Data-Structures for Multisensory Information Processing in an Embodied Machine-Mind

Romi Banerjee and Sankar K. Pal, *Life Fellow, IEEE*

***Abstract*—The real-world is a medley of multisensory information and so are our experiences, memories, and responses. As embodied beings, we respond to the information endogenously and in ways derived from self-defining factors. Thus inspired, we attempt formalization of data-structures to facilitate generation of system-bespoke comprehension-granules of the real-world. The conceptualized structures encapsulate multisensory inputs (sourced from the real-world or memories), intrinsic and deliberate emotions, messages (bearing intermittent process-results, queries, multimodal data, etc.) across system modules and memory units, and sensorimotor responses to the inputs. The structural-schematics are anthropomorphic. These variable-length constructs are theoretically platform-independent, support genericity across data-modality and information-inclusion, and provide for representation of novel sensory-data. An epigenome-styled header node for the afferent data-units provides for the activation of intuitive "fight-flight" behavior. The documentation includes a flow-graph, depicting the translation of information across the data-structures and the different ways of thinking while interpreting a real-world scene or a mind-generated event. Applicability of the structures has been analyzed in the context of comprehension in an embodied mind-machine framework and other similar architectures. Studies herein target contribution to the design of generally intelligent man–machine symbiotic systems.**

***Index Terms*—Autonomous mental development, data-structures, embodied cognition, generally intelligent man–machine systems, multimodal information processing, self-conscious systems, thinking machines.**

I. INTRODUCTION

THIS paper is an elucidation on the schematics of active data-structures in the mind of embodied cognitive systems—those possessing general intelligence, self-awareness, qualia (feeling of "experiencing" [1]–[3] the real-world) and commonsense, are capable of cross-domain contemplation, comprehending, and interacting in the multisensory real-world. The structures described here focus on: a) Encasing all possible features of real-world multisensory inputs and memories with their respective visceral, deliberated or self-conscious system-emotions, b) Conveying multimodal information across functional modules and memory units of the system, and c)Transportation of instruction-codes to the sensorimotor units for activation of motor and affective responses (instinctive and contemplated) to the inputs.

The conceptualization attempts a "paradigm-shift" from fixed-format structures [4], [5] and semantics-only formats [4], [6] to variable-length, "hybrid" [4] schematics which bind real-time sensory inputs, qualia and semantics. These formats are theoretically platform and modality independent; support representation of structured (new instances of experienced data-types) and unstructured (novel data-types) data; broadcasting, multicasting and unicasting of inter-modular messages; and consolidation of objective and subjective information. Furthermore, an epigenome-inspired header unit in the afferent data-structures actuates feedforward traces toward intuitive, procedural system-behavior. Besides the basic modalities of vision, audition, olfaction, touch, and taste, the structures cover secondary senses like temperature, kinesthetics, flavor, space, and time. The design draws inspiration from functions of human brain components, particularly the angular gyrus, amygdala, hippocampus, multisensory neocortex and limbic system, and epigenomes.

This paper begins with a succinct introduction to theories inspiring the formalizations (Section II), followed by a detailed description of the constructs and demonstration of a sample operation scenario (Section III), analysis of the strengths and issues of the design (Section IV), pertinence to some cognitive architectures (Section V), and a summary of the conceptualization (Section VI). Through incorporation of factors denoting the system's sense of self, together with those characterizing objective and semantic-properties of events, the proposed structures envisage extension of the capabilities of existing formats [4], [5], [7].

II. THEORY

A. Epigenomes

The human genome represents the entire DNA-assembly and is key to the "uniqueness" of an individual. While the DNA encodes instructions for protein-generation, leading to various cell-functions, the epigenome [8], [9] is composed of chemical compounds and proteins that can attach to DNA and direct actions, such as the switching on or off of genes and protein-production control in specific cells. These do not alter

Manuscript received March 3, 2017; revised November 16, 2017, February 13, 2018, and March 6, 2018; accepted March 14, 2018. Date of publication March 16, 2018; date of current version September 7, 2018. The authors acknowledge the DAE Raja Ramanna Fellowship of the Government of India. Professor Sankar K. Pal also acknowledges the J.C. Bose Fellowship (grant-in-aid) of the Government of India. *(Corresponding author: Romi Banerjee.)*

R. Banerjee is with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, and is also with the Department of Computer Science and Engineering, University of Calcutta, Kolkata 7000098, India (e-mail: rm.banerjee@gmail.com).

S. K. Pal is with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, India (e-mail: sankar@isical.ac.in).

This paper has supplementary downloadable multimedia material (these are the Appendix sections referred to in the article) available at http://ieeexplore.ieee.org provided by the authors.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2018.2816744

DNA sequences but affect gene expression, arise from natural processes of development and tissue differentiation, and can be altered by environmental exposures or disease.

From the above fact, we reason that if an epigenome-inspired header-node (*H*) of a data-structure (*S*) were to bear pointers to critical instructions and summaries of information in *S*, where *S* encoded afferent sources (e.g., real-time multisensory data) of complex, "slow" deliberative system responses, *H* would actuate primary 'fast' system reactions which in turn would influence the total "expression" or "interpretation" of *S* without modifying *S*. This would particularly be useful in conditions where "thinking on one's feet" is of consequence. Appropriate feed-forward and feed-back channels of data and response-exchanges would be necessary to balance impulsive and rational system-behavior [10].

B. Multisensory Integration and Perception in the Human Brain

The real-world is multisensory and we are constantly bombarded by a mélange of sensory information. We construe the world around us as an amalgamation of multisensory, kinesthetic, and abstract cues. The brain seamlessly consolidates signals "not only from the body but, in some of its sectors, from parts of itself that receive signals from the body" [11], into an adaptive, coherent perceptual symbol [12] of the world; using attention to limit the number of stimuli being consciously processed at any given time [13]. Sensory integration continually modifies our perception of the world and consequently our responses too. It contributes to arousal of the sense of self [11], [14], qualia [15], and emotions [16]. Besides the spatial, temporal and physical properties of stimuli [17], subjective factors [11], [18] (like interests, emotions, expectations, combinatorial plausibility, and context) influence perception.

Neuroscientific evidence [19] highlights the rich interconnectedness and interactions across sensory and emotion areas of the brain, and the consequent problem of categorizing anything as "purely visual" or "purely auditory" or "purely anything." Multimodal memories are stored in distributed networks across the cortex [20], integrated into a coherent whole in the hippocampus and then linked to the neocortex for long-term storage [21]. Embodied simulation [22] forms the bedrock of empathy—the intersubjective sharing of experiences while preserving the sense of otherness. The same neural structures involved in our own bodily self-experiences are reactivated [23] during memory-retrieval and prereflective understanding of behaviors, some mental states and emotions of other individuals.

As embodied [24], [25] organisms, we are hard-wired to respond to stimuli-features (e.g., size, color, motion, sounds, and expressions), with an emotion. The amygdala receives rudimentary visceral signals and processed sensory information, while the angular gyrus integrates them [26]. The limbic system (amygdala, hippocampus, thalamus, hypothalamus, cingulate gyrus, and basal ganglia), brain-stem arousal centers, basal forebrain, and bidirectional neocortex-connections flavor stimuli-inputs with emotional significance that trigger body-states and shape cognitive processing. Primary responses influence the system's flight or fight behavior, and trigger complex secondary emotions (somatic-markers [11]) as functions of environment-semantics. This latter stage commences with conscious "internal deliberations and reflections" [1], [27] over multimodal "images" [11], forming (ideally) rational and adaptive, thought processes. In the nonconscious level, the prefrontal cortex endogenously responds to bespoke embodiments of emotional responses and events, aroused by these images. Sensorimotor response-instructions are sent to the associated organ-structures through the cerebellum or cerebral cortex [28]. The detection and experience of affects orients attention to the more relevant or consequential stimuli [29], [30].

Traditional views assumed that cortical integration of disparate multisensory information was handled by specialized, higher-order association areas of the neocortex. However, recent neurobiological data [33], suggest that much, if not all, of the neocortex is multisensory; targeting robust and efficient perception of the complex world. Multisensory neurons in the superior colliculus are fired more if multimodal signals reach the cell within a window of 500 ms, and activity in the posterior superior temporal sulcus increases for related bimodal auditory–visual inputs. High-density electrical mapping studies in humans depict multisensory convergence as an outcome of the full range of anatomical connections (feedforward, feedback and lateral) [34] available in brain circuitry.

In addition to the basic sensory perceptions, concepts of space and time [17], [20], [35]–[37] are fundamental to real-world understanding. The two basic forms of mental spatial representations [38] are: a) Egocentric (Self-to-Object Mapping) – a sense of space where one has a more accurate sense of neighboring objects than distant ones; and b) Allocentrism (Object-to-Object Mapping) – an abstract sense of space used to plan routes. The brain must also account for speed disparities between and within its various sensory channels to resolve temporal associations between features such that object-wholeness and event-concurrency are perceived. These spatio-temporal representations of the environment rely on a confluence of sensory (e.g., vision, olfactory, and kinesthetics) information. Feature-integration [18], [39] is a fundamental challenge for the nervous system.

