Architectural Design of Mind and Brain from an Evolutionary Perspective

Jerald D. Kralik
Department of Bio and Brain Engineering,
Korea Advanced Institute of Science and Technology (KAIST),
Daejeon, 34141, South Korea
Jerald.Kralik@gmail.com

Abstract

To contribute to the development of a standard model of mind and brain, I present the general model I have been developing based on an interdisciplinary review of the literature. More specifically, I hope to show that evolutionary considerations are necessary to produce an accurate model of the mind as a whole. The basic architecture consists of three general design dimensions: a basic action control circuit (1st dimension), specific content domains (2nd dimension), and different levels of action-control circuits (3rd dimension), with the levels based on the evolutionary history and spatial layout of the brain, as well as increasing computational sophistication. It is hoped that this architecture will provide a complementary perspective and additional information to help develop a common standard model.

Introduction

Because of the highly integrative nature of brain function, progress toward a true understanding of the human mind and brain will be limited without a clearer specification of the main behavioral control systems and their dynamic interactions. That is, we need to specify the general cognitive architecture. This lack of a comprehensive theoretical model has also limited the achievements in artificial intelligence and autonomous systems (i.e., robotics), especially for intelligence beyond expert systems, i.e., artificial general intelligence. I have thus undertaken a review of the psychology, cognitive science, systems neuroscience, and evolution (especially behavioral biology, biological anthropology, and evolutionary neuroscience) literature to build the general model of mind and brain presented here. In particular, an evolutionary perspective provides a theoretical framework to best understand the overall findings, as well as insights into characteristics that are otherwise typically neglected or underdeveloped in models of mind and brain: e.g., levels of action control circuits (discussed below).

Taken together, the literature suggests three broad dimensions by which the cognitive architecture is organized: a general action control circuit (1^{st} dimension) that is utilized by systems dedicated to specific content domains (2^{nd} dimension) and at different levels of processing complexity (3^{rd} dimension).

In what follows, I first describe each main dimension of the general model, then I present the complete model, and then discuss how the model relates to the four main architectural features of (1) structure and processing, (2) memory and content, (3) learning, and (4) perception and motor control (Laird et al., in press).

The Cognitive Architecture of Mind & Brain

1D: Action Control Circuit

The fundamental unit of the general model is the well-known *perception-cognition-action* cycle or *system*, which corresponds to the control and execution of a *deliberate act* (Laird et al., in press). Before presenting the entire action control circuit, I will describe key segments that comprise it. Inputs that drive the circuit originate from two main sources: those internal and external to the organism, which are symbolized as "{S_{EI}}" in the model, with the brackets indicating that stimuli are composed of sets of component features (see Fig. 1A).

The internal stimulus input consists of signals from the body based on needs for growth, homeostasis and survival. The external stimuli are transduced by the sensory systems and then perceptually processed. The goal of perceptual processing is to sift through the otherwise intractable amount of potential information and derive a representation of the relevant input. This is achieved via the use of current knowledge in Long-Term Memory (LTM) and affective gating that determines whether a given input is potentially relevant to the organism's goals (Gazzaniga et al., 2013; LeDoux, 1996, 2003; Kahneman, 2011). Perceptual processing seamlessly moves into cognitive processing to obtain the input representation (i.e., the neural activity that represents the input). Fig. 1B depicts a simple example representation of an organism's environment, in which the

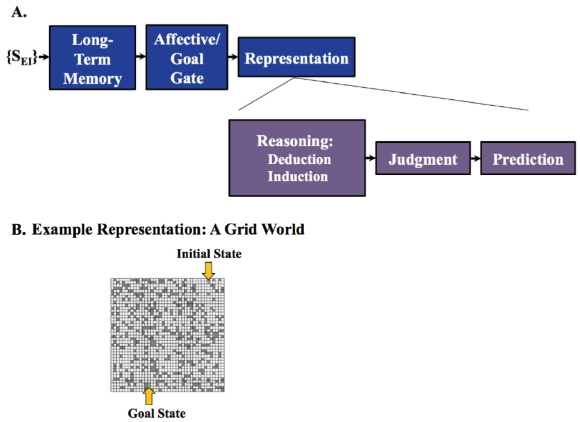

Figure 1. First components of the action control circuit. ***A****. Stimuli (both external and internal) are input to long-term memory with affective gating determining whether they are potentially relevant to the organism's goals. A representation of the environment is then produced via processes such as reasoning, judgment and prediction.* ***B****. An example representation of an organism's environment, with the individual in a particular location (initial state), the goal in a different location (goal state), and other locations containing obstacles (darker squares) to avoid.*

