PERMIS'2001

White Paper

Measuring Performance and Intelligence of Intelligent Systems[1]

Table of Contents

1. Is Testing of Intelligent Systems different from Testing of Non-intelligent Systems?

Testing of performance pertains to evaluation of the potential and actual capabilities of a system to satisfy the expectations of the designer and the users via exploration of its functioning. This includes determining how well the system performs its declared "job," how efficiently and effectively it does so, how robust it is, and so forth. The "job" and expected performance must therefore be defined at the outset. Efficiency is defined as how well the system does things right, effectiveness is defined as how well the system does the right thing , and robustness is defined as "the degree to which a system … can function correctly in the presence of invalid inputs or stressful environmental conditions." [Finklestein, 00] When we are talking about the measure of intelligence, we are not referring to comparing the system to humans to determine whether it is indeed intelligent. We are not looking for and are not interested in a nouveau Turing test.

Furthermore, the tests under consideration are not meant to be broad-based general evaluations of the system's knowledge or the full spectrum of its capabilities. In particular, we are not striving to ascertain whether a system has common-sense generic knowledge applicable to general-purpose problem solving. The system being evaluated has a given sphere of responsibility and known abilities and tasks that it is able to undertake under its specifications.

[1] This paper is written by E. Messina, A. Meystel, and L. Reeker.

Comments regarding the testing of intelligent versus non-intelligent systems are not meant to underestimate the difficulty of testing non-intelligent systems. Testing robustness, efficiency, and even functionality of *non-intelligent* software systems is difficult enough [Tsai, 99]. Since the software execution can follow a myriad of combinations of paths through the code, it is impossible, in typical practice to exhaustively test all the possible combinations. In non-deterministic real-time systems, the problem is compounded by the uncertainty in the execution times of various processes, the sequence of events, asynchronous interrupts, etc [Butler, 93].

In general, the evaluation of intelligent systems (IS's) is broader than testing of non-intelligent systems (NIS). A system that has intelligence should in general be able to perform under a wider range of operating conditions than one that does not have intelligence. In fact, it should learn from its experiences and either improve its results within the same operating conditions or extend its range of acceptable conditions. What does this mean? Let's look at the main elements typically found in an intelligent system: Behavior Generation, Sensory Processing, and World Modeling (Knowledge Representation) [Meystel, 00].

2. Behavior Generation

Dealing With General and/or Incomplete Commands

An IS is given a job to do (task, mission, set of commands). *The job definition for IS is expected to be less specific than in an NIS*. A system with intelligence ought to have the capability to interpret incomplete commands, understand a higher level, more abstract commands and to supplement the given command with additional information that helps to generate more specific plans internally. The IS should understand the context that the command is given in. For example, instead of telling a mobile robot to go to a specific location in world coordinates "GO_TO(X, Y)," the command could be "Go to the window nearest to me." The robot should understand what a window is and know that it needs to find one which is the minimum distance away from me and move to that location. It also has a nominal proximity that it maintains from the goal location. Notice, that the command did not determine how close the robot needs to get to the window. It is expected that the robot knows where to stop the motion in similar cases, or the distance from the window should allow for convenient performance of other, or consequent movements.

Ability to Synthesize the Alternatives of Decisions and to Choose the Best One

There was time, when the processes of decision making and planning were understood and reproduced as choosing from the preprogrammed lists and menus. This time has passed. Now, it became clear that most of the decisions should be synthesized on line. It becomes increasingly clear that most of the planning procedures require searching. It was discovered that the advantages of search algorithms can be achieved when the space is represented and search is organized in a multiresolutional fashion. (See Meystel, 98).

Ability to Adjust Plans, Reschedule, and Re-plan

All job definitions interpretable by IS should be more abstract than would be given to an NIS. The command may encapsulate multiple individual actions, but it is the IS's business to figure that out. A mobile robot could be told to get the necessary signatures for a document. (This assumes that electronic signatures on the document are not an option.). The robot would have to understand which signatures are necessary (for example, if this is for a purchase, the purchase amount dictates what level of management needs to sign off), locate the individuals, interact with them to ask for their signature, and perform the intricate physical maneuvers necessary to present the document for signature. The individuals might not be in their office, hence the robot may have to search for them in alternative locations or try to arrange to meet them at some other time (re-scheduling). If someone is out of the office, the robot will have to decide whether to get the signature from someone else with equivalent signature authority or wait until the original person returns. Contrast this type of behavior with explicit instructions where the individuals and their locations are precisely given. If one of the individuals is not available, a non-intelligent robot would have to consult its human supervisor about how to proceed next.

The ability to adjust plans (re-plan) when the original ones are no longer valid is another crucial aspect that must be considered. It is one thing to create very elaborate plans to carry out a task (and the plans may even be derived from high level, abstract commands), but it is another matter to be able to deal with situations that are not as anticipated. Therefore, the intelligent system must be tolerant of changes as it is executing its plan and be able to react to the changes. In the bureaucratic robot introduced above, the change may occur if the vice president refuses to sign until he is given more information. The robot would then create another set of plans for itself to address the request, going to the originating individual to get background information or to the web to print out the specifications of the system being purchased along with alternatives that were not chosen. It would return to the vice president and present the information, and proceed to reintroduce the document to be signed. Obviously, all of this requires using appropriate architectures of knowledge representation, in particular, appropriate ontologies, as discussed in the subsequent sections.