Drawing on: 1) the importance of multisensory stimuli-acquisition and processing to function as an embodied, empathetic, and cognitive system in the real-world, and 2) the nature of associations between stimuli, affects, and perceptions, from the preceding paragraphs, this paper documents our attempts at formalization of data-structures aiming to support multimodal stimulus granulation, representation and intermodular exchanges in cognitive frameworks. Intuitively, these structures require the underlying system-architecture to:

1) comprise unisensory, multisensory and motor units that: a) capture and process multimodal inputs from the external and internal worlds of the system, and b) exhibit and record sensorimotor behavioral attributes;
2) create, store, retrieve, and process multisensory memories, knowledge, and experiences, that are inclusive of objective and subjective (qualia, feelings, etc.) properties, and are connected as networks of intuitive, commonsense, procedural, declarative, semantic and episodic knowledge;

3) be instinctive, self-conscious, adaptive, and capable of handling novel stimuli, information, or experiences;
4) possess functional-units to emulate cognitive processes like:
    a) embodiment, leading to spatio-temporal multisensory feature binding into dynamic "allocentric and egocentric" [38] maps of the real-world;
    b) activation, operations on, and comprehension of primary and secondary emotions of the self and peers;
    c) real-time estimation of fixation points and saccades, attention modulation, and arbitration, as functions of perceptions, importance, interest, and novelty of the problem [40];
    d) syntax, semantics, and context disambiguation;
    e) intuition, deliberation, reflection, self-reflection, creative contemplation, and counterfactual planning; feedforward and feedback channels to activate and assess instincts and reasons;
    f) mapping between feelings and linguistic expressions;
    g) bind experiences into seamless timelines; facilitate ease of mental traversal into the past, present, and future visualizations, and emulation of subjective time [41].

The machine-mind architecture, outlined in [31] and [32], being conceptually compliant with the aforementioned prerequisites, has been considered as the ground architecture for the proposed schematics.

C. Machine-Mind Framework

The machine-mind framework [31], [32] has been conceptualized as an embodied, self-aware, modular, data-driven, self-regulated, adaptive, causal, and social system. It is largely based on Minsky's "Society of Mind" [2] and "Emotion Machine" [1], and endeavors emulation of thinking (across all the different layers of an active mind [1]–[3]) and understanding for social, empathetic, and symbiotic relationships with other machine and human peers. Its modules encapsulate real-world multisensory inputs pertaining to an event (E), and activate and process associated affects, experiences, and knowledge to produce a network of hypergraphs of concept-associations toward perception of E. It operates through incremental cycles of: 1) problem discovery and decomposition; 2) goal identification; 3) intercomponent exchanges (memories, knowledge, and partial results); 4) frame-manipulation trials and executions; 5) temporal trajectory adjustments; 6) prospective visualizations; 7) affective processing; 8) solution analyses; 9) improvisation; 10) interpretation-alignment with interests and motivation; 11) information/solution-consolidation; and 12) long-term memory updates and reinforcements. Fig. 1 illustrates framework components and interconnections, and Table I its functional details.

TABLE I
MACHINE-MIND ARCHITECTURE COMPONENTS AND FUNCTIONS

Category	Macro-agency	Sub-components	Major functions
Agency	Sensory Gateway (**SG**)	*Primary:* Vision (**V**), Audition (**A**), Olfaction (**O**), Taste (**Ta**), Tactile (**Tc**) *Secondary:* Temperature (**Te**), Pain (**P**), Kinesthetics (**K**), Balance (**B**)	Medium of 2-way communication with the external world – sub-agencies receive and respond to real-world inputs
		Self (**Sf**)	Manifestation of machine-proprioception and sense of self; emotion, qualia and affect activation
	Deducer (**De**)	Syntax (**Sy**)	Syntax resolution
		Semantic (**Se**)	Semantics resolution
		Recall (**Re**)	Maps problems or sub-problems to relevant memories; context resolution; knowledge reinforcement
		Creative (**Cr**)	Deliberation, reflection, imagination, improvisation, MIQ
		Summary (**Su**)	Memory consolidation; strategic comprehension evaluation to censor, suppress or encourage lines of thought
	Manager (**M**)		System-administrator; activation and execution of 'involuntary' essential system functions (resource-arbitration, self-appraisal, etc.)
Memory	Long-term structures	Lexicon (**L**)	System-vocabulary (words, facts, imagery, metaphors, etc.)
		Answer-Library (**AL**)	Repository of experiences [(*context_parameters*, *problem, solution_strategy, result, reasons*) tuples]
		Concept-Network (**CoN**)	Hypergraphs of multimodal abstract and concrete concepts
		Commonsense-Network (**ComN**)	Hypergraphs of commonsense and intuitive behavior;
	Working-memory	Log	Record of time-stamped, reason-annotated agency-activities, system-states, inter-agency messages and partial-results; triggers agencies, generates error signals, facilitates backtracking and counterfactual reasoning
		Frame-Associations (**FA**)	Categorized into *global* and *local* (per sub-agency) types; recollections are placed in the former and relevant sections thereof are transferred into the latter for deliberations by sub-agencies; *local* **FA** of **SG** sub-agencies store real-time sensory memories; *local* **FA** of **De** and **M** sub-agencies serve as their resident solution test-grounds before global suggestions of (*subproblem*, *solution*, *reason*) tuples through **Log**; sub-agencies may share sections or all of its *local* **FA** with other agencies.
	System-memory constructs	Working-Set (**WS**)	Set of pointers to *global* **FA** networks being referenced within a narrow time-window (of the order of seconds).
		Active-Frames (**AF**)	Set of pointers to *global* **FA** networks being referenced within a broad time-window (of the order of minutes); **WS** ⊆ **AF**.
		Passive-Frames (**PF**)	Set of pointers to *global* **FA** networks, which were pruned out of **AF** due to irrelevance or lack of use; support rapid 'on-demand' re-placement of frames into *global* **FA**.

[These agencies collaborate to realize crucial cognitive processes such as:
1. **Su, M:** Co-ordinate inter-agency activity; use heuristics and approximation schemes to prevent combinatorial explosions of thoughts; fixation-delineation.
2. **Se, Sf,** secondary **SG** sub-agencies: Create a sense of space and time (real-time and subjective time [41]).
3. **Re, Cr, Su, Se, Sf:** Use indicators of conscious thoughts and system-states to arouse subjective event-interpretations; resolve cognitive-biases; emulate self-motivation and innate mental-rewards.
4. **Re, Cr, Su:** Constitute the 'Difference-Engine' [2].
5. **Se, Sy, Su, Cr:** Emulation of creativity, improvisation; handle novel information.
6. **L**, **CoN** and **AL** elements are incorporated into **ComN** after prolonged reinforcement
7. **WS**, **AF**, **PF**, intuitively, reflect the system's current region of interest or attention.]

Given an E at *time*, the sensory gateway (SG) and self (Sf) units of the framework are endogenously activated. The actuated SG subunits extract and record symbols from E and visceral Sf-responses in their *local* frame association (FA) unit, over some fixation-duration (Δ_{time}) determined by manager (M) and summary (Su). This triggers iterations of commonsense, intuitive, learned, deliberative, reflective, and self-reflective actions by deducer (De)—beginning with spatio-temporal feature-binding of the inputs and instinctive interpretations, followed by incremental processing of pertinent long-term multimodal memories and knowledge, and ideas and conceptualizations arising out of De's operations. Agency-operations could involve multiple data-types (e.g.,

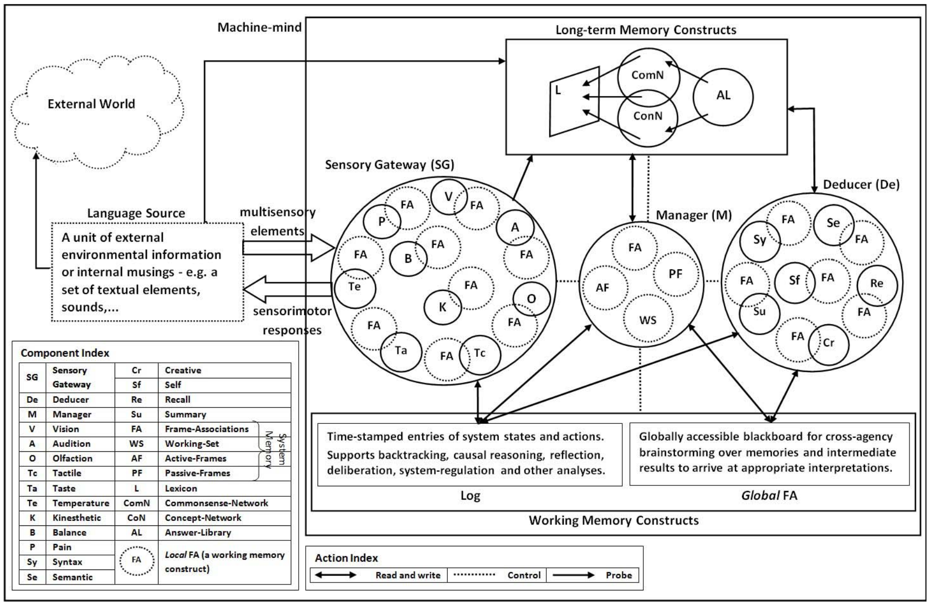


Fig. 1. Machine-mind framework [31], [32]. Illustration of the functional components and interactions thereof. The framework describes a modular system: the SG extracts multisensory inputs from the external world and transports system responses to the same, and the M and the De provide core support and semantic interpretation functionalities, respectively. The long-term multimodal memory units are dynamic, continually updated, repositories of instinctive reactions, commonsense, autobiographical memories, concepts, and knowledge. The *local* scratch-pad FA units of the modules support in-memory operations, while the *global* FA serves as the common blackboard for deliberations across interacting components. Actions are recorded in the Log to facilitate reflection, backtracking, and counterfactual planning. Procedural responses, never-ending learning, self-appraisal, creative contemplation, and self-consciousness are other properties of the framework (refer Section II-C and Table I for details).

text, imagery, and auditory memories for a sentence in *E*). De continually assesses the status (familiarity, semantics, context, affects, interests, relevance, understanding, etc.) of the problem (words, sentences, memories, ideas, visualizations, etc.) under process and opportunistically monitors fixation-saccades [42] and fixation-points, suggests interpretation-strategies and outcomes (hypotheses, predictions, theories, etc.) en route to comprehension of *E*. (See Appendix B in the supplementary material and [31], [32] for details.)