individual is in a particular location (initial or current state), its goal is in a different location (goal state), and various other locations contain obstacles that must be avoided when navigating through the world to attain the goal. Depending on the level of cognitive processing (3rd dimension described below) this may include more explicit reasoning processes, but typically requires some form of judgment about the inherently uncertain and incomplete input (Kahneman, 2011). Moreover, given the dynamic nature of the environment, representations are not static, and thus must be constantly updated. However, even then it is often not sufficient to simply track the inputs received. Rather, organisms need to use the inputs to derive predictions to anticipate subsequent states. That is, even before decision-making or action selection based on the state of the environment, the relevant state must be predicted from a sequence of states. Prediction, then, can be considered one of the main tasks of the organism, and in particular, the main 'goal' of the perceptual side of processing that produces the representation of the world that is required to select the proper action (Sutton & Barto, 1998; Glimcher & Fehr, 2014).

Once a representation is sufficiently formed the individual or agent can determine what to do. This typically entails a decision-making process that requires a *valuation* process, to determine the expected payoff from outcomes resulting from each choice option, and then the decision, ultimately leading to *action selection*, determined by a comparison of the expected values of the choice options (Fig. 2) (Glimcher & Fehr, 2014).

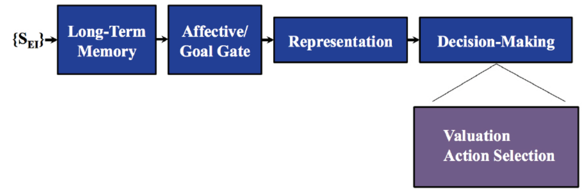

Figure 2. Once a representation is formed the individual determines what to do via a ***decision-making process*** *that selects an action based on a comparison of the expected values of the choice options.*

Once the action is taken, an outcome occurs, which then influences subsequent actions as feedback. Fig. 3 depicts this feedback error as being generally broadcast, which has been well established, especially in neuroeconomics with respect to dopamine neurons that have been shown to carry an error signal (Glimcher & Fehr, 2014). Finally, although evidence shows that this general feedforward sequential circuit drives motor control, all components of the circuit appear to have bidirectional arrows to the other components, attesting to the highly integrative nature of brain processing. These arrows have been largely omitted to clarify the main structure of the cognitive architecture, with the exception of the recurrent connection from decision-making to representation, reflecting the large body of evidence that the decision-making process itself can help to produce the most relevant and accurate state representation (Holyoak & Morrison, 2012).

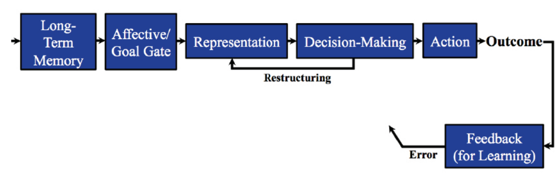

Figure 3. Once the action produces an outcome, feedback is broadcast back through the circuitry. The recurrent connection from decision-making to representation shows that the decision-making process itself can help to produce the most relevant and accurate state representation.

In sum, the fundamental unit of the general model is the 'perception-cognition-action' cycle that drives deliberate acts (Laird et al., in press). Multiple iterations that together achieve the organism's goal can also be conceptualized as *problem-solving* via the execution of an action policy.

