APPENDIX A

THE PRELIMINARY DISCUSSIONS

In this part, the excerpts from the Preliminary Discussion of the Advisory Board Members are given. Several months of active exchange helped to clarify many issues that precipitated into the Workshop Agenda and its panel discussions. The discussion was conducted by Alex Meystel.

DEFINING INTELLIGENCE

LET US CLARIFY WHAT THIS PHENOMENON IS

Dear Advisory Board Member,

As a working definition of intelligence, we use the following statement (proposed in 1991 by one of the Advisory Board members):

"Intelligence is the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system's ultimate goal.

Another member of the Advisory board stated:

"We regard as "intelligent system with autonomy" only a system that can function in a self sustained manner, i.e. has information from the World of what is going on, updates its representation of the World, checks it with the goal, evaluates the situation, develops behavior that is appropriate in this situation and executes (actuates) this behavior, and again, receives the information from the world, and so on."

Finally, an opinion was voiced by the third member of the board that:

"The intelligence is incorporated in the mechanisms of inferring decisions and/or self-generating new rules from existing ones in combination with external data. These properties might exist as a potentiality, they should not be associated with really successful functioning."

In other words, we have three platforms proposed about intelligence:

No. 1: success in achieving goal means "intelligence"

No. 2: self-sustaining functioning is "intelligence"

No. 3: abilities to infer and learn mean "intelligence," nothing else matters!

What do you think about this trichotomy?

A. Meystel

NOT EVERYTHING IS ADDRESSED IN THE WHITE PAPER…

1. You didn't come anywhere near covering the spectrum of philosophical views of intelligence (just start to read the mind/body literature!) I would scale back your analogies to human intelligence and testing to something more pragmatic. Your a-y classification of measurable characteristics goes in the right direction but seems too constrained by existing systems and ways of representing information.

2. You don't talk much about learning which is a critical characteristic of intelligence. It's there, but it's primarily implicit.

3. I would define intelligence relative to a domain of application. Even in the human cases there are people who are "car intelligent" but "literature ignorant" - different domains, different abilities. Also in the human domain you have different types of intelligence (Gardner's 7, Sternberg's 3, etc.) - do you want to try something similar in the autonomous systems?

4. What's the goal of these metrics? Are they do be used in a TREC type environment?

John Cherniavsky, NSF March 16, 2000

INTELLIGENCE AND THE REQUIREMENT OF BEING SELF SUSTAINED

Interestingly enough, we regard as "intelligent system with autonomy" only a system that can function in a self sustained manner, i.e. has information from the World of what is going on, updates its representation of the World, checks it with the goal, evaluates the situation, develops behavior that is appropriate in this situation and executes (actuates) this behavior, and again, receives the information from the world, and so on.

No matter whether this is a human, a robot, a manufacturing system, an e-commerce system - it is a full cycle of activities oriented toward being "self sustained".

I wonder whether a behavior can be regarded as "intelligent" without this component of being used by a "self sustained" creature.

Maybe, this is why some of our intelligence-oriented endeavors fail: we consider just a component and do not take in account the whole creature? If this is correct, one should be suspicious of any set of criteria that are not derived from the goals and means of the overall functioning.
Would you agree with this? (I hope, I am not intrusive!)

From the letter by Moderator

These are difficult and controversial questions. Which is why I like limited tests such as TREC or the DARPA speech understanding. Other researchers (Gelernter for example) dismiss the notion of intelligence (human that is) as nonsensical in machines and would feel the search for intelligence metrics as doomed from the beginning.

John Cherniavsky March 20, 2000

… These are difficult and controversial questions. Which is why I like limited tests such as TREC or the DARPA speech understanding. Other researchers (Gelernter, for example) dismiss the notion of intelligence (human that is) as nonsensical in machines and would feel the search for intelligence metrics as doomed from the beginning.