The following section describes our attempts at defining data-structure schematics for real-time multisensory data handling in the machine-mind and similar cognitive architectures. These structures aim to support encapsulation of real-world multimodal data and consequent system-visceral responses, intermodular exchanges of such data, instructions, and queries, and transportation of system-responses to the external world through SG-like units.

III. PROPOSED WORK

In this section, we describe our formalizations of generic formats of data-structures to aid emulation of multisensory perception in machine-mind-like cognitive systems toward bespoke, sympathetic real-world comprehension, and interaction. These structures have been envisioned to provide for representation, processing, and carriage of multimodal information—arising from the external and internal worlds of the system—across modules and environments. Though this paper does not elaborate on the schematics of long-term memory units in these architectures, the studies herein provide crucial pointers to probable structures for the same. Section III-A describes the schematics, Section III-B outlines their operation principle in connection with the machine-mind architecture in [31] and [32], and Section IV discusses features and issues of the conceptualization.

A. Categories of Data-Structures

Data-structure configurations influence: 1) type and properties of data that can be handled; 2) the system-instruction set; and 3) nature and execution of algorithms. In the context of embodied cognitive architectures like the machine-mind [31], [32], these structures necessitate facilitating the execution of complex cognitive processes involving multimodal data-types across all layers [1]–[3] of thinking. Key points on multisensory integration and perceptions (with reference to Section II) that the structures have considered are:

1. Allocentric and egocentric maps of a region of interest (*R*) are functions of proprioception and spatio-temporal feature binding of objects in *R*.
2. Perception of an object or an event arises from its attribute-values, aroused qualia and memories (commonsense, semantic knowledge, episodic experiences, etc.).
3. Perception is an incremental process [42], beginning with primary visceral, instinctive interpretations followed by secondary, contextual, deliberative, and self-conscious contemplation.

Inspired thus, the defined structures bear placeholders for perceptions of multisensory attributes of objects in *R*. These are integrated across space and time into complex granules that trigger associated memories, interrogative mentalese, intra-system exchanges and operations, toward rational responses to the inputs. An epigenome-inspired essential summary of real-world data activates instinctive reactions, influences context-predictions and prompts secondary "deliberative" responses—affecting sensory-input interpretations, analogous to modulating gene expression. Primary properties of the defined structures are:

1) Hybrid format—to support representation and processing of structured, unstructured, and novel real-time data.
2) Variable-length data-placeholders and messages, in sync with unpredictable amounts of real-time operational data—a conscious shift away from unimodal [7], sparse, fixed-format structures [4], [5], and their ensuing memory-space wastage and loss of information; intuitively involves complex hardware requirements.
3) Structures encapsulate both objective (e.g., stimuli feature-values) and subjective (e.g., system emotions) information.
4) Platform-independence to facilitate use across cognitive frameworks and hybrid-system interactions.

The following sections elaborate on the structure-categories and their properties. The formats have been illustrated as regular expressions.

Here, *time* = fixation start-time, $\triangle_t$ = fixation-duration, E = event/environment, and ' ' = delimiter between two values. The regular expression format index – let, A = string variable; a = string character; if $A = ab$ and $B = bc$, then:

1) AB = concatenated A and $B = abbc$.
2) $A|B = A$ (Boolean OR) B.
3) () = grouping or parantheses; e.g., gr(e | a)y = {grey, gray}.
4) A^+ = One or more occurrences of A; e.g., $A^+ = a^+ = \{a, aa, aaa, \ldots\}$; $(a\ b)^+ = \{(a\ b), (a\ b)(a\ b), (a\ b)(a\ b)(a\ b), \ldots\}$.
5) A^* = Zero or more occurrences of A; e.g., $A^* = a^* = \{\emptyset, a, aa, aaa, \ldots\}$; $(a\ b)^* = \{\emptyset, (a\ b), (a\ b)(a\ b), (a\ b)(a\ b)(a\ b), \ldots\}$.
6) $A^?$ = Zero or one occurrence of A; e.g., $A^? = a^? = \{\emptyset, a\}$.

1) Afferent Structures: These are multifield constructs encapsulating primary and secondary multisensory information describing the system's perception of the external or internal world and its visceral responses thereto. Fields include placeholders for stimulus-feature values, perception-time, spatial coordinates of the stimulus-source, and innate system responses. The structures are spawned in the following levels:

1) *Level_1:* Comprise raw stimulus-features (e.g., shape, color, orientation, movement, sound, and smell)—exuding from source(s) in the system's egosphere [38], [43], [44]—acquired per sensory unit, through [*time*, Δ_t]. These flow through the limbic system, triggering consequent primitive emotions (e.g., "frightening" and "spine-tingling"). This data is stored as a spatio-temporal granule (refer Fig. 2) in the local memory of the sensory units (e.g., the *local* FA of the SG-agencies in [31] and [32]). **Format** (per sensory organ in SG-like modules): (*sensory_unit_id* (*fixation_point_id duration*) ((*source_spatial_coordinates*)$^+$ (*feature_name feature_value visceral_emotion*)$^+$)*), where *visceral_emotion* = (*affect strength valence*)$^+$.
2) *Level_2:* These are level_1 structures grouped spatially and temporally into multisensory feature-sets, per object, in the region of interest. These are instantiated after (ideally, in parallel with) preliminary feature-binding of level_1 structures and are available for processing after a [*time*, Δ_t] timeframe. **Format** (per object, approximately identified by spatial coordinates): (*epigenome_header* (*fixation_point_id duration*) (*source_spatial_coordinates*)$^+$(*sensory_unit_id* (*feature_name feature_value visceral_emotion*)$^+$)$^+$), where *epigenome_header* = (*header_id fixation_point_id* (*source_spatial_coordinates*)$^+$ (*sensory_unit_id* (*feature_name feature_value visceral_emotion feature_pointer*)$^+$)$^+$*summary*).[1]

The epigenome-inspired (*epigenome_header*) field of a level_2 afferent structure (S) is a critical summary of S—sensory information that the system requires being "actively" or "consciously" [11], [18] aware of. It encases the system's instinctive responses to "high-impact" [11], [18] sensory-information enclosed in S and pointers to corresponding data-fields therein. Intuitively, in these responses, the *visceral_emotion* would (exactly or synonymously) have *strength* $\in$ ["definitely," "moderate-high"], and/or opposite *affect* and *valence* senses implying inherent "surprise," "sarcasm," "alarm," "awe," or similar emotions (e.g., *affect* = "good" and *valence* = "−"). For a multisensory feature (F), constituent unimodal responses with *strength* $\in$ [moderate-high, "very low"] would be included if these were ratified across modalities, such that the cumulative *visceral_emotion* for F gets strengthened [e.g., total effect of multisensory food_temperature in Fig. 7].

The *epigenome_header* actuates an initial feedforward [34] pulse through the procedural-memory and associated modules to trigger preliminary interpretations of the sensory-inputs and activate predefined responses. This is particularly effective when responding to perilous events—subtracting delays out of the rigors of in-depth processing. These interpretations influence contextual inferences [45] and calibration of expectation-windows for relative arrival times for stimuli (e.g., clapping of hands, fall, and thud) such that event-simultaneity is perceived [20]. This field underlines the observation that "*everything we do is an approximation of sophistication*" [46]. Just as the epigenome encodes the gene-expression context (refer Section II-A), the header sets the context for afferent-data interpretation.

Typically, the level_2 structures flow through system modules, activating object-recognition and associated memories in the process. For a given (*time*, E), a 3-D network (akin the ego-sphere [43], [44]) of spatially connected level 2 afferent structures of perceived objects in E, forms a mental-map (allocentric and egocentric) [38] of E in the machine-mind. Temporal connections across structures extracted over consecutive [*time*, $\triangle_t$] time-frames symbolize environment dynamics. Incremental integration of interpretations of such maps contributes to the system's concept of the general context of E and influences sensorimotor responses. Fig. 2 illustrates a sample translation from level_1 to level_2 afferent structures.