3. Sensory Processing

Choosing the adequate set of sensors

The system receives signals from the real world through whatever sensors it may have. Note that a system may inhabit a software world, in which case "sensing" involves perceiving what exists external to itself, even if that is additional pieces of software. It must determine how to interpret the sensed signals in order to accomplish its tasks: the required actions are not prescribed in advance. Multiple sensors may be necessary and the system must be able to fuse information from them, collecting them into a registered, meaningful world model. Different sensors may give conflicting reports reports due to different interpretations of the world given their sensing modalities. Sensors may fail in certain circumstances or give insufficient information. The intelligent system should determine that it needs to utilize an additional or different sensor or process the signals it has differently. For example, it may be using a range sensor and a CCD camera as it navigates a house. It may hypothesize that instead of facing a wall or door, it may be confronted with a curtain hung in a doorway. In this case, it may need to apply additional or different processing algorithms in order to see if it can discern fabric (or something soft) from a planar, rigid surface. It may have to utilize a tactile sensor, if one is available.

Recognizing the unexpected

A system with intelligence (IS) ultimately must understand what its sensors are discerning. It must perform all of the requisite sensor or image processing to identify items in its environment to the level appropriate to the task. The requirements to processing will vary, depending on the situation and task. It may need to distinguish between certain types of tall weeds if it is an off-road vehicle, and it can drive only through certain leafy plants (not woody ones), or it would look unintelligent if it skirts around patches of tall grass. However, if it is a civilian car that should stay on roads, it probably doesn't need to identify what type of vegetation is growing on the side of the road, just that it is vegetation and not likely to jump out into the middle of the road. It will be directed by the behaviors to look for specific objects it may need in order to localize itself or find the object it is to act on. For example, it may look for a specific intersection as it navigates around a city or it may try to find a specific tool. The system's perception algorithms will have to be tolerant of a wide variation in the location and appearance of objects. Not all chairs look alike. A wrench may be on the floor or on a table, in a random position. Contrast this with non-intelligent systems that have limited tolerance for variations in their surroundings or in the objects with which they interact.

Dealing with unknown phenomena

The intelligent system will have to perceive entities and objects as it encounters them. It will classify and recognize items in its field(s) of view. It may classify a portion of the space in front of itself as a chair, or may have to deal with this as with an unknown object that might be interpreted as an obstacle. The sensory processing system, in conjunction with the world modeling system, must therefore know what it doesn't know about, and determine whether it needs to focus attention on the unknown in order to classify and identify. This ability to recognize the functional implications of unknown objects should be one of the major properties of IS. It is not impossible (in the future) to integrate multiple perceptions of an unknown object in various situations and eventually label it and deal with it as with a regular "known" object. Movements of unknown blobs can be interpreted with implication to possible planned maneuvers of the robot under consideration.

Multiresolutional Sensory Processing

The intelligent system will have to perceive entities and objects as it encounters them. However, sensory processing typically would require considering representation at multiple level of resolution. In all cases it provides for efficient computing. It is possible to demonstrate that this would correspond to the multiresolutional systems of knowledge representation (multiresolutional ontologies) and multiresolutional systems of decision making (multiresolutional planning) [Messina, 00].

4. World Modeling

Knowledge Representation

In most intelligent systems, an internal model of the world and/or a long-term knowledge store are utilized as a part of the overall knowledge representation system (KR). The long-term knowledge store (repository, or knowledge base) contains fairly invariant information, such as street maps or machining rules. An enabling aspect of the system's intelligence is the a priori knowledge it has and knows how to use. The internal model of the world is used to formulate a subset of KR that would allow the robot for planning expeditiously the required responses to the environment and situation. The sensory processes (discussed above) update and populate the current world model. The model might not be a single, monolithic one, but should rather comprise a set containing different types of information and/or different representations of perhaps the same information. The long-term knowledge may have to be merged with the in situ generated knowledge. For instance, the local sensors detect a road and some landmarks, such as buildings (using the knowledge base maps). The knowledge base supplies the name of the road, which is kept in the current world model.

The locally sensed information is obviously more current than that in the long-term store. Therefore, it must supercede what is in the knowledge base if there's a conflict. If a road has been closed, the system will plan around it and should, if appropriate, update the long-term maps. Obviously, these processes of updating our knowledge of the world belong to different levels of granularity, require different scale for interpretation and serve for supporting different resolutions of planning. It becomes a commonplace that most of intelligent systems either have or can be substantially improved by using multiresolutional systems of representation (including multiresolutional ontologies).