2D: Content Domains

Converging evidence strongly suggests specialized networks for processing specific types of content (Table 1). Perhaps the best example is the study of reward processing in the brain, which has characterized multiple brain regions and the brain circuitry dedicated to valuation of stimuli such as food and fluids for decision-making (Glimcher & Fehr, 2014). Another clear example of specialized content processing is the "social brain network" comprised of regions such as STS (biological motion), FFA (face processing), TPJ (predicting other's beliefs, goals, and intentions: i.e., theory of mind), ventral premotor cortex (observing other's intentional actions), and MPFC (processing self and others) (Gazzaniga et al., 2013). Evidence also shows that social stimuli are considered primary rewards (Gazzaniga et al., 2013). Regarding environmental threats, it has long been established that brain circuits involving the amygdala, for example, monitor the environment for potential threats and danger (LeDoux, 1996, 2003).

This content-specific architecture is also anticipated by an evolutionary-neuroscience perspective. Extensive research in physiological neuroscience has led researchers such as Swanson (2000) to consider the hypothalamus as the first central nervous system region (moving rostrally from spinal cord to cortex) to control a complete, purposeful behavioral act, such as ingesting food. Indeed, hypothalamic nuclei appear dedicated to the major fundamental functions of survival, including ingestion, defense, mating, and basic dominant-subordinate social behaviors. Given that this framework constitutes the main architectural design for the first behavioral control center in the brain, one might suspect that it would provide the foundation for future evolutionary advances, including the archetypical telencephalon that evolved with vertebrates (i.e., cortex and subcortical structures, many of which make up the limbic system). This line of reasoning would also suggest that even neocortex may be organized around these fundamental functions. And in fact there is such evidence, for example, in premotor and primary motor cortex, where Graziano and colleagues found evidence for a mapping of these regions based on evolutionary functions (e.g., reaching and grasping, i.e., procurement, and defense) (e.g., Graziano, 2016). Indeed, findings in affective neuroscience are revealing the extensive influence of regions mediating affect and emotion, leading one of their most prominent researchers to conclude that cognitive neuroscience should consider these fundamental functions as truly constituting the underlying design principle of the brain (LeDoux, 2003).

Taken together, the second overarching architectural design principle is that the action-control circuitry is organized into specialized content domains based on the main fundamental functions of organisms identified by evolutionary theory and neuroscience (Table 1).

Content Domains	***Challenges***
1 Foraging & Ingestion	Obtaining Food & Fluids
2 Physical	Physical environment & Tool use
3 Defense	Predators & other (e.g., snakes) threats
4 Mating	Intrasexual competition & Intersexual mate choice
5 Social	Cooperation & Competition

Table 1. Content domains and their key challenges.

3D: Levels of Control Circuits

In this section I describe the evidence for levels of action control circuits, and it is important to first clarify the use of "levels" in this architecture. As opposed to *levels* based on scale, here they are based on the evolutionary history and spatial layout of the brain, as well as increasing computational sophistication (c.f., Laird et al., in press). I first define the levels more broadly in terms of level *types*. Given that the characteristics of each specific level are still being actively investigated, this broader characterization may prove most useful for a current version of a standard model of mind and brain. I also describe important related features such as working memory and the general concept of foreground versus background processing. Then, to provide a clearer understanding of what the specific levels would be, I present my model of these that includes one level of Type 1, two of Type 2, and three each for Types 3 & 4, for a total of nine levels of action control circuits.

Level Types 1-4: The Algorithmic Mind

The general concept of levels of cognitive systems in the brain has become prevalent, leading to a wide acceptance of at least two general classes of systems: i.e., a dual-system conception of mind and brain. This view is vividly portrayed by prominent cognitive psychologists such as with Kahneman's (2011) *Systems 1* and *2*. Nonetheless, most researchers admit that there are likely multiple systems, but it has been difficult to reach consensus on exactly what that number is, leaving two general systems to capture the two extremes of fast, associative and perceptually based processing versus slower, more deliberate and abstract reasoning-based processing (for reviews: Kahneman, 2011; Evans & Stanovich, 2013). Two key dimensions that characterize these systems are (1) the level of abstraction, regarding the type of content processed, and (2) the type of processing. For the former, cognitive science has well characterized levels of abstraction from highly concrete, perceptual and specific instances of stimuli (e.g., specific objects) to categories to concepts (including fully imagined ones such as forces, spirits, or minds) to fully abstract concepts (such as variables in mathematics) (Gazzaniga et al., 2013). For processing types, there is strong evidence for three general classes of cognitive systems: innate, associative- and reasoning-based, with the former two systems processing relatively content-specific stimuli, and reasoning systems processing content-specific or abstract stimuli (Wynne & Udell, 2013; Tomasello & Call, 1997; Holyoak & Morrison, 2012; Gazzaniga et al., 2013).