2) Intermodule Messages: Intermodule messages represent granules of information exchanged between system

---

[1] The substring *(sensory_unit_id (feature_name feature_value visceral_emotion feature_pointer)*$^+$*)*$^+$ constituting *epigenome_header* is an essential subset of the substring *(sensory_unit_id (feature_name feature_value visceral_emotion)*$^+$*)*$^+$ comprising the level_2 afferent structures.

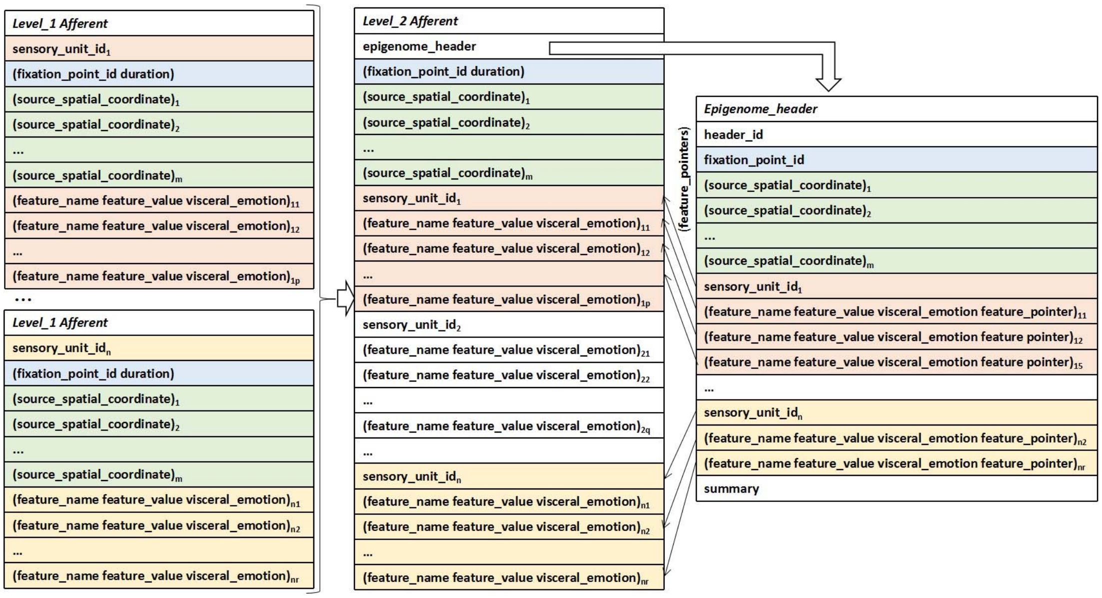


Fig. 2. Afferent structural schema and levels. Depiction of spatio-temporal feature-binding (through color-mapping across levels) for a single multimodal source.

components (modules and memory). These exchanges could be in the order of afferent-data, retrieved memories, intermittent process-queries and results, or system instructions. **Format**: (*message_id* (*#message_sets* (*pointer*)$^+$) *system_time* (*fixation_point_id*)$^?$ *source_module message*), where, *message_id* relates *source_module* requests to *destination_module* responses; *#message_sets* = number of messages packed into the module; *message* = ((*destination_module*)* (*information_code information*)$^+$)$^+$; *information_code* indicates the "type" (*query* OR *data* OR *instruction*) of information-exchange; *information* = (*query* | *data* | *instruction*); (*pointer*)$^+$ = set of pointers to individual message sets; for afferent-data exchanges, *data* = level_2 afferent granule (G) and *fixation_point_id*$^?$ = placeholder for the corresponding *fixation_point* code in G.

A set of consecutive intermodule messages represent snap-shots of the system's active internal environment (∼ "thoughts in action"). A *source_module* could broadcast, multicast or unicast messages across *destination_module*s. The *source_module* is the default recipient of data-responses to its query-messages. Fig. 3 presents the generic intermodule message format.

3) Efferent Structures: Efferent structures enclose sensorimotor commands—translations of results of system processes on stimuli (sourced from the external world or system-mind), into physical responses thereto. These encode instinctive and deliberated responses (e.g., fight or flight and saccadic adjustments) in the form of intermodule "instruction" messages for transportation to destination sensorimotor units. While intermodule messages carry linguistic equivalents of instructions (I) (e.g., "want to laugh"), their efferent consequents link "action words" in I to their physical interpretations (e.g., codes to actuate the limbic system and motor elements to execute "laughing") (refer Fig. 4 for a pictorial representation of the efferent structures). **Format** (per efferent instruction): (*efferent_id* (*message_id destination_module instruction*) (*sensorimotor_unit_id action*)$^+$).

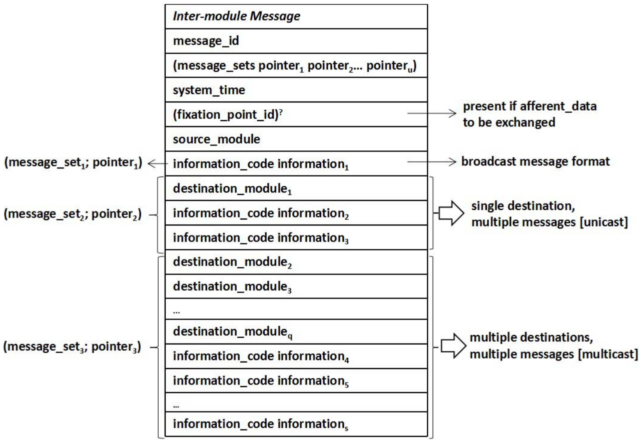


Fig. 3. Intermodule message schema.

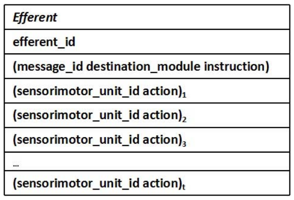
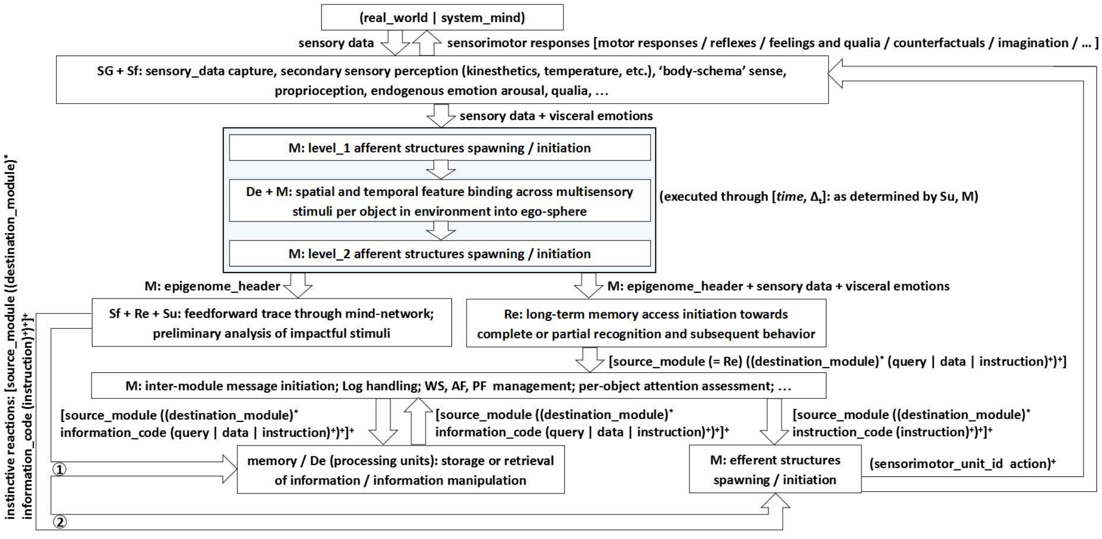


Fig. 4. Efferent structural schema.

Section III-B and Fig. 5 describe the nature of data translation across the defined structures.

B. Operation

A real-world scene is a semantically coherent collection of objects. This, in addition to the state of the self [1], [3], equips the system with contextual cues. The context determines where objects may occur, their inter-relationships and interpretations—consequently inducing anticipations, arousing



Note:
Pathway descriptions:
1. Instinctive reactions contribute to deliberations, reflections and self-conscious contemplation
2. Instinctive reactions short-circuited to efferent structures for prompt translation into 'fight or flight' hardwired sensorimotor responses

Fig. 5. Instantiation of the data-structures in the machine-mind architecture [32] (agency-codes follow format in Fig. 1, refer Section II-C and Table I for a brief on the machine-mind working principle). Level_1 afferent structure-granules are instantiated at the onset of a fixation-point (*time*) and are assimilated into level_2 spatio-temporal afferent granules at the end of a fixation-duration ($\triangle_t$). The *epigenome_header* triggers a feedforward trace through the procedural repository, results of which activate hardwired efferent codes (pathway-2), and contribute to consequent deliberative, self-conscious efferent responses (pathway-1). The machine-mind components exchange data, instructions, and queries through the intermodule messages. The **M** instantiates the constructs and the De is in charge of their processing into rational sensorimotor instructions.

associated emotions, and affecting general comprehension. Perceptual processing of a real-world scene entails deriving meaning of the medley of multimodal data inundating the sensory units of the system—processing captured stimuli in sync with aroused memories, leading to sensorimotor responses.

