Alternative Paths to Intelligent Systems

Jens Pohl, Ph.D.

Executive Director, Collaborative Agent Design Research Center (CADRC)
California Polytechnic State University (Cal Poly)
San Luis Obispo, California, USA

Abstract

This paper examines the three prevalent approaches to Artificial Intelligence (AI), namely symbolic reasoning systems, connectionist systems, and emergent systems based on the principles of the subsumption theory. Distinguished by their top-down and bottom-up mechanisms all three approaches have strengths and weaknesses. While the logical reasoning approach is precise and well supported by mathematical theories and procedures, it is constrained by a largely predefined representational model. Connectionist systems, on the other hand, are able to recognize patterns even if these patterns are only similar and not identical to the patterns that they have been trained to recognize, but they have no understanding of the meaning of those patterns. The subsumption approach appears to overcome many of the weaknesses of the other two approaches in theory, but there is concern that it may not scale to more complex real world applications.

The author points out that in addition there are weaknesses that all three AI approaches share, namely inability to deal with exceptions, lack of mechanisms for analogous comparisons, and very primitive conceptualization capabilities at best. It is noted that the human agent performs decidedly better in these areas.

The paper concludes with the proposition that only a hybrid approach holds sufficient promise to meet the full expectations of intelligent systems. It is further suggested that this hybrid approach should include the contributions of the human agent as an integral component of the intelligent system, in most cases.

Keywords

Actuator, agents, Artificial Intelligence (AI), connectionist systems, context, data, embodied, emergent, information, intelligence, neural network, neurode, ontology, representation, sensor, situated, subsumption, symbolic reasoning, synapse.

Computation and Data-Processing

The need for devices with computational capabilities that exceed manual processes by several orders of magnitude, in terms of speed, was driven largely by mathematicians and physicists. During the first half of the 20th Century it was not uncommon for persons with doctorate degrees to spend weeks on the tedious solution of large sets of simultaneous equations for solving partial differential equations. These mathematical solutions were required for the preparation of tables that served as essential practical aids for many military and navigational purposes.

In this respect even the first, by today's standards, slow and clumsy electronic machines did not disappoint their creators. The ENIAC (Electronic Numerical Integrator And Computer) that was completed in 1946 at the University of Pennsylvania under the guidance of John Mauchly and J. Presper Eckert was able to perform a set of calculations that would have taken conventional calculating machines 40 hours, in 20 seconds. Even though it operated at only 100,000 pulses per second[1], this represented a hundred fold increase in calculation speed over the conventional electro-mechanical calculator technology and a more than one thousand fold increase in calculation speed over manual calculations (Goldstine and Goldstine 1982, Rojas and Hashagen 2000).

However, it soon became apparent that apart from computation there was another growing manual task that was in need of assistance, namely data-processing. Much of the data came in the form of textual data items that needed to be stored, sorted, and analyzed.

In more recent years, with advances in data storage technology accompanied by considerably decreasing costs, there has been an enormous increase in the amount of data stored and processed by computers. This has led to a bottleneck because while computers are able to store and process data as rapidly as they are able to compute numbers, they are unable to interpret the meaning of the data being processed. This essential task was left to the human users, because only they understood the context within which the data was being generated and destined to be used (Figure 1).

With rising expectations that the ability to store and process ever greater amounts of data should lead to better quality and faster planning and decision-making capabilities, the human user increasingly fell behind. This has become particularly apparent in the military intelligence and homeland security communities. High visibility examples include the World Trade Center catastrophe of 11 September 2001. As soon as the initial emergencies had been more or less dealt with, both the general public and the press media asked the obvious questions: Why wasn't this tragedy prevented? Were there not signs that the attack was about to take place? How could the intelligence agencies have been unaware that something sinister was being planned? The almost immediate truthful response was: *There is so much data out there that our intelligence analysts are often unable to see the forest for the trees*. In other words, the intelligence analysts were simply overwhelmed by the continuous flow of data through the many connected and also unconnected intelligence networks.

Based on human expectations and an increasing demand for *actionable information*[2] the pressure is mounting for computers to be able to not only store and manage data, but also to be able to interpret the meaning of the data. In particular, to be able to automatically detect and interpret the changes in data that occur as a direct result of events in the knowledge domain to which the data pertains.

