Modeling medical diagnosis using a comprehensive cognitive architecture

Stephen Strain, MD, MS[1] and Stan Franklin, PhD[2]

[1]Department of Biomedical Engineering, University of Memphis, Memphis, TN, USA
[2]Department of Computer Science, University of Memphis, Memphis, TN, USA

ABSTRACT

Medical diagnosis is accomplished by a set of complex cognitive processes requiring the iterative application of Charles Sanders Peirce's three modes of inference: abduction, deduction, and induction. Previous research in computational modeling of medical diagnosis has had only limited success by defining sub-domains which offer a computationally tractable problem. However, the aspect of diagnostic reasoning requiring intelligence lies in the extraction of a well-structured problem from an ill-structured one—the very difficulty obviated by expert systems which begin with a well-structured problem and lack the robustness of the human clinician. We propose a design for an agent based on the LIDA architecture, a model of cognition designed for implementations requiring deliberation and learning, which utilizes a psychologically and neurologically inspired cognitive cycle. The proposed agent, MAX (for "Medical Agent X") will be equipped to comprehend clinical data in the context of its perceptual ontology and learned associations, and to construct, evaluate, and refine by investigation a differential diagnosis from that data which progressively reduces the dimensionality of its search space with each iteration. Furthermore, the agent will appropriately modify its own ontology with experience and supervised instruction, in a manner inspired by traditional medical education.

Keywords: Medical Diagnosis; Artificial Intelligence in Medicine; Cognitive Architecture; Clinical Decision Support System;

1. INTRODUCTION

Medical diagnosis is a complex cognitive process requiring specialized incarnations of the cognitive faculties of perception, attention, knowledge, understanding, deliberation, and decision. The elements of diagnostic thinking will be briefly described, with especial attention to the differential diagnosis (DDx), a list of hypotheses—sorted by etiology, likelihood, and/or urgency—which is used to organize diagnostic deliberation and to strategize lines of investigation for confirming or ruling out diagnoses of clinical interest. Patient response to empiric medical therapy plays a crucial role in diagnosis; for the present purpose, therapeutic interventions will be considered solely with regard to their role in establishing a diagnosis. Previous work in the computational modeling of medical diagnosis is reviewed, and the formidable challenges inherent in this endeavor are enumerated.

It will be argued that the LIDA model of cognition [1] is an appropriate architecture for modeling medical diagnosis, in that it lends itself well to single-cycle action selection and multicyclic reasoning in the setting of previously existing and/or learned knowledge, and utilizes a neuro-plausible mechanism for attention and consciousness. Furthermore, a design is proposed for a medical diagnosis agent--hereafter referred to as MAX, for 'Medical Agent X'. MAX employs cognitive structures based on the event models of Zacks and colleagues [2] to represent the entities and elements of diagnostic reasoning MAX also bases the processes of diagnostic investigation and decision making on behavior streams in a manner inspired by Drescher [3] and Maes [4] in an agent embodying the LIDA architecture.

The proposed MAX will be a software agent that interacts with an electronic medical record (EMR) database serving as the documentation system of a health-care center which might include outpatient, inpatient, rehabilitative, assisted-living, surgical and/or special care facilities. MAX will have access to all clinical data in the EMR system, including but not limited to admission histories; physician and nursing progress notes; physician orders; and pharmacy, radiology and laboratory records.

MAX will engage one patient record at a time in order to observe and learn, deliberate diagnostically, and

participate in decision-making when sufficiently trained. During such an engagement, the contents of the EMR may change, as health-care providers enter clinical findings, diagnostic conclusions, and treatment decisions into the documentation system. MAX's proposed course of training parallels that of traditional general medical training in North American medical schools and residencies. It would begin as an observer with a limited fund of knowledge; with time interacting more and more with human healthcare providers, with the aim of eventually participating in diagnostic decision-making and decision review. Thus its abilities would sequentially resemble those of the medical student, intern, senior resident, and attending physician. It will possess the ability to perceive and understand clinical data, to generate hypotheses regarding the nature of an illness, and to propose lines of inquiry and/or therapeutic interventions. Ultimately, it should possess sufficient natural language capability to enable it to respond meaningfully to questions of the sort posed by a staff physician or medical students and residents, and to understand and learn from instruction by human health care experts.

MAX will interact with other health-care workers by means of a graphical user interface (GUI) displaying the current focus of its attention, the contents of its deliberations, and any relevant queries, suggestions, or alerts. With sufficient training, the agent would possess the ability to contact human health-care providers appropriately in a time critical manner.

A comprehensive computational specification of this general design scheme for MAX is beyond the scope of the present paper, and will provide ample material for additional research. For the present, the discussion is confined to illuminating the way in which MAX will transform clinical data into diagnostic assertions.