In machine-mind like cognitive frameworks, the afferent structures transport real-world sensory data to the system core modules, whereupon these units exchange information, queries, intermittent results, and instructions through intermodule messages leading to appropriate efferent sensorimotor codes. Attention modulation (a function of interest, importance and relevance) across objects in a region of interest, is the key to handling the deluge of sensory information (refer [40] and Appendix C in the supplementary material). Fig. 5 is a pictorial representation of the information-flow through the defined data-structures in the machine-mind framework-agencies (see Section II-C).

IV. RESULTS AND DISCUSSION

Drawing from Figs. 1 and 5, Figs 6 and 7 illustrate the theoretical instantiation and flow of data during comprehension of a sample multisensory real-world scene by an embodied machine-mind like architecture. The execution assumes that the system's attention is fixated on a blue-bowl of reddish-yellow, aromatic, steaming food, relegating the rest of the environment to the background (Section II-B enumerates some framework requisites). Key features of results are as follows:

1) Processes in Figs. 6 and 7 present two out of many possible responses to the presented scenario. These trajectories were selected from a compilation of possible action-suggestions from 20 individuals (ten females and ten males in the age-group 28–32 with similar educational backgrounds). These subjects were individually presented the same blue-bowl of steaming curry and were asked to make notes of their behavioral-visualizations, thoughts, and feelings in natural language. These notes were compared and the most common responses (15 out of 20) were theoretically encoded in terms of data-structure-actions and annotated with the type of contemplation ("deliberation," "reflection," etc.) involved.
2) The trajectory in Fig. 6(a), (b), and (d) illustrates deliberative thinking. Fig. 6(a)–(c) and Fig. 7(a)–(c) depict impulsive responses overcome by "reflections" and "self-conscious emotions" in the latter stages. The instinctive behavioral strain was envisioned by 15 human-subjects, five of the remaining chose to be deliberative, and two though traced the instinctive response in detail; mentioned deliberative possibilities as well.
3) Fig. 7(a) portrays cross-modal affirmation of judgments ("semisolid," "hot," and "savory") made in Fig. 6(c).
4) Figs. 6(b) and 7(b) highlight an instinctive "urge to taste," triggered across multiple modalities, and the system's consequent "stronger" deliberated/instinctive motivation toward "tasting"; overriding "panic" due to "steam."
5) Arousal of primary (vision, olfaction, touch, and taste) and secondary (temperature, flavor, and kinesthetics) senses have been illustrated.
6) The machine-mind M and De agencies trigger and operate on the data-structures, respectively.
7) The *epigenome_header* summarizes vital attribute-information, where:
   a. *visceral_emotion strength* ∈ ["absolute," moderate-high]: $(\text{color}_1 \text{ reddish-yellow (delighted high } +))_{11}$ in Fig. 6(a);

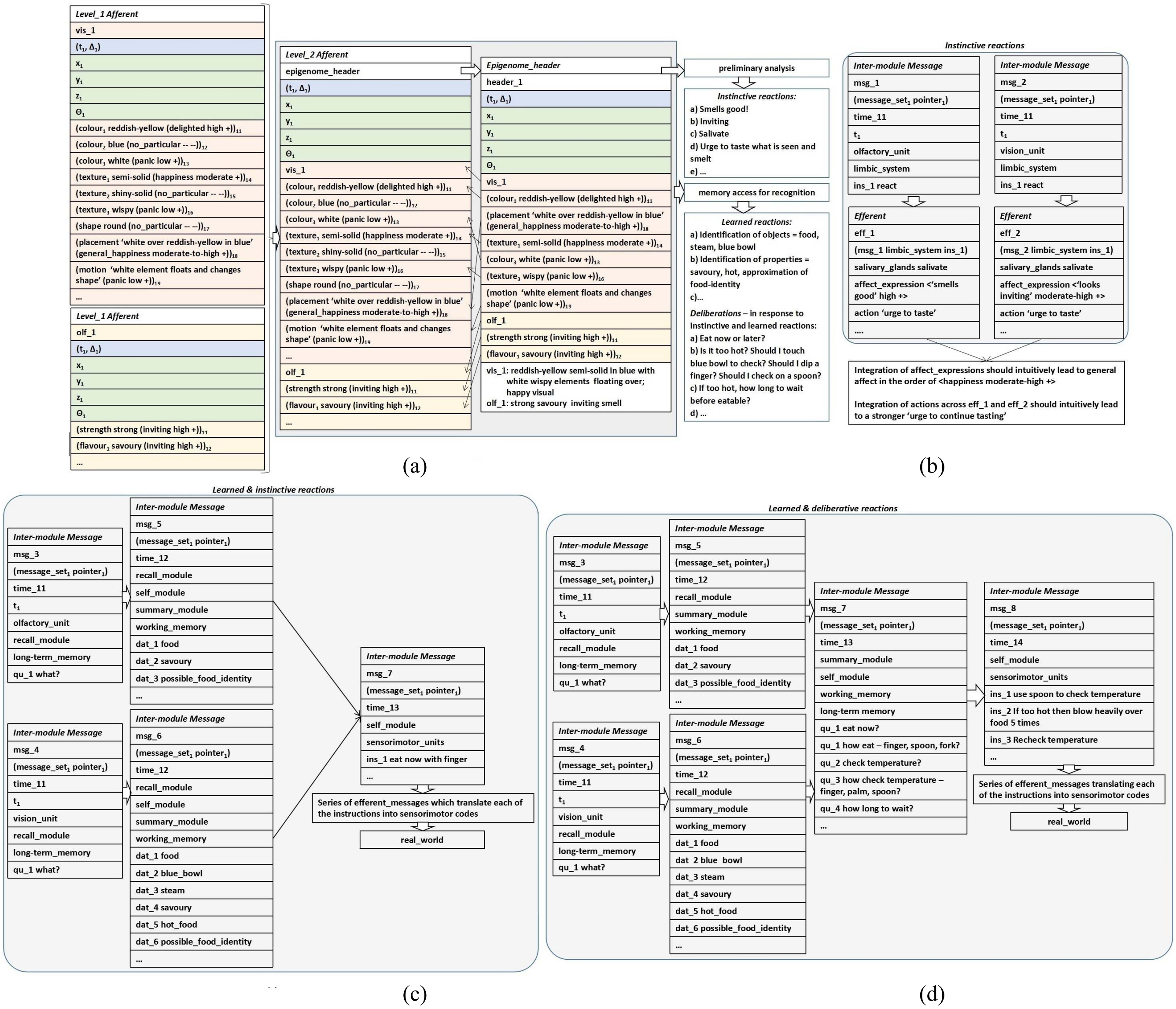


Fig. 6. Part I. Stage 1 results of comprehension of the given scene–"a blue-bowl of steaming reddish-yellow savory food." (a) Afferent-input granulation of preliminary sensory stimuli. (b) Intermodule messages and efferent constructs symbolizing instinctive reactions, triggered through pathway-2 of the machine-mind, by the *epigenome_header*. Instantiations for: (c) learned and instinctive reactions and (d) deliberative reactions. (c) and (d) Consequences of pathway-1 operations in the machine-mind.

  b. contradicting (*affect*, *valence*) sense-combinations inherently denote alarm: (color$_3$ white (panic low +))$_{13}$ in Fig. 6(a);

  c. *visceral_emotion* of a multisensory feature is reinforced after summation of *strength*s of constituent unimodal *visceral_emotion*s: (food_temperature hot (painful, moderate-to-high, +)) endorsed across vision, touch and taste in Fig. 7(b).

8) Figs. 6(c)-(d) and 7(c) illustrate the system linking sensory data ("white wisps" - in Fig. 6(a)) to their semantics ("smoke" and "hot food").

9) Intuitively, in stage 2 of comprehension, the afferent touch and taste data would attract greater attention, considering their novelty and induced curiosity about the scene [47], than vision and olfaction.

Analysis of properties of the defined structures, in accordance with our studies of perceptions in Section II-B:

1) The structures intend to support representation of any multisensory real-world scenario.

2) While the structures follow a generic schema with respect to the constituent fields, the number of subunits per field at *time* is proportional to the information available at *time*. The structures are thus of variable-lengths. This property encourages flexibility in the number of messages that can be packed into a single structure. Not only does this resolve sparse-representation issues in fixed-format structures [4], [5], it also supports modal-genericity [refer Figs. 2, 6(a), and 7(a)] and representation of novel afferent-stimuli. Ensuing complexities are countered by the pointer-field—bearing pointers to individual subunits—in the schema.

