their antecedent clues, animals can behave proactively, instead of merely reactively, to enhance survivability.

The basal ganglia (BG) are widely viewed as the center of this *reinforcement learning* (RL) in mammalian brains [7]. Sketched in Figure 3, the BG are large subcortical structures that receive convergent inputs at the striatum from many cortical areas. The striatal cells appear to function as a layer of competitive context detectors, since a) each neuron receives inputs from circa 10,000 cortical neurons, b) their electrochemical properties are such that they only fire if many of those inputs are active, and c) they have inhibitory connections to other nearby striatal neurons. The basal ganglia appears to involve many parallel loops, the great majority of which involve the prefrontal cortex (PFC)[12, 9].

Striatal modules consist of striosomal cells surrounded by matriosomal cells. The former send outputs to the Substantia Nigra (SN) either directly or indirectly via the STN. Conversely, the matriosomes send signals to the pallidum, again directly and indirectly. In Figure 3, notice that the direct paths are inhibitory, while the indirect are excitatory. Several researchers [1, 10] characterize the BG as a combination of actor and critic, with the matriosomes and pallidal neurons as the actor's input and output ports, respectively, while the striosomes and substantia nigra demarcate the critic.

Essentially, the BG map contexts to other contexts, where each context may contain current sensory inputs, expected sensory inputs and/or intended actions. When a context-detecting matriosome fires, it inhibits a few downstream pallidal neurons. In contrast to the striatum, the pallidum consists of low-fan-in neurons, most of which are constantly firing and thereby inhibiting their downstream counterparts in the thalamus. When a striatal cell inhibits a pallidal neuron, this momentarily disinhibits the corresponding thalamic neuron, which then excites a cortical neuron, often in the PFC. The cortical excitation links back to the thalamus, creating a positive feedback loop that sustains the activity of both neurons, even though pallidal disinhibition may have ceased. Thus, the striatal-pallidal actor circuit momentarily gates in a response whose trace may reside in the working memory of the PFC for seconds or minutes.

The PFC is the highest level of motor control. Its firing patterns influence activity in the pre-motor (PMC) and motor (MC) cortices, while the MC sends signals to muscles via the spinal cord. In addition, sustained PFC activity provides further context for the next round(s) of striatal firing and pallidal inhibition that embody context detection and

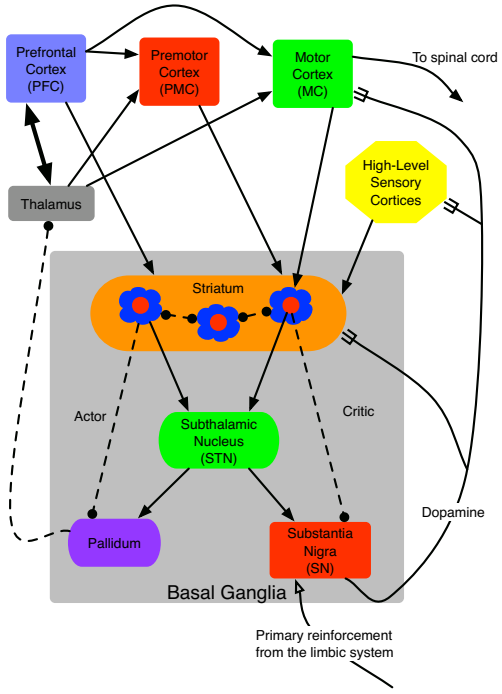

Figure 3: Basic topology of the basal ganglia and its main inputs. Solid lines with arrows denote excitatory links, while dashed lines with circular heads are inhibitory. Forked heads denote diffuse neuromodulator (i.e. dopamine) transmission.

action selection, respectively. Also, a BG-activated context may embody a desired future state, thus triggering an inverse model in the cerebellum to produce a recommended action. Via parallel recurrent looping of the BG and cerebellum, sequences of desired states and proposed actions are generated.