Research especially in behavioral biology has characterized many basic innate processes (i.e., instincts) in nonhuman animals (Wynne & Udell, 2013). Although rigorous learning theory work in the laboratory has suggested that even presumably 'hardwired' abilities have a strong learning component (such as learning which foods to eat) (Glimcher & Fehr, 2014), genetic studies nonetheless point to a strong genetic component (Knopik et al., 2016). With respect to content specificity, the more innate systems respond to relatively specific releasing stimuli.

The set of innate systems are classified here as Level Type 1, and as illustrated in Fig. 4, Type 1 is displayed as a basic action control circuit (with "thinking" representing *representation* and *decision-making*, and with action execution, outcome, and feedback omitted for clarity). Type 1 is also displayed as the lowest level type, following the general structure and evolution of the brain.

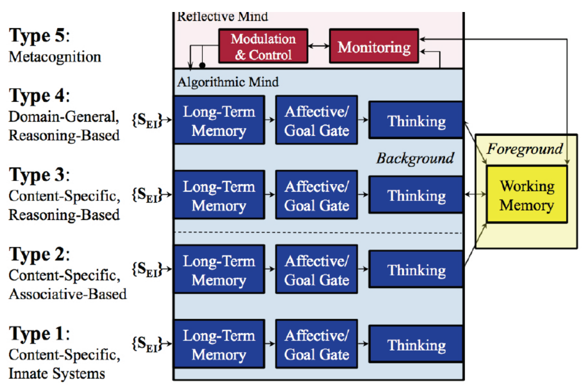

Figure 4. The 3rd model dimension (levels of action control circuits) contains four algorithmic-based ***level types****, representing progressively higher levels of computational abilities, and a fifth level type that represents metacognition, which helps orchestrate overall processing. Working memory participates in both algorithmic and reflective processing helping to produce seamless orchestration and more optimal solutions. Foreground and background processing represent degrees of awareness, and the dashed line represents parallel (below) versus relatively more serial (above) processing among the specific systems of each type.*

For associative (or, more generally, statistical) processes, 'animal learning' research in particular has amassed a large body of evidence for different types of associative learning processes, most notably Pavlovian and instrumental reinforcement learning (Glimcher & Fehr, 2014; Sutton & Barto, 1998). Currently, one of the most active areas of research in neuroeconomics has focused on these reinforcement learning processes, providing more detailed evidence for these systems and showing that they occupy different levels in the brain (Glimcher & Fehr, 2014). With respect to content specificity, although these systems have significant generalization ability, they are normally driven by more surface-level perceptual features relating to specific types of reward grounded in the fundamental content domains. These associative processes are classified as Type 2 (Fig. 4).

Higher-level cognition research, on the other hand, has focused more on reasoning-based processes, which offer multiple complementary advantages to trial-and-error associative learning, including bypassing the effort, time, and errors required for learning, as well as utilizing knowledge for higher degrees of prediction accuracy, planning and control. In particular, *causal* reasoning appears to be a central cognitive process in humans, with the basic ability apparently genetically underpinned, attesting to its importance (Holyoak & Morrison, 2012).

With respect to content specificity, reasoning systems can process both content-specific and more abstract stimuli. Following research by Kahneman, his colleagues and others, it is suggested that lower reasoning systems tend to process more perceptually based stimuli, leading to a "WYSIATI" (i.e., "what you see is all there is") characterization; while higher systems tend to process more abstract, conceptual stimuli (Kahneman, 2011; Evans & Stanovich, 2013). Although this progression from a concrete to abstract ontogeny structure may prove to be more of a continuum with the specific levels themselves, it is nonetheless useful to capture these major distinctions, and I have thus classified the reasoning systems as Level Types 3 and 4 (Fig. 4).