3) Comprehension is an "incremental-developmental [42]" process, involving feedforward and feedback iterations (refer Figs. 5–7); results could be conceptual,

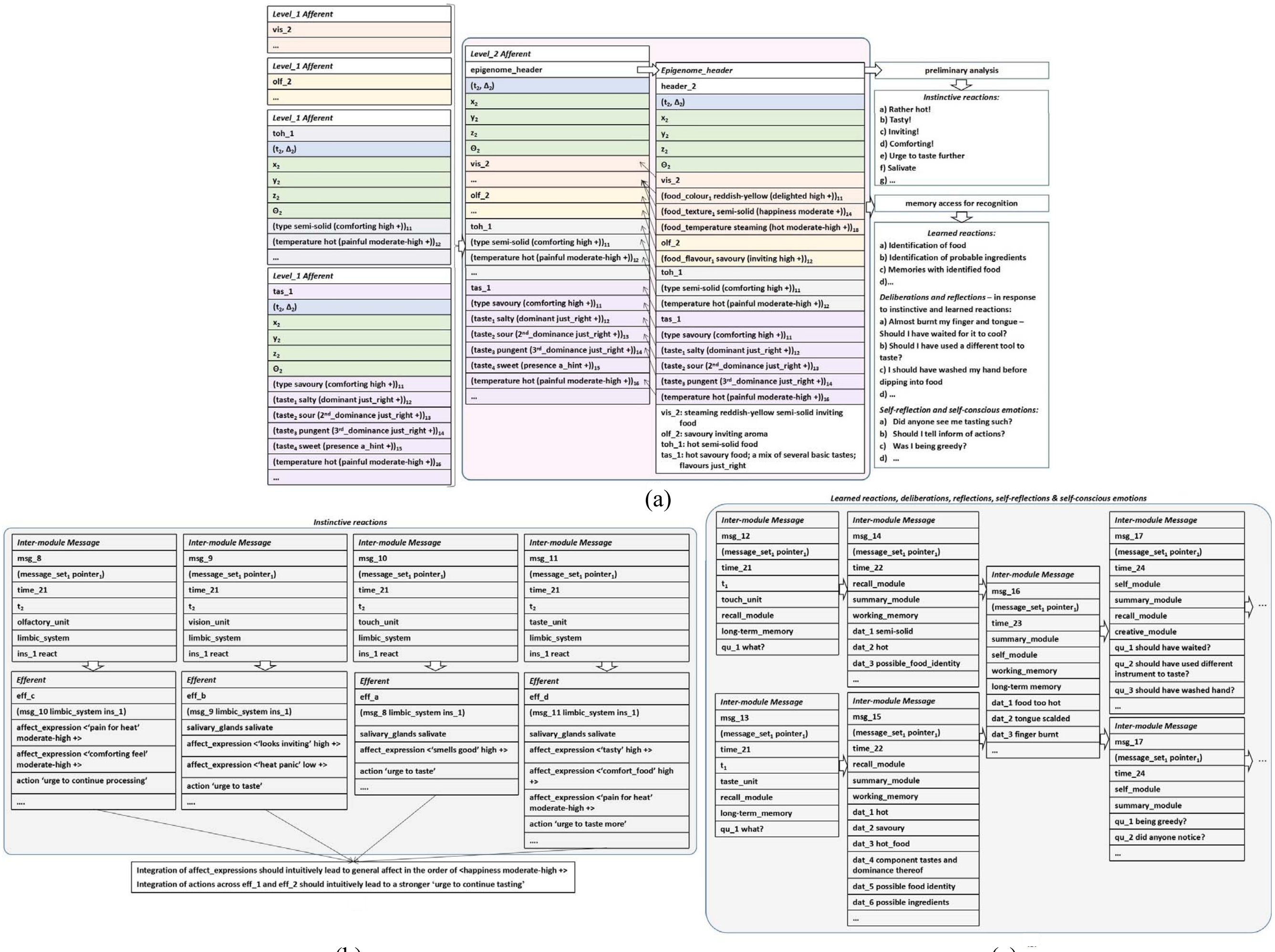


Fig. 7. Part II. Stage 2 inputs and system responses after Fig. (c). (a) Greater attention being delegated to the blue-bowl of steaming food, all the sensory units are now solely directed upon the multisensory source of information, and subsequent afferent-input granulation. Instantiations for: (b) instinctive reactions (pathway-2 machine-mind operations) to the *epigenome_header* and (c) higher realms of contemplation (pathway-1 machine-mind operations).

well-established, speculative, robust, fragile, public, or vestiges of silent introspection.

4) The *epigenome_header* initiates a feedforward strain through the system-mind, actuating a preliminary examination of the captured stimuli, activating hardwired, procedural responses, and/or deliberative thinking (refer Fig. 5).
5) The afferent and the efferent structures emulate A-Brain [1] behavior of the system-mind, while the intermodule messages cater to B-Brain behavior [1] (refer Appendix A in the supplementary material for notes on A and B-Brain concepts).
6) Perception of an object is an assimilation of the objective and subjective properties of its constituent real-time stimuli and active memories. The afferent structures encase details of stimuli exuding from objects that the system pays attention to. Properties of the information captured depend on "quality (active or passive)" and "degree of consciousness" or "intensity of attention" [40] paid. Taking cue from the "cocktail party effect" [48], more pronounced be the novelty, innate emotional impact or objective importance of a stimuli, greater is the attention paid, and consequently more is the information (or finer aspects) encapsulated (refer Appendix C in the supplementary material).
7) Translation of intermodule_message_instruction–to–efferent_structure is analogous to source_code–to–object_code conversions by compiler systems [49]. Just as one high-level source code instruction triggers a series of microcodes, an efferent instruction (e.g., move hand) activates a series of sensorimotor codes (e.g., contract or expand muscles, etc.).
8) In Fig. 5, pathway-2 triggers hardwired somatic markers [3] in response to environmental stimuli, pathway-1 facilitates deliberation between pernicious and useful automated somatic markers. The latter forces the system to exhibit restraint, given a flood of emotions (e.g., panic), that might hinder rational behavior.
9) Data transfer across the schematics cover all the different layers of cognition [1]–[3] (refer Figs. 5–7).
10) Space complexity at *time* for one:
    a. afferent structure—is proportional to the number of stimuli, and features of sensory information per stimulus, captured at *time*;

    b. intermodule message—is proportional to the number and size of constituent atomic messages;
    c. efferent structure—is proportional to the length of the microcode equivalent of a sensorimotor instruction(s).

11) Time complexity at *time* for instantiation of one:
    a. afferent structure—is proportional to the number, complexity, and degree of novelty of stimuli captured at *time* (the novelty of a stimulus, *st*, is a function of the rarity of the system's experience concerning *st* and the endogenous affects aroused);
    b. intermodule message—is proportional to the number and complexity of atomic messages that require being packed into it;
    c. efferent structure—is proportional to the complexity of the sensorimotor instruction(s).
12) The defined structures are by no means exhaustive, but they do encourage thinking on thinking. Some questions that arise in this regard are as follows. How do multisensory neurons function? Is there a universal code that runs through sensory information that facilitates such? Is this code the reason behind arousal of interpretations and emotions for instrumental music? Would engineering multisensory neuromorphic [50] sensors be useful? What leads to the sense of experiencing and consciousness? Would these structures evolve with time (akin the neocortex [33])?
13) Some technical and philosophical issues concerning realization of the conceptualized data-structures and cognitive frameworks, and future work are as follows:
    a) Testing the ability of the definitions to represent different scenarios of various complexities—greater the flexibility and lower the information-loss, stronger is the design. In this respect, synthesis of procedures to translate afferent data into adequate "representations [51]" of real-world patterns is a challenge.
    b) Methodologies to support continued accommodation of novel attributes, emotions and ideas to encapsulate attributes of the ever-changing natural world, without affecting robust performance statistics.
    c) Methods to instantiate spatio-temporal feature integration [39] and subsequent instantiation of the *epigenome_header* and resultant hardwired responses in real-time (approximately in the order of speeds of average humans).
    d) Mechanisms to map sensory data and visceral emotions to their linguistic equivalents and qualia (e.g., linking the sensation of "being happy" to the meaning of the word "happiness").
    e) Writers, artists, scientists, and inventors learn several representations of words as well as invent new ones through refinement of basic word-senses and cultural-exposure. Techniques to emulate the same in artificial cognitive systems could lead to artificial creativity.
    f) Semaphore-inspired procedures [52] to arbitrate conflicting instructions to the same module at the same time, such that system-stability is maintained.
    g) Intuitively, a non-von Neumann architecture—e.g., a hybrid dataflow [53] and in-memory system [54] could support construction the system.
    h) Perceptions derive from agglomeration of granules [55] of comprehension across different modalities, contexts, and layers of thoughts. These granules being intuitively imprecise conglomerations of facts and affects in complex hybrid formats—as is evident in the formulated structures, computational processes (an algebra or logic) to handle such data is crucial.

V. RELEVANCE TO OTHER COGNITIVE FRAMEWORKS

The machine-mind framework [31], [40] being conceptually equipped with modules to handle real-world objective and subjective data (refer Sections II-B and II-C), the purpose of the structures has been investigated through its projected use therein. However, these have been envisaged to be sufficiently generic—in terms of hybrid data-type support, platform-independence, variable number of fields and attribute-placeholders depending on information available at the time of instantiation (external attention [56]) and internal attention modulation [56], and enclosure of objective and subjective data.