Of critical importance to the philosophical underpinnings of SEAI, the PFC is also the highest level of **cognitive** control. Hence, PFC activation patterns can affect both motor activity and thought processes. It serves as a blackboard where many neural regions can log indicators of current activity, which may then serve as inputs to other regions. Our conscious awareness of a motor sequence may depend upon PFC activation. As motor sequences become more automatic, their control is believed to shift from a PFC-dominated BG circuit to a loop in which the thalamic outputs go directly to the motor cortex [7]. So these *compiled* sequences are no longer accessible to conscious thought, although they still govern behavior. Conversely, more cognitive activity sequences, such as mental arithmetic, appear to depend upon BG-PFC loops [13].

The situation-action rules housed within the BG may comprise significant portions of our common sense understanding of body-environmental interactions, whether consciously accessible or not. Our smooth execution of both motor and cognitive tasks requires healthy BG. Major BG ailments, such as Parkinson's and Huntington's disease, cause significant cognitive impairments along with the physical deterioration [12, 2].

However, the source of Parkinson's disease, and the BG's key to reinforcement learning, resides in the critic circuitry (see Figure 3). Here, dopamine (DA) signals from the Substantia nigra (SN) influence the synaptic plasticity of the regions onto which they impinge. In unfamiliar situations, the SN fires upon receiving stimulation from various limbic structures, such as the amygdala (the seat of emotions), which triggers on painful or pleasurable experiences. The ensuing dopamine signal encourages the striatum to remember the context that elicited those emotions. Due to the biochemical temporal dynamics [10], the striatal neurons that become biased (i.e., learn a context) are those that fired circa 100 ms **prior** to the emotional response. Hence, the BG learns a context (C) that **predicts** the reinforcing situation (R): $C \Rightarrow R$. Furthermore, the topology of the critic network enables these predictions to recursively regress backwards in time, such that long sequences, $C_1 \Rightarrow \ldots C_n \Rightarrow R$, are learned. These sequences have obvious utility in both motor activities and cognitive processes such as planning.

Again, regarding the goals of SEAI, note that the sequences generated by the basal ganglia can be anything from a series of motor acts during cross-country skiing to the words of a song to the the steps of long division. The motor and cognitive sequences may be processed in parallel, non-interacting tracts, but they involve the same type of neural machinery.

In general, the basal ganglia are driven more by complex internal contexts than by immediate sensory feedback. Some of these contexts can function as plan segments (in an abstract sense) in that they involve enduring activation patterns in the PFC, where they can have strong effects upon the cerebellum, premotor and motor cortices, as expected of plans, goals, intentions, etc.

From this perspective, the learning differences between the two regions make perfect sense. Results of cerebellar decisions are immediate and short-lived, so frequent feedback (via the climbing fibers) is appropriate for assigning credit to the most recent choices. In contrast, the basal ganglia's context choices can have broader temporal and spatial consequences, so immediate feedback is of less utility than an occasional (dopamine) reinforcement that provides a more holistic evaluation.

4.3 The Hippocampus

Often viewed as the center of long-term memory formation, the hippocampus (HC) resides in the temporal lobe and receives inputs from a wide variety of cerebral regions. As shown in Figure 4, the HC and surrounding areas perform a drastic compression (via high convergence) of information between the neocortex and area CA3, and a complementary expansion (via divergence) on the return path through CA1 and Subiculum. The topology of the HC proper is a main loop with several shortcuts from EC to CA3 and CA1.

Only CA3 contains extensive recurrence, with each neuron connected to approximately 4% of the others. This indicates that CA3 performs associative learning by standard Hebbian means: neurons that fire together wire together. The high convergence from a diverse array of neocortical areas onto CA3 hints of the holistic nature of these patterns.

The hippocampus' importance to long-term memory formation is well-established, as is the fact that memories reside in the HC only until off-loaded to the cortex for more permanent storage. In rats, individual neurons in CA3 and CA1

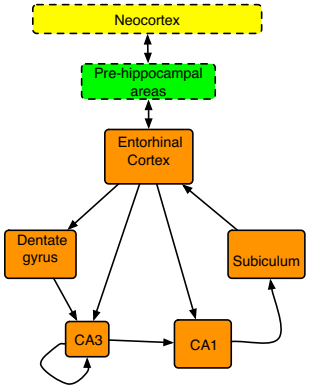

Figure 4: Basic topology of the hippocampus (solid boxes) and surrounding areas (dashed). Box widths (very roughly) illustrate relative sizes of neural populations in each area. All connections are excitatory.

are known as *place cells*, since they fire only when the animal is at a particular spot, while in monkeys, they are called *view cells*, since they fire when the primate merely looks at such a location [16, 5]. These findings have motivated a cottage industry of ANN models of HC-based navigation (see [5] for an overview). Many of these involve implicit predictive knowledge in CA3 and CA1, wherein place cells fire before the animal arrives at the corresponding location. Others posit CA3 as the site of predicted situations and CA1 as the site of real situations (via direct inputs from EC). Then, CA1-CA3 mismatches drive learning in CA3, yielding better predictions in the future.

