Engineering Computer Architecture Models for the Evolution of Interiorized Creator-in-a-Box Roles for Conscious Robots, Self-Aware Robots, Automata, Machines and Artifacts (CR / SARAMA)

John-Thones Amenyo
Department of Math & Computer Studies
York College, City University of New York (CUNY)
Jamaica, NY 11451
{*jta*}*@york.cuny.edu*

1. Introduction

In the pursuit of the engineering of cognitive informatics, non-biological artificial consciousness, machine consciousness for robots, machines, automata and other artifacts, (henceforth, "robots" or "cognitive robots"), [Chella & Manzotti 2007], [Haikonen 2008], it is a useful research strategy to avoid and by-pass endless, interminable discussions, skepticism and controversies [Chalmers 1995], regarding the issue whether it possible or even feasible for engineered or human-constructed machines to ever be conscious, have a sense of self or self-awareness, and be capable of subjective experience, or phenomenal experience, (henceforth, "self conscious minds").

Instead of attempting to duplicate, replicate, reproduce or emulate natural (human) self conscious minds, one can adopt the alternative research agenda to merely use biological metaphors to build biologically inspired robots, automata and machines that are deliberately engineered to exhibit and manifest self-like, consciousness-like and self-awareness attributes and qualities, which may be termed quasi-consciousness, equivalent consciousness, or even pseudo-consciousness, etc. As a form of self-imposed engineering discipline, it is required that such artifacts and robots be potentially biologically-plausible or biologically-credible. Thus, the research stance is that the technology-based non-biological artifacts merely share a few fundamental principles of operation and functioning, with the biological counterparts; instead of the artifacts slavishly copying, replicating or reproducing the biological organisms or organs.

As an alternative to a pre-occupation with terminology and definitions, one's focus shifts to relative or comparative advantages, usefulness and functional roles of self, consciousness, self-awareness, etc. that can be justifiably mimicked or imitated in artificial or synthetic contexts. These functional roles and utility can be studied by addressing the advantages in evolution of self conscious minds and supporting or related (biological) phenomena and processes.

This paper introduces the concept of artifact-as-organism as the appropriate correspondence of engineered, cybernetic artifacts and machines with biological organisms and their biological or natural evolution. The concept of artifact-as-organism is useful because such "*organisms*" can be regarded as being subject to laws, principles, processes and mechanisms, similar to those operating to affect the evolution of biological organisms. The related Creator-in-a-Box concept is then described, to establish support for evolution or "mobility-in-wide-sense" (diachronic evolution, change or dynamics, for short term and long term survival, sustainability, longevity and persistence). The Creator-in-a-Box role is advocated as the primary and essential function that engenders the need and presence of manifestations of self conscious minds. Operationally plausible characterizations of self consciousness minds are then advanced as embodiments and manifestations that can be used to support the Creator-in-a-Box concept in autonomous (and possibly mobile) cognitive robots. The paper then concludes with a successive, hierarchical refinement of a supporting PMSCIO-computational architecture model, [Bell & Newell 1971], that can

implement the Creator-in-a-Box for robots, artifacts, etc., (PMSCIO P – Processor, M – Memory , S – Switching / Communications, C – Control, Coordination, Cybernetics, IO – Input & Output).

2. Concepts of Artifact-as-Organism & Creator-in-a-Box

One can regard robots as useful technological artifacts, built and used for specific purposes and teleological tasks. From the viewpoint of evolution or diachronic progression, what is significant and relevant is not any specific, individual or particular robot, but rather the multi-generational family, lineage or genealogy of robots and their evolution or progression through time. The diachronic and evolutionary view about a robot is only complete and comprehensive if the associated generations of user (consumer / operator), as well as the designer (creator / builder / maker / engineer) roles are also taken into account, in addition to the individual robots in the evolving family.

Each generation of robot in a family, together with the associated user/operator and designer/creator/engineer can be regarded as an organism, called *artifact-as-organism*, conceptually on par with a biological organism, (see Fig. 1). It should be noted that the user/operator, designer/creator can be individual (single) or institutional (organized distributed) agents or agencies.

If any individual robot, artifact-as-organism is built, operated and used as an autonomous, self-reliant system, then the automated or automatic embodiments of the user/operator and designer/creator/engineer roles have to be bundled, enclosed, encapsulated and associated with the robot, so as to accompany it at all times. The bundled accompanying embodiments can be regarded as "interiorizations" or "internalizations" of theses roles, in place of the conventional situation of a wired and/or wireless interlinking of the robot to the external operator and/or creator, with interactions via tele-presence and tele-operation. The internalized, (logical or physical / material) manifestation or embodiment of the user/operator and creator/make roles is termed the *Creator-in-a-Box* concept, (see Fig. 1).

An artifact-as-organism can be considered as being subject to processes and mechanisms similar to those operating or occurring in biological evolution. Of great interest here is the emergence and evolution of "self" and "consciousness" in biological organisms, (in particular humans), and how its principles can be used to inform the synthesis of self and consciousness, (or self-like and consciousness-like attributes and qualities) in robots and robot-like (cybernetic) artifacts.

In brief, technological artifacts-as-organism can be considered as living or alive and evolving, and necessarily self-reproducing in the sense of artifact replacements by migrations, upgrades, transitions, re-engineering, etc. A useful metaphor is that biological organisms evolve (their bodies, as well as the brain, mind, self, consciousness and self-awareness) to support and help cope with the struggle for long term survival of the individual, the family, clan, tribe, group, race, the species, and even Life-itself. Thus, the technological artifact-as-organism may also evolve the corresponding equivalents of controls, emotions, motivations, memory, intelligence, mind, self, consciousness and self-awareness, in order to support or help cope with the artifact-as-organism's struggle for long-term survival.

3. Self, Consciousness, Self-Awareness, and Self-Conscious Mind

In the proposed CR/SARAMA computational architecture model, Self is characterized as the logical separation (and possibly separate physical manifestation and embodiment) of all or some of the roles that are directly concerned with the *organism-as-a-whole*, that is, the embodiment of holistic concerns, aspects and roles.

A virtual Self (logical and un-embodied), is the collection of all the roles of a population, society, organization, structure or complex of functional agents, actors, participants, automatisms that are geared to or focused directly on the concerns and aspects of the population, as-a-whole, rather than being concerned with individual members, sub-populations, sub-systems or sub-structures. For example, a (behavior, action, activity, awareness, sensory processing) role that is concerned with the (short-term or long-term or beyond-current-lifetime / lifecycle) survival of the organism-as-a-whole is part of the (virtual or logical) Self.