Broadly, given any framework (for a review see [56]), the afferent constructs bind real-time multisensory environment-descriptors (inclusive of attributes of inanimate and animate constituents, background features, ambient temperature, etc.) and consequent system-visceral arousals, the *epigenome_header* triggers instinctive behavior and sets the stage for contextual inferences, the intermodule messages support exchanges (unicast, multicast, and broadcast) of data, instructions and queries across modules, memory-units and agents, and the efferent structures transport "low-level" [57] sensorimotor-response codes to the respective system-units. The definitions, theoretically, provide structural support for emulation of self-conscious thinking, reflection, deliberation, instinctive behavior, scope for "lifelong" [58] inclusion of novel attributes, and qualia. In this section, we analyze the relevance of the defined constructs in some contemporary cognitive frameworks.

In Em-One [59] and the accountable layered system in [60], the afferent structures would help capture bespoke multisensory details of the region of interest, the intermodule messages would support inteactions between the "critics", and the efferent messages could be used to transport action-commands to the sensorimotor units.

ATLANTIS [52], could utilize our afferent structures to capture multisensory information for its "controller unit," whereupon the "controller," "deliberator," and "sequencer" would interact via intermodule messages and the sequencer would encode system-instructions into efferent response-sequences.

ADAPT [61] models perception as an "active," context-sensitive process. Its "sensory schema" corresponds to our afferent structures and the "motor schema" to the efferent constructs, the *epigenome_header* would activate "procedural schemas," and the intermodule messages would facilitate intercomponent exchanges leading to activations of "deliberative schemas."

In ROBOEARTH [62], the afferent structures would support representation of the "kinematic" and "semantic" models, and these in turn would activate the "software components" for object recognition. The intermodule message format would be part of the "communication module," assisting establishment of inter- and intra-robot module interactions. The "action recipes" would be encoded in the efferent constructs.

In the autobiographical memory architecture [58], the afferent structures (through instantiations and details thereof, depending on grades of attention and consciousness) would support real-time encasing of multimodal data across different levels of abstraction. Neither the *epigenome_header* nor the efferent structure is of much use, since this is not an action-based framework. The intermodule messages support intra- and inter-robot connectivity.

In the DAC architecture [63], [64], the afferent structures would support emulation of the "world and self" representations of an agent, the *epigenome_header* relates to the "reactive layer," the efferent structures the "action or behavior codes" and the intermodule messages for exchanges between functional components and databases. Other architectures like those in [65] and [66] can be envisioned to make use of the constructs in identical ways.

VI. CONCLUSION

This paper catalogues our efforts toward the formalization of data-structures to support bespoke multisensory real-world representation, comprehension and symbiotic interaction between embodied systems. Shifting away from fixed-format [4], [5] and solely semantic [4], [6] representations, the constructs build on the system's sense of self and its derivatives, theory of mind, commonsense, and domain, procedural and declarative knowledge. The afferent structures transport multisensory inputs to core system-modules, intermodule messages facilitate exchange of data, queries, and instructions across systems components, and the efferent structures translate process-results into sensorimotor codes. Drawing from the human brain, these variable-length schematics enable platform and data-modality independence, flexibility in the number of intermodular messages that can be compacted into one unit, and representation of novel sensory data-types. An epigenome-styled [8], [9] header-node in the afferent structures regulates activation of system-intuitive and reflective behavior. The average space and time complexity of the structures, at any instant, are functions of the amount and type of information available. The structures have been theoretically tested for appropriateness in an embodied machine-mind framework [31], [32] during comprehension of a real-world scenario, and application in some existing cognitive frameworks. Out of many others, some critical prerequisites that the realization of these constructs calls for are understanding the composition of modal signals and generation of consequent visceral responses and qualia in the natural world, and processes to emulate, map, and operate on affects and multimodal data—through space and time. These constructs are envisioned to be applicable to generally intelligent embodied cognitive architectures and motivate reflections on "thinking."

REFERENCES

[1] M. L. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, NY, USA: Simon and Schuster, 2006.

[2] M. Minsky, *The Society of Mind*. New York, NY, USA: Simon & Schuster, 1986.

[3] A. Damasio, *Self Comes to Mind: Constructing the Conscious Brain*. London, U.K.: Vintage, 2012.

[4] F. E. D. S. Webber, "Semantic folding theory and its application in semantic fingerprinting," Vienna, Austria, Cortical.io, White Paper, Nov. 2015.

[5] C. Clifton, H. Garcia-Molina, and D. Bloom, "HyperFile: A data and query model for documents," *VLDB J.*, vol. 4, no. 1, pp. 45–86, 1995.

[6] J. F. Sowa, "Conceptual Graphs," in *Handbook of Knowledge Representation*, F. van Harmelen, V. Lifschitz, and B. Porter, Eds. London, U.K.: Elsevier, 2008, pp. 213–237.

[7] K. Trentelman, "Survey of knowledge representation and reasoning systems," Defence Sci. Technol. Org., Edinburgh, SA, Australia, Rep. DSTO-TR-2324, 2009.

[8] B. Weinhold, "Epigenetics: The science of change," *Environ. Health Perspectives*, vol. 114, no. 3, pp. A160–A167, 2006.

[9] S. Ebrahim, "Epigenetics: The next big thing," *Int. J. Epidemiol.*, vol. 41, no. 1, pp. 1–3, 2012.

[10] K. E. Cullen, "Sensory signals during active versus passive movement," *Current Opin. Neurobiol.*, vol. 14, no. 6, pp. 698–706, 2004.

[11] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY, USA: Avon, 1995.

[12] L. W. Barsalou, "Perceptual symbol systems," *Behav. Brain Sci.*, vol. 22, no. 4, pp. 577–660, 1999.

[13] T. J. Sejnowski, H. Poizner, G. Lynch, S. Gepshtein, and R. J. Greenspan, "Prospective optimization," *Proc. IEEE*, vol. 102, no. 5, pp. 799–811, May 2014.

[14] V. S. Ramachandran, *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes us Human*. New York, NY, USA: W. W. Norton, 2010.

[15] D. Balduzzi and G. Tononi, "Qualia: The geometry of integrated information," *PLoS Comput. Biol.*, vol. 5, no. 8, 2009, Art. no. e1000462.

[16] B. de Gelder, J. Vroomen, and G. Pourtois, "Multisensory perception of emotion, its time course and its neural basis," in *The Handbook of Multisensory Processes*, B. E. Stein, C. Spence, and G. Calvert, Eds. Cambridge, MA, USA: MIT Press, 2004, pp. 581–596.

[17] N. P. Holmes and C. Spence, "Multisensory integration: Space, time and superadditivity," *Current Biol.*, vol. 15, no. 18, pp. R762–R764, 2005.

[18] C. O. Callaghan, "Grades of multisensory awareness," *Mind Lang.*, vol. 32, no. 2, pp. 155–181, 2017.

[19] A. Chatterjee, "Disembodying cognition," *Lang. Cogn.*, vol. 2, no. 1, pp. 79–116, 2010.

[20] D. M. Eagleman *et al.*, "Time and the brain: How subjective time relates to neural time," *J. Neurosci.*, vol. 25, no. 45, pp. 10369–10371, 2005.

[21] J. F. Danker and J. R. Anderson, "The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding," *Psychol. Bull.*, vol. 136, no. 1, pp. 87–102, 2010.

[22] V. Gallese and S. Ebisch, "Embodied simulation and touch: The sense of touch in social cognition," *Phenomenol. Mind*, vol. 4, pp. 269–291, 2013.

[23] T. W. Buchanan, "Retrieval of emotional memories," *Psychol. Bull.*, vol. 133, no. 5, pp. 761–779, 2007.

[24] J. Schulkin, *Bodily Sensibility: Intelligent Action*. New York, NY, USA: Oxford Univ. Press, 2004.

[25] P. J. Marshall, "Embodiment and human development," *Child Develop. Perspectives*, vol. 10, no. 4, pp. 245–250, 2016.

[26] M. L. Seghier, "The angular gyrus: Multiple functions and multiple subdivisions," *Neuroscientist*, vol. 19, no. 1, pp. 43–61, 2013.

[27] R. T. Hurlburt, C. L. Heavey, and J. M. Kelsey, "Toward a phenomenology of inner speaking," *Consciousness Cogn.*, vol. 22, no. 4, pp. 1477–1494, 2013.

[28] H. Imamizu *et al.*, "Human cerebellar activity reflecting an acquired internal model of a new tool," *Nature*, vol. 403, no. 6766, pp. 192–195, 2000.

[29] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn. Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.

[30] R. M. Shiffrin and W. Schneider, "Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory," *Psychol. Rev.*, vol. 84, no. 2, pp. 127–190, 1977.

[31] R. Banerjee and S. K. Pal, "Text comprehension and the computational mind-agencies," *Nat. Comput.*, vol. 14, no. 4, pp. 603–635, 2015.

[32] R. Banerjee and S. K. Pal, “A machine-mind architecture and Z*-numbers for real-world comprehension,” in *Pattern Recognition and Big Data*, A. Pal and S. K. Pal, Eds. Singapore: World Sci., 2017, pp. 807–844.