So how can an electronic computing device automatically reason about data, recognize patterns, and perform any of the tasks that we normally associate only with human intelligence? This is a question that has been a subject of intense study by small groups of researchers for the past 50 years. Initially this relatively small Artificial Intelligence (AI) community of researchers had

---

[1] This is equivalent to a clock speed of 0.1 MHz, compared to a fairly standard laptop computer today (2007) operating at over 2 GHz (i.e., 2,000 MHz).

[2] The term *actionable information* has been coined by the military to describe data that have been interpreted, filtered, and are ready to form the basis of human decision-making sequences leading to actions.

high hopes of being able to produce systems that could match human intelligence in the short period of time of a few years. It turned out that these expectations were far too optimistic. The disappointment was so great that for some time AI research fell into disrepute. Funding sources dried up and research in this field was considered by many to be unproductive and ill advised. This over-reaction was unfortunate, because it is now generally recognized not only that there is a real need for intelligent systems but also that such systems will eventually become reality. In fact, it is even argued by some that there is a gradual merging of biology and technology, a convergence that is an essential part of the intellectual evolution of the human species (Kurzweil 1999).

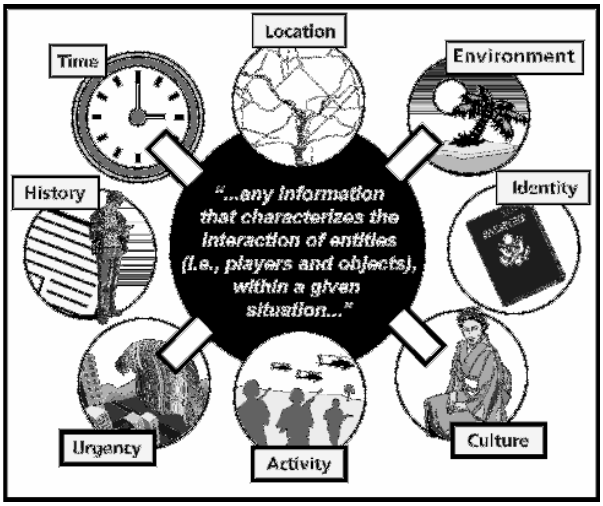

Figure 1: Typical components of context

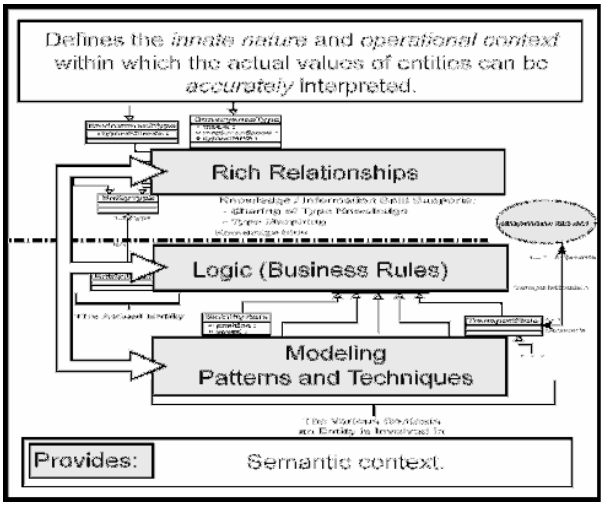

Figure 2: Ontology representation

Artificial Intelligence Approaches

Fundamentally, there have been two principal approaches to the development of intelligent systems, generally referred to as the *top-down* and the *bottom-up* approaches. In the top-down approach researchers have tried to emulate the rationalistic reasoning capabilities of the human cognitive system. In other words, these efforts have been focused on constructing virtual representations of real world situations and conditions in computer software. The strategy is that these virtual information models will then provide some measure of context to software modules with built-in reasoning engines.

The bottom-up approach has taken an entirely different path. The proponents of this approach have attempted to simulate the biological basis of the human nervous system. Referred to as the *connectionists*, they set out to construct mathematical engines to simulate the neurons and synapses that are an integral component of all forms of intelligent life. In this way they were able to achieve some measure of pattern recognition and matching that appears to be the foundation of human intelligence.