2. MEDICAL DIAGNOSIS

Diagnosis is partially a variety of the cognitive faculty of recognition: perceiving learned patterns in novel clinical data and categorizing the patterns appropriately. It includes the process of clinical data acquisition, and thus invokes the additional question of how much data, and which data in particular, to acquire. Some data present immediately—a patient is unconscious, or breathes with great difficulty. Other facts must be acquired by deliberate investigation, ranging from a simple question—for instance, "Is there a family history of colon cancer?"—to an expensive and potentially dangerous test, such as the appearance of the brain on a computerized tomography (CT) scan. A wide array of judgments, some quite subtle, need be made to avoid data collection entailing excessive cost in multiple domains simultaneously, including the interpersonal, the temporal, the physiologic, the financial, and the computational.

Since the problem posed by diagnosis is such an open one, the ideal diagnostic approach successively transforms the problem space into a more and more bounded one during acquisition of the clinical data. Although medicine is often accused of being an art rather than a science, diagnosis in fact parallels the scientific process in its application of Charles Sanders Peirce's three modes of inference [5]: Abduction is employed to formulate diagnostic hypotheses; deduction confirms or disproves them; and induction results in the fund of expert knowledge known as "clinical intuition."

Clinical data are typically acquired in a stereotypic pattern, which has evolved over the history of modern medicine, frequently referred to as the "History and Physical," or "H&P." The least expensive data--the historical data obtained by questioning the patient, his or her family, other healthcare providers present, and the available medical record--are acquired first, and, after being organized into stereotypic elements (see Table 1), are collectively known as the medical history or *anamnesis*. After, or perhaps during, the taking of the medical history, the physical exam is conducted, which essentially consists of a gross examination of the patient's body, aided by stethoscope, otoscope, ophthalmoscope and so on. By the time more expensive tests are ordered, the physician has already begun to formulate diagnostic categories in which to confine further search for the diagnosis, and these categories form the basis of the investigation strategy which shapes future choices about which particular tests to order, when to conclude the data acquisition, and so on. The reduced search space that comprises the chosen categories is often referred to as the differential diagnosis.

2.1. Differential Diagnosis

Dorland's Medical Dictionary defines differential diagnosis as "the determination of which one of two or more diseases or conditions a patient is suffering from, by systematically comparing and contrasting their clinical findings."[6] However, in clinical practice, the expression "the differential diagnosis"—or more simply "the differential"—can refer to the list of diagnoses that the diagnostician considers as possibly consistent with the currently known clinical data. When an attending asks a resident, "What is your differential?" he means to ask, "Given what you know about the patient so far, what diseases do you consider likely to be responsible for what is happening?" When used in this sense, as a list of possibilities, the term is preceded by the definite or indefinite article (or another specifier), and is often abbreviated "DDx." Thus, for the sake of disambiguation, the authors shall adopt the convention of

Table 1: Typical divisions and categories of clinical findings included in the History and Physical (H&P).	
Medical History	Chief Complaint. History of Present Illness. Past Medical History. Past Surgical History. Social History. Medications & Allergies. Review of Systems.
Physical Exam Findings	Vital Signs. General appearance. Head, Eyes, Ears, Nose & Throat (HEENT). Heart, Lungs & Thorax (Chest). Abdomen. Genitourinary (GU). Neurological System (Neuro). Extremities. Skin.
Test Results	EKG. Oxygen Saturation. Blood and Urine Lab Studies. Radiologic Studies.

using "the differential diagnosis," "the differential," or "the DDx" to refer to the specific list of diagnostic possibilities under consideration, while "differential diagnosis" or "the process of differential diagnosis" will signify the diagnostic thought process in general.

Generally speaking, the differential is ordered by likelihood and/or clinical priority, and it can list very specific conditions such as "pulmonary embolism secondary to deep venous thrombosis of the left lower extremity," or broad causal categories known as etiologies.

A brief example will serve to clarify the latter case. A diagnosis of shock is made, or at least considered, whenever there is a sufficiently low blood pressure, or a marginally low blood pressure accompanied by signs of physiologic distress. Proper treatment of shock requires ascertainment of the etiology of the shock, and this determination often proceeds from an initial differential diagnosis of hypovolemic, cardiogenic, neurogenic, or infectious processes as possible etiologies. By comparing clinical findings to learned typical patterns, and observing patterns in the disease's progression and response to therapy, the relative likelihood of each hypothesis is assessed in an intuitive fashion.

A more extended example will illustrate the complexities of the process of differential diagnosis. When a patient presents to a physician with a chief complaint of chest pain, several previously learned lines of diagnostic inquiry instantiate seemingly in parallel, most often with some degree of diagnostic focus on ruling out a cardiac event. With regard to any complaint of pain, specific information about its site, radiation, character, severity, provoking or relieving factors, onset, duration, progression, and so forth is sought. Focused history is taken to ascertain specific facts which are known to impact most significantly the patient's risk for cardiac disease--age, gender, family history of cardiac disease, smoking history, hyperlipidemia, diabetes, hypertension, previous history of heart attack, and others. The assessment of urgency must begin simultaneously—is the patient currently experiencing chest pain or shortness of breath? Are the vital signs within normal limits? If the patient's risk factors suggest the possibility of coronary artery disease, or the vital signs are unstable, an EKG, chest x-ray and cardiac enzyme studies may be ordered during the early moments of the patient's presentation. Meanwhile, especially in the emergency room where the urgency of the presentation is more likely to be elevated, the physical exam may well begin concurrently with the above diagnostic investigations.

