Dimensions of Intelligent Systems

Dr. Gary Berg-Cross
Knowledge Strategies Division, SLAG, Inc.
Potomac, Maryland 20854, USA

1. Abstract

As intelligent systems have become more fully functional and commonly available, questions about their capabilities and relationship with humans have increased. This paper builds on the IS requirements ideas of Messina et al [2001] to explore middle ground between anthropomorphic approaches like the Turing test that rely on similarity to human behavior in an "imitation games" and the narrowness of tests of chess mastery. I contrast a system like Deep Blue which has a very fixed environment in which it performs to more complex types such as Associate technology. Deep Blue, I argue, is an example of system whose performance is expert, but whose competence is fragile and it may not satisfy extended definitions of competence and performance intelligence that we measure in dynamic environments. A clinical protocol system is used to explore the basic functional capabilities and knowledge. Beyond symbol processing and the knowledge level are grounded reactive intelligent within more of an environmental/systems perspective. I build on grounded systems to discuss the use of goals using, learning-based systems and & multi-modal logics that characterize "realistic" intelligent systems. It is argued that such characterizations will evolve as IS design matures into grounded intelligence and situated, rational agent systems. In the future belief models and measures of rational coherence might be used, as basic approaches to facilitate intelligent system performance in dynamic environments.

Keywords: IS, Intelligent Systems, Turing Test, Cognitive Model, situated cognition, BDI, Deep Blue, constructionism

1: Introduction

Investigation of artificial intelligence system capabilities now has a long history with notable discussion stemming from the original Turing test with many modern elaborations. Motivating contests exist for passing a test such as Turing's (Loebner Competition) as well as prizes for tasks in chess and various robot competitions (RoboCup, Office Navigation, Trash Pickup etc.). However, something like a Turing Test, imitation of human conversation, seems too difficult if we take it to logical conclusions, while success in something like chess as demonstrated by the triumph of Deep Blue seems too narrow an achievement to feel that we have made general progress on truly intelligent systems. Since the intersecting concepts of performance, intelligence and systems are complex, success with Intelligent Systems may be aided by focus on tasks of intermediate challenges. I will discuss several examples from the medical realm to illustrate the challenge and the state of affairs. From the example of patient safety performance I build a framework of intelligence perspectives or dimensions to organize the discussion. It has some clear idea of successful performance. For diagnosis we can ask if it is correct given ratings by an expert panel. Several IS diagnosis systems have been implemented and indeed, by expert ratings, they perform better than a typical physician, yet they have little penetration in the healthcare industry. Why? One reason is that the measure of correctness is isolated and doesn't look at the total picture including cost and maintenance issues. The system's knowledge can be difficult to maintain and further systems have difficulty fitting into the working environment, an issue I discuss more under the topic of Associate Systems.

Still another factor concerns the issue of patient safety and system error. There are 4 categories by which human performance is judged in relation to patient safety [Marx, 2001]. These are:

1. Human/system error
2. Negligent conduct
3. Reckless conduct
4. Knowing violations

We can see that judging human skill quickly gets beyond mere performance when assigning these categories[1]. Human/system error is a judgement that the system's performance was "inadvertent" and other than intended. We make such errors every day with minimal consequences and so might systems. The 2nd category, negligence, is a more culpable behavior and in healthcare is generally assigned when an individual has been harmed by a failure to exercise skill, care and learning expected of a provider [Marx, 2001]. Thus, we quickly leave a purely behavioral domain and enter one with concepts like "learning' skill and intentions. There are several distinct architectural levels that can be distinguished meaningful beyond the "what" of behavior that include why (a knowledge and cognitive level), how (functional /symbol) level and what descriptions. This leads to higher dimensions for judging behavior outlined in the four dimensions or views as shown in Table 1. At the

[1] For current implementations we could agree that ISes aren't going to make category 4 safety errors until we have systems whose intentions are explicit!

foundational level of behavior we can discuss the most obvious aspects of performance. When we talk about a system at this level the intent is not to go beyond the dimension of its behavior. But already there are many issues here, such as emerged from earlier attempts at "Behavioral Psychology". Our patient safety example illustrates the problems in principle. A second dimension, shown in Table 1 is a functional approach that is typically couched in a stimulus, information processing, and response model.

Dimension/Perspective	Characterization
1. Behavioral/What	Performance assumes not other system/dimensional knowledge. Such behavioral descriptions list the observable behavior exhibited by a system when it is being applied or executed. This model of system behavior (i.e., a series of episodes of the system's activities) relies on observation.
2. Symbolic/Functional Architecture (How)	Traditional information processing of symbols. The functional model describes an (implemented) representational and computational commitments/architectural primitives.
3. Environmentally Reactive (External why)	Intelligent behavior is a coherent response to environmental challenge. To do this it may functionally involve goals and world models. The knowledge level description describes a system in terms of the knowledge of the world and some principles that are applied when using that knowledge.
4. Goal-oriented and Intentional (Internal why)	Agent-orientation to rationalize intelligence at a belief/goal/ intention level. At this level we have refined principles of rationality.