Symbolic Logic Systems

This top-down approach recognizes that *context* is a most likely prerequisite for all rational thought processes such as: the interpretation and filtering of data; the recognition of useful and

actionable information; and, the problem solving processes that are used in planning, evaluation, and decision-making endeavors. Accordingly, the proponents of this approach established conventions and languages such as the Unified Modeling Language (UML) and the Object Constraint Language (OCL) to form the necessary building blocks for constructing complex virtual models of real world knowledge domains (Booch et al. 1997, Larman 1998).

Initially, these representational models were referred to as object models because they defined real world entities as objects with attributes and behavioral characteristics. The objects were then associated with each other to simulate the relationships that allow human reasoning to determine the meaning and significance of changes in data that result from events in the real world. As these representational models and the methods for their construction became more sophisticated an increasing number of the reasoning capabilities could be moved from the reasoning engines (i.e., agents) into the representation (i.e., the virtual model). Today, these virtual models are more aptly referred to as ontologies because they do, true to their dictionary definition, attempt to represent all of the concepts and notions in a particular knowledge domain.

An ontology is an information model, rich in relationships, which describes the context of a real world situation or problem domain such as the management of goods movement across international borders, or the design and manufacturing process of an engineering product, or the command and control decision-making environment of a military battlespace (Figure 2). It is not limited to the modeling of physical entities such as buildings, roads, persons, activities, climatic factors, weapons, and vehicles. An ontology can also model abstract concepts such as threat, privacy, consumability, mobility, stature, and time. Some of these entities are of course easier to model than others. For example, the behavioral characteristics of an aircraft can be quite easily described in terms of attributes such as cruising speed, range, dimensions, maximum payload, and so on. Even the relationships between an aircraft and its landing and take-off requirements, its refueling and crew needs, and its maintenance schedule can be comprehensively modeled with a high degree of accuracy. However, to build a comprehensive model of the cultural characteristics of a nation or even an organization is a much more difficult undertaking. The reason for this difficulty is not necessarily due to the shortcomings of the available representational tools and methodologies, but rather the inadequacies of our own human understanding of these human behavioral characteristics.

The reasoning engines or software agents that are designed to navigate these somewhat simplistic virtual models are commonly in the form of *if-then* rules (i.e., if certain conditions or antecedents are reflected in the current state of the model then certain inferences are automatically made). These inferences are normally expressed as actions that are immediately executed by the agent. In this way software agents are able to autonomously communicate with other software agents or human users, monitor events by tracking data changes, retrieve data from external sources, request and provide specialized services, pursue interests and objectives, and accomplish at least low level learning tasks.

While this may seem impressive, and is certainly orders of magnitude more powerful than rote data-processing without context, it is still quite limited in comparison with human cognitive capabilities. This becomes particularly apparent when we consider human capabilities in the areas of intuition and common sense.

More specifically, difficulties with this top-down approach are routinely encountered in several areas. It would appear that the entire mechanism of antecedents (conditions) and conclusions

(actions) is overly simplistic. Often there are too few matching conditions to unambiguously determine which rule most clearly fits a particular situation. The ability to create during execution a new rule that combines those elements of two or more rules that fit the situation certainly exists in theory but is difficult to implement effectively in practice. On the other hand, there are often too many matching conditions. The available recourses in this case are by no means foolproof. Some of the more common alternatives include the implementation of a priority scheme to select one rule, or invoking some other mechanism such as case-based reasoning to select the most appropriate rule, or to create a new rule that combines the most desirable features of all the rules that fit the situation, or perhaps it may be more useful to pursue multiple alternative paths in parallel.

In summary, the advantages of the top-town approach are basically threefold. First, the formal symbolic logic on which this approach is based provides clarity and verifiable precision. Second, there is the availability of well established mathematical theories and procedures, and third, the similarity to the human reasoning process is obviously very attractive to us human beings.

However, on the downside there are also some serious disadvantages that apply to this AI approach. Foremost among these is the intrinsically static nature of the representation model. It forces the top-down approach to adhere to a largely predefined and explicit representation of objects, the behavioral characteristics of those objects, and the relationships among the objects. Further advances in methods that will allow ontologies to be extended, shared, and merged during execution are urgently needed and likely to become available in the foreseeable future. While the availability of such methods will greatly increase the power of ontology-based systems, there will remain the constraints imposed by a predefined and strictly ordered view of a world that in reality is subject to continuous change due to the interactions of the elements in its domain.