Since particular memorable situations may only occur once, episodic memory formation must involve a snapshot mechanism. However, [15] points out that forcing a new pattern into an associative network, via extensive synaptic modification (corresponding to a high learning rate in an artificial neural network), can corrupt pre-existing memories. Instead, new patterns should be repeatedly presented to the network in interleaved fashion, with only small synaptic changes each time. The HC acts as the trainer for the neocortex by a) temporarily caching snapshots of episodes via very fast Hebbian associative learning in CA3, and then b) repeatedly *re-presenting* these patterns to the cortex until it too encodes the associations [16, 15].

5. THE BARRIER TO REUSE

One striking difference between the cerebellum and BG versus the cortex (and CA3) is the high density of excitatory recurrent intra-layer connections in the latter. These support associative memories in which a) partial patterns can be completed via spreading activation, and b) stable attractor firing patterns can emerge and persist. This stability seems prerequisite to the focus of attention underlying conscious cognition.

Conversely, the basal ganglia and cerebellum consist of thousands of parallel tracts through a series of layers which have inhibitory intra- and inter-layer recurrent collatorals. These areas are designed to map cortical contexts onto (motor or cognitive) acts, but not to hold patterns active. Using these two regions, one can sing a song perfectly but cannot,

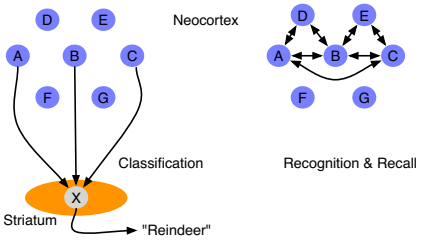

Figure 5: Synaptic connections for classification versus recall and recognition.

without the aid of external media (e.g. paper and pencil), recall and compare the 7th and 23rd lines. Each word or phrase of the song may be stored in the cortex, but extraction is mediated by the combination of preceding cortical context (declarative) and basal gangliar wiring (procedural). There are similar restrictions on common-sense skill knowledge for writing, riding, throwing, etc. Essentially, the glue that holds pieces of complex cortical patterns together (across time) resides in a procedural bottleneck - only a few striatal neurons fire simultaneously - such that the whole pattern cannot be activated and analyzed as one stable cortical image. Hence, our causal knowledge may involve disjoint abstract cortical snapshots whose combinations generally evade conscious contemplation.

Convincing evidence of the strong barrier between declarative and procedural knowledge comes from patients with hippocampal damage (i.e., amnesiacs), who can learn a wide variety of complex procedural tasks as well as normal patients, but who have no awareness of what they have learned [17]. For example, they may learn to classify a set of training cases, but afterwards, they cannot *recognize* any of the cases.

As shown in Figure 5, classification often occurs in the striatum of the basal ganglia, where striatal cells learn invariants among similar contexts via a competitive learning process. However, categorization can occur very quickly, without the formation of a stable, consciously-accessible pattern in the neocortex. In the figure, assume 2 examples of reindeer have feature combinations ABCDE and ABCFG, where all features have detector neurons in the neocortex. Compare the cortical-striatal connections needed to classify a reindeer - assuming invariant features ABC (left) - to the intra-cortical connections needed to complete and stabilize a memory of example ABCDE during recognition or recall (right). The neocortical connections manifest declarative knowledge in an associate memory and require a hippocampus for proper synaptic tuning.

In this analysis, one cannot view declarative and procedural memories as the sole provinces of cognition and sensorimotor behavior, respectively. For example, one of the procedural tasks on which amnesiacs and controls perform equally well involves dynamic staffing of a simulated sugar factory to achieve an optimal production level . This clearly qualifies as cognitive, since it involves non-trivial arithmetic reasoning. But, somewhat surprisingly, it lacks a strong declarative component and thereby remains fully accessible to amnesiacs [17].