Table 1 Perspectives of Intelligent Systems (What, how and why)

Prior PERMIS conferences in 2000 and 2001 provide a broad discussion on the testing of Intelligent System (IS) based both on behavioral performance, including efficiency and effectiveness measures drawing on the expectations of designers, as well as functional capabilities including robustness and learning capabilities etc. One way of pursing the question has been to take performance measures of non-intelligent systems and to attempt to add measures for intelligence [Messina et al 2001]. The simplest way to frame this has been to discuss the main elements found in IS. Messina et al [2001] propose several elements that make up a functionally intelligent architecture including:

- behavior generation to deal with incomplete commands (e.g. interpret commands, supplement instructions),
- synthesize alternative behaviors and adjust plans;
- adjust sensory processing to deal with the unexpected and unknown; and
- represent the world using an updatable, long-term stores of knowledge, including commonsense notions.

Such a listing fits the now classical Input, Process, Store and Output information processing type of model of intelligence such as shown in Figure 1. In such robotic models, perception and behavior are treated as separate, front end functions and "cognition" which, goes on in the processor and memory functions controls the perceptual and effector functions. This is a popular concept of intelligence as basically cognition [Newell, 1982]- the capacity to construct and manipulate symbolic representations, i.e. "approximate models" that are mapped to the environment and determine "appropriate" action. This is also called a knowledge-oriented view, since a system's knowledge is a way to describe behavior (e.g. synthesizing, adjust plans etc.). Chong and Berg-Cross [1990] provide an example of such work to understand the types of errors that ISes might make. Although models differ widely in terms of how sophisticated their concepts of knowledge, cognitive process and learning are Messina et al [2001] provide a useful base list several of functional requirements for testing cognitive systems e.g.. tests to measure the ability to fuse data from multiple sensors, including the resolution of conflicts. One of the goals of this paper is to follow up this approach by applying this criteria and additional characteristics raised in higher levels (3 and 4) discussed later or one advanced systems to illustrate assessment. A medical protocol planning system is used later in the paper to illustrate this.

There are many alternative ways of distinguishing the third distinction of intelligence. I call this Interactionist following a view of intelligence that see it, like knowledge, as open to interpretation and always relative to others things that provide an environment [Clancy 1989]. Steele [1995], for example, follows this view and sees self sufficiency in such interactions, which is called agenthood, as the basis for intellect. In this view we judge behavior as intelligent to the extent that it sustains an agent in an environment. Mail delivery robots are intelligent if they can successfully interact with and navigate a mail environment, especially if it is complex, uncertain and changes often. The key measurement is not a specific behavior but is described in terms of the quality of a result relative to this environment and what it presupposes. That result can be categorized as "success" in terms of a context. This is the same level of analysis as provided in the patient safety example previous discussed. Hence calling a system intelligent or its behavior intelligent, or in error, is based on an external, human/observe judgement as shown in Figure 1 [Van de

Velde 1995] rather than being a structural or even functional part of a "system". This shift in view includes the notion that knowledge is situational and cannot be viewed as self contained. Instead, it is inherently coordinated with an environment as situated context. If we buy this interactionist view of intelligence we quickly see the connection to a Turing Test (TT) which embeds system intelligence in a human context [(Saygin et al. 2000]. The Turing test is behavioral and interactionist but has a naive anthropomorphism implicit in a human "imitation" games. “If one were offered a machine purported to be intelligent, what would be an appropriate method of evaluating this claim? The most obvious approach might be to give the machine an IQ test . However, good performance on tasks seen in IQ tests would not be completely satisfactory because the machine would have to be specially prepared for any specific task that it was asked to perform. The task could not be described to the machine in a normal conversation (verbal or written) if the specific nature of the task was not already programmed into the machine. Such considerations led many people to believe that the ability to communicate freely using some form of natural language is both an essential attribute of an intelligent entity and a confirming test of underlying competence.

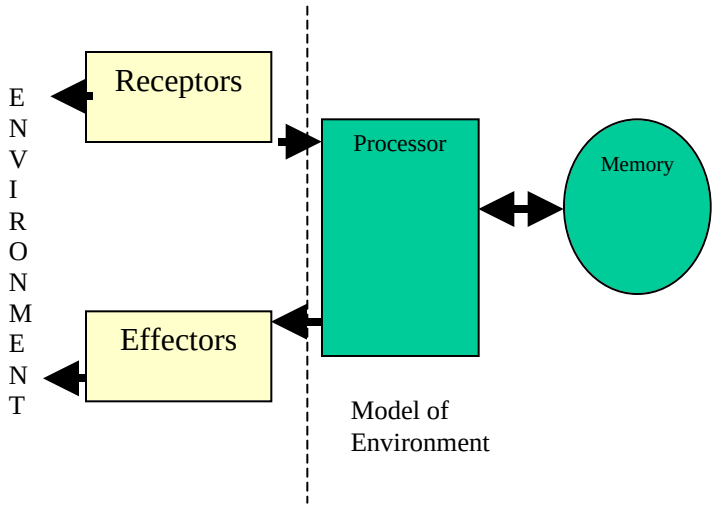

Figure 1 Model of Information/Symbol Processing

But philosophers like Searle with his Chinese Room argument challenge the Turing Test and its natural language exchange as a basis for assigning intelligence. This group maintains that the judgement of cognitive phenomena cannot be solely on the basis of observed input-output behavior. It is worth pointing out in passing that many strong AI critics like the Interactionist view, especially when it touches on the relative merits of symbolic learning and connectionist learning for implementing intelligence. Many find that Stevan Harnad’s [1990] hybrid model to grounding symbols in the analog world with neural nets is a useful approach. The issue of the behavioral problems of the naive Turing test are taken up later in the paper in the context of Deep Blue’s intelligent behavior.