Differential diagnosis delineates a framework for this clinical problem. In demographic subsets where it is a significant threat, the possibility of a cardiac etiology in a chest pain presentation must be addressed not only by data that confirm or disprove the cardiac hypothesis, but by investigation which examines other potential etiologies for chest pain. Might the pain be related to chest wall trauma, esophageal reflux, or pulmonary or pleural inflammation? Might it be referred pain from an intraabdominal pathology? Might a mass effect from a malignant tumor be involved, or a pulmonary embolism? Might it be psychosomatic or factitious?

While ruling out myocardial ischemia secondary to coronary arteriosclerosis often dominates the diagnostic approach to chest pain due to the high prevalence of this disease in identified populations, other diseases of cardiac etiology, such as pericarditis, dissecting aortic aneurysm, or myocardial trauma, may need to be considered[1]. Moreover, each alternative etiology in the differential diagnosis may require similar specification, to an extent dictated by the clinical endpoint desired in each particular case. In a general sense, this endpoint is ill defined, and might be thought of as the point at which a treatment plan adequate to the clinical needs of the moment is indicated. The meaning of the term "adequate" is determined by the judgment of the chief clinician, and can depend upon the stability of the patient, the scope of expertise of the particular clinician, and the

[1] Pericarditis is an infection of the pericardium, the fibrous sac encasing the heart; dissecting aortic aneurysm is a grave condition that occurs when the endothelial lining of the arterial trunk begins to tear away from the underlying wall; myocardial trauma refers to mechanical injury to the heart wall.

particular health care setting. For instance, after a cardiovascular event is satisfactorily ruled out, a complete diagnostic work-up to determine the cause of chest pain in a stable patient would generally not be appropriate in an emergency department, and when indicated, such a workup would be conducted by an internist or family practitioner, but not a cardiologist.

2.2. Computational Modeling of Diagnosis

There has been considerable on-going debate about the psychology of diagnostic reasoning, and the intersection of this debate with modeling efforts has resulted in a proliferation of computational approaches. Early psychological models postulated that diagnosis proceeds by hypothetico-deductive reasoning—that is, generation and testing of hypotheses in a goal-oriented, backward-chaining fashion—but later research has established that as domain knowledge and expertise are acquired, a data-driven pattern recognition strategy becomes dominant except in situations of unusual complexity or uncertainty [5,7,8]. It is likely that a significant number of real-world diagnostic problems have complexity sufficient to prevent them from falling neatly into one approach or the other—both are used in parallel in a fashion which somewhat resembles a bidirectional search algorithm. [8]

Early work in diagnostic modeling has explored the use of propositional logic and probability theory to develop formal systems for expressing the relationships between diseases and symptoms that comprise medical knowledge. Logically, medical knowledge can be seen as an entailment, which infers the possibility of disease states from symptoms—"The effect of medical knowledge is to eliminate from consideration disease complexes that are not related to the symptom complex presented" [9]. Probabilistically, medical knowledge can be viewed as the conditional probabilities of disease states given symptoms, and the diagnostic problem becomes a Bayesian calculation of the most likely diagnosis accounting for the given symptoms [9]. Parmigiani states, "One approach . . . is to develop a statistical prediction model $p(y|x,\theta)$ for the true disease status y given patient characteristics x and parameters θ" [10].

While this approach has a certain intellectual appeal, from a practical perspective it can result in a combinatorial explosion of even simple problems. Moreover, basic assumptions of the mathematical model, such as independence of conditional probabilities, and exclusivity and exhaustiveness of the diagnostic categories, are not the case in general [11]. Certainly mathematical modeling obviates the problem of representing a causal relationship between disease and symptom—this is not necessarily an advantage from a computational or a clinical point of view. Medical tradition suggests that knowledge of the pathophysiology of disease provides a useful heuristic in the solution of the diagnostic problem, and "think aloud" research studying the psychology of medical cognition reveals this heuristic to be one widely used among human practitioners [8].

Debate also continues regarding the structure of long-term memories of diagnostic knowledge and experience. Researchers at Stanford University implemented knowledge-based reasoning as a system of production rules in MYCIN, an expert system application that assigns clinical data from infectious disease cases to diagnoses with an accuracy comparable to that of human experts [12]. A case-based reasoning approach has been proposed, in which interpretation of new data is linked to specific past experiences, which serve as exemplars to provide scripts, contexts, and expectations. [7,8] Frames, which index data and processes relevant to a particular diagnostic condition, form the backbone of the Present Illness Program (PIP), developed collaboratively at MIT and Tufts University. [13]

The purpose of PIP[2] was to elicit a history and determine a diagnosis in a patient with edema. Within this very narrowly focused sub-region of internal medicine, PIP performed with fair accuracy compared to human physicians. In order to lay groundwork for the elaboration of some of MAX's design specifications for implementation in LIDA, it will be instructive to examine PIP's design in greater detail.