Also, due to its reliance on an explicit representation the top-down approach cannot easily deal with exceptions (Minsky 1990). Yet, in the real world we are often reminded of the importance of the exceptions to the general rules. This is perhaps an unfair criticism of the top-down approach because we really have not devised any formal mechanisms for dealing with exceptions. Statistical methods were developed to establish norms and deviations from these norms. Fuzzy logic provides a set of mathematical methods for establishing the certainty of inferences. However there is no equivalent mathematical method available for identifying patterns derived from the confluence of exceptions.

Another drawback of the top-down approach is its apparent inability to support the formulation of analogies. An example of such an analogy capability is the ability to conduct conceptual searches in a distributed database management system (DBMS). A traditional DBMS typically supports only factual searches. In other words, users and applications must be able to define precisely and without ambiguity what data they require. In complex problem situations users rarely know exactly what information they require. Often they can define in only conceptual terms the kind of information that they are seeking. Also, they would like to be able to rely on the DBMS to automatically broaden the search with a view to *discovering* information. This suggests a need for the ability to formulate search strategies based on incomplete definitions. It should be possible to infer from rather vague information requests and knowledge of the problem context, a set of executable query procedures that will lead to the discovery of analogous information domains.

Finally, formal information models are unable to represent the wealth of information and knowledge that allows us human beings to exercise common sense. This became a serious criticism of early AI research and still remains today one of its greater weaknesses.

Connectionist Systems

The objective of the bottom-up connectionist approach is to emulate the biological basis of the human nervous system. The human brain contains over 100 billion computing elements (i.e., neurons), which communicate throughout the human body through nerve fibers with over 100 trillion interconnections (i.e., synapses). Some neurons have only a few synapses and others may have thousands. This network of neurons is responsible for all of the human phenomena that are referred to as thought, memory, emotion, and cognition.

The principal capability of the human brain appears to be related to the recognition and processing of patterns. This human pattern matching capability applies to speech communication, the recognition of persons and objects, the performance of tasks, and reasoning. The connectionist approach to emulating this pattern matching capability depends on the construction of a network of interconnected nodes that are capable of sending signals to each other. The strength or numerical values of these signals are based on the cumulative application of mathematical weighting functions.

The nodes can be thought of as artificial neurons and are referred to as *neurodes*. Each neurode is essentially implemented as a mathematical *transformation function*[3] that associates a given level of input signals with a particular level of output. The neurodes are generally connected in multiple layers that include an input layer, one or more intermediate or *hidden* layers, and an output layer. Each input neurode is connected to each neurode in the next layer. These connections or artificial synapses can be unidirectional or bidirectional. Typically, due to the many interconnections, each neurode receives a large number of input signals. These input signals are accumulated by the neurode until they exceed a threshold value, at which time the neurode will send an output signal to other connected neurodes (Figure 3).

When implemented in software executing on a digital computer, each input layer neurode receives a starting value (signal) between 0.0 and 1.0. If the cumulative input values exceed a predefined threshold value then the neurode will *fire* and send identical output values (signals) to each second-layer neurode. Each of these second-layer neurodes will multiply the value (signal) received by a weighting factor. These values are summed and as soon as the combined value exceeds the threshold value the neurode will *fire*, and so on.

The advantages of the connectionist approach are very different from the advantages of the top-down approach and intuitively attractive to our human viewpoint. Apart from their apparently elegant mathematical formulation, neural networks perform quite well even with incomplete input and can recognize conditions that are similar but not identical. Moreover, any particular neural network can normally be trained to recognize several kinds of patterns. For example, the same neural network can be trained to recognize most (if not all) letters of the alphabet. Neural networks have been successfully applied to perform quite complex pattern matching tasks, such as navigating a car on a road with other traffic at speeds above 50 mph.

[3] The *fractional Fourier transform* (FRFT) is a linear transformation that is often used in pattern matching neural networks.