Following Dennett's [1987] philosophical formulation, the fourth level explicitly considers intentions and belief, such as our patient safety . In philosophy the so called intentional stance position serves as a convenient, abstract way of talking about intelligent systems, allowing us to predict and explain their behavior without having to understand or describe how the cognitive mechanism actually works. Cognitive theorists and modelers have elaborated this in terms of cognitive structures that are capable of performing cognitive processes which in turn use those structures. Developers have to make many architectural decisions to actually implement such a philosophy. In the main this is the view a useful or “realistic” agent whether it is to be realized as "intelligent" software or as an autonomous robot. Such practical agents range from information gathering and trading agents to autonomous vehicles and may include real time physical capabilities, good for dangerous situations beyond the human central nervous system capacity.

The remainder of the paper is as follows. Section 2 walks through an example of an intelligent system applied to clinical guidelines to illustrate both the behavioral and functional criterion of Messina et al [2001] as well as the use of plans and goal reasoning in a system. Section 3 discusses the intelligence of Deep Blue using an interactionist robotics and unified cognitive architecture perspective. Section 4 returns to the Turing test and expands the concept to bridge to more useful concepts of judgement of intelligence and in particular the models that employ beliefs and intentions. Section 5 summarizes major findings, proposed initiatives arising from this view, dusts off the old concept of Associate Systems and proposes motivational competitions to enliven the field.

2. Gauging the Performance of Intelligent Clinical Protocol System

One of the most prolific areas of AI research has been in the medical realm which provides more representative measures of intelligence than chess performance and where there already exist regular monitoring efforts to judge success. A survey of the entire field is well beyond the scope of this paper, and even sub-areas, such as

diagnosis are too diverse to cover easily. Instead I look at one well researched area - the support provided by intelligent system for clinical protocols/guidelines and a system that has been designed to aid people in developing and using clinical protocols also call guidelines. Clinical protocols have been developed over the last 12 years provide a quality standard of care for such things as diagnostic and therapeutic procedures, typically based on the consensus of experts. Protocols are increasingly in widespread use and The American medical Association's Directory of Practice Parameters listed over 1,500 several years ago. However, there is little sound data to judge effectiveness across medical practice and intelligent processing has been researched to automate, support and improve guideline-oriented medical care [Musen et al 1996]. One reason for selecting a clinical protocol system is that task analysis of the field has been conducted [Sharhar et al 1998]. along with system development such as by the Stanford Medical Informatics work with EON [Musen et al 1996]. That work includes support of tasks for:

- determining the applicability of a guideline for a given patient,
- generating recommendations for therapeutic interventions and lab tests via a protocol
- tailoring the recommendation to the context of the current patient situation and stage of protocol execution,
- monitoring the application of the protocol guideline,
- assessing the effectiveness of the guideline.

I use the Messina et al [2001] list of requirements for testing Intelligent systems along with the performance evaluation properties (I would call functional capabilities) for ISes in non-numerical domains list and Musen et al [1996] to illustrate several of EON's interesting features. EON but is built from general, purpose software components and has been applied to protocol-based care in domains as diverse as oncology, hypertension, AIDS and diabetes.

Requirement 1 & 13: "interpret high level, abstract, and vague commands and convert them into a series of actionable plans" and " to understand generic concepts about the world that are relevant to its functioning and ability to apply them to specific situations".

EON's main task is a general one, generate an acceptable plan given the current clinical situation and relevant guidelines. It must determine a patient's eligibility and refines abstract plans to fit the situation. A relevant property is its ability to deal with general and abstract Information. EON's designers recognized that it needs to be a general problems solver like a physician . Thus it deals with both detailed, patient data in the medical record and database and abstract protocol specifications. It infers higher-level, interval-based concepts using time-stamped, patient data. Conceptual abstractions are a major feature of the EON approach and the PROTÉGÉ II system is used to build the EON KBs emphasizing the use of conceptual abstractions to define problem-solving behaviors independently from the programming logic.

Also relevant is the ability to "deduce particular cases from general ones". EON contains time abstraction, general medical ontology (e.g. concepts and relations between prescription, drug regime, medication, clinical trial etc.) as well as disease and patient specific knowledge/information. It uses an episodic skeletal plan refinement method on very general time concepts to instantiate a patient guideline plan. An illustration of this is shown in Figure 2. This deduces particular plans from abstract information in combination with very specific information.

Figure 2 EON Skeletal Plan and Refinement Process

Requirement 2, 3 , 15 & 14: "to autonomously make decisions as it is carrying out its plans" and to "re-plan while executing its plans and adapt to changes in the situation" and " work with incomplete and imperfect knowledge by extrapolating, interpolating, or other means " and " deal with and model symbolic and situational concepts as well as geometry and attributes".

While EON is not an independent agent it's processing includes setting sub-goals as part of its plan refinement. An observer would see it going far beyond the original specification in several steps:

1. Identify and propose a starting standard, abstract hierarchical plan

2. Instantiate the plan based on situation and decomposition and to allow execution (time constraints etc.)
3. Identify problems that might make application of this plan (practical drug admin challenges, side effects, etc.)