Frames stored in long-term memory compose PIP's knowledge base, containing typical findings, numerical scores to compute likelihoods from findings, causal links to and from other frames, and rules for excluding or confirming the diagnosis represented by the frame.[3] PIP's long-term memory module resides alongside a short-term memory that in some ways resembles LIDA's Workspace (see sections 3.1-3.2 below). Frames have the capability to scan short-term memory for specific findings which "trigger" the particular frame. If a frame is triggered by a finding, it is brought into short-term memory, and rendered "active." Per Pauker, "This process is synonymous with forming an hypothesis." [13]

The frames are linked in a network intended to represent associations and causal connections between diseases. Frames that are a single link away from an active frame become "semi-active"—they are brought "near" the short-term memory, and can now be activated

[2] The Present Illness Program should not be confused with the Probabilistic Information Processing System, also referred to by the acronym PIP, and used in early studies of Bayesian modeling of medical diagnosis. See Lusted (1968) [14], pp. 163-64.

[3] In Pauker et al. (1976) [13], the phrases "long-term memory" and "short-term memory" are technical terms referring to modules in the PIP program; the reader should not infer a connection with concepts referred to by the same name in the cognitive science literature.

by additional triggers. In effect, certain findings, when present in the patient's history, will always trigger the frame into long-term memory; certain other findings will only act as triggers if the frame is in the semi-active state. One side effect of this mechanism is to limit the number of diagnostic possibilities under consideration at any given time.

Each hypothesis in short-term memory is tested by calculating its "goodness of fit" with the known findings. If sufficiency criteria are met, the diagnosis becomes elevated to the level of a fact, unless subsequent findings contradict it, in which case a conflict resolution process ensues. The numerical scores from the frame are used to calculate scores representing the extent to which the hypothesis explains the case at hand, and the proportion of the hypothesized condition's prototypical findings that are present in the current findings. At the end of scoring, the active hypotheses are ranked.

Starting with the highest-ranked hypothesis, PIP selects questions to ask the user, in order to acquire additional information to "improve its understanding of the clinical problem." It begins its inquiry with questions about the classic findings of the disorder as specified in the frame. Each positive finding is characterized according to a set pattern of attributes such as location, severity, duration, etc. The validity of the findings is checked, and then the hypotheses are re-evaluated in light of the newly acquired findings. The above process then repeats in a cyclic manner.

Within the narrow domain of a presentation of severe edema in an adult patient, PIP performs assessments that correspond fairly well with those of human practitioners. However, it lacks a means of determining an algorithmic endpoint—it has no way of representing degree of diagnostic refinement, nor does it possess any mechanism to match diagnostic goals to clinical demands—"the diagnosis need to be only as precise as is required by the next decision to be taken by the doctor." [11] All its clinical knowledge is hardwired, and no learning is possible [13], so clinical experience is unable to improve performance. It is unable to structure its information acquisition and exhibits behavior that, while it often converges to the correct diagnosis, does not necessarily follow a sensible train of thought in doing so, nor can it "exploit the possibility of deferring its own decision with a deliberate eye to waiting for disease evolution." [11]

Nonetheless, PIP lays some important groundwork for aspects of the medical reasoning problem. In response to initial clinical data, MAX will be capable of evoking diagnostic hypotheses that are associated with prior clinical knowledge in a manner reminiscent of PIP's frames, will evaluate these hypotheses on the basis of their likelihood and their ability to account for the data, and will select lines of inquiry designed to advance the "clinical understanding" towards greater diagnostic certainty. Problems of learning, comprehension of and response to clinical demands, and development of context-appropriate strategies will be deferred—for the present purpose it will suffice to expound upon the LIDA architecture sufficiently to convince the reader that MAX's design can accommodate the eventual development of responses to these admittedly formidable challenges.

3. THE LIDA ARCHITECTURE

The LIDA ("Learning IDA") architecture is a continuation of the work embodied in IDA ("Intelligent Distribution Agent"), which successfully integrates modules and processes modeling perception, episodic memory, comprehension, consciousness, deliberation and volitional decision-making in order to perform the duties of a task reassignment personnel officer for the US Navy. IDA implements a deliberative process which incrementally modifies its own semantic representations in a cognitive cycle (see below), references and applies Naval policies as appropriate, and employs limited natural language processing to understand and write email communications with sailors regarding preferences and availabilities for the reassignment options it decides to offer.