A physically manifested Self is an actual separation and separate or distinct materialization (substantiation, incarnation, incorporation, corporeality) of the concerned-

with-organism-as-a-whole roles. Its primary diagnostic feature is that it can be distinguished from the inevitable inter-coordination and inter-association structures and infrastructures that are required to organize the distributed and inchoate population of the primordial functional roles into a cohesive, coherent, integrated and fused union, confederation or purposive (goal-oriented) organization or "organism".

An alternative viewpoint is that the Self includes all those roles in the overall inter-coordination and integration infrastructures which are concerned with the organism-as-whole. These roles are (logically) separated from the rest of the inter-coordination and integration roles; for example, dealing with pair-wise, multi-cast, multi-conference, team and various other subsystem constellations, coalescing, coalitions, fusions and combinations of the basic functional roles.

Thus, a physically embodied self can also be regarded an embodiment of separated centralization of organism-as-a-whole roles. The use of the term centralization here is somehow misleading because of the implicit sense of (physical) co-localization or sympatry, spatial concentration and compactness of the participating functional roles or agents (in one place, site, local, unit, module, cell, molecule, circuit, network, at one spatial scale, etc). It is possible to implement ("geographically" or physically distributed, but logical) virtual centralization by means of inter-coordination and inter-signaling infrastructures. The self roles are here described as logically centralized because they are separated out, logically isolated, and explicitly identified and differentiated from all other roles.

Awareness is modeled as the logical separation (and possibly separate physical manifestation and embodiment) of the roles that are concerned with the *organism-in-context*, (with respect to the external, ecological environment and ambience).

Thus, Self-Awareness can be specified as the (logical and/or physical) separation and embodiment of the functional roles that support both Self and Awareness. Intelligence is equated with the Mind.

As part of the overall architectural refinement, the embodiment of Self is to be supported and implemented by both "Consciousness" and the "Unconscious," (that is, by the "conscious self" or "self-conscious mind," and by the "unconscious self").

There are several, in fact numerous, formulations and discussions and characterizations about the concerns, aspects and roles of (biological or human) Consciousness, (see for example, [Chalmers 1996]). For example, aspects, concerns and roles of Consciousness are typically considered to include, degree of wakefulness; attention, attentional focus and control; self, self-identity, unity, continuity, individuality; awareness, self-awareness, verbal reporting, conscious awareness, phenomenal awareness; subjective experience or phenomenal awareness, also called qualia; awareness of emotions, especially pain and pleasure; imagination; self-awareness of one's own thoughts, concepts, intentions; inner speech or silent speech (talking oneself internally); inner (visual) imagery; awareness of self-in-context in the past, present and future.

Unfortunately, these characterizations are never diagnostic, in the sense of providing a clear differentiation or distinction between conscious vs. unconscious operations and processes.

In the CR/SARAMA computational architecture model, the Unconscious mode is taken to mean all distributed and parallel aspects of roles in operations and interactions. Thus, Consciousness is taken to be the complement, or the separation and separate embodiment of serial and sequential aspects of roles, as well as parallel-to-serial conversions, see Fig. 2.

This distinction is clearly diagnostic (in that it provides differentiating characterizations). In biological organisms, the Unconscious evolves first and is indeed the "kitchen sink", being concerned with operational roles that occur in parallel, and since the population of basic elementary roles is distributed, unconsciousness is in turn parallel and distributed. In biological organisms, Consciousness evolves from this in that it is a manifestation of explicitly separating out and supporting operational roles that are explicitly necessitated (arranged, organized, ordered) to operate serially or sequentially (totally linear order or chaining).

Since the Unconscious is designated the "kitchen sink," (that is everything else, the democratic party, except for death), it is taken to also include each and all of the following shades or variants of "consciousness": unconscious, sub-conscious, semi-conscious, pre-conscious, post-conscious, super-conscious, hyper-conscious, ultra-conscious, meta-conscious,

virtual conscious, vicarious conscious, simulated conscious, etc.)

A significant role for the integration of Self and Consciousness is the occurrence and opportunity for serialized and sequential (inner), spontaneous and automatic initiators, originators and sources of (top-down, macro-scale to micro-scale) inherited attributes that are concerned with the organism-as-a-whole. According to the current model, the principal use of these downward mereological flows is to support Creator-in-a-Box roles and concerns in the evolution-for-survival.

To recapitulate, the Self is separation (possibly with separate physical embodiment and materialization) of functional roles directly concerned with organism-as-a-whole, using a population of basic functional roles. The Conscious Self or Self-Conscious Mind consist of the separation and separate embodiment of explicitly serialized or sequential operation of (some) Self roles. The Unconscious or Unconscious Mind is the rest of Self roles that engage in or embody parallel role operations. All the roles can be distributed, and portions can also be concurrent, particularly, with regard to the use of shared-resource, resource-sharing, multi-access conflict resolution mechanisms.

In the engineering model being advocated, the relevance of Self (and the associated support roles of the Unconscious and Consciousness) is that Self can be used to support the implementation of the Creator-in-a-Box roles for autonomous robots, automata, machines and artifacts.

4. Architectural Model

The length requirements imposed on this paper preclude showing considerable detail of the successive refinement of the computational architectural model engineered to support the above characterizations. Fig. 3 and Fig. 4 show two stages of the architectural refinement. Details are reported elsewhere.

Succinctly, the highlights of the CR/SARAMA architecture can be characterized as being multi-level, hybrid reducible / irreducible integration or fusion (irredex); it is based on Bell-Newell PMSCIO functional roles, organized as cognitive cybernetic loops or pathways, at all scales.

Equivalents of mental-cognitive-affective-conative features or facilities are explicitly provisioned and engineered, instead of assuming they will occur as emergent properties of complex organizations of micro-scale building blocks. For example, there are portions of the architecture that are explicitly provisioned to act as the Consciousness, Self-aware subsystems, etc, see for example, Fig. 3. In the top-down (hierarchical) refinement of the architecture, at each level, in addition to the reduction to lower level micro-scale modules, there still remains an explicit non-reducible or irreducible residue for each macro-scale functional role. This approach to componentization (sub-structuring or sub-organization) results in an irredex architecture.