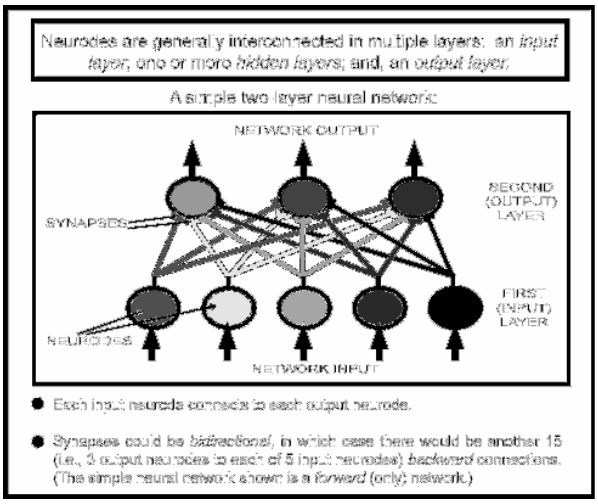

Figure 3: Neural network components

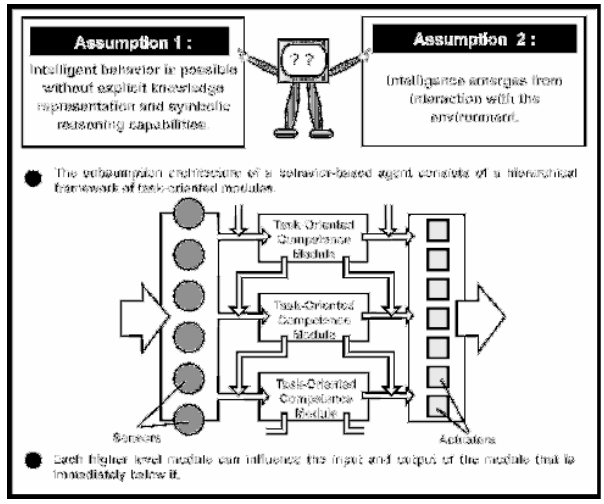

Figure 4: Subsumption architecture

However, the connectionist approach also has some intrinsic disadvantages. First and foremost, there is absolutely no *understanding* within the network of the meaning of the pattern that has been recognized. All that the neural network has been able to achieve is to generate a set of very similar output values (signals) every time it receives a different but also similar set of input signals. For example, if a neural network has been trained to recognize the alphabetic character *H* then it will recognize this letter with reliability only if it is presented to it in the same surroundings. Let us assume that the letter *H* was located at the bottom left side of a computer screen when the network was trained. Then there is no guarantee that the network will recognize the same letter *H* if it is moved to the top right side of the screen. In fact, if there are other competing images around the *H* in one or both locations it is likely that the network will fail to recognize the *H* pattern in the new location.

The above example suggests quite correctly that there is in fact little theoretical understanding of exactly how the mathematical representation leads to the pattern matching capability of the neural network. Attempts to gain such an understanding are confounded by the fact that the knowledge within the internal nodes of the hidden layers is not readily accessible. Also, the weighting coefficients that modify the input values (signals) in each node cannot be changed like software can change the content of the memory cells in a digital computer in the top-down approach.

Finally, it is of course very difficult to explore alternatives with neural networks. There is no reasoning capability because the network simply matches patterns. It has no understanding of the semantics of those patterns and is therefore unable to build on its initial recognition achievement with higher level tasks. Of course such tasks could be performed by external top-down systems that map the output of the neural network to a symbolic representation for reasoning purposes.

Emergent Systems

A bottom-up approach that is quite different from the connectionist methodology was first proposed in the 1980s by Rodney Brooks, who has been leading an AI research effort in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology (MIT) for the past

30 years. Brooks argues that the top-down approach is fundamentally flawed for at least three reasons (Brooks 1991). First, most of the activities performed by humans on a daily basis are routine and do not involve problem-solving or planning[4]. Second, complete models of real world environments are impossible to construct because they are subject to change. Intelligence can emerge from subcomponents interacting with each other and the environment through sensory mechanisms. Accordingly, Brooks has been conducting most of his research with mobile robots that are able to interact with the environment in which they exist. Third, agents can have beliefs and goals without actively reasoning about high level semantics. This does not necessarily imply that robots should not reason about their environment, but rather that the semantic representation that is required for reasoning will need to be built by the robot during its interactions with the environment.