A property of the system is the ability to reschedule and replan and adjust the plan to updated situations [Messina et al 2001]. It also might be said to "recognize the unexpected" in that the guidelines project out a path and it will replan if deviations occur such as reduce AZT if anemia develops or side effects develop. Similarly it deals with incomplete information routinely. It typically does not have the entire attribute value set to begin with and generates queries of relational DBs that are processed into patient history. EON deals with situations of time, but distance and geometry, such as robotic path concerns are not part of its knowledge base (KB).

Requirement 4-9: to "register sensed information with its location in the world and with a priori data " and " fuse data from multiple sensors, including resolution of conflicts " and " to handle imperfect data from sensors, sensor failure or sensor inadequacy for certain circumstances" and " to direct its sensors and processing algorithms at finding and identifying specific items or items within a particular class" and to "focus resources where appropriate" and " to handle a wide variation in surroundings or objects with which it interacts" .

EON does none of this. It is not robotic. An IS is often robotic in having sensor and/or effectors but many are decision supports. Adding a robotic element to an IS is discussed later in the context of interactions and a Total Turing Test.

Requirement 10-12: to "deal with a dynamic environment" and "map the environment so that it can perform its job" and " update its models of the world, both for short-term and potentially long-term".

As noted EON deals with changes in the patient situation as well as changes to guidelines and phase of care. However, it's function does not result in model-mapping of the situation as might be implied here. Machine learning approaches which do this routinely such as embodied in the SOAR architecture are discussed later in the paper.

Requirement 15: " to predict events in the future or estimate future status".
EON does provide projections to allow comparison such as for the T-Helper implementation of Eon for AIDS - what is the situation 172 hours after symptomatic treatment.

Requirement 16: "ability to evaluate its own performance and improve".
EON does not have such ability.

3. Deep Blue's Brand of Intelligence and Grounding in the Interactionist Perspective

Chess was long seen as an extreme test of human intelligence and an excellent domain for IS [Levinson, 1991]. Chess performance is easy to monitor because success and skill categories are well defined. Studies of experts have been conducted to construct cognitive models which have been more broadly applied (uncertainty management and problem space pruning for example). However, as long as a dozen years ago computer success at chess was largely based on brute force computation using alpha-beta minimax search with selective extensions IS [Levinson, 1991], rather than elegant knowledge structures or complex processing strategies - the intelligent parts of a cognitive model. This was necessary to achieve effective time performance - conventional AI techniques were too slow for real-time response and chess is very much a time bound game. In the late 90s Deep Blue achieved its victory and it's processing capabilities are well known but the victory raises some interesting issues. Foremost is, do we consider Deep Blue intelligent? Behaviorally the answer has to be yes. If we take strictly behavioral views of intelligence in chess we may list the behavioral pattern without making any claim of an agent's cognitive level. Also by a judgement of interacting with its environment it is successful. By performance measures Deep Blue is intelligent, but this seems unsatisfactory on several other levels. It is grounded in the main chess objects and how they behave, but this is trivial, a game of simple rules. Deep Blue doesn't match up against the Messina et al [2001] criteria. It has minimal sensing capabilities, no commonsense knowledge to speak of, no ability to fill in knowledge. An interesting sidelight is that the IBM team found that while chess suggestions from experts were useful, they could not always be relied upon to aid Deep Blue's evaluation function - the essential process. What the team wound up doing is to create a "knowledge-free" machine using just available on-line chess databases to give the system a statistical experience base. That is, the final knowledge base had learned via working through a lifetime of chess games - essentially an grounding in chess reality, leading to its expertise, but the resulting knowledge base lacks the more abstract knowledge such as is often assumed underlies intelligence.

Now it is true that we typically are not talking about ISes with the full range of human ability, but in many cases we are talking about sensori-motor capacity and the ability to distinguish things in the world. Broader world knowledge is more typically learned by robots as summarized by Harnad [1993]in his discussion of a revised test of intelligence he calls the Total Turing Test (TTT):

> Well, in the case of the Turing Test (TT), there *was* more we could ask for empirically, for human behavioral capacity includes a lot more than just pen-pal (symbolic) interactions. There is all of our sensori-motor capacity to discriminate, recognize, identify, manipulate and describe the objects, events and states of affairs in the world we live in (the same objects, events and states of affairs, by the way, that our thoughts happen to be about). Let us call this further behavioral capacity our *robotic* capacity. Passing the TTT would then require indistinguishability in both symbolic and robotic capacity.

We can see this direction as also having been taken by real-time robots to handle problems such as:

- Symbol grounding to the real world (easy in chess, but not elsewhere)
- RT signal interpretation and planning under time constraints
- Situatedness issues - how is behavior adjusted to dynamic situations?

Subsumption architectures have been one attempt around these. [Brooks 1986] The main assumptions behind these attempts include:

- No attempt to construct an full, central symbolic model of the environment
- Behaviors are not controlled by a central executive looking at master plan, but may have a network of behaviors that may excite or inhibit each other.
- Sensor interpretation, planning and execution are not separated. Rather they are organized around modular competencies.
- Complex behavior is not programmed in but emerges from dynamic interactions between the environment and the component behaviors.