The irreducible residues at each level can be used to support intermediate adaptor based, flexible and dynamic, re-configurable, ad-hoc, inter-level (embedded, nested or mereological) network architecture overlays, in addition to the direct peer-to-peer network architecture and infrastructures established at each level or scale. An irreducible residue can also be used to control, regulate and inter-coordinate all of its participating components. For example, one can establish a full, complete and pervasive potential interconnectivity among all the modules at any particular scale. In actual operation, specific inter-module connections and associations can be explicitly suppressed, inhibited, deactivated or attenuated.

The irredex organization is also consistently applied to the architectural evolution and progression through time. Thus, older, simpler, generic and cruder versions of manifestations of functional roles operate concurrently with the more advanced, sophisticated, complex and specialized realizations of the same roles. Although it is out of the scope of this paper, it would seem that the irredex architecture can be used to attempt a resolution to D. Chalmers' "hard" problem(s) of consciousness, namely (the irreducibility of) subjective experience, phenomenal awareness and qualia, [Chalmers 1995].

There can be multiple, separate but concurrent expansions, refinement or reduction from any macro-level to the next micro-level or sub-level. In grammar-theoretic terms, the mereological and inclusion relations of the reducible/irreducible expansions of the architecture, result in multi-level, multi-scale attributed structures, just like attributed parse trees or syntax trees of the attribute grammar of a language; instead of being a replacement / substitution rewriting, translation, denotational reduction or compilation system. The irredex

architecture allows a natural top-down flow the influence, interference and modulation of inherited attributes from the macro-levels to the micro-levels, as well as the bottom-up flow of influence and modulation of synthesized attributes from the micro-scale to the macro-scale levels.

At all scales of the architecture, all the modules or nodes (0-spaces, q-complexes) and inter-module inter-associations and inter-couplings (1-spaces, p-spaces) are annotated or tagged with E | M | m valuation attributes, (E | M | m = Emotion | Motivation | memory). Each E| M|m attribute consists of three parts: a) the pre-E|M|m that influences or modulates the associated (cognitive or computational) functional role; b) the post-E|M|m that is generated, produced or induced by the functional role as the outcome co-product, by-product, record, archive, documentation, reporting or journaling (memoization), and c) the post-E|M|m → pre-E|M|m transformation or translation adaptor / intermediate, or the co-E|M|m. As a further refinement, each of E | M | m role can be irredex to short-term, meso-term and long-term sub-roles.

Thus, memory (or memory types), emotional and motivational valuations are explicitly and inherently distributed and parallel in the architecture. There are numerous places at all levels where inflows converge on a functional role, unit or module, thus necessitating a need for (shared-resource, resource-sharing) multi-access conflict resolution (MACR). The MACR mechanisms, processes and strategies can cognitively be equated with attention functional roles. Thus, the architecture inherently supports multiple, parallel and distributed loci of attention processes.

Once again, it is important to carefully distinguish between the logically separation and provisioning for a (mental) cognitive-affective-conative feature in the architecture, from a physical realization, packaging or materialization supporting the feature in an operational system. A neuron-biologically inspired realization, packaging and layout of each core feature will consist of a distributed infrastructure vs. a spatially compact organ of co-located elements in one physical area or region of space. A useful metaphor to visualize the entangled and inter-mixed distributed realization is the following. Consider each (core) feature as having a finite collection of distinct colors associated with it. The organization implementation elements ("cells"), ("balls", "blobs", "clumps"?), of a feature, (0-spaces / nodes, 1-spaces / edges, links, p-spaces / hyper-edges, q-complexes / subsystems), are each assigned various colors from the feature color set. All the implementation elements of all the features are then thoroughly intermixed and superimposed, to form a colloid or emulsion composite. Formally, the neuron-biologically inspired distributed implementation means that each feature is space-filling, omnipresent, ubiquitous and pervasive in the operational architecture. The implementation model also accommodates the use of redundant subsystems, as well as alternative realizations (diversity, alternative / parallel universe) at any scale, in order to support fault-tolerance and high availability. The net result is that the architecture implementations can be called colloidal, entangled multiple, co-existing and coincident arrays, nested arrays, graphs, hyper-graphs, networks, polyhedra, algebraic topological complexes and infrastructures, (compare alloys, soups, concrete and paint mixtures).

The architecture also adopts a multi-paradigm approach to knowledge, information, and data (KID) representations and manipulations. From the viewpoint of semiotic doctrine of signs of C.S. Peirce, it supports symbolic, iconic and indexical KID structures. From the viewpoint of formal structuralism, for example, as advocated by J. Piaget, the architecture supports synchronic and diachronic structures, including topological, mereological / inclusion, mereo-topological, spatial, algebraic (equivalence, similarity, differentiation, variation, gradation), ordered (partial, total), ordered-lattice, ordered-temporal, and mapping or morphism KID structures. For manipulations, each KID structure is considered as an attributed structure, that is, suitably tagged, annotated or labeled with (multiple) metadata structures.

In their manipulations, each KID structure is considered as structure-as-concept or structure-in-context, meaning that each KID structure is modeled or represented as having the following components: the structure itself; sub-structures and super-structures (involving its mereological and inclusion relations); pre-structures and post-structures (involving its causal and order relations); co-structures, pro-structures and anti-structures (involving its topological relations); and meta-structures (involving its algebraic and valuation annotation relations). Finally, all the attributed KID

structures are manipulated using push / pull co-processing (comparison, contrast and resolution) mechanisms: push – inflow signals / KID as tests, cases, probes, criticisms, refutations, confirmations; pull – theory, model, knowledge, expectation, anticipation.

An interesting case to use in understanding the CR/SARAMA architectural principles is to attempt the engineered evolution of a (face) shaving robot.

5. Discussion

Diachronically, the evolutionary tendencies of biological organisms and technological artifacts, (and therefore artifacts-as-organisms), are somewhat different and shed interesting light on the issues of the emergence of "self" and "conscious self-awareness." Biological evolution and technological evolution approach the occurrence, construction or emergence of self and consciousness from opposite ends.

Biological evolution proceeds from the parallel to the serial, that is, from an organized, distributed population of roles and active agents, mainly biological cells, with separation and centralization of some roles such as the nervous system and the brain emerging from this to support mobility. The current dominant thinking is that the biological mind, self and consciousness emerge as further separation and embodiment of centralization, to support the nervous system and the brain, (Biological Evolution: Parallel x Distributed → Serial x (possibly Centralized/Single)).