Multiple types of information

The intelligent system must be able to utilize a variety of types of information about the world in which it is functioning. If it is mobile, it must understand 2D or 3D space and have an adequate representation that enables it to move to the desired location efficiently while avoiding obstacles. It may need to take into consideration aspects beyond simple support surface (terrain or floor) geometry and obstacles. The type of terrain and traversability characteristics may be important as it determines which way it can go and how difficult it will be. So, for instance, if maintaining line-of-sight with a communications station may be necessary, the IS must be able to model the world so that it can perform the supporting computations to plan its movements.

Commonsense knowledge

An intelligent system should be able to have generic models available that guide it as it interacts with the world. This is as opposed to non-intelligent systems, where the environment is constrained to fit within the system's expectations (limited knowledge about *what is possible*). Although all possible situations cannot be predicted, the system should be prepared to many of them by a sub-store of *commonsense knowledge*. For example, the system may have to recognize and model stairs and elevators if it needs to go between floors. Not all stairs have the same geometry or configuration. It must know how elevators work, if that is appropriate to its job, namely, how to call an elevator, determine that one is available going in the right direction, selecting the floor, waiting until the right floor is reached and the door is open, etc. There is a general model of how to use an elevator, but there is tremendous variability in the actual elevator experience. The intelligent system has to be able to map between the generic and the specific.

Knowledge Acquisition: Updating, Extrapolating, and Learning

The updating of all sub-stores is conducted as the new information arrives. This information is frequently incomplete as far as satisfying the documents and models used by IS. An intelligent system must also be able to fill in gaps in its knowledge. If a moving object appears behind a robotic vehicle, the vehicle notes that it has an unknown entity that must be identified. Is it an emergency vehicle that must be given the right of way or an aggressive driver? It has to extrapolate or interpolate based on what it knows and what it discovers. All these knowledge acquisition activities require taking into account the uncertainty about what it does know. When driving down a road, if it is about to crest a hill, it cannot see the road beyond the hill. Rather than stopping, it should be able to assume that the road continues, and extrapolate based on the local geometry to forecast where the road exists even if it can't see it.

Related to this is the concept of predicting what will happen in the future. A machine tool that has a model of tool wear should forecast when a particular cutter will need to be replaced. A mobile vehicle will have to estimate its own trajectory and that of others with which it could potentially collide. The multiresolutional planning processes use various horizons of anticipation (larger at lower resolution and smaller at higher resolution.

The ability to anticipate will be amplified by learning new phenomena and control rules from experience. An intelligent system should become better at performing its job as it learns from its experiences. Therefore, one aspect that should be part of the testing or evaluation is the evolution and improvement in the system's functioning. The IS should have an internal measure of success as it performs its job. It can use the measure to evaluate how well a particular approach or strategy worked. Just as humans build expertise and become more efficient and effective at doing a certain job, the intelligent systems should have some means of improving their performance as well.

Requirements for Testing Intelligent Systems

Based on the discussion above, there is an initial set of requirements for testing intelligent systems that arise. The tests should therefore be designed to measure or identify at least the following abilities:

1. to interpret high level, abstract, and vague commands and convert them into a series of actionable plans
2. to autonomously make decisions as it is carrying out its plans
3. to re-plan while executing its plans and adapt to changes in the situation
4. to deal with imperfect sensors
5. to register sensed information with its location in the world and with a priori data
6. to fuse data from multiple sensors, including resolution of conflicts
7. to handle sensor failure or sensor inadequacy for certain circumstances
8. to direct its sensors and processing algorithms at finding and identifying specific items or items within a particular class
9. to focus resources where appropriate
10. to handle a wide variation in surroundings or objects with which it interacts
11. to deal with a dynamic environment
12. to map the environment so that it can perform its job
13. to update its models of the world, both for short-term and potentially long-term
14. to understand generic concepts about the world that are relevant to its functioning and ability to apply them to specific situations
15. to deal with and model symbolic and situational concepts as well as geometry and attributes
16. to work with incomplete and imperfect knowledge by extrapolating, interpolating, or other means
17. to be able to predict events in the future or estimate future status
18. the ability to evaluate its own performance and improve

Most of the items on the list allow for a numerical evaluation. However, non-numerical domains play a substantial role in evaluating intelligence and performance of IS.

5. Performance Evaluation in Non-numerical Domains

This theme focuses upon the aspects of intelligent system performance that are not directly quantifiable, but which should be subject to meaningful comparison. An example of an analogous aspect of human performance is the term "intelligent" itself. The notion of quantifying intelligence has always been controversial, even though people regularly use terms that ascribe some degree of intelligence. Terms ranging from smart, intelligent, or clever to dumb, stupid, or idiotic, with all sorts of degrees between, express people's judgments. But of course, these are often arbitrary judgments, without any basis for comparison or consistency of application. The notion of IQ, based on the widely used tests, was intended as a means of providing some consistency and quantification, but is still controversial.