In general, brain imaging reveals high activity in procedural areas such as the basal ganglia and cerebellum during pure thinking tasks.

Similarly, sensorimotor tasks surely require the long-term storage of certain explicit contexts, particularly of detailed sensory cues that help initiate tasks, after which the brain may run in a purely procedural auto-pilot mode. Consider a predator-prey example. It does not behove prey to launch into a full-fledged escape movement, thus giving away their presence and location, based merely on a reflexive reaction to a low-level stimulus. Rather, a positive identification of a predator should proceed any dramatic action, and such recognition often involves declarative memory.

5.1 One-Way Passage

Although GOFAI could not successfully tunnel through the barrier between declarative and procedural, human brains apparently can. Skills begin as conscious, declarative activities involving the frontal cortices and often become *compiled* into faster, but less flexible procedural routines in the basal ganglia, cerebellum and motor cortex. Is it unreasonable to imagine traffic in the other direction: from procedural to declarative?

Using the old computer metaphor of the brain (and considering that the compilation from high-level languages to machine language is many-to-one, hence non-invertable), one could argue that the mapping between neuronal patterns in subconscious sensory and motor areas and corresponding high-level patterns in the frontal cortices is one-to-many and thus nondeterministic.

However, this analogy misses one critical difference between computer programming and human skill learning: the nature of testing and debugging. In computer science, the programmer determines the overall task, judges errors, and modifies the code. Conversely, in animal skill learning, conscious activity determines the general context to which a procedural module is exposed, but the body and environment provide the detailed inputs. Error detection and correction may be conscious or unconscious, but the exact location and detailed nature of the changes is unconscious and intrinsic to the procedural circuits. These synaptic changes are based on specific local information plus vague global broadcasts indicating only that *something* significant just occurred. Thus, the forebrain cannot monitor procedural learning to any useful degree. Only by *observing* overt behavioral improvements can the conscious PFC confirm the success of its training regimes. This decentralized adaptation is critical to sensorimotor learning in the wild, and SEAI follows this emergent paradigm for good reason.

Basically, evolution designed brains this way. Higher-level cortical areas arose to *enhance* procedural activity, not to understand or explain it. Unfortunately, this provides a disconcerting precedent for SEAI.

6. PROSPECTS FOR SEAI

Given its direction of approach, SEAI may have an even harder time crossing the barrier than did GOFAI. Although natural evolution shows that brains evolved to control sophisticated sensorimotor activity, and that cognition arose by borrowing that same machinery, there is little evidence of direct reuse of procedural knowledge for declarative purposes. But, in the very least, by understanding the neural processes of sensorimotor control, we should have a good

biologically-based start in building cognitive controllers by exaptive means.

Since by definition SEAI systems must first crawl before they can contemplate, the designed or evolved architectures will naturally be optimized for sensing and acting, and as nature reveals, this entails a daunting cognitive impenetrability of the acquired common sense. Conceivably, transfer may occur via the environment: the agent could perform actions, observe results, and inductively form declarative causal representations. Observing other agents would also work, but clearly, the somatosensory feedback involved in ones own activity has powerful effects on both declarative and procedural memory formation. For example, when a particular movement causes pain, we often do consciously attend to the situation and learn explicit heuristics.

Societial approaches, as explored in [19], seem promising, since if animats must evolve to both perform tasks and to transfer behavioral tips, then their brains will not necessarily become optimized for solipsism. The demands of communication may force an early coupling between declarative and procedural knowledge such that common sense could in fact transfer directly from procedural to declarative realms.

Also, Squire and Zola [18] observe that subjects with functioning hippocampi form a parallel, auxiliary declarative representation (of no immediate performance benefit) while doing a purely procedural task. SEAI systems might utilize a similar mechanism, wherein both competitive classifiers and associative memories (see Figure 5) are tuned during sensorimotor adaptation,

Another direction involves tasks such as the sugar-factory controller, which are cognitive but largely procedural. These provide an interesting variant of minimally-cognitive tasks [3]: minimally *declarative* cognitive tasks, i.e., hard thinking tasks that are not representation hungry.