As Harnad [1993] says,

> Real transduction is in fact *essential* to TTT capacity. A computational simulation of transduction cannot get from real objects to either robotic performance or symbolic performance (not to mention that motor interaction with real objects also requires the output counterparts of transducers: effectors). This is the requisite nonarbitrary argument for the special status of transduction that we did *not* have in the case of parallelism (or silicon). In addition, there are other things to recommend transduction as an essential component in implementing cognition. First, most of the real brain is either doing sensory transduction or analog extensions of it: As one moves in from the sensory surfaces to their multiple analogs deeper and deeper in the brain, one eventually reaches the motor analogs, until finally one finds oneself out at the motor periphery. If one removed all this sensorimotor equipment, very little of the brain would be left, and certainly not some homuncular computational core-in-a-vat that all this transduction was input *to*. No, to a great extent we *are* our sensorimotor transducers and their activities, rather than being their ghostly computational executives.

There are now examples of unified robotic systems that include a commitment to symbolic manipulation as well as sensory transduction and organized motor responses which might satisfy the TTT, should we want to engage in it. One classic one is ICARUS [Langley et all 1991], which is made up of the standard 3 major components. However, architecturally the sensory buffer proves input to a mapper to the conceptual level, there is a reactive action planner to identify appropriate actions for world situations and there is a mapper of the action to an action "scheme" that drives the effectors that connect to the motor buffer.
The innovation here is that all 3 components use a similar representation (hierarchical probabilistic concepts) and reasoning is driven by a set of heuristics on the classifications in the hierarchy. Thus we have an index of world objects from perception, plans and more schemes each of which is defined by attribute-value pairs with a conditional probability of an attribute having a particular value given membership in a particular class. As noted by van de Velde [1995] by having consistent representation we get natural integration between perception, planning and action. Learning in inherent in the architecture since an instance is sorted down a hierarchy by class selection[2]. When a class is selected the probability distribution of attributes is updated using Classit's incremental function [Gennari et al 1989] or a

[2] One might use other approaches like Grossberg's adaptive resonance theory (ART) to organize clusters as a variety of levels and use the vigilance parameter to adjust the degree of clustering.

new class is created. This integration provides a natural symbol/concept grounding, which is not built in by a knowledge engineer. There need not be world view, rather, like reactive robots success is built through interacting with the world. It goes two steps better than reactive robots in that it creates:

1. manipuable concepts for planning as it goes along.
2. builds coordinated/coupled processes that are closed as well as defined by their organization and the action dynamics these processes imply.

This is also fundamentally different from the traditional I-P-O architecture and the separation between perception and action is largely gone. Both are reactions to environmental changes and employ a planning cognition to preserve or reach some state. Thus, with such an architecture I could speak of a motor performance with plans and percepts underlying it as an integrated thing. More importantly, I might expect the IS to be able to justify its actions by means of a trace of the activated concepts and attribute values. It is a step, but not necessarily as large a step, as we have with Deep Blue to imagine a robust performance over a range of domains.

4. Goal- Oriented Agent Intelligence with Beliefs, Desires and Intentions

In our final dimension I take a large step towards an intelligent performance that might meet some of the judgmental characteristics involved in patient safety. As previously noted categories of patient safety involve a judgement of the intentionality to do wrong. Like the Turing test it implicitly rests on human ability to predict and understand the behavior of “others” in complex interactions. This may lead to misjudging the intelligence of a system. As Hayes and Ford [1995] argue, the Turing Test is fundamentally flawed for two reasons: it is a basically poor experimental design, and it tests for the wrong thing. It is the wrong design because while it seems "unashamedly behavioristic and operationalistic", yet it is based on hidden assumptions rising from on our naïve psychology. Similarity to human behavior is just not a sensible criterion for intelligence. Our social experience provides an implicit, observer bias to assign mentality and intentions to the system in a test and many would argue that typical human use reasoning techniques haven't found their way into typical intelligent systems. E.g., humans use extremely complicated, temporally extended mental images and associated planned intentions to reason. It is the goal of the final dimension to build in such capabilities so that such judgements of IS could be justified.

In the prior section we briefly discussed goals within the ICARUS architecture, but the dynamics of these was not detailed. There is a large body of work to actively incorporate such planning about goals and belief abilities into ISes/intelligent agent architecture as more than reactive systems. ICARUS fits into this class of "cognitive architectures" as does SOAR [Newell, 1990]. SOAR, like ICARUS is interactionist in the sense that the task environment determines the possible structure of problem spaces. It is goal oriented in that problem solving is built around control knowledge that selects goals and sub-goals as it searches a problem state. To understand its behavior we have to look at its functional architecture and ontological commitments to knowledge and goals. Globally problem solving involves search, which is controlled by a context tree which might consist of a 4-tuple of: goal, problem space, state and operator. At any instant we may loosely say that this 4-tuple object is what such an IS actively "knows". The cognitive decision cycle (shown in Figure 3) consists of two phases to manages the context tree by determining what slot attributes should be changes. In the elaboration phase long-term knowledge is represented by production rules fire and those that fit the situational pattern fire in parallel until now more changes in the 4-tuple object occur. In this process new "preferences"[3] for a part of the 4-tuple context object may arise. In the second phase preferences are evaluated via a decision procedure. The result is a new contextual object better than any others. If such an object fails based on preferences one of fours types impasse is reached - tie. No-change, reject impasse or conflict impasse. All impasses are solved by the same goal- search process used in the cycle. Thus the system has a unified approach to problem solving around goal -- based learning that uses environmental results expressed in the problem statement and state as a factor.