Technological evolution, so far in human history, has proceeded in the other direction, from serial (single, centralized) to parallel (multiple, distributed) organizations of roles and moieties (Technological Evolution: Serial x Central/Single → Parallel x Distributed). The initial artifacts, for example computers, because of prohibitive technological resource costs and other limitations, are typically single, spatially compact entities. They typically operate, or are operated using serial algorithms, procedures and centralized management, administration and control schemes. Later on, the designers and builders struggle to create functionally effective, cohesive and coherent parallel, distributed, and complex organizations.

Thus, in a sense, in biological evolution, the (parallel) Unconscious comes first and the (serial) Consciousness struggles to separate and centralize itself and emerge out of the former, (Biological Evolution: Unconscious → Conscious). In contrast, in technological evolution the (serial and centralized) Consciousness can easily be constructed and fashioned, while there is a struggle to establish the (parallel) Unconscious organization from it, (Technological Evolution: Consciousness → Unconscious). Consider for example the concerns and challenges of distributed computing, parallel processing, ad hoc computing, ubiquitous computing, pervasive computing, multi-agent computing, network computing, cluster computing, Internet computing, cloud computing, grid computing etc.

The Creator-in-a-Box concept can be regarded as an extension of the stored-program concept in computer architecture, conventionally credited to J. von Neumann. On the one hand, the stored-program concept is concerned with control and coordination of (originally, uniprocessor) computational resources. On the other hand, the Creator-in-a-Box is an online, real-time, enclosed, encapsulated, incorporated, onboard, embedded, or carry-along embodied role that provides life cycle support, survival and evolution of the artifact-as-organism.

Although it is harder to contemplate how an embedded Creator-in-a-Box plays any role in biological organisms such as human beings, a good starting point is to consider "innate curiosity," "imagination," speculation, conceptualization, "cognitive innovation" "imaginative dreaming" and "exploring instinct" as playing the role of proxies (stand-ins, delegations) for creator / designer / maker and user of biological organisms. The use of such faculties and processes leads to active, spontaneous (self-initiated and self-directed) exploration, navigation, traversal of envisioned spaces and worlds, browsing, engagement, participation, involvement, inspection, examination, testing, evaluation, assessment, signification, recognition, vicarious trials, critical decision-making, "elimination of errors," problem-solving with respect to the external, ecological, social, political and cultural environments, which are all roles expected to be played by the Creator-in-a-Box. Thus, there is no need to appeal to a ghost-in-the-machine or homunculus.

6. Related Work

According to K. Popper [Popper & Eccles 1977], the Unconscious is concerned with routine, automatic operations; while Consciousness is concerned with non-routine, non-automatic, deliberate, deliberative operations. The Parallel vs. Serial characterizing differentiation suggested above between the Unconscious vs. Consciousness clearly supports Popper's dichotomy.

Although a more competent and detailed comparative treatment is required, it would seem that the several ideas exploited here for the architectural refinement, intersect with those of D. Dennett regarding the mind and consciousness, [Dennett 1991]. Namely, a) the use of *evolution* for explanation vs. for engineering design; b) the emphasis on *parallelism* of brain processes vs. of computational infrastructures; and c) a critical and principal operational characteristic of *consciousness* being its *serial* (narrative) nature.

The concerns of the CR/SARAMA model intersect with those of building *integrated cognitive architectures*, [Franklin 2007], [Gray 2007], [Sun 2004], in the research areas variously named as machine consciousness, artificial consciousness, cognitive robotics, embodied intelligence, computational intelligence, situated intelligence, (multi-) agent architecture, universal or general intelligence, executable or computational cognitive modeling, strong AI and artificial general intelligence (AGI), [Goertzel & Wang 2007], [Goertzel & Pennachin 2005].

A primary ultimate goal in each of these areas is to build (integrated) computational architectures for formulated models of the mind, intelligence or consciousness. Not surprisingly therefore, just like the CR/SARAMA architecture, each concrete, published proposal of a cognitive architecture typically involves a single level or multiple, interlocking levels of cognitive cybernetic loops, pathways or circuits, that is, cyber-cognitive pathways of sensing/perception → cognitive (+ affective + motivational) processing / manipulation, selection / mapping / control → action, behavior. Using the MVC pattern of software engineering as a metaphor, a cyber-cognitive loop consists of the View (V) → Model (M) → Control (C). Using the Bell-Newell PMSCIO computer architecture model as a metaphor, a cyber-cognitive loop consists of one of the following M-supported cybernetic pathways: I → C/S → O; I → IP → C/S → O; I → C/S → OP → O; I → IP → C/S → OP → O.

In psycho-cognitive-pragmatic terms, in more detail, each cyber-cognitive loop or pathway consists of an organized and directed infrastructure that includes the manifestations of the functional roles for various ways of creating, producing, manipulating, consuming and utilizing (external and internal) knowledge, information and data (KID) structures, patterns, schemas, representations:

a) **Input (I)** and (**post**) **Input Processing (IP):** Of the encounter of situations, contingencies and conditions from the environment, via (sensing, sensations, sensory processing, monitoring, tracking, observation and detection perception, decoding, decryption, translation, understanding, comprehension, diagnosis, prognosis, understanding, interpretation, diagnosis and prognosis prediction, forecasting, evaluation, assessment, judgment, classification, categorization, elaboration, analysis;

b) **Output (O)** and (**pre-**) **Output Processing (OP)**: Actual execution, implementation, discharge and performance of action programs, schemes, patterns, and scripts.

Additionally, action and behavioral control, regulation, coordination, activation, deactivation, planning, command, instruction, decision, choice, selection, desire, intention, application, execution, motion, mobility, locomotion, kinesthetic, kinetics, kinematics, dynamics, change, expression, behavior, operation, action, activity, transactions, reactions, pro-actions, interactions, social interactions, exploration, navigation, space or space-time traversal, expression and gestures, communication, coordination, cooperation, collaboration, competition, contest and conflict Induced or provoked changes and modifications; proposed changes and modifications to roles, goals, structures, internal inter-relations and external interfacing inter-relations; control and regulation: plans, strategies, patterns and schemes for attenuation, inhibition, restriction, constraint, confinement, prohibition of un-satisfactory, undesirable, unwanted or unacceptable states, conditions, processes, relations; amplification, support, augmentation, helping or aiding to achieve desirable, acceptable and satisfactory outcomes.;

c) Internal manipulations, **Processing** (**P**), **Control**, **Coordination**, **Communications /**

Switching (**C/S**): internal cognitive, affective, conative and motivational processing; problem-solving, particularly with regard to changes, modifications, violations of expectations and anticipations, such as errors, faults, failures, threats, dangers; ineffectiveness in accomplishment of functional roles; inefficiencies in resource utilization; changes in situations and ambient conditions; quality of role performance; exploitation and leveraging of opportunities and resources; mapping into action programs; reasoning; inference (deduction, induction, abduction); intuition, valuation Structures that can be manipulated, stored and recalled.