So how might we do measurements for machines of the virtues that we associate with intelligence? First, we have to encapsulate the notion of what we mean by intelligence a little better. From the previous section one can see that the following properties are tacitly considered to pertain to intelligent systems:

- the ability to deal with general and abstract information
- the ability to deduce particular cases from the general ones
- the ability to deal with incomplete information and assume the lacking components

- the ability to construct autonomously the alternative of decisions
- the ability to compare these alternatives and choose the best one
- the ability to adjust the plans in updated situation
- the ability to reschedule and re-plan in updated situation
- the ability to choose the set of sensors
- the ability to recognize the unexpected as well as the previously unknown phenomena
- the ability to cluster, classify and categorize the acquired information
- the ability to update, extrapolate and learn
- being equipped with storages of supportive knowledge, in particular, commonsense knowledge

Then we need to find consistent measurements of what we consider to be the characteristics for each item on the list. We want these characteristics, like characteristics of software system performance quality in general, to provide us with goals to strive for in developing systems.

Ideally, the characteristics of value would be even more than engineering goals. They would be theoretical constructs in a “science of the artificial” [Simon, 69] – in this case, the science of Artificial Intelligence, or (being more specific) in the science of knowledge representation. As with other scientific fields, the constructs would be used in models (generally called scientific theories when they have been combined with a means of generating hypotheses and the hypotheses have been tested enough that the models are widely trusted). Some theoretical constructs may be easily judged from behavior of systems (“surface constructs”), but as in natural sciences, they might also be deeply hidden from view, within very complex models (“deep constructs” [see Reeker, 00]). In general, the depth of the construct is determined by the level of resolution accepted in a particular representation. In a multiresolutional system of knowledge representation, each level of resolution can be characterized by a particular “depth of the construct.” These phenomena find their implementation in Entity-Relational Networks of words that are organized in the multiresolutional hierarchies of ontologies [Meystel, 01].

From the standpoint of human cognition, the components of intelligence are hidden deeply in the models of Cognitive Science (an interdisciplinary part of Psychology, which is also a developing science). This is one reason that IQ is still controversial: The model that backs up the measures is not complete. But it has nevertheless been possible to endow IQ with some consistency that *ad hoc* descriptions do not have. This is because there is some consistency in measurement and some predictive value in terms of future human behavior. We would like this to be true for measures of intelligence in artificial systems, too, and it may turn out that we have a distinct advantage over the cognitive scientists. This advantage is that we can, so to speak “get into the heads” of intelligent artifacts more readily than we can with humans.

Ontologies and Reasons for Comparing Them in Intelligent Systems

How do we proceed to compare intelligent systems in these non-numerical areas? As a beginning, it is suggested that we look at what is the core of an intelligent system (maybe of a human as well as an intelligent computer program) – the way in which a system conceives of the world external to itself, the internal representation of *what is* and *what happens* in the world. This is what has come to be called an *ontology* in recent years. Ontologies are closely connected to a number of basic constructs that are highly relevant to the performance of an intelligent system. They are clearly of importance in planning, making decisions, learning, and communicating, as well as sensing and acting. An *ontology* is used in a computer program along with a *logic*. The “control” or dynamic aspects of that logic may embedded in the computer program itself, or it may be in a special program that manipulates a knowledge base of logical formulas, or a database manipulation system.

Whether an ontology is used within a computer program (or even the requirements statement of a planned computer program), a database (and its associated programs), a knowledge based system, or an autonomous artificially intelligent system, the ontology is indeed an informational core. As the architecture of the knowledge repository, the ontology (ontologies) are multigranular (multiresolutional, multiscale) in their essence because of multiresolutional character of the meaning of words [Rieger, 01]. In integrating systems, the presence of a shared ontology is what will allow interoperability. The term can be applied to the world-view of a human, too (in fact, is derived from a human study) though it may be easier to elicit it from the machine, as remarked above. (A fact related to the “knowledge acquisition bottleneck”.) Thus it is an aspect of intelligent behavior that we may be able to compare from one system to another and correlate with the more general notion of intelligence in a system.

Returning to the best attempts to date to measure human intelligence, it is worth noting that a human's individual ontology might be explanatory for human intelligence, so it is not surprising that there are indirect measures of ontologies on IQ tests and achievement tests. They may give us an idea as to how to proceed with this aspect of an intelligent system. To measure the breadth of the person's intelligence, is it useful to ask if some people have "broader" ontologies than others. That is, do they cover more areas, or more subjects, or more aspects, or more details. Should we expect that these broader ontologies will manifest themselves in, say, a scholastic aptitude test (which in turn correlates with IQ)? Does the "broader" ontology testifies for the *breadth* of intelligence? Would that broader ontology influence the ability of the intelligent system (including humans) to make better decisions? For people, the answers seem to be "yes". It is tempting to imply that for machines, as well.

Undoubtedly some people have ontologies that make more adequate, at least more accurate distinctions among different activities and objects that are present in the world (we can call this a "deeper" ontology"). That makes it possible for them to reason with more precision. In other words, the breadth and the depth of the ontology entails more powerful knowledge representation system. So the evaluation of ontologies is, to some extent at least, not unreasonable in gauging human cognitive performance. Is it a reasonable measure for machines? If so, how is the measure to be utilized? These are questions to be examined at PERMIS'2001.