SOAR is a pioneering effort which continues, but beside being goal driven we may also introspect about intents. Having a system aware of its performance was listed as an IS feature by Messina et al [2001]. SOAR has a step in that direction, but we wouldn't be comfortable speaking of its intention to "monitor itself”. Such intentions built into a system might satisfy professional guidelines for patient safety.

Intentions have been added to agent architectures based on Bratman's [1987] theory of human, rational behavior,

[3] Preference types are fixed simply at elements such as feasibility, exclusivity, desirability, necessity (require, prohibit), termination for desirability of alternative objects occupying slots in the context tree.

which formalized the ideas of Belief, Desire and Intention (BDI). BDI logics, such as developed by Rao and Georgeff [1998] , are multi-model symbolic logic[4] used in implementations to make agents more "realistic.

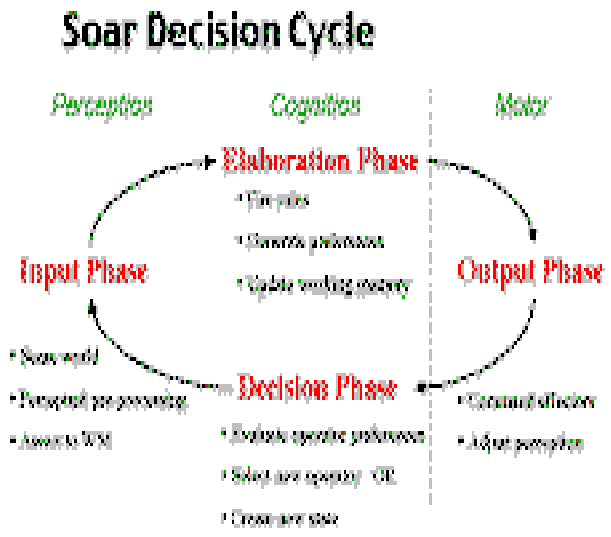

Figure 3. SOAR decision cycle

Realistic reasoning separates deciding what to do for, how to do it (planning), something like was the case in classical information processing systems but with a higher type of reasoning added. Both deciding and planning can be computationally expensive and an agent needs a strategy on when and how to drop an intentional commitment (its not possible or not feasible etc.). Because of this, implementations, while realistic are not typically practical. I take up a proposal to make these both practical and realistic in my final section. Modal style logics have been most widely explored for styles of intentional inference such as reasoning about time and belief. Agent propositional "event" knowledge is qualified as a temporal truth AKA belief, so deduction can both exploit features implicit in such qualification along with the context of the proposition. To say in the past, something true in the past is still today true in the past, asserts a global property implicit in the interpretation of the modality past. An analogous property of belief is positive introspection: if we believe something, we believe we believe it. So each modality implicitly carries a particular sort of inference and the use of different modalities allows the applied logician to make distinctions on the sort of inference. The benefit is economical notation, akin to natural language in the way detail is encapsulated in context, but at the cost of the functional transparency of extensional mathematics. Before closing it is worth noting that BDI implementations are typically more interested in autonomous agents, rather than developing AI systems for specialized purposes such as a particular medical application. Can these be brought together in a tractable way?

[4] A standard of classical truth functional propositional and first order predicate logic.

5. Summary and Directions towards Practical & Realistic Intelligence

Having reached the BDI level with deliberative and planning competence, let's discuss some of the possible directions for making ISes successful and measuring their performance. As we saw with the definitions of patient safety and in the naïve Turing Test our judgments involve concepts like intent which are not typically designed in a functional architecture. However as ISes become increasingly advanced, we can imagine drivers to systems with all the Messina requirements s well as an additional set to improve human –IS cooperation. Such cooperation and working relationships were envisioned by DARPA's AssociateTechnology program of the late 80s and early 90 [Berg-Cross 1991] in order to help:

- Handle increased amount of detail (bookkeeping)
- Remove bias
- Broaden the experience base by combining knowledge
- Provide easier visualization and
- Unify joint action

In an associative relationship human and IS share goals and tasks and communications. Interactions are structured /designed to work in mutually cooperating ways. The quality of IS decision and control depends greatly on the quality of information generation on its interfaces. In such interfaces we have the problems of situational awareness and situated cognition. A healthcare provider working with a traditional medical application is only modestly supported in adapting to changing demands and contexts. Working with an Associate IS makes it simpler in a way, but like any introduction of a new agent we have the problem of coordination - perception about patient status and treatment effect, clinical path, relevant medical procedures. Successful associate coordination would involve at least 3 things besides the raw information

1. Abstract knowledge
2. Situated dialog
3. BDI based dialog

As we saw in SOAR and EON an approach to the coordination problem is to combine abstraction with state space representation of tasks. Abstraction, rather than exhaustive reduction, deals with complexity by searching for global relational properties that exist somewhat independently of the knowledge elements whose behavior they govern. The models of abstraction need to be coordinated. Strong knowledge engineering and ontology tools, for situational knowledge specification are needed for this along with suitable browsers/query tools and displays.