LIDA generalizes IDA's cognitive cycle into a complete model of cognition, incorporating perceptual, episodic, and procedural learning in both selectionist and instructionalist modes, and, like IDA, draws on the Baars' Global Workspace Theory to formulate its mechanism for attention and consciousness. [15]

LIDA's cognitive cycle proceeds in an iterative manner through a set of computational modules that process data from the (internal and external) environment, but operate in parallel on LIDA's internal representations (see Figure 1). While many subprocesses have an asynchronous character, the cycle's integrity is preserved by a pair of 'bottlenecks' which enforce an overall seriality, the conscious broadcast of the Global Workspace and the action selection of the Behavior Net, each of which occurs exactly once per cycle.

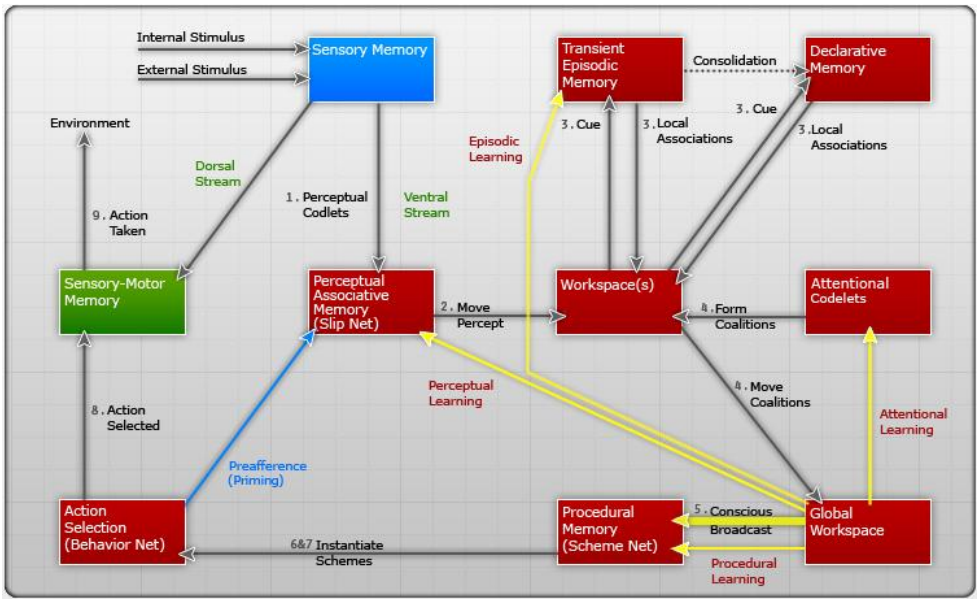

Figure 1: The LIDA cognitive cycle.

In addition to imposing seriality, the bottlenecks assert an hypothesis about how minds work. Thus, while LIDA seeks to accomplish an engineering goal by providing a framework for the construction of agents—such as IDA or MAX-- it approaches this goal along a route which also advances its science goal of proposing a set of hypotheses about cognitive function which is supported by current neurobiological and psychological understanding.

3.1. The Modules of LIDA and Their Processes

The modules of LIDA are as follows:

- Sensory Memory – creates a scene from sensory data.
- Perceptual Associative Memory (Slipnet) – integrates sensory primitives into representations of higher-order primitives, objects, actions, events, and known relations.
- Workspace (preconscious working memory) – receives and holds percepts and local associations for processing by codelets; contains various codelets which create new internal structures to represent perceptions, contexts, and goals, supply activation to internal structures as appropriate
- Transient Episodic and Declarative Memories – content-addressable, associative memories cued by and instantiated into the contents of the Workspace. Transient Episodic Memory (TEM) decays within hours or a day. [16, 17] Declarative Memory includes both Autobiographical and Semantic Memory, and, depending upon the strength of affect and the number of times it recurs in consciousness, may last the lifetime of the agent. Taken together, these memory modules are referred to as Episodic Memory.
- Global Workspace – the "spotlight of consciousness"; within each cognitive cycle, attention codelets monitor the Workspace for items of interest and compete for the right to determine the contents of the conscious broadcast, a single, limited set of informational content, from the Global Workspace to all other modules. This functionality of the Global Workspace constitutes one of the two 'bottlenecks' of the LIDA cognitive cycle.
- Procedural Memory (Scheme Net) – a collection of data structures called schemes, variably activated by the contents of consciousness, which represent a belief about the likely result of a primitive or composite action if taken in a particular context. If a scheme sufficiently activated, a behavior stream encoding its action is instantiated into the Behavior Net.
- Action Selection (Behavior Net) – contains undecayed behavior streams instantiated from the Scheme Net on current and previous cognitive cycles; during each cycle, and after the conscious broadcast, a single action is selected on the basis of activation which flows from situational contexts and currently instantiated goal structures as well as activation which flows between behaviors according to successor, predecessor and conflictor links. This module is LIDA's second 'bottleneck.'
- Sensory-Motor Memory – a repository of code which selects a specific algorithm for execution of the action selected by the Behavior Net. Sensory-Motor Memory receives direct, unconscious input from Sensory Memory to facilitate the execution of the action. This pathway of sensory information resembles that of the dorsal stream in the human visual system.