A Human View of Ontology

In this subsection, we would like to push a view of the human ontology further, with the purpose of expanding the analogy to intelligent systems. For better understanding, it is presented in an anecdotal and subjective view, so it is worded in the first person.

As a human, I use my ontology (and actually, the whole system of my knowledge representation) to label, categorize, characterize, and compare everything -- every object, every action. If I learn something, it is because that label of this thing is put in my knowledge representation (KR) system, and eventually in my ontology. If I know about anything, it is because of its attributes, bounds, and relationships specified in the Entity-Relational Network (ERN) of my knowledge representation (KR) where the ontology resides. The more clearly specified things are in that ontology, the more I understand those things. I do not have to bring all of that understanding to my conscious attention all the time, as it would be a distraction. So, I access the ontology directly as I need to make the decision, or to communicate ideas and receive them from others. This allows me to note resemblances and make comparisons.

My Knowledge Representation (KR) system (to which the ontology is the part of) reflects reality to the extent that it helps me to deal with the world external to myself in a way that enables good decisions and accurate predictions. If it does not, I should be able to change it so that it better reflects reality, by learning that enriches the ERN of KR. That is one way one's theorymust depend on one's experiences. The experiences themselves depend on actions that I have taken, sensory information I have absorbed and communications I have received and understood. My ontology is therefore unique to me, since my experiences, and maybe the ways I learn vary from those of any other person. I discover new ideas and made new distinctions in ways that I do not fully understand and they become a part of my ERN ON KR nested one within another.

The relationship between the ontology and direct experiences of a sensory nature, coupled with activity and what it accomplishes is a part of the property called *grounding* which is a part of the process of *symbol grounding* [Harnad, 90]. When I learn language or learn the external world, this constantly extends my symbol grounding, since information might be conveyed that affects the ontology. There may be innate tendencies that provide symbol grounding, such as the fact that we can store information and access it and have a sense of sequence, but it is not our specific purpose to inquire about these.

The rational interpretation of things communicated to me (or discovered by me) is affected by and affects my ontology. I may encounter "raw" pains, perceptions, and emotions that are not fully understood, but even these may be refined and contextualized by my ON. If I am to successfully communicate to others, I must encode, in a shared language, things that are in my ontology and shared to at least some degree in the ontologies of those to whom I am communicating. Questions, context, and conversations help to facilitate this sharing.

Decisions that lead to a high probability of success in dealing with the external world can only be made in the light of my KR-based understanding of the facts surrounding the decision. If I do not have alternative actions characterized in my ontology, I cannot compare them, and therefore cannot consider them in rational decision processes. If my ontology does not reflect reality, I will make irrational and perhaps unsuccessful decisions. Complex decisions involve problem solving, and I must be able to access methods for solving problems.

The issue of such methods as part of ontologies is developed more deeply in a paper authored by Chandrasekaran, Josephson, and Benjamins [see Chandrasekaran, 99]. There it is pointed out that a decision-making system requires both a subject matter ontology and a problem solving method ontology. It is possible – and may be needed - to imagine even a larger ontology of activities.

If I am to learn, I generally have to use my ontology in the learning process. Maybe natural linking mechanisms in sensory processes can be brought to bear in certain learning tasks, so I can learn a path through the woods or a list of words in seemingly built-in ways. This *rote learning* I can improve upon by relating items within my ontology. If I am to classify items, I need to do so based on attributes, which are in my ontology. If I am to search my memory, I need to do so based on shared attributes, related activities, and other sorts of relationships. If I am to learn by reinforcement, I need to associate the reinforcements with actions, objects, features, bounds, and relationships. If I am going to transfer learning from one task to another, I need to use my ontology to find mappings from one action or object to another.

Objects in my ontology can be composed of other objects. An action may involve many objects (with their attributes, bounds and relationships) and other actions that somehow get “hooked together”. An object may be defined by attributes that include defining actions.

Measuring Non-Numerical Aspects of Intelligent Systems Related to Ontologies

Can we exploit the idea of the human ontology above as a “core” of intelligence to characterize and compare intelligent behavior is machines based on a machine’s ontology, built-in of acquired? Like a human, a machine may have sensors connected to subsystems of sensory processing. The machine may be able to take certain actions that provide grounding for the ontology. If it can learn, perhaps it can extend its ontology. How can we characterize that ontology in a way that will allow us to characterize the machine’s capabilities? How can we characterize its ability to change the ontology? If it has an ability to communicate to other machines or people, how does this ability add to its capabilities (and to its ontology)? These are some of the ideas to be explored in PERMIS’2001.

6. Evaluation: Mathematical and Computational Premises

Consider a general situation: there is a set of goals (G_1,…, G_n) and a set of IS (or intelligent agents) to achieve these goals. Different intelligent systems, or agents might have different goals, or they might put different weights on the various goals. Further, they might be better or poorer at pursuing those goals in differing contexts. That is, they might have different components of intelligence (I_1, I_2,..I_s) and these would be more or less important in the different contexts (C_1,…, C_q) that should also be known.