The major differences between the various proposed cognitive architectures in the published literature are primarily about the a) Various functional roles that are emphasized and modeled / implemented in detail; b) The choice of KID representation and attendant manipulations. KID representation schemes include: first-order logic, logical symbolic / schema-based / attribute-based, probabilistic, probabilistic logic, biological neural, artificial neural; c) Explicit incorporation of multiple levels of cybernetic loops with distinct functional roles, such as reflex loop, deliberation / reflection loops, and meta-cognition (self-awareness and meta-self awareness) loops; d) Explicit incorporation of various models of **Memory**, (**M**), including short-term memory (STM), active working memory (AWM), long-term memory (LTM), procedural / skill / script / algorithmic memory, semantic / declarative / conceptual memory, episodic / personal experience memory; subjective experience, phenomena awareness memory, as well as mechanisms for searching, accessing, retrieval and reasoning / inference with information from the various kinds of memory (repositories and depositories) and storage structures; e) Explicit incorporation of valuation roles for **Emotions**, **Feelings**, **Moods** and inner dispositions; f) Explicit incorporation of valuation-drive roles for **Motivations**, urges, obsessions, wishes, desires; g) Incorporation of various forms of **Learning**, in order to cope with and adapt to uncertainty and change in the environment, as well as system or artifact's own goals and objectives; and h) Computational details of the mappings, inter-couplings and co-dependencies of the various functions roles, such as: External → Input; Input → Output; Input → Input Processing; Input → Memory; Input Processing → Memory; Memory → Output Processing; Memory → Output; Output Processing → Output; Output → External.

An exhaustive enumeration of the published cognitive architectures is beyond the scope of this paper. Some major examples integrated cognitive architecture frameworks include those of **ACT-R** [Anderson 2007], [Anderson 1993], [Anderson & Lebiere 1998]; P. **Haikonen** [Haikonen 2003], [Haikonen 2007a], [Haikonen 2007b]; **LIDA** [Ramamurthy et al. 2006], [Franklin 2002], [Franklin et al. 1998], [Franklin 1995]; M. **Minsky** [Minsky 2006]; **SOAR** [Laird 2008], [Rosenbloom et al. 1993], [Newell 1990], [Laird et al. 1987], ; and A. **Sloman** [Sloman 1999], [Sloman & Chrisley 2003], [Sloman & Chrisley 2005], [Sloman et al. 2005].

Haikonen has attempted to build artificial neural network based modules for use in architectural implementations. However, neither Minsky's nor Sloman's model have yet been specified to an engineering level of detail, whereby they can be implemented on practical hardware platforms.

Other published significant cognitive architectures include: **BDI** [Bratman 1987]; **CLARION** [Sun 2006], [Sun 2001], [Sun & Peterson 1998], [Sun et al. 2001]; **EPIC** [Kieras & Meyer 1997]; **ICARUS** [Langley & Choi 2006]; **MicroPsi** [Bach 2003a], [Bach 2003b]; **Novamente** [Goertzel & Pennachin 2005a], [Looks et al. 2004]; **Polyscheme** [Cassimatis et al. 2004], [Cassimatis 2002]; **THEO** [Mitchell 1990].

Currently, integrated computational architectural models explicitly based on conventional computer architectures, such as parallel computers, multi-core and many-core machines, for implementing cognitive robotics, are missing in the published literature.

There also exists considerable ongoing research on computational models of individual aspects of machine-related or artificial consciousness and self-awareness, such as consciousness, attention, memory, imagination, emotions, etc. The Global Workspace Theory (**GWT**) of B. Baars, [Baars 1988], [Baars 1997], including variants, is the most commonly used computational architecture model for consciousness.

The significant differences of the CR/SARAMA architecture from the extant ones consists of the following: a) Explicit provisioning of any necessary or desired

equivalent of a mental or cognitive feature, instead of reliance on emergent properties; b) The irredex architecture organization at all scales; c) The targeting of multi-core, many-core (poly-core), parallel and distributed computer architectures as the ultimate platforms for the realization and implementation of the executable architecture; d) Coping with changes and uncertainty via explicit engineered stages of evolution of artifacts (transformations, upgrades, transitions, mutations, augmentations, migrations, replacements, subsumptions, usurpations, etc.), instead of relying on (automatic) learning mechanisms and processes; even learning mechanisms and processes can be subject to evolution; and e) Engineering-as-evolution approach to the multi-stage or multi-phase realization of any and every equivalent of chosen mental, cognitive, affective, motivational and conative features and roles, explicitly recognized and adopted in the architecture. For example, Consciousness is implemented as multi-stage evolved structures.

The CR/SARAMA effort also intersects with the concerns and goals of Autonomic Computing Systems (**ACS**), for example, **Project Joshua Blue** [Ganek & Corbi 2003], [Adams et al. 2002], [Adams et al. 2001], [Alvarado et al. 2001]. The ACS is typically modeled as having the knowledge-driven or knowledge-based cybernetic loop of Sensor → Monitor → Analyze → Plan → Execute → Effector. Additionally, the ACS initiative is attempting to build large-scale software systems that have the following engineering attributes: autonomy, fault-tolerance, self-support, self-awareness, self-modifying, self-motivating, self-managing, self-monitoring, self-healing, self-protecting, self-optimizing, and self-regulating. The CR/SARAM initiative, by using the concept of Creator-in-a-Box, seeks to augment the above "self-" concerns and roles with others, such as: self-organizing, self-reorganizing, self-maintaining, self-repairing, self-upgrading, self-correcting, self-assembly, self-enhancing, self-improving, self-remediation (auto-remediation), self-learning, self-control (auto-control), self-developing, and self-evolving.

7. Summary and Conclusions

To make progress, as a starting point of engineering artificial consciousness, one may want to be merely biologically inspired, rather than attempting to emulate or duplicate biological phenomena and manifestations of consciousness, self and self-awareness.

The paper provided a careful characterization of artifact-as-organism that allows one to exploit biological evolution as a fruitful metaphor for the emergence of artificial consciousness. This approach gives rise to the concept of Creator-in-a-Box role as the primary role that can be discharged by Self, Consciousness and Self-Awareness in robots, automata, machines and artifacts.