3.2. The Cognitive Cycle of LIDA

The cognitive cycle [17,18] can be conceptualized as occurring in three phases: 1) Understanding; 2) Attention; and 3) Action. During understanding, the

LIDA agent "makes sense of the world" by assembling a Current Situational Model (CSM) in the Workspace with various components representing percepts, remembered associations, actions, events, contexts, goals, and relationships between objects. The processes of Sensory Memory, Perceptual Associative Memory, and Episodic Memory might collectively be referred to as sensation, perception, and recall respectively. Each of these occur in parallel with the Workspace model-building which comprises understanding, and these processes do not necessarily stabilize into a fixed pattern within a single cognitive cycle.

Attention codelets form coalitions of selected portions of the CSM and move them into the Global Workspace, where a competition ensues. On the basis of an activation value that summarizes relevance, novelty, importance and urgency, the Global Workspace selects a coalition to which to attend and broadcasts a representation of the coalition's contents to all of the modules of the agent.

In response to the information in the broadcast, mechanisms for learning occur in many of the modules. In Perceptual Associative Memory, new nodes and links may be created, and existing links are modified as appropriate. Events in the broadcast are encoded in Transient Episodic Memory, and may, later and offline, also be stored in Declarative Memory. In Procedural Memory, the broadcast can trigger the formulation of new schemes, and existing schemes can be reinforced or inhibited.

The broadcast is of prime influence on the competition for action selection in the Behavior Net. A single action is selected and then passed to Sensory Motor Memory for execution at the close of each cycle. [19]

3.3. LIDA's Suitability for Modeling Diagnosis

The flow of activations in Maes' Behavior Network endows LIDA's action selection mechanism with a dynamic character that offers a tuneable balance between goal orientation and opportunism, and thus provides the flexibility, adaptability, and robustness needed in a medical situation where unanticipated events can radically and rapidly reorganize clinical priorities. [4, 20, 21] Goals are not required to be rigidly defined as propositional sentences in LIDA, which allows them to be coarsely or finely tuned to specific contexts and interleaved with other priorities, since multiple goal structures can be instantiated in parallel in the Workspace. The LIDA agent can be designed with capabilities for creating syntactically valid queries to previously designed expert systems and/or databases, and for "understanding" the responses of such systems by incorporating them meaningfully into its situational model. In addition to the benefits bestowed by its architectural attributes, LIDA's multiple learning modes create possibilities for both supervised learning and experience to continually enhance an agent's performance. Supervised learning and performance do not have to occur in separate phases of an agent's operation—in LIDA, they may occur in parallel.

4. THE MEDICAL DIAGNOSIS AGENT

To create a broad-brushed design scheme for the prototype version MAX 1.0, we examine a model diagnostic problem. There are two aspects of the diagnostic process: 1) What are the facts? 2) What do they mean? The first question suggests actions related to information acquisition; the second requires pattern recognition to the extent such knowledge is available given the training level of the agent, and possibly some degree of hypothetico-deductive problem solving.

Assume that some findings from a patient data record are brought into the Workspace as a percept or set of percepts. For instance, a 55 year-old male patient is in acute distress, complaining of chest pain and shortness of breath for an hour, provoked by heavy exertion during yard work on a hot day. These findings trigger local associations in Declarative Memory. Some are past cases stored in Autobiographical Memory, some are "book knowledge" stored in Semantic Memory, perhaps in a structure endowed with functionalities relevant to diagnostic problem solving, like PIP's frames.

The links between cases and frames in Episodic Memory (i.e., TEM and DM) may possess a stratification similar to Minsky's level-banding, in which the associations of an object in memory are layered according to the levels of abstraction at which they are removed from the cueing object. [22] While it may be useful to augment the fundamental LIDA memory architecture in order to add PIP frame style functionalities, the level-banding effect may be achievable with unadorned Sparse Distributed Memory, which employs a concept known as Hamming distance to identify spheres of associative proximity in a space of high-dimensional vectors. [23]

By whatever mechanism they arrive, the local associations are scanned for diagnoses, which in subsequent cognitive cycles are "recognized" as such by nodes in PAM and imported into the Workspace again as percepts. These diagnostic percepts then trigger new associations. All of this happens over multiple cognitive cycles.

As an aside, it should be noted that--depending upon the experience and/or hardwired knowledge of the agent--there may also exist nodes in PAM which correspond to diagnostic categories and receive activation from certain findings—for instance, shortness of breath, in combination with the appropriate set of findings and patient parameters, might activate a node for "possible

congestive heart failure" in a trained agent, but would not be expected to do so in a naïve one. If such nodes can be created by MAX as a result of learning from clinical experience and/or training through dialogue with a human practitioner, one might observe the same progression in MAX from backward-chaining to forward-chaining predominance that is observed in human medical experts.