Taken together, the evolutionary, psychological and neuroscience evidence points to levels of systems in the brain that each can control behavior. Furthermore, although these levels may prove to be more of a low-to-high continuum (including with each component: e.g., the affective/goal gate comprised of drive to motivation to affect to emotions to goals), nonetheless, such a computational architecture necessitates mechanisms of arbitration that determine which system(s) (i.e., levels) will actually guide behavior. And substantial evidence suggests that there is indeed an explicit cognitive-control system, which I have labeled more generally as Level Type 5, *Metacognition* (Gazzaniga et al., 2013). Fig. 4 also utilizes the distinction of more direct "algorithmic" information processing (Types 1-4) versus this final type of "reflective" processing (Evans & Stanovich, 2013; Holyoak & Morrison, 2012).

Metacognition & Working Memory

Regarding competition and cooperation among the different action control systems, there is evidence for multiple types of mechanisms, including direct competition (especially among more closely situated systems) and more direct top-down control (e.g., control of one or more below) (Striedter, 2005). Nonetheless, substantial evidence points to explicit cognitive control, especially in the highest brain regions, such as regions of the prefrontal cortex (e.g., anterior cingulate and lateral prefrontal) (Gazzaniga et al., 2013). To provide this control, evidence shows that these metacognitive or reflective processes include monitoring, evaluation, planning, and mental simulation, both with respect to knowledge (metamemory) and processing (Gazzaniga et al., 2013). More specifically, there is evidence for explicit control of each component in the action-control unit: e.g., control of attention, emotion, thinking, and action selection. Yet there is also evidence for more direct metacognitive influences in algorithmic-level information processing, and this more direct involvement appears to be highly mediated via *working memory* (Gazzaniga et al., 2013). Fig. 4 thus represents working memory in yellow to signify characteristics of both reflective and algorithmic processes.

Awareness is represented in the model via *background* and *foreground*, with much of awareness appearing to be mediated by working memory, although it likely leaks into the highest algorithmic levels and other components of metacognition (Gazzaniga et al., 2013). Not only because of the inherent interest in consciousness, it is also important to distinguish foreground and background based on research that strongly suggests that much of even high-level cognitive processing appears to take place beneath awareness (Holyoak & Morrison, 2012). Finally, the dashed line in Fig. 4 represents parallel versus relatively more serial processing (Kahneman, 2011; Evans & Stanovich, 2013; Striedter, 2005; Gazzaniga et al., 2013).

The Actual Action Control Systems: Levels 1-9

As stated, the level *types* designation represents broader classes of action-control systems that have been well established. However, it is useful to provide a clearer view of the specific levels. Although the actual number and main characteristics of each are less clear, there is considerable pertinent evidence. And it is here, in fact, where an evolutionary perspective has again provided substantial insights. To that end, I present my model of the main specific levels (Table 2). An important caveat is that I am not focusing on when each ability originated, but rather, when it was more clearly prominent.

As already discussed, Levels 1-3 are in fact well established. For the 4th level, nonhuman animal cognition research has shown that multiple species exhibit evidence for some significant degree of causal reasoning ability, at least with respect to self (whether self-self or self-others: i.e., primary & secondary relationships) (Wynne & Udell, 2013; Tomasello & Call, 1997). At the same time, this research has led to the view that a major dividing cognitive ability between primates (including humans) and other species involves causal reasoning about tertiary relationships, such as between two other individuals (besides self) or a tool's functional component and the object being manipulated (Tomasello & Call, 1997). Such evidence has led to my Levels 4 and 5.