Operationally useful characterizations of roles, aspects and concerns of phenomena and manifestations of Self, Consciousness, the Unconscious and Self-Awareness, etc. have been provided within the paradigm.

One advantage is that these teleological characterizations can readily be translated into computational and architectural models that can be used to bring cognitive robotics closer and closer to engineering realization in modern digital systems.

Acknowledgements

The author is grateful to the B. Goertzel for suggesting making tighter and more detailed connections of the current work with a larger number of other cognitive architectures, and who also provided useful references and Web links.

References

[Adams et al. 2002] Adams, S.S., Alvarado, N., Burbeck, S. & Latta, C., Bootstrapping semantics in an autonomic computing system. *Fourth International Workshop on Computational Semiotics for Intelligent Systems, Joint Conference on Information Systems (JCIS)*, Chapel Hill, NC (2002).

[Adams et al. 2001] Adams, S. S., Burbeck, S., Alvarado, N., & Latta, C., Project Joshua Blue: Common sense via common experience. Anchoring symbols to sensor data in single and multiple robot systems. 2001 *AAAI Fall Symposium.* AAAI Press, (2001).

[Alvarado et al. 2001] Alvarado, N., Adams, S. S., Burbeck, S. & Latta, C. Project Joshua Blue: Design considerations for evolving an emotional mind in a simulated environment. Emotional and intelligent II: The tangled knot of social cognition. *2001 AAAI Fall Symposium*, AAAI Press, (2001) pp. 1 – 2.

[Anderson 2007] Anderson, J.R., Cognitive Architecture. Chapter 1 of Anderson, J.R., *How Can the Human Mind Occur in the Physical Universe*? Oxford University Press, New York, NY (2007), pp. 1 – 43.

[Anderson 1993] Anderson, J.R., *Rules of Mind*, Lawrence Erlbaum Associates, Mahwah, NJ (1993).

[Anderson & Lebiere 1998] Anderson, J.R. & Lebiere, C., *The Atomic Components of Thought.* Lawrence Erlbaum Associates, Mahwah, NJ (1998).

[Baars 1988] Baars, B., *A Cognitive Theory of Consciousness*. Cambridge University Press (1988).

[Baars 1997] Baars, B., *In the Theater of Consciousness*. Oxford University Press, New York, NY (1997).

[Bach 2003a] Bach, J., The MicroPsi Agent Architecture, *Proceedings of ICCM-5, International Conference on Cognitive Modeling*, Bamberg, Germany, (2003), pp. 15 – 20.

[Bach 2003b] Bach, J., Connecting MicroPsi Agents to Virtual and Physical Environment, *7th European Conference on Artificial Life*, (2003), pp. 128 – 132.

[**Bell & Newell 1971**] Bell C. G., A. Newell, *Computer Structures: Readings and Examples*. McGraw-Hill, 1971.

[Bratman 1987] Bratman, M.E., *Intention, Plans and Practical Reason.* Harvard University Press, Cambridge, MA (1987).

[Cassimatis 2002] Cassimatis, N., *Polyscheme; A Cognitive Architecture for Integrating Multiple Representation and Inference Schemes.* Ph.D. Dissertation. MIT, Cambridge, MA (2002).

[Cassimatis et al. 2004] Cassimatis, N.L., Trafton, J.G., Bugajska, M., & Schultz, A.C., Integrating Cognition, Perception and Action through Mental Simulation in Robots, *Robotics and Autonomous Systems*, Vol. 49, (2004), pp. 12 – 23.

[Chalmers 1995] Chalmers, D.J., Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, Vol. 2., No. 3, (1995), pp. 200 – 219.

[Chalmers 1996] Chalmers, D.J., The *Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press. New York, NY. (1996).

[Chella & Manzotti 2007] (eds.) Chella, A. & Manzotti, R., *Artificial Consciousness*, Imprint Academic, Exeter, UK, (2007).

[Dennett 1991] Dennett, D., *Consciousness Explained*, Little, Brown & Co., New York, NY (1991).

[Franklin 2007] Franklin, S., A Foundational Architecture for Artificial General Intelligence, in [Goertzel & Wang 2007], op. cit. pp. 36 – 54.

[Franklin 2002] Franklin, S., Conscious Software: A Computational View of Mind, In (eds) Loia, V. & Sessa, S., *Soft computing agents: new trends for designing autonomous systems*. Physica-Verlag, Heidelberg, Germany (2002).

[Franklin et al. 1998] Franklin, S., Kelemen, A. & McCauley, L., IDA: A Cognitive Agent Architecture, *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, (1998), pp. 2646 – 2651.

[Franklin 1995] Franklin, S., *Artificial Minds*, MIT Press, Cambridge, MA (1995).

[Ganek & Corbi 2003] A. G. Ganek, A.G. & Corbi, T.A., The Dawning of the Autonomic Computing Era, *IBM Systems Journal*, Vol. 42, No. 1 (2003), pp. 1 – 5.

[Goertzel & Pennachin 2005] (eds.) Goertzel, B. & Pennachin, C., *Artificial General Intelligence*, Springer Verlag, New York, NY (2005).

[Goertzel & Pennachin 2005a] Goertzel, B. & Pennachin, C., The Novamente AI Engine, In [Goertzel & Pennachin 2005], *op. cit.*

[Goertzel & Wang 2007] (eds) Goertzel, B. & Wang, P., Advance of Artificial General Intelligence, IOS Press, Amsterdam, Holland (2007).

[Gray 2007] Gray, W., *Integrated Models of Cognitive Systems*, Oxford University Press, New York, NY (2007).

[Haikonen 2003] Haikonen, P., *The Cognitive Approach to Conscious Machines*, Imprint Academic, Exeter, UK (2003)

[Haikonen 2007a] Haikonen, P., *Robot Brains: Circuits for Conscious Machines*. Wiley & Sons, New York, NY (2007).

[Haikonen 2007b] Essential Issues of Conscious Machines. *Journal of Consciousness Studies*. Vol. 14, No. 7, (2007), pp. 72 – 84.

[Haikonen 2008] (ed.) Haikonen, P., *Nokia Workshop on Machine Consciousness*. Helsinki, (August 2008).

[Kieras & Meyer 1997] Kieras, D. & Meyer, D.E., An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction. *Human-Computer Interaction*, Vol. 12, (1997), pp. 391- 438.

[Laird 2008] Laird, J., Extending the SOAR Cognitive Architecture. *Proceedings of the First Conference on Artificial General Intelligence*, IOS Press, Memphis, TN, (2008).