After the settling of this activity, let us say that the Workspace now contains diagnostic constructs that include acute myocardial infarction (aka acute MI, a heart attack), pulmonary emboli (aka PE, blood clots to the lung), and community-acquired pneumonia (CAP).[4] These ideas have now been assembled in the Workspace as a result of learning and/or hardwired knowledge, but some of them may not yet have been conscious. Attention codelets begin building coalitions, each seeking to advance one of the identified diagnostic possibilities as a conscious hypothesis by broadcasting messages of the form "Let's consider <condition> as an explanation for <findings>."

When broadcast from the Global Workspace, such messages set in motion—over several cognitive cycles--the activation of schemes and streams, the actions of which (if selected for execution) will create an event model for diagnostic deliberation in the current situational model and update the differential diagnosis structure within this model by adding or removing a diagnosis from the list, re-ranking the list, or elevating a diagnosis to established status.

Each diagnosis in the DDx will evaluate its own ability to account for known findings, seek findings in Sensory Memory and/or PAM which strengthen or weaken its likelihood, request investigations from healthcare workers to generate such findings if absent, and trigger local associations in proportion to its likelihood. These functionalities might or might not require access to consciousness, or might require it to varying degrees, depending upon the level of expertise of the agent. Local associations triggered by a diagnosis in DDx might be for possible complicating conditions, or for possible therapeutic interventions.

For example, let's assume that over multiple cognitive cycles, MAX assembles a DDx containing each of the above diagnostic percepts as hypotheses, ranked in the following order: acute MI, CAP, and PE. Each diagnosis is associated with Workspace variables, data structures and/or links which represent or point to its relative likelihood, its possible complications and their gravity, findings which typify the diagnosis or suggest it as a possibility, findings necessary and sufficient for establishing it as fact, and learned lines for investigating the diagnosis, if any. Links associated with various groupings of findings within the diagnostic frame structure might indicate appropriate therapeutic courses of actions, or processes for determining them. The diagnostic hypotheses and their associated data are bundled into an event model representing a diagnostic deliberation event in the Workspace.

If no additional clinical data are added to the patient record during diagnostic deliberation, MAX will eventually exhaust available lines of hypothetico-deductive reasoning and reach a point in its deliberation at which a "watch and wait" strategy will be selected. In addition to having assembled the above DDx, it will have also made requests for information and considerations or recommendations for therapy. A GUI will be implemented to communicate the content of MAX's deliberations, requests and therapeutic recommendations to human healthcare providers,

A MAX agent with the training level of a first-year resident might eventually issue the following considerations, recommendations and requests for our sample case:

CONDITION GUARDED – LIKELY ADMISSION TO CARDIAC UNIT

- **Acute MI** –History is very typical for MI. Check blood pressure, heart rate, respiratory rate, oxygen saturation, EKG, cardiac enzymes. Determine major cardiac risk factors: smoking history, hypertension, diabetes, family history, hyperlipidemia. Check current medications and if any, determine their indications. Administer oxygen per nasal cannula, and aspirin unless contraindicated. Assess severity of pain, consider trial of nitroglycerine or morphine pending results of EKG and enzyme studies. Watch for cardiogenic shock, respiratory failure.
- **CAP**—Ask for history of recent cough, other signs and symptoms of infection, history of night sweats, past history of pneumonia, TB exposure, positive PPD. Check temperature, blood pressure, heart rate, respiratory rate, oxygen saturation. Check breath sounds on chest exam; head, eyes, ears, nose and throat for signs of infection. Consider sputum culture; blood cultures for fever > 102. Consider antibiotics pending history and exam results. Watch for septic shock, respiratory failure.
- **PE** – Ask for history of deep venous thrombosis [DVT, a.k.a. leg clots], leg injury, long trip [a risk factor for DVT]. Check breath sounds on chest exam, leg exam for varicosities, peripheral edema, cords. D-dimer studies.

As new information is entered into the database by healthcare workers (i.e. doctors, nurses, lab technicians, etc.), MAX will generate new likelihoods and may add or remove requests. For instance, if the vital signs are

[4] Many other possibilities might come to mind, but for simplicity's sake, the present discussion is limited to these three.

stable, EKG and enzymes are negative, and there are some risk factors for cardiac disease, MAX may decide that the patient still needs overnight observation to rule out cardiac disease, but might remove "Watch for cardiogenic shock, respiratory failure."

5. CONCLUSIONS

LIDA extends the already implemented IDA with a Scheme Net procedural memory module. LIDA offers great promise in providing a computational framework which can progressively extract bounded, well-structured clinical problems from open-ended, ill-structured ones; incorporate training and experience into its ontology; communicate effectively with healthcare providers and expert systems; and rapidly adapt to a sudden change in priorities.