Flexible and intelligent dialog is also needed. It must be situated to support coordination. Also as noted by McRoy et al [1999] part of the problem is that right now human-human communication is different than human IS. Systems provide large amounts of information and have difficulty handling feedback about it. They are not focused to human style communication, which makes coordination difficult and limits the trust of ISes. In the same way that some anthropomorphic judgements are projected on systems, so biases against their capabilities also may exist. Human interaction tends to be more incremental and feedback is interpreted in terms of beliefs and goals. For most interaction it is very important to know that beliefs, goals and intentions are shared and that beliefs about the dialog can be made explicit.[5]

I have used alternative views of the Turing test and Deep Blue's performance with four dimensions of intelligence to illustrate the difficulty of gauging success. My speculation is that we are far enough from success on any one dimension from achieving the type of success needed to master patient safety problems such as inherent in medical protocols. Obtaining a systematic view of system intelligence in complex, dynamic environments has rarely been a high-priority objective because obtaining realistic reasoning is computatuionally expensive. I think however it is time to consider some challenges (grand and otherwise) for ISes. Such challenges that fall between TT and chess. They could be small but interesting such as the annual robot building competition sponsored by AAAI or the self-replication contest held during the ICES98 conference. The object of this was to demonstrate a self-replicating machine, implemented in some physical medium, e.g., mechanical, chemical, electronic, etc. Larger efforts might be organized around work to integrate levels such as shown in Figure 4. Essentially there is a bottom up area taking the idea that intelligence is a coherent, grounded response to environmental challenge. What has been missing here is adequate understanding of the tasks implied in the environmental complexity going from the more predictable in chess to the more dynamic. Setting up a suitable lab and test environment would be a good area for NIST leadership.

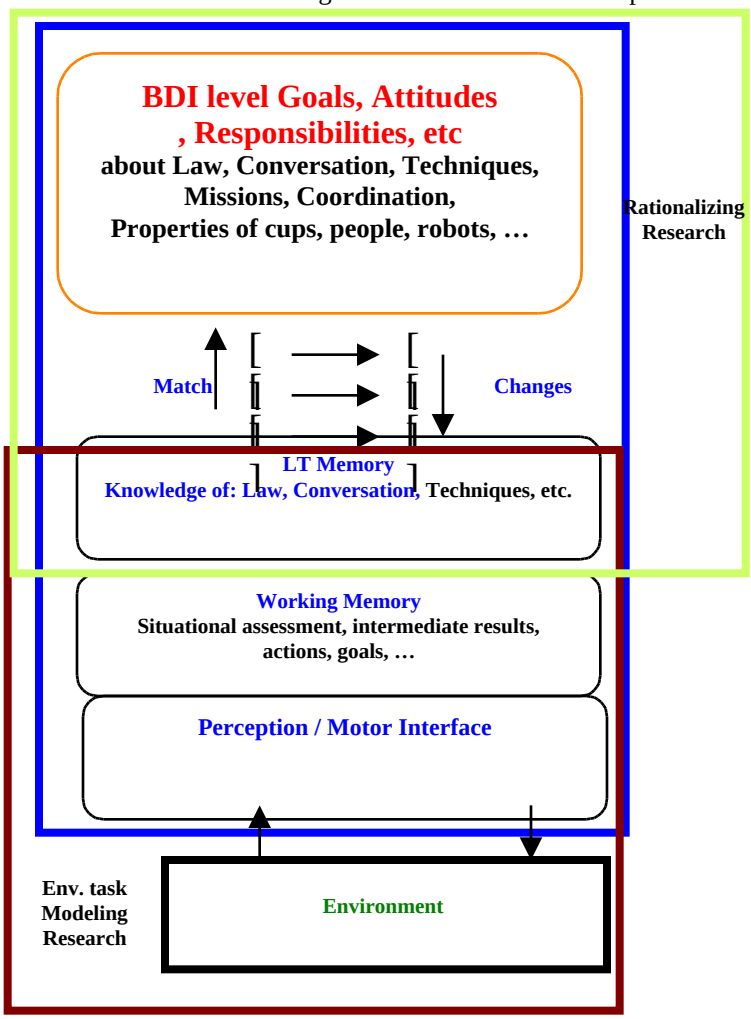

Figure 4. Environmental task modeling and Rationalizing Research: Areas of Integration

At the top level we have rationalizing research to have challenges in certain areas. The goal of this level of work is to be able to support the level of professional and legal responsibility such as we saw implied in patient safety. Henry Hexmoor and Gordon Beavers [2002] consider such needs from an agent perspective and propose to extended the intentional notions of Belief, Desire, and Intention (BDI) to include social "properties" of Value[6],

[5] Vicente and Jens Rasmussen [1990] pursue a related idea of ecological interface which includes a degree of abstract and BDI.

[6] Values are understood as principles that govern the agent's behavior and which the agent will attempt to uphold as end-goals

Obligation, and Norm (VON)[7]. These 3 provide a level), of predictability, accountability, and responsibility to an IS's high-level goal behavior (Figure 5) which allow us to talk of an associate system.