[Laird et al. 1987] Laird, J.E., Newell, A. & Rosenbloom, P.S., SOAR: An Architecture for General Intelligence, *Artificial Intelligence*, Vol. 33, (1987), pp. 1 – 64.

[Langley & Choi 2006] Langley, P. & Choi, D., A Unified Cognitive Architecture for Physical Agents, *AAAI-2006*, (2006), pp. 1469 – 1474.

[Looks et al. 2004] Looks, M., Goertzel, B. & Pennachin, C., Novamente: An Integrative Architecture for Artificial General Intelligence. *AAAI Fall Symposium Series* (2004).

[Minsky 2006] Minsky, M., *Emotional Machine: Common Sense Thinking, Artificial Intelligence and the Future of the Human Mind.* Simon & Schuster, New York, NY (2006).

[Mitchell 1990] Mitchell, T.M., Becoming Increasingly Reactive (Mobile Robots), Proceedings, *8th National Conference on AI* (*AAAI-90*), Vol. 2. (1990), pp. 1051 – 1058.

[Newell 1990] Newell, A., Unified Theories of Cognition. Harvard University Press, Cambridge, MA (1990).

[**Popper & Eccles 1977**] Popper, K.R. & Eccles, J.C., *The Self and Its Brain*, Springer-Verlag, New York, NY (1977).

[Ramamurthy et al. 2006] Ramamurthy, U., D'Mello, S.K., & Franklin, S., LIDA: A Computational Model of Global Workspace Theory and Developmental Learning, *BICS 2006, Conference on Brain Inspired Cognitive Systems*, (2006).

[Rosenbloom et al. 1993] (eds.) Rosenbloom, P., Laird, J. & Newell, A., *The SOAR Papers: Research on Integrated Intelligence*. MIT Press, Cambridge, MA (1993).

[Sloman 1999] Sloman, A., What Sort of Architecture is Required for a Human-like Agent? In (eds.) Wooldridge, M. & Rao, A., *Foundations of Rational Agency*, (1999).

[Sloman & Chrisley 2003] Sloman, A. & Chrisley, R., Virtual Machines and Consciousness, *Journal of Consciousness Studies*, Vol. 10, No. 4/5 (Apr./May 2003), pp. 133 – 172.

[Sloman et al. 2005] Sloman, A., Chrisley, R. & Scheutz, M., The Architectural Basis of Affective States and Processes. In (eds.) Fellous, J-M. & Arbib, M.A., *Who Needs Emotions*? Oxford University Press. (2005).

[Sun 2007] Sun, R., The Challenges of Building Computational Cognitive Architectures. In (eds.) Duch, W. & Mandzuik, J., *Challenges in Computational Intelligence*, Springer-Verlag, Berlin, Germany, (2007).

[Sun 2006] Sun, R., The CLARION Cognitive Architecture: Extending Cognitive Modeling to Social Simulation. In (ed.) Sun, R., *Cognition and Multi-Agent Interaction,* Cambridge University Press, New York, NY (2006).

[Sun 2004] Sun, R., Desiderata for Cognitive Architectures, *Philosophical Psychology*, Vol. 17, No. 3 (2004), pp. 341 – 373.

[Sun 2001] Sun, R., *Duality of the Mind.* Lawrence Erlbaum Associates, Mahwah, NJ (2001).

[Sun & Peterson 1998] Sun, R., & Peterson, T., Autonomous Learning of Sequential Tasks: Experiments and Analyses, *IEEE Transactions on Neural Networks*, Vol. 9, No. 6, (1998), pp. 1217 – 1234.

[Sun et al. 2001] Sun, R., Merrill, E. & Peterson, T., From Implicit Skills to Explicit Knowledge: A Bottom-Up Model of Skill Learning, *Cognitive Science*, Vol. 25, No. 2, (2001), pp. 203 – 244.