John Cherniavsky April 3, 2000

KNOWLEDGE IS NOT SUFFICIENT TO QUALIFY FOR INTELLIGENCE

Alex,

I'm a big fan of Allen Newell's definition of intelligence (see "Unified Theories of Intelligence"), which is essentially that the intelligence of a system/agent is its ability to use its available knowledge to select appropriate actions to achieve its goals.

Available knowledge is important because a system that does not have knowledge available is just ignorant, whereas a system that has knowledge but doesn't use it is stupid (not intelligent). This also provides an entry for learning because available knowledge can be construed as what should be expected to be learned from experiences with the world.

Selecting actions is critical because intelligence without action is meaningless.

Goals are critical because action without purpose is also meaningless.

There is another dimension as to the generality of the intelligence of a system/agent based on the breadth of knowledge it can acquire, encode, and use, the breadth of actions it has available, and the breadth of goals it can attempt. All of these are related to the types of environments that the agent will be successful in.

This definition fits pretty well with 1, although the emphasis in the above definition is on selecting actions that the system thinks will achieve its goals (and not on some probability). I also think that adding uncertain environments in unnecessary. Chess has no uncertainty and requires intelligence.

Autonomy is probably another dimension, but is related to the generality of intelligence - what environments the system/agent can be successful in.

Hope this helps. These are interesting issues.

John Laird April 3, 2000

FINDING THE UNIFIED TECHNIQUE WILL INCREASE EFFICIENCY OF TESTING

The problem with all these contests is that the measures are very task oriented and thus, specialized. Each system is approached individually and its individual performance is measured.

Our intention is to standardize the measures of performance of intelligent systems so that one could judge the level of intelligence of the system separately from its present concrete application. It is not as far-fetched as it might seem. Indeed, the problem of software reuse will release huge amounts of funds because we will stop developing "new" pieces of software just because the application is different. (And this is just one of many justifications for standardizing the measures of performance of intelligent systems.

Look, if we determine that the success of functioning depends on a particular set of abilities: for example, the ability to search in the large set of data, or the ability to generalize upon action rules, or upon object descriptions, or depends on other abilities, then, we will be able to recommend a particular intelligent control system, or a particular pattern discovery system, or another system which contributes to the overall intelligence, for a variety of applications, where these abilities are critical.

Consider the examples that you've mentioned in your letter: data mining software, information retrieval software, speech understanding software. Obviously, they contain reusable sub-systems. These subsystems cluster information, search for patterns, recognize the anticipated pattern, and perform several more typical tasks. These typical tasks performance could be evaluated in terms of objective measures of the particular abilities (that are components of their intelligence and determine the level of their intelligence).

But we do not know anything about their level of intelligence in terms of objective measures of the set of their abilities. All we know is how this or that software package was working with a concrete task assignment, and this does not allow to say anything how good it might be for another task assignment. Its subsystems for control, recognition, etc., might be quite dumb and I should avoid their reuse. Or they might be very powerful and I should look forward to their reuse!

As you've suggested earlier, I studied TREC results as much as they are available. This is a great work. But it does not allow me to make a judgment of how much intelligence these systems rely upon and/or demonstrate. Maybe, they are very powerful, and I can use them for dealing with large knowledge bases of the intelligent mobile vehicles, or for information processing in the autonomous vehicle computer vision system. However, I do not have this information and must invent these systems anew.

Maybe, my questions about defining intelligence are more pragmatic in their essence than it seems. Maybe, by asking you I am asking a right person. Maybe, your vision and experience will be extremely helpful in PROPER FORMULATION of these really new problems.

Alex Meystel

FOLLOWING THE BIOLOGICAL MODELS MIGHT BE HELPFUL

[From the very beginning of this discussion, clear division of participants in several clusters affected the style and the first results of the exchange. Participants belonging to different clusters had different initial premises of "how are we supposed to approach our thinking about intelligence." The groups belonging to the "different schools of thought" were using slightly different vocabularies and probably different underlying models for describing functioning of "intelligences." Two very prominent groups can be mentioned including:

a) researchers thinking about intelligence in the terms of biological models where processes of evolution meant to be a component, and another – thinking in terms of computational intelligence

The question was:

In other words, we have three platforms about intelligence:

No. 1: success in achieving goal means "intelligence"

No. 2: self-sustaining functioning is "intelligence"

No. 3: abilities to infer and learn mean "intelligence," nothing else matters!