For the next levels, the strongest evidence for grade shifts in ability between great apes and humans derives from an examination of the competences of our hominin ancestors, and in particular, the stone tool technology (Nowell & Davidson, 2010; Relethford, 2013; Striedter, 2005). The major broader general change within the hominin lineage is an increase in brain size (both overall and relatively), mostly notably neocortex, and more particularly, parts of higher-order cortex such as prefrontal. This size increase also was chiefly driven by new cortical fields similar to the previous ones, rather than, for example, higher neural density or increased number of cortical layers (Striedter, 2005). Additional fields in general suggest greater *elaboration* with respect to levels of abstraction, enabling a hierarchical organization of information and processing, leading to further abilities, including a greater time horizon, comprehension of larger event complexes (such as an entire foraging excursion), and subgoal or subroutine processing in problem-solving, language, etc., leading to recursion (Gazzaniga et al., 2013; Holyoak & Morrison, 2012). The greater elaboration also enables more extensive cross-referencing, providing heightened generalization abilities as well as greater higher-level access to details. In fact, the ability to access a greater number of details appears to be a particularly important advance. This advance includes not only adding more perceptual dimensions, but hidden and eventually truly imagined ones. More specifically, objects become combination of parts and the "glue" that holds them together. Thus, rather than knowledge organized around objects and their properties, a new orientation occurred that essentially flipped this structure such that objects may be seen more as a function of the glue and parts. This change appeared to solidify throughout the hominin lineage leading to *Homo sapiens*. Put simply, a leading continued advance across the lineage is best described as *reductionism*. Moreover, an intriguing and enigmatic property of the human brain's evolutionary trajectory is the apparent retaining (rather than replacement or revision) of the prior abilities. Thus, each major grade shift (i.e., level) is maintained in the model.

Table 2 provides a summary of the major grade shifts. The final stage (Level 9), culminating in the cultural explosion with the origin of *Homo sapiens*, suggests the most heightened degree of reductionism, including a reasonably complete (yet 'folk') separation of the concepts of matter and forces, as well as a greater delineation of the similarities (generalization) and differences (specialization) across instances. This final level of competence would provide the ability to comprehend, for example, analogies (i.e., relationships of relationships), which has been proposed as a critical ability separating people from other animals (Holyoak & Morrison, 2012).

Type	*Level*	*Cognitive Ability*	*Description*	*Proposed LCA*
1	1	Innate Processes	Instincts	Prior to vertebrates
2	2	Associative Learning I	Pavlovian	Prior to vertebrates
2	3	Associative Learning II	Instrumental	First vertebrates
3	4	Causal Reasoning I	Primary & Secondary relationships	First mammals
3	5	Causal Reasoning II	Tertiary relationships	Primates
3	6	Reductionism I	Partially hidden parts & causal factors	*H. habilis*
4	7	Reductionism II	Smaller imagined parts & causal factors	*H. ergaster/erectus*
4	8	Reductionism III	Substances & more abstract causal factors	*H. heidelbergensis*
4	9	Reductionism IV	Elemental units & Forces	First *H. sapiens*

Table 2. Levels of control circuits with key cognitive ability and proposed evolutionary establishment via last common ancestor (LCA).

In sum, the specific levels are presented here to provide a more complete model based on an analysis of the literature across many relevant fields, although it is clear that further research is necessary to establish these specifics more definitively.

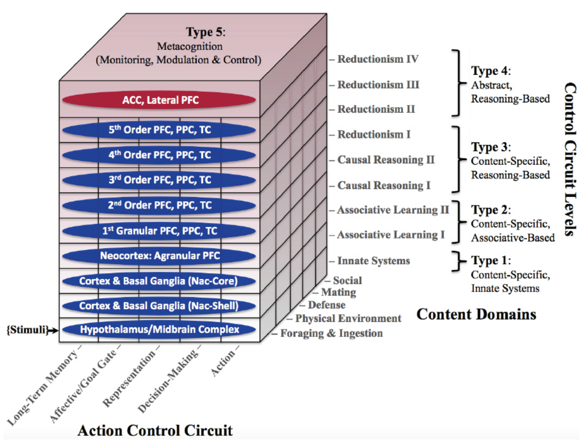

Figure 5. Complete architectural design model, with three main dimensions: (1) the basic action control circuit, (2) content domains (that progressively dissolve across levels), and (3) levels of action-control circuits; together with central brain regions implicated for each system: ACC (anterior cingulate cortex), PFC (prefrontal cortex), PPC (posterior parietal cortex), TC (temporal cortex), NAc (nucleus accumbens). 1st granular and Nth order PFC fields expand out from agranular PFC and motor cortex. Stimuli are input into every level but omitted for clarity. Action outcomes and subsequent learning signals are also omitted.