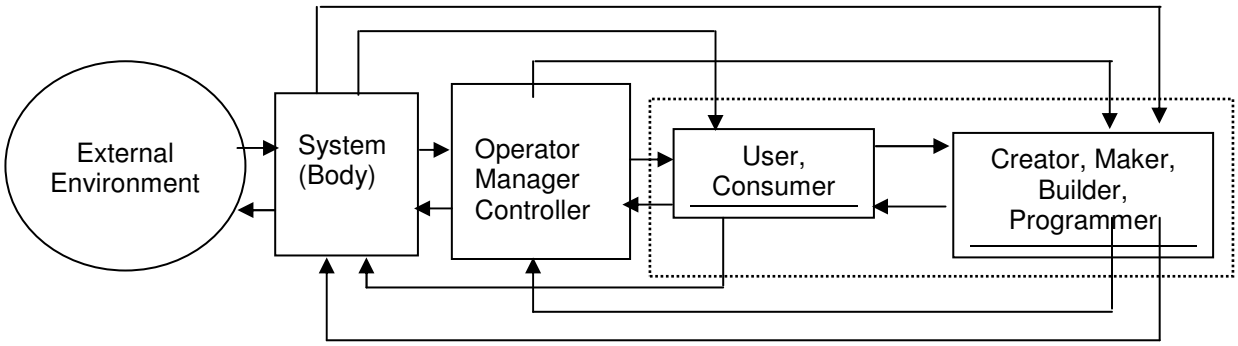


Figure 1: Creator-in-a-Box Architecture Model

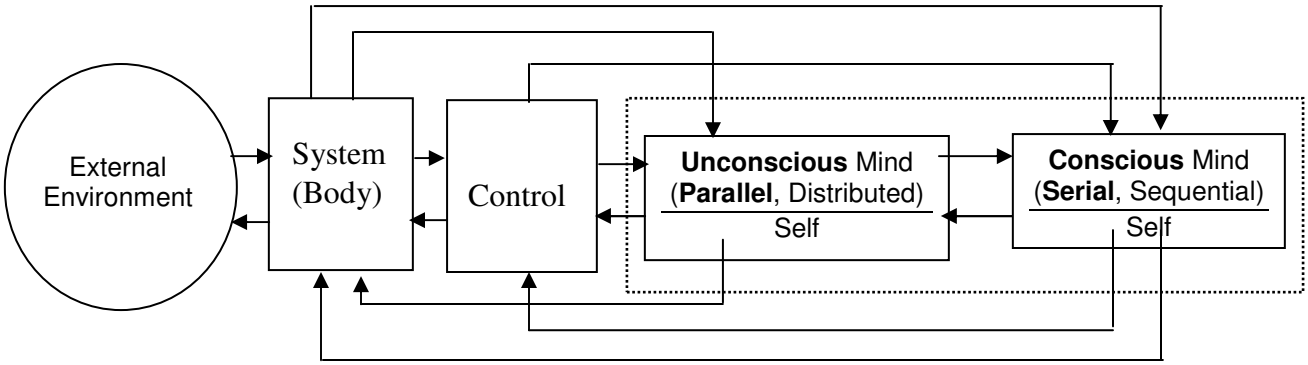
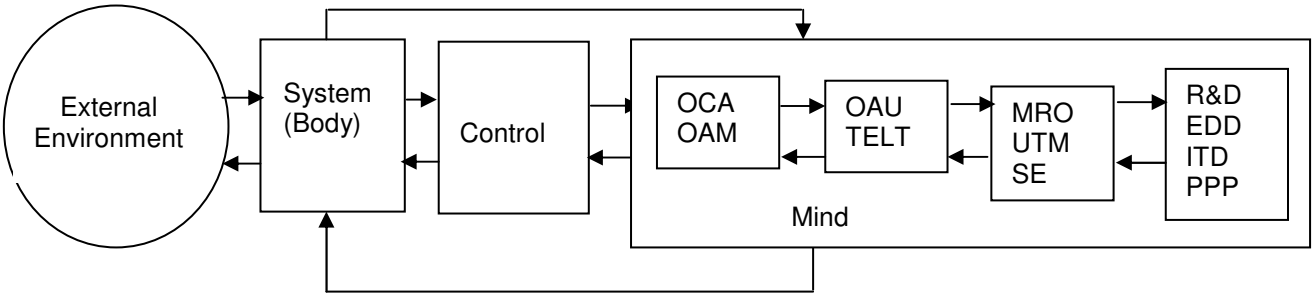


Figure 2: Consciousness / Unconscious Computational Model



OCA – Operations, Control, Admin; OAM – Operations, Management, Admin; OAU – Oper., Admin., Use
TELT – Teaching, Education, Learning, Training; MRO – Maintenance, Repair, Operations; SE – Service Eng.
R&D – Research & Dev.; UTM – Upgrades, Transitions, Migrations; EDD – Engineering, Design, Dev.
ITD – Installation, Testing, Deployment; PPP – Purchasing, Procurement, Provisioning

Figure 3: Evolution & Life Cycle Support (LCS) Roles of the Creator-in-a-Box Concept

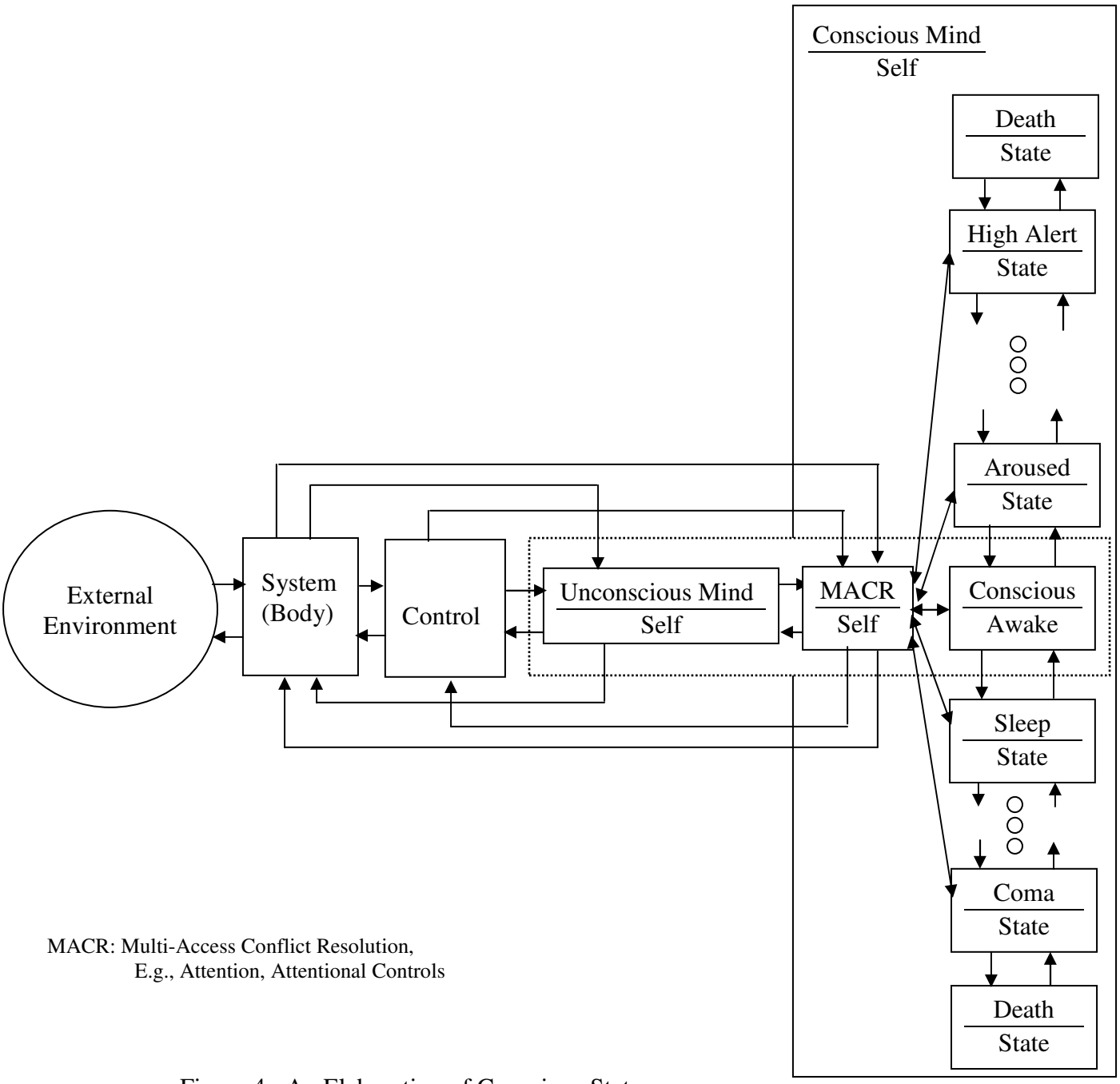


Figure 4: An Elaboration of Conscious States