What do you think about this trichotomy?

Model 3 is the weakest - doesn't distinguish intelligence from the performance of any existing high-grade adaptive control system.

Model 1 is better - but it doesn't specify that a system must be enabled to create its own subgoals in the context of the ultimate goal prescribed by the agent that built it and released it, and to evaluate 'appropriateness' and its own 'success' by criteria of its own design. These functions in biointelligence are subsumed under intentionality. An intelligent device must have this.

Model 2 is best of the 3 - incorporates the action-perception cycle that characterizes biological systems, which is the mechanism of intentional action, but fails to address the complementary property of assimilation, by which organisms construct and maintain a fully integrated life-long store of information through learning through actions into the World, or the mechanisms of reafference by which biosystems determine the information that is to be taken from the World, as the basis for making their decisions.

Walter Freeman March 31, 2000

INVARIANCE OF INTELLIGENCE

DEFINING SOMETHING MIGHT BE A POWERFUL THING (SOMETIMES)

May 16, 2000

Dear Dr. Kanayama:

1. Let us try to discover what is the substance of our argument; then, we will try to address it. You quoted my understanding of the problem: defining the intelligence of the system so that you could judge how it affects functioning of the vehicles? Then, you are trying to explain that you do not see the problem of finding how the intelligence affects functioning, but rather you see the problem in making these systems function well by equipping them with a custom made intelligence.

Please, understand that the problem that I have formulated is not INSTEAD of the problem that you are solving, it is a DIFFERENT problem. The solution of architectural problem should make easier searching for a solution of your problem. You will see it from positions 2 and 3 of my letter.

This is what you wrote:

First I want to define the problem I work on. Only after then, you are able to evaluate the performance of the system as a problem solver. Possible problems for groups of autonomous unmanned vehicles could be: playing soccer, playing football, clearing a land-mine field, clearing a devastated city area by a tornado, chasing a fleeing prisoner, standing-off against a criminal with hostages, driving themselves in a row in a highway, placing themselves in a museum to watch if someone hurts the masterpieces, serving people in a reception with drinks and hors d'oeuvres, line-dancing, ballet-dancing, fighting against an enemy, and so forth.

A specific team of autonomous unmanned vehicles may be good at one of the problems, but may not be good at another.

I would like to compliment you on an impressive list of possible intelligent robots that will surround us very soon.

2. Now, let me ask you two questions:

a) Can you see that all of these proficient robots (each in its domain) will have something in common architecturally?

b) Can you imagine that one might be interested in selling the box of "system's intelligence" to all your companies that manufacture these skillful, agile, and cunning creatures?

The whole issue is hidden in your belief that any system must be its own specifications oriented, while I believe that the system's intelligence is beyond particular specifications.

It is the INVARIANCE for all these intelligent robots that can be manufactured separately, have capabilities of all of the above robots, will be more reliable and cost less.

3. I would like to remind you one important thing. About 40 years ago, the systems for control and automation of metalcutting machines were designed and manufactured individually. Depending on the skills that a particular machine was supposed to demonstrate, we were used to design very sophisticated electrical schemata, and the machines were successfully functioning.

Then, people realized that all of these machines could be controlled from the same stereotypical "intelligence" and this is how the CNC systems emerged. Programmable controllers turned out to be another solution for the problem. Now, each complicated automated machine is controlled from the same PC computer that embodies its intelligence. The technology develops as usual: from individualized custom-made solutions to a typical architecture.

The same will happen with intelligent systems as soon as we understand the nature of intelligence better.

Moderator

May 25, 2000

WHICH SIDE OF THE ARGUMENT...

Dear Advisory Board Members:

I found this question in the recent mail:

Prof,

The notion that rocks have consciousness is just as counterintuitive as your and Albus' notion that a thermostat is intelligent.