This dependence on the context determines that agents might be good at one set of matters, but bad in others. The agent might be good at trying and learning about recognizing new objects in the surrounding world, but poor at doing anything risky. It is typical for humans to have a portfolio of "intelligences" as well as “goals.” It would give some value to all the different goals, and would have some value to each dimension of intelligence. One agent might be characterized as an explorer, while another is very good in performing repetitive routines. Which agent should be evaluated as a preferable one? Obviously, this would depend on the goal and the context. An unequivocal answer might be impossible at a single level of resolution because the true result depends on the distribution of the types of agents and the contexts that the groups of agents find themselves in. Thus, the “intelligences” as well as “goals” might require representing them as a multiresolutional system.

Multiresolutional Vector of Intelligence (MVI)

What should be measured to evaluate intelligence? The Multiresolutional Vector of Intelligence (MVI), and the level of success of the system functioning when this success is attributed to the intelligence of the system. The need to construct a MVI and determine their success emerges in many areas. It is not clear whether “success” is (or should be) correlated with “reward” and “punishment.”

What constitutes the appropriate scope and levels of details in an ontology is practically driven by the purpose of the ontology. The ability to dynamically assume one level of detail among many possible details is important for an intelligent system. It might depend on the purpose of a system. In that sense the long term purpose of the system is different from its short term or middle term goals. Clearly, the long term purpose and the multiple term goals are goals belonging to different levels of resolution and should be treated in this way. This brings us back to the measures of intelligence through success: is intelligence to be measured by the ability of a system to succeed in carrying out its goals? Can the highly successful functioning at one level of resolution co-exist with the lack of success at another? Are the “successes “nested” or independend one from another?

Evaluation of intelligence requires our ability to judge the degree of *success* in a multiresolutional system of multiple intelligences working under multiple goals This means that if success is defined as producing a summary of the situation (a generalized representation of it), the latter can be computed in a very non-intelligent manner especially if one is dealing with a relatively simple situation. Indeed, in primitive cases, the user might be satisfied by composing a summary defined as “list the objects and relationships among them” i.e. a subset of an entity-relational network (ERN). On the other hand, the summary can be produced intelligently by generalizing the list of objects and relationships to the required degree of quantitative compression with the required level of the context related *coherence*. Thus, *success* characterizes the level of *intelligence* if the notion of *success* is clearly defined.

The need in determining levels or gradations of intelligence is obvious: we must understand why the probability of success increases because somebody is supposed to provide for this increase, and somebody is supposed to pay for it. This is the primary goal of our effort in developing the metrics for intelligence. The problem is that we do not know yet is the basis for these gradations and are not too active in fighting this ignorance. What are these gradations, how should they be organized, what are their parameters that should be taken in account? We can introduce parameters such that each of the parameters affects the process of problem solving and serves to characterize the faculty of intelligence at the same time.

Multiresolutional Architecture of Ontology is a part of the Multiresolutional Vector of Intelligence. The following list of 25 items should be considered an example of the set of coordinates for a possible Multiresolutional Vector of Intelligence (MVI):

(a) memory temporal depth
(b) number of objects that can be stored (number of information units that can be handled)
(c) number of levels of granularity in the system of representation
(d) the vicinity of associative links taken in account during reasoning of a situation, or
(e) the density of associative links that can be measured by the average number of ER-links related to a particular object, or
(f) the vicinity of the object in which the linkages are assigned and stored (associative depth)
(g) the diameter of associations ball (circle)

The association depth does not necessarily work positively, to the advantage of the system. It can be detrimental for the system because if the number of associative links is excessively large the speed of problem solving can be substantially reduced. Thus, a new parameter can be introduced

(h) the ability to assign the optimum depth of associations

This is one more example of recognition that should be performed, in this case, within the knowledge representation system. Obviously, the ability “h” is tightly linked with the ability of IS to deal with incomplete commands and descriptions (see Section 1).

Functioning of the behavior generation module, for example, evokes additional parameters, properties and features:

(i) the horizon of extrapolation, and the horizon of planning at each level of resolution

(j) the response time
(This factor should not be confused with a horizon of prediction, or forecasting which should combine both planning and extrapolation of recognized tendencies).
(k) the size of the spatial scope of attention
(This corresponds to the vicinity of the associative links pertinent to the situation in the system of knowledge representation)
(l) properties and limitations of the aggregation and decomposition of conceptual units.

The latter would characterize the ability to synthesize alternatives of decisions and choosing one of them (see Section 1).

The following parameters of interest can be tentatively listed for the sensory processing module:
(m) the depth of details taken in account during the processes of recognition at a single level of resolution
(n) the number of levels of resolution that should be taken into account during the processes of recognition
(o) the ratio between the scales of adjacent and consecutive levels of resolution
(p) the size of the scope in the most rough scale and the minimum distinguishable unit in the most accurate (high resolution) scale

It might happen that recognition at a single level of resolution is more efficient computationally than if several levels of resolution are involved. A more fine system of *inner* multiple levels of resolution can be introduced at a particular level of resolution assigned for the overall system. The latter case is similar to the case of unnecessarily increasing the number of associative links during the organization of knowledge.