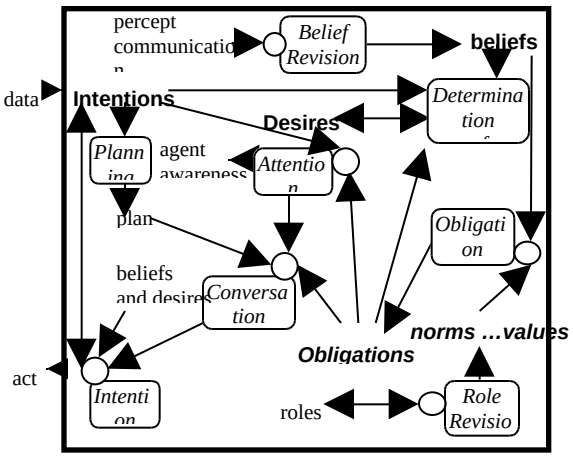

Figure 5 Intra-agent concepts
[Hexmore and Beavers, 2002]

> "Our legal system holds the owners of software agents responsible for the actions of those agents, therefore, agents capable of considering their responsibilities could offer some protection to the owner of the agent. Such software agents might be agents involved in electronic commerce, automated teller machines, proxy email agents, or robot assistants. Likewise in a command and control situation, a commander is responsible for the actions of the agents under his/her control and therefore would have greater confidence in responsible agents capable of considering the repercussions of their actions."
> Henry Hexmoor and Gordon Beavers [2002]

Additional efforts would be needed to standardize the abstract knowledge in ontologies necessary to achieve generality. The standard upper ontology (SUO) is one effort that might provide a start, but it should be tested by application to a several domains such as healthcare. My speculation is that situated BDI dialog will be needed to achieve higher reliability, verifiability, and precision of expression to support reasoning about other agents and extensive human-agent interaction. Given a few rounds of such work, it might be time to consider a true Turing test.

[7] Norms yield default behaviors that the agent is expected to observe whenever the agent finds itself in a situation to which the norm applies [Henry Hexmoor and Gordon Beavers , 2002]

References

Berg-Cross, G. Issues for Multiagent-Based Associative Planning Systems, *DARPA conference on Associate Technology*, George Mason University, 1991

Bratman, ME. Intention, Plans, and Practical Reason, Harvard University Press, 1987

Rodney A. Brooks. *A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation*, 2(1):14--23, March 1986.

Clancy, W. J. "The frame of reference problem in the design of intelligent machines." In van Lehn, K. (ed.) *Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition*, pp. 357-423. Hillsdale, NJ: Lawrence Erlbaum. 1989.

Gennari, J., Langley, P. and Fisher, D., "Models of Incremental concept formation." *Artificial Intelligence* (40) 11-61, 1989

Harnad, S. "The Symbol Grounding Problem". *Physica D* 42: 335-346. 1990

Harnad, S. Grounding Symbols in the Analog World with Neural Nets. *Think* 2: 12-78 , Special Issue on "Connectionism versus Symbolism" D.M.W. Powers & P.A. Flach, eds. , 1993.

Hayes, P.J. & Ford, K.M. (1995) Turing Test Considered Harmful, Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-95), pp. 972-977, Montreal.
Langley P., McKusick K.B., Allen J.A., Iba W.F., Thompson K. (1991): "*A Design for the ICARUS Architecture*" Procs. of the AAAI Spring Symposium on Integrated Intelligent Architectures.

Hexmoor, Henry and Beavers , Gordon, In search of simple and responsible agents, NASA *Workshop on Radical Agents*, Jan. 2002

R. Levinson, F. Hsu, T. A. Marsland, J. Schaeffer, and D. E. Wilkins. Panel: The Role of Chess in Artificial Intelligence Research. In Proc. of the 12th *IJCAI*, pp. 547-552, Sidney, Australia, 1991.

Marx, D. Patient Safety and the 'Just Culture': A Primer for Health Care Executives. Columbia University Medical Event Reporting System for Transfusion Medicine, April

2001. Available online at http://www.mers-tm.net/support/Marx_Primer.pdf

McRoy, S., Ali, S. Restificar, A. & Channarukeul, S. :Building Intelligent Dialog Systems" , *Intelligence*, Spring 1999

Messina,E., Meystel, A. and Reeker, L. Measuring Performance and Intelligence of Intelligent Systems, paper at *PERMIS* 2001.

Mark A. Musen, Samson W. Tu, Amar K. Das, Yuval Shahar: A Component-Based Architecture for Automation of Protocol-Directed Therapy. *AIME* 1995: 3-13

Allen Newell: The Knowledge Level. *Artificial Intelligence* 18(1): 87-127 (1982)

Newell, A. Unified Theories of Cognition, Harvard U Press, 1990

A. S. Rao and M. P. Georgeff. *Decision procedures for BDI logics*. Journal of Logic and Computation, 8, 293--342, 1998.

Shahar, Y.; Miksch, S.; Johnson, P.: The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines, *Artificial Intelligence in Medicine*, 14, pp. 29-51, 1998.

Ayse Pinar Saygin, Ilyas Cicekli, and Varol Akman, Turing Test: 50 Years Later, *Minds and Machines*, Vol. 10, No. 4., (2000), pp: 463-518

Steels L. (1995) The Biology and Technology of Intelligent Autonomous Agents. NATO ASI Series F, vol 144. Springer-Verlag, Berlin.

Van de Velde, Walter, Cognitive Architectures - From Knowledge Level To Structural. In Steels The Biology and Technology of Intelligent Autonomous Agents. 1995.

Vicente, K. and Rasmussen, J. , The ecology of human–machine systems II: Mediating "direct perception" in complex work domain. *Ecological Psychology 2*, 3, pp. 207–249. 1990