Brooks first introduced the notions of the *subsumption* theory in 1986 (Brooks 1986). The term *subsumption* derives from the tight coupling between emergent intelligence and the real world environment. The subsumption architecture of a behavior-based agent (robot) consists of a hierarchical framework of task-oriented modules (Figure 4). Taking input from sensors and providing output to actuators, each higher level module can influence the input and output of the module that is immediately below it. Triggered by its sensors a robot agent dynamically builds a temporal model of the real world that surrounds it. Objects and relationships are primarily relevant as they are sensed and only secondarily important within the larger context of a more complete model of the world (i.e., the kind of model that is a fundamental prerequisite of the top-down approach). The subsumption theory is based on four major pillars, namely that robot agents are situated, embodied, intelligent, and emergent.

A robot agent is *situated* because it continuously refers to its sensors rather than an internal model of the world. Responding quickly to its sensor inputs, the robot is forced to build a temporal model of its surroundings relative to itself rather than an external framework. For example, the robot would refer to a sofa as the obstacle that is right now to its right, rather than *object 7*, which is a sofa. In other words, the robot agent is required to learn about its environment by interpreting its real world experiences with little (if any) initialized knowledge. Clearly, this approach is particularly appropriate in a dynamically changing world in which the past state of the world provides little reliable information about the current and future states.

The notion of *embodiment* is based on the fact that the physical presence of the robot agent forces it to potentially deal with all issues that its sensors and actuators are capable of processing. The assumption is that only an *embodied* agent can be validated in terms of its autonomous capabilities and its intelligence. Therefore, timely perception and action in preference to strategic planning and problem-solving are likely to be the most challenging behavioral capabilities of such an agent. However, this also suggests that some degree of redundancy will be necessary because such agents are likely to be vulnerable to sensor malfunction.

Brooks argues that robot *intelligence*, like human intelligence, is largely a function of the degree of complexity of the environment rather than its own internal complexity. Since intelligence is determined by the dynamics of interaction with the environment, it is often difficult and not necessarily useful to draw a distinction between intelligence and successful environmental interaction. Therefore, similar to human evolution where the development of perception and

[4] This view is shared by Jeff Hawkins whose Hierarchical Temporal Memory (HTM) theory is based on his research of the functions of the neocortex (Hawkins and Blakeslee 2004).

mobility capabilities took much longer to develop than reasoning capabilities, the intelligence of a robot agent depends more on its dynamic interaction capabilities than its reasoning capabilities.

Finally, the intelligence of a robot agent *emerges* through the interaction of components. These components are best focused on behavior producing tasks than functional information processing tasks. Accordingly, the components of these behavior-based robots are designed to produce the required environmental interaction and mobility capabilities collectively. It is therefore difficult to identify the seat of intelligence of a robot agent because the intelligence is the result of the interaction of many contributing capabilities. In other words, higher level intelligent behavior *emerges* from lower level behavioral capabilities through a process of repetitive learning.

In summary, subsumption or behavior-based systems are reactive rather than proactive systems, whose planning interests and capabilities are driven largely by unexpected needs. Their strength lies in the fact that without resorting to any central symbolic representation they are capable of: making predictions and forming expectations about their environment; developing plans that relate to their immediate needs; and, formulating and implementing goals.

At the same time, it is reasonable to question whether the bottom-up subsumption systems will be able to scale to more and more complex real world environments. This will most likely depend on the ability of such robot agents to learn from their interactions with the environment in which they operate. Will they be able to respond with sufficient speed to their sensor inputs to progressively develop a level of intelligence that is several orders more sophisticated than their foundational sense and response mechanism? Will they be able to build an experience-based pattern identification and problem-solving capability?

The Path Ahead

Clearly all three of the AI approaches are facing major obstacles. The symbolic reasoning approach depends on a largely predefined virtual model of the real world. While some aspects of the real world environment are fairly static, there are other features that are subject to continuous change as the players in the real world interact with each other and their environment. For example, the furniture in virtual any building space will be moved around due to the various interactions of the building occupants. Even more apparent chaos will exist at certain times in a manufacturing plant or supermarket. Since these changes cannot be adequately predefined there is a need for the virtual model to adapt as the real world environment changes. The ability to extend ontologies during execution is absolutely necessary and will certainly be helpful, however, the first sign that something has changed in the environment will likely come from the agents in the environment (i.e., from the bottom up). Therefore, there is a need for these agents to be able to identify and adapt to the changes before the virtual representation model can be corrected from the top down.