If successfully implemented, MAX will prove useful for diagnosis verification, medical error detection and prevention, medical care optimization, as well as support for providers of medical services. A proposed initial trial for MAX is as a Triage Information Agent (TIA) which assists an emergency department triage nurse with her duties, and would offer a potential reduction of over 30% in emergency physician services required to determine patient disposition while maintaining quality of care. MAX would accomplish this by providing a more effective approach to pre-ordering laboratory and radiologic tests needed for the physician to complete the diagnostic assessment than that of a human ER nurse. [24]

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Amy Curry, Dr. Eugene Eckstein, and Dr. David Russomano for their invaluable assistance in honing the arguments presented herein.

REFERENCES

[1] S. Franklin and F. G. Patterson, Jr., "The LIDA Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent," Integrated Designs and Process Technology, IDPT-2006, San Diego: Society for Design and Process Science, 2006.

[2] J. Zacks, N. Speer, K. Swallow et al., "Event Perception: A Mind-Brain Perspective," *Psychological Bulletin*, Vol. 133, No. 2(2007), 273-293.

[3] G. Drescher, *Made Up Minds: A Constructivist Approach to Artificial Intelligence*, Cambridge: MIT Press, 1991.

[4] P. Maes, "How to Do the Right Thing," *Connection Science*, Vol. 1(1989), No. 3, 291-323.

[5] V. Patel, J. F. Arocha, and J. Zhang, "Thinking and Reasoning in Medicine," in Keith Holyoak, *Cambridge Handbook of Thinking and Reasoning*, Cambridge, UK: Cambridge University Press, 2005.

[6] *Dorland's Illustrated Medical Dictionary*, 28th ed., Philadelphia: W. B. Saunders Co., 1994.

[7] A. S. Elstein and A. Schwarz, "Clinical problem solving and diagnostic decision making: selective review of the cognitive literature," *BMJ*, Vol. 324 (March 23, 2002): 729-732.

[8] J. P. Kassirer, J. B. Wong, and R. I. Kopelman, *Learning Clinical Reasoning*, 2nd ed., Baltimore: Williams & Wilkins, 2010.

[9] R. S. Ledley and L. B. Lusted, "Reasoning Foundations of Medical Diagnosis," *Science*, Vol. 130(3366), July 3, 1959, pp. 9-21.

[10] G. Parmigiani, *Modeling in Medical Decision Making: A Bayesian Approach*, New York: John Wiley & Sons, 2002.

[11] P. Szolovits and S. G. Pauker, "Categorical and Probabilistic Reasoning in Medical Diagnosis," *Artificial Intelligence*, Vol. 11(1978), pp. 115-144.

[12] B. G. Buchanan and E. H. Shortliffe, eds., *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Reading, Mass.: Addison-Wesley Pub. Co., 1984.

[13] S. G. Pauker, G. A. Gorry, J. P. Kassirer, and W. B. Schwartz, "Towards the Simulation of Clinical Cognition: Taking a Present Illness by Computer," *The American Journal of Medicine*, June 1976, 60:981-995.

[14] L. B. Lusted, *Introduction to Medical Decision Making*, Springfield: Charles C. Thomas, 1968.

[15] B. J. Baars, *In the Theater of Consciousness: The Workspace of the Mind*, New York: Oxford University Press, Inc., 1997.

[16] Martin A. Conway, "Sensory–perceptual episodic memory and its context: autobiographical memory," *Philosophical Transactions of the Royal Society of London, Series B (Biological Sciences)*, Vol. 356(2001), 1375–1384.

[17] S. Franklin, B .J. Baars, U. Ramamurthy, and M. Ventura [2005], "The Role of Consciousness in Memory," *Brains, Minds and Media*, Vol. 1, pp. 1-38.

[18] B. J. Baars and S. Franklin, "How conscious experience and working memory interact," *Trends in Cognitive Science*, Vol. 7(2003), 166–172.

[19] S. Franklin (2009), personal correspondence, September 29, 2009.

[20] A. Negatu and S. Franklin (2002), "An action selection mechanism for 'conscious' software agents," *Cognitive Science Quarterly, Vol. 2*(2002), 363-386. Special issue on "Desires, goals, intentions, and values: Computational architectures" with guest editors Maria Miceli and Cristiano Castelfranchi.

[21] S. Franklin, U. Ramamurthy, S. K. D'Mello, et al., "LIDA: A Computational Model of Global Workspace Theory and Developmental Learning," in *AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches*, Arlington, Virginia: AAAI, 2007.

[22] M. Minsky, *The Society of Mind*, New York: Simon & Schuster, 1985.

[23] P. Kanerva, "Sparse Distributed Memory and Related Models," in M. H. Hassoun, ed., *Associative Neural Memories: Theory and Implementation*, New York: Oxford University Press, 1993, pp. 50-76.

[24] S. Franklin and D. Jones, "A Triage Information Agent (TIA) based on the IDA Technology," *AAAI Fall Symposium on Dialogue Systems for Health Communication*, Washington, DC: American Association for Artificial Intelligence, 2004.