Of course, in the end, SEAI may simply reach the opposite bank of a wide chasm that GOFAI discovered 20 years ago: knowing how and knowing about are non-interchangeable. However, one clear advantage of the low road to AI is that charting the neural processes for sensorimotor control provides powerful leverage for understanding cognition, since both appear to use similar cerebral machinery.

Summing up, SEAI appears motivated by an implicit belief that bottom-up sensorimotor agents will eventually scale to general intelligences, since worldly experience is the only way to achieve GOFAI's Achille's heel, common sense. However, nature has stumbled upon a formidable barrier between procedural and declarative faculties, and SEAI, whether biologically-inspired or not, may be forced to follow many of nature's moves in design space and thus must settle for systems that, like ourselves, *have* common sense but cannot always analyze or articulate it.

7. REFERENCES

[1] A. Barto, *Adaptive critics and the basal ganglia*, in Models of Information Processing in the Basal Ganglia, J. Houk, J. Davis, and D. Beiser, eds., Cambridge, MA, 1995, The MIT Press, pp. 215–232.

[2] M. Bear, B. Conners, and M. Paradiso, *Neuroscience: Exploring the Brain*, Lippincott Williams and Wilkins, Baltimore, MD, 2 ed., 2001.

[3] R. Beer, *The dynamics of active categorical perception in an evolved model agent*, Adaptive Behavior, 11 (2003), pp. 209–243.

[4] R. Brooks, *Cambrian Intelligence: The Early History of the New AI*, The MIT Press, Cambridge, MA, 1999.

[5] N. Burgess and J. O'Keefe, *Hippocampus: Spatial models*, in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed., The MIT Press, Cambridge, MA, 2003, pp. 539–543.

[6] T. Deacon, *The Symbolic Species: The Co-evolution of Language and the Brain*, W.W. Norton and Company, New York, 1998.

[7] K. Doya, *What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex?*, Neural Networks, 12 (1999), pp. 961–974.

[8] G. Edelman and G. Tononi, *A Universe of Consciousness*, Basic Books, New York, NY, 2000.

[9] J. Houk, *Information processing in modular circuits linking basal ganglia and cerebral cortex*, in Models of Information Processing in the Basal Ganglia, J. Houk, J. Davis, and D. Beiser, eds., Cambridge, MA, 1995, The MIT Press, pp. 3–9.

[10] J. Houk, J. Adams, and A. Barto, *A model of how the basal ganglia generate and use neural signals that predict reinforcement*, in Models of Information Processing in the Basal Ganglia, J. Houk, J. Davis, and D. Beiser, eds., Cambridge, MA, 1995, The MIT Press, pp. 249–270.

[11] D.-A. Jirenhed, G. Hesslow, and T. Ziemke, *Exploring internal simulation of perception in mobile robots*, in Proceedings of the 4th European Workshop on Advanced Mobile Robots, Arras, Baerveldt, Balkenius, Burgard, and Siegwart, eds., 2001, pp. 107–113.

[12] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*, McGraw-Hill, New York, NY, 2000.

[13] G. Lakoff and R. Nunez, *Where Mathematics Comes From*, Basic Books, New York, 2000.

[14] R. R. Llinas, *i of the vortex*, The MIT Press, Cambridge, MA, 2001.

[15] J. Mcclelland, B. McNaughton, and R. O'Reilly, *Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory*, Tech. Rep. PDP.CNS.94.1, Carnegia Mellon University, Mar. 1994.

[16] E. Rolls and A. Treves, *Neural Networks and Brain Function*, Oxford University Press, New York, 1998.

[17] L. Squire and E. Kandel, *Memory: From Mind to Molecules*, Henry Holt and Company, New York, 1999.

[18] L. Squire and S. Zola, *Structure and function of declarative and nondeclarative memory systems*, Genetic Programming and Evolvable Machines, 93 (1996), pp. 13515 – 13522.

[19] L. Steels, *Intelligence with representation*, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 361 (2003), pp. 2381–2395.

[20] D. Wolpert, R. C. Miall, and M. Kawato, *Internal models in the cerebellum*, Trends in Cognitive Sciences, 2 (1998), pp. 338–347.