The connectionist systems can recognize patterns but have no understanding of what they have recognized. In other words, they can detect non-literal similarities but need some other mechanism to determine what action (if any) should be taken in response to the detected situation. The strengths and weaknesses of the symbolic reasoning and connectionist systems appear to be somewhat complementary. For example, whereas logical reasoning systems are very limited in their ability to detect non-literal similarities, neural networks can recognize conditions that are similar but by no means identical. Also, if a neural network recognizes a pattern then a

symbolic reasoning system working in conjunction with the network may be able to determine the meaning and significance of the condition within the context of its virtual representation model.

The subsumption approach embodies several very powerful notions that overcome some of the theoretical objections to both the symbolic reasoning and connectionist approaches. Its key constructs of *situated*, *embodied*, and *emergent* are intuitively obvious. Also, the configuration of sensors, task-oriented competence modules, and actuators in the subsumption architecture is seductively simple in theory. However, can it be implemented in practice in systems that are useful beyond mere demonstrations or toys? In other words, is this theoretically very promising approach scalable in practice?

There are other serious shortcomings that apply to all three AI approaches. First, there remains the problem of dealing with exceptions. It would appear that the representational models of the top-down approach must by their very nature always be inclusionary in character by adhering to the principles of consistency, regularity, and conformity. The subsumption approach perhaps comes closest to dealing with exceptions because the agent robots do not necessarily draw a distinction between norm and exception until they start to overly rely on their experience. Second, there do not appear to be any effective mechanisms for dealing with analogous comparisons in computers, and third, only very primitive conceptualization capabilities such as case-based reasoning have been demonstrated in symbolic systems to date.

So, what is the path ahead? It does not appear that any one approach is sufficiently strong to meet the full expectations of intelligent systems. A hybrid approach is likely to be necessary. However, it would be well to consider the human user not only as a necessary but also as a beneficial contributor in most intelligent system environments. If we consider technology as an enabler of human capabilities then it would seem appropriate that the strengths of the human should form an integral part of the intelligent environment. In the same way that at least some of the strengths and weaknesses of the top-down and bottom-up AI approaches complement each other, the human strengths in the areas of analogous thought and conceptualization complement the weaknesses of all three AI approaches.

References

Booch G. J. Rumbaugh, and I. Jacobson (1997); ‘The Unified Modeling Language (UML) User Guide’; Addison-Wesley, Reading Massachusetts.

Brooks R. (1991); ‘Intelligence Without Representation’; Artificial Intelligence, Volume 47, (pp.139-159).

Brooks R. (1991); ‘Intelligence Without Reason’; A.I. Memo No. 1293, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, Massachusetts.

Brooks R. (1990); ‘Elephants Don’t Play Chess’; Robotics and Autonomous Systems, Volume 6, (pp 3-15).

Brooks R. (1986); ‘A Robust Layered Control System for a Mobile Robot’; IEEE Journal of Robotics and Automation, RA-2, April, (pp. 14-23).

Goldstine H. and A. Goldstine (1982); 'The Electronic Numerical Integrator and Computer'; in *The Origins of Digital Computers: Selected Papers*, Springer-Verlag, New York, New York (pp. 359-373).

Hawkins J. and S. Blakeslee (2004); 'On Intelligence'; Times Books, Henry Holt and Company, New York, New York.

Kurzweil R. (1999); 'The Age of Spiritual Machines: When Computers Exceed Human Intelligence'; Viking, Penguin Group, New York, New York.

Larman C. (1998); 'Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design'; Prentice Hall, Upper Saddle River, New Jersey.

Minsky M. (1990); 'Logical vs. Analogical or Symbolic vs. Connectionist or Neat vs. Scruffy'; in Winston P. *Artificial Intelligence at MIT: Expanding Frontiers*, Volume 1, MIT Press, Cambridge, Massachusetts.

Minsky M. (1982); 'Why People Think Computers Can't'; AI Magazine, Volume 3, No. 4, Fall.

Rojas R. and U. Hashagen (2000); 'The First Computers: History and Architecture'; MIT Press, Cambridge, Massachusetts.