Spatio-temporal horizons in knowledge organization as well as behavior generation are supposed to be linked with spatio-temporal scopes admitted for running algorithms of generalization (e.g. clustering). Indeed, we do not cluster the whole world but only the subset of it which falls within our scope. This joint dependence of clustering on both spatial relations and the expectation of their temporal existence can lead to non-trivial results.

One should not forget that generalization (the ability to come up with a “gestalt” concept) is conducted by recognizing an object within the chaos of available spatio-temporal information, or a more general object within the multiplicity of less general ones. The system has to recognize such a representative object, event, or action if they are *entities*. If the scope of attention is too small, the system might not be able to recognize the entity that has boundaries beyond the scope of attention. However, if the scope is excessively large, then the system will perform a substantial and unnecessary job (of searching and tentatively grouping units of information with weak links to the units of importance).

Thus, any system should choose the value of the horizon of generalization (that is the scope of the procedure of *focusing of attention*) at each level of resolution (granularity, or scale).

All of these parameters characterize the realities of the world and the mechanisms of modeling that we apply to this world. These parameters do not affect the user’s specifications of the problem to be solved in this system. The problem is usually formulated in the terms of hereditary modeling that might not coincide with the optimum modeling, or with the parameters of modeling accepted in the standard toolbox of a decision-maker.

The problem formulated by a user often presumes a particular history of the evolution of variables available for the needs of the intelligent system. Simultaneously, the user requests a particular spatio-temporal zone within which the solution of the problem is desirable. However, the input specifications often do not require a particular decomposition of the system into resolution levels and the intelligent system of CSA is free to select it in an “optimal” way. In other cases, the user comes up with already existing decomposition of the system that appeared historically and must not be changed (like the organizational hierarchy of a company and/or an Army unit). Sometimes, it is beneficial to combine both existing realistic resolution levels and the “optimal” resolution levels implied by the optimum problem solving processes.

The discrepancy between these decompositions requires a new parameter of intelligence

(q) an ability of problem solving intelligence to adjust its multi-scale organization to the hereditary hierarchy of the system, this property can be called "a flexibility of intelligence"; this property characterizes the ability of the system focus its resources around proper domains of information.

In the list of specifications of the problem the important parameters are

(r) dimensionality of the problem (the number of variables to be taken in account)
(s) accuracy of the variables
(t) coherence of the representation constructed upon these variables
For the part of the problem related to maintenance of the symbolic system, it is important to watch the
(u) limit on the quantity of texts available for the problem solver for extracting description of the system
and this is equally applicable for the cases where the problem is supposed to be solved either by a system developer, or by the intelligent system during its functioning.
(v) frequency of sampling and the dimensionality of the vector of sampling.

Most of the input knowledge arrives in the form of stories about the situation. These stories are organized as a narrative and can be considered *texts.* In engineering practice, the significance of the narrative is frequently (traditionally) discarded. Problem solvers use knowledge that has been already extracted from the text. How? Typically, this issue is never addressed. Now, the existing tools of text processing allow us to address this issue systematically and with a help of the computer tools of text processing

Finally, the user might have its vision of the cost-functions of his interest. This vision can be different from the vision of the problem solver. Usually, the problem solver will add to the user's cost-function of the system an additional cost-function that would characterize the time and/or complexity of computations, and eventually the cost of solving the problem. Thus, additional parameters:

(w) cost-functions (cost-functionals)
(x) constraints upon all parameters
(y) cost-function of solving the problem

This contains many structural measures. We need to trace back from an externally perceived measure of "success" or intelligence to a structural requirement. E.g, the construction codes specify thickness of structural members, but these dimensions are related to the amount of weight to support – the performance goal is the lack of building collapse.

Important properties of the Intelligent Systems are their ability to learn from the available information about the system to be analyzed. This ability is determined by the ability to recognize regularities and irregularities within the available information. Both regularities and irregularities are transformed afterwards into the new units of information. The spatio-temporal horizons of Intelligent Systems turn out to be critical for these processes of recognition and learning.

Metrics for intelligence are expected to integrate all of these parameters of intelligence in a comprehensive and quantitatively applicable form. Now, the set $\{VI_{ij}\}$ would allow us even to require a particular target vector of intelligence $\{VI_T\}$ and find the mapping $\{VI_T\} \rightarrow \{VI_{ij}\}$ and eventually, to raise an issue of design: how to construct an intelligent machine that will provide for a minimum cost (C) mapping

$$[\{VP_T\} \rightarrow \{VI_{ij}\}] \rightarrow \min C.$$

By the way, has this ever been done for the systems that are genuinely intelligent? Of course, this question is not related to design, just to measurement.