[33] S. L. Pallas and Y. Mao, “The evolution of multisensory neocortex,” in *New Handbook of Multisensory Processes*, B. E. Stein, Ed. Cambridge, MA, USA: MIT Press, 2012, pp. 627–642.

[34] J. J. Foxe and C. E. Schroeder, “The case for feedforward multisensory convergence during early cortical processing,” *Neuroreport*, vol. 16, no. 5, pp. 419–423, 2005.

[35] E. Macaluso and J. Driver, “Multisensory spatial interactions: A window onto functional integration in the human brain,” *Trends Neurosci.*, vol. 28, no. 5, pp. 264–271, 2005.

[36] A. S. Nandy, J. F. Mitchell, M. P. Jadi, and J. H. Reynolds, “Neurons in macaque area V4 are tuned for complex spatio-temporal patterns,” *Neuron*, vol. 91, no. 4, pp. 920–930, 2016.

[37] S. Soto-Faraco and A. Kingstone, “Multisensory integration of dynamic information,” in *The Handbook of Multisensory Processes*, B. E. Stein, C. Spence, and G. Calvert, Eds. Cambridge, MA, USA: MIT Press, 2004, pp. 49–68.

[38] R. L. Klatzky, “Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections,” in *Spatial Cognition—An Interdisciplinary Approach to Representation and Processing of Spatial Knowledge*, C. Freksa, C. Habel, and K. F. Wender, Eds. Berlin, Germany: Springer-Verlag, 1998, pp. 1–17.

[39] C. O. Callaghan, “Perception and multimodality,” in *Oxford Handbook of Philosophy and Cognitive Science*, E. Margolis, R. Samuels, and S. P. Stich, Eds. New York, NY, USA: Oxford Univ. Press, 2012.

[40] R. Banerjee and S. K. Pal, “A computational model for the endogenous arousal of thoughts through Z*-numbers,” *Inf. Sci.*, vol. 405, pp. 227–258, Sep. 2017.

[41] V. Arstila and D. Lloyd, Eds., *Subjective Time: The Philosophy, Psychology and Neuroscience of Temporality*. Cambridge, MA, USA: MIT Press, 2014.

[42] T. A. Harley, *The Psychology of Language: From Data to Theory*, 3rd ed. New York, NY, USA: Psychology Press, 2008.

[43] R. A. Peters, K. A. Hambuchen, and R. E. Bodenheimer, “The sensory ego-sphere: A mediating interface between sensors and cognition,” *Auton. Robots*, vol. 26, no. 1, pp. 1–19, 2009.

[44] S. Lallee *et al.*, “Towards a platform-independent cooperative human robot interaction system: III an architecture for learning and executing actions and shared plans,” *IEEE Trans. Auton. Mental Develop.*, vol. 4, no. 3, pp. 239–253, Sep. 2012.

[45] M. I. Coco, F. Keller, and G. L. Malcolm, “Anticipation in real-world scenes: The role of visual context and visual memory,” *Cogn. Sci.*, vol. 40, no. 8, pp. 1995–2024, 2016.

[46] P. Doolittle, *How Your ‘Working Memory’ Makes Sense of the World*, Ted Talks, New York, NY, USA, 2013, accessed: Mar. 14, 2018. [Online]. Available: https://www.ted.com/talks/peter_doolittle_how_your_working_memory_makes_sense_of_the_world

[47] J. Gottlieb, P.-Y. Oudeyer, M. Lopes, and A. Baranes, “Information-seeking, curiosity, and attention: Computational and neural mechanisms,” *Trends Cogn. Sci.*, vol. 17, no. 11, pp. 585–593, 2013.

[48] B. G. Shinn-Cunningham, “Object-based auditory and visual attention,” *Trends Cogn. Sci.*, vol. 12, no. 5, pp. 182–186, 2008.

[49] A. V. Aho, R. Sethi, and J. D. Ullman, *Compilers: Principles, Techniques, and Tools*. Reading, MA, USA: Addison-Wesley, 1986.

[50] C. Mead, “Neuromorphic electronic systems,” *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

[51] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[52] E. Gat, “Integrating planning and reacting in a heterogeneous asynchronous architecture for controlling real-world mobile robots,” in *Proc. Conf. Artif. Intell.*, San Jose, CA, USA, 1992, pp. 809–815.

[53] Arvind and D. E. Culler, “Dataflow architectures,” *Annu. Rev. Comput. Sci.*, vol. 1, pp. 225–253, Jun. 1986.

[54] A. Sebastian *et al.*, “Temporal correlation detection using computational phase-change memory,” *Nat. Commun.*, vol. 8, no. 1115, pp. 1–10, 2017.

[55] W. Pedrycz, “Granular computing: An introduction,” in *Future Directions for Intelligent Systems and Information Sciences*, N. Kasabov, Ed. Heidelberg, Germany: Physica, 2000.

[56] I. Kotseruba, O. J. A. Gonzalez, and J. K. Tsotsos, “A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications,” *arXiv:1610.08602*, 2016.

[57] S. Zhang, M. Sridharan, M. Gelfond, and J. Wyatt, “Towards an architecture for knowledge representation and reasoning in robotics,” in *Proc. 6th Int. Conf. Soc. Robot.*, vol. 8755. Sydney, NSW, Australia, 2014, pp. 400–410.

[58] M. Petit, T. Fischer, and Y. Demiris, “Lifelong augmentation of multimodal streaming autobiographical memories,” *IEEE Trans. Cogn. Develop. Syst.*, vol. 8, no. 3, pp. 201–213, Sep. 2016.

[59] P. Singh, “EM-ONE: An architecture for reflective commonsense thinking,” Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2005.

[60] B. Morgan, “A substrate for accountable layered systems,” Ph.D. dissertation, Program Media Arts Sci., School Archit. Plan., Massachusetts Inst. Technol., Cambridge, MA, USA, 2013.

[61] D. P. Benjamin, D. Lyons, and D. Lonsdale, “ADAPT: A cognitive architecture for robotics,” in *Proc. 6th Int. Conf. Cogn. Model.*, 2004, pp. 337–338.

[62] M. Tenorth, A. C. Perzylo, R. Lafrenz, and M. Beetz, “Representation and exchange of knowledge about actions, objects, and environments in the RoboEarth framework,” *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 643–651, Jul. 2013.

[63] P. F. M. J. Verschure, “Distributed adaptive control: A theory of the mind, brain, body nexus,” *Biologically Inspired Cogn. Archit.*, vol. 1, pp. 55–72, Jul. 2012.

[64] S. Lallee and P. F. M. J. Verschure, “How? Why? What? Where? When? Who? Grounding ontology in the actions of a situated social agent,” *Robotics*, vol. 4, no. 2, pp. 169–193, 2015.

[65] A. Baranès and P. Y. Oudeyer, “R-IAC: Robust intrinsically motivated exploration and active learning,” *IEEE Trans. Auton. Mental Develop.*, vol. 1, no. 3, pp. 155–169, Oct. 2009.

[66] P.-Y. Oudeyer, “What do we learn about development from baby robots?” *WIREs Cogn. Sci.*, vol. 8, nos. 1–2, 2017, Art. no. e1395.

Romi Banerjee received the B.Sc., B.Tech., and M.Tech. degrees in computer science and engineering from the University of Calcutta, Kolkata, India, in 2007, 2010, and 2012, respectively, and where she is currently registered for the Ph.D. degree.

She is a Senior Research Fellow with the Center for Soft Computing Research—a National Facility, Indian Statistical Institute, Kolkata. Her current research interests include cognitive machines, artificial general intelligence, developmental cognitive neuroscience, and the philosophy of the mind.

Ms. Banerjee is a Student Member of the Association for the Advancement of Artificial Intelligence, and a member of the Women in Cognitive Science and the Cognitive Science Society.

Sankar K. Pal (M’80–SM’84–F’93–LF’15) received the Ph.D. degrees from the University of Calcutta, Kolkata, India, and Imperial College London, London, U.K.

He joined the Indian Statistical Institute, Kolkata, in 1975, as a CSIR-SRF, where he became a Full Professor in 1987, a Distinguished Scientist in 1998, the Director in 2005, and founded the Machine Intelligence Unit and the Center for Soft Computing Research. He was with the University of California at Berkeley, Berkeley, CA, USA, the University of Maryland at College Park, College Park, MD, USA, NASA JSC, Houston, TX, USA, and U.S. Naval Research Laboratory, Washington, DC, USA. He held several visiting positions in Italy, Poland, Hong Kong, and Australia. He has co-authored 20 books and over 400 research publications in the areas of pattern recognition, machine learning, image/video processing, data mining, Web intelligence, soft computing, bioinformatics, and cognitive machines.

Prof. Pal was a recipient of national/international awards, including the S. S. Bhatnagar Prize in India, the Padma Shri Award in India, and the Khwarizmi International Award in Iran. He has been serving as an IEEE CS Distinguished Visitor since 1987. He is/was on the editorial boards of 20 journals including some IEEE TRANSACTIONS. He has visited 45 countries as a keynote/invited speaker. He is a fellow of TWAS, IAPR, IFSA, and all four National Academies for Science/Engineering in India. He is currently a J. C. Bose National Fellow and a DAE Raja Ramanna Fellow.