The Complete Architectural Design Model in 3D

With the three general design dimensions described — a basic action control circuit (1st dimension), specific content domains (2nd dimension), and levels of action-control circuits (3rd dimension) — Fig. 5 assembles them into the complete architectural design model, together with central brain regions implicated.

Discussion

Here I briefly discuss how the model contacts the main design components outlined by Laird, Lebiere, and Rosenbloom (in press). In my model, *structure* derives from the three main dimensions (the basic circuit structure, content domains, and the organization of control circuits into types and levels). *Processing* is captured by the specific process modules represented by the boxes, as well as the arrows representing the sequence flow, with greater detail omitted: such as the submodules and subprocesses within each module, and the massive interconnections among them. Nonetheless, the architectural design is revealed once one sufficiently zooms out from this greater detail.

For *memory*, I have used two general classes: long-term and working memory. However, when further fleshed out, each well-known type of memory has a corresponding place: e.g., declarative versus procedural, semantic versus episodic (Gazzaniga et al., 2013). *Content* is in fact a main feature of the model, providing a basic organizational structure (i.e., the 2nd dimension as ontologies based on the fundamental functions of the organism). The evolutionary life cycle of organisms (and corresponding hypothalamic mediation) underpins this organizing principle.

For *learning*, I have focused on associative processes (Pavlovian and instrumental reinforcement learning), however, other types of learning fit readily into the model, such as more general statistical learning and higher-level learning such as via instruction (Gazzaniga et al., 2013).

The details of *perceptual* and *motor* processing reside within the general modules, as the model has been developed to be consistent with them. Regarding details, the same is true for all of the main processes represented in the model (such as memory and affective processing).

Conclusion

Progress accelerates when overarching theory guides investigation. It is my hope that the architectural model of mind and brain presented here will provide a complementary perspective and additional information to help in the development of a common standard model.

References

Evans, J. S. B. T., & Stanovich, K. E. 2013. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3): 223–241.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. 2013. Cognitive neuroscience: the biology of the mind. WW Norton & Company.

Glimcher, P. W., & Fehr, E. 2014. Neuroeconomics: Decision making and the brain. Oxford: Academic Press.

Graziano, M. S. 2016. Etholocical Action Maps: A Paradigm Shift for the Motor Cortex. *TICS*, *20*(2): 121–132.

Holyoak, K. J., & Morrison, R. G. (Eds.). 2012. The Oxford Handbook of Thinking and Reasoning. Oxford University Press.

Kahneman, D. 2011. Thinking, Fast and Slow. NYC: Farrar, Straus & Giroux.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). Behavioral genetics. New York: Worth Publishers.

Laird, JE, Lebiere, C, & Rosenbloom, PS. in press. AI Magazine.

LeDoux, J. E. 1996. The Emotional Brain. NYC: Simon Schuster.

LeDoux, J. E. 2003. Emotion Circuits in the Brain. *Annual Review of Neuroscience*, *23*(1): 155–184.

Nowell, A., & Davidson, I. 2010. Stone Tools and the Evolution of Human Cognition. Boulder: University Press of Colorado.

Relethford, J. H. 2013. The Human Species. NYC: McGraw-Hill.

Striedter. 2005. Principles of Brain Evolution. Sinauer.

Sutton & Barto. 1998. Reinforcement Learning. MIT Press.

Swanson, L. W. 2000. Cerebral hemisphere regulation of motivated behavior. *Brain Research*, *886*(1-2): 113–164.

Tomasello, & Call. 1997. Primate Cognition. Oxford Univ. Press.

Wynne & Udell. 2013. Animal Cognition. Palgrave Macmillan.