The Tools of Mathematics

The following areas of mathematics should be considered belonging

The following tools are known from the literature as proven theoretical and practical carriers of the properties of intelligence:

- Using Automata as a Generalized Model for Analysis, Design, and Control
- Applying Multiresolutional (Multiscale, Multigranular) Approach
  1. Resolution, Scale, Granulation: Methods of Interval Mathematics
  2. Grouping: Classification, Clustering, Aggregation
  3. Focusing of Attention
  4. Combinatorial Search
  5. Generalization
  6. Instantiation
- Reducing Computational Complexity
- Dealing with Uncertainty by
  1. *Implanted compensation at a level (feedback controller)*
  2. *Using Nested Fuzzy Models with multiscale error representation*
- Equipping the System with Knowledge Representation
- Learning and Reasoning Upon Representation
- Using bio-neuro-morphic methodologies
- General Properties of Reasoning

  —Quantitative as well as qualitative reasoning
  —Generation of limited suggestions, as well as temporal reasoning
  —Construction both direct and indirect chaining tautologies (inferences)
  —Employing non-monotonic as well as monotonic reasoning
  —Inferencing both from direct experiences as well as by analogy, and
  —Utilizing both certain as well as plausible reasoning in the form of

    1. *Qualitative Reasoning*
    2. *Theorem Proving*
    3. *Temporal Reasoning*
    4. *Nonmonotonic Reasoning*
    5. *Probabilistic Inference*
    6. *Possibilistic Inference*
    7. *Analogical Inference*
    8. *Plausible Reasoning: Abduction, Evidential Reasoning*
    9. *Neural, Fuzzy, and Neuro-Fuzzy Inferences*
    10. *Embedded Functions of an Agent: Comparison and Selection*

Each of the tools mentioned in the list allows for a number of comprehensive embodiments by using standard or advanced software and hardware modules. Thus a possibility of constructing a language of architectural modules can be considered for future efforts in this direction.

The Tools of Computational Intelligence

Proper testing procedures should be associated with the model of intelligence presumed in the particular case of intelligence evaluation. It seems to be meaningful to compare systems of intelligence that are equipped with similar tools. In this section we introduce the list of the tools that are known from the common industrial and research practice of running the systems with elements of autonomy and intelligence. It is also expected that these tools can be used as components of the intelligent systems architectures. Thus, they might help in developing and applying types of architectures that will be used for comparing intelligence of systems.

Learning.

We have separated this into an independent sub-section because of the synthetic nature of the matter. Learning is the underlying essence of all phenomena linked with functioning of an intelligent system. It uses all mathematical and computational tools outlined for all other subsystems. In the machine learning community, the attention is paid to three metrics: the ability to generalize, the performance level in the specific task being learned, and the speed of learning. From the intelligence point of view, the ability to generalize is the most important since the other two capabilities dwell on the ability to generalize. Systems can do rote learning, but without generalization, it is impossible, or at least very difficult to apply what has been learned to future situations. Of course, if two systems were equivalent in their ability to generalize, with the same resulting level of performance, then the one which could do this faster would be better.

7. References

[Butler, 93] Butler, R.W., and Finelli, G.B. The infeasibility of quantifying the reliability of life-critical real-time software. IEEE Trans. Softw. Eng 19, 1 (Jan. 1993), 3–12.

[Chandrasekaran, 98] Chandrasekaran, B., J. R. Josephson and V. R. Benjamins, "Ontology of Tasks and Methods," 1998 Banff Knowledge Acquisition Workshop. [Revised Version is cited above. It appears as two papers "What are ontologies and why do we need them?," *IEEE Intelligent Systems*, Jan/Feb 1999, 14(1); pp. 20-26; "Ontology of Task and Methods," *IEEE Intelligent Systems*, May/June, 1999.]

[Finklestein, 00] Finkelstein, R., "A Method for Evaluation the "IQ" of Intelligent Systems," Proceedings of the 2000 Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD, August, 2000.

[Harnad, 90] Harnad, S., "The Symbol Grounding Problem," *Physica* D, 42, 1990, p.335-46

[Meystel, 00] A. Meystel, Messina, E., "The Challenges of Intelligent Systems," Proc. of the IEEE Int'l Symposium on Intelligent Control, Patras, Greece, 2000

[Meystel, 98] Meystel, A., "Robot Path Planning," *The Encyclopedia of Electrical and Electronic Engineers*, Wiley, New York, 1998

[Meystel, 01] Meystel, A., "Principles of Word-based Knowledge Representation and Knowledge Processing," in ed. P. Wang, *Computing with Words*, Wiley, 2001, pp. 329-346

[Reeker, 00] Reeker, L., "Theoretical Constructs and Measurement of Performance and Intelligence in Intelligent Systems," Proceedings of the Workshop on Performance Metrics for Intelligent Systems, August 14-16, 2000, NIST

[Rieger, 01] Rieger, B. B., "Computing Granular Word Meaning: A Fuzzy Linguistic Approach in Computational Semiotics," in ed. P. Wang, *Computing with Words*, Wiley, 2001, pp. 147-208

[Simon, 69] Simon, H. A. *The Sciences of the Artificial*. Third edition, Cambridge, MA, MIT Press, 1996. [First edition published 1969].

[Tsai, 99] Tsai, W., Vishnuvajjala, R.; Du Zhang, Verification and validation of knowledge-based systems, IEEE Transactions on Knowledge and Data Engineering 11, 1(Jan.Feb., 1999), 202-212.