Which side of this argument are you on?

Cheers, Mike

Help me to answer this question. As a moderator, I would suggest to read Hans Moravec's letter about John Searle's review of Ray Kurzweil, April 8, 1999 (see http://www.frc.ri.cmu.edu/~hpm/project.archive/general.articles/1999/NYRB.990325.html).

Excerpts from this letter are given below.

Which side are you on, indeed?

Letter re. John Searle's review of Ray Kurzweil, April 8, 1999

Subject: Re: "I Married a Computer" by John R. Searle, April 8, 1999

To the Editor, New York Review of Books:

In the April 8 NYRB review of Raymond Kurzweil's new book, John Searle once again trots out his hoary "Chinese Room" argument. So doing, he illuminates a chasm between certain intuitions in traditional western Philosophy of Mind and conflicting understandings emerging from the new Sciences of Mind.

Searle's argument imagines a human who blindly follows cleverly contrived rote rules to conduct an intelligent conversation without actually understanding a word of it. To Searle the scenario illustrates machine that exhibits understanding without actually having it. To computer scientists the argument merely shows Searle is looking for understanding in the wrong places. It would take a human maybe 50,000 years of rote work and billions of scratch notes to generate each second of genuinely intelligent conversation by this means, working as a cog in a vast paper machine. The understanding the machine exhibits would obviously not be encoded in the usual places in the human's brain, as Searle would have it, but rather in the changing pattern of symbols in that paper mountain.

Searle seemingly cannot accept that real meaning can exist in mere patterns. But such attributions are essential to computer scientists and mathematicians, who daily work with mappings between different physical and symbolic structures. One day a computer memory pattern means a number, another it is a string of text or a snippet of sound or a patch of picture. When running a weather simulation it may be a pressure or a humidity, and in a robot program it may be a belief, a goal, a feeling or a state of alertness. Cognitive biologists, too, think this way as they accumulate evidence that sensations, feelings, beliefs, thoughts and other elements of consciousness are encoded as distributed patterns of activity in the nervous system. Scientifically-oriented philosophers like Daniel Dennett have built plausible theories of consciousness on the approach.

Searle is partway there in his discussion of extrinsic and intrinsic qualities, but fails to take a few additional steps that would make the situation much clearer, but reverse his conclusion. It is true that any machine can be viewed in a "mechanical" way, in terms of the interaction of its component parts. But also, as Alan Turing proposed and Searle acknowledges, a machine able to conduct an insightful

conversation, or otherwise interact in a genuinely humanlike fashion, can usefully be viewed in a "psychological" way, wherein an observer attributes thoughts, feelings, understanding and consciousness. Searle claims such attributions to a machine are merely extrinsic, and not also intrinsic as in human beings, and suggests idiosyncratically that intrinsic feelings exude in some mysterious and undefined way from the unique physical substance of human brains.

Consider an alternative explanation for intrinsic experience. Among the psychological attributes we extrinsically attribute to people is the ability to make attributions. But with the ability to make attributions, an entity can attribute beliefs, feelings and consciousness to itself, independent of outside observers' attributions! Self-attribution is the crowning flourish gives properly constituted cognitive mechanisms, biological or electronic, an intrinsic life in their own mind's eyes. So abstract a cause for intrinsic experience may be unpalatable to classically materialist thinkers like Searle, but it feels quite natural to computer scientists. It is also supported by biological observations linking particular patterns of brain activity with subjective mental states, and is a part of Dennett's and others' theories of consciousness.

Elsewhere Hilary Putnam and Searle independently offered another kind of objection. If real thoughts, feelings, meaning and consciousness are found in special interpretations of the activity patterns of human or robot brains, wouldn't there also be interpretations that find consciousness in less traditional places, for instance (to use their examples), in the patterns of particle motion of arbitrary rocks or blackboards? Putnam, once a champion of the interpretive position, found this implication impossibly counterintuitive, and turned his back on the whole logical chain. To Searle, it simply bolsters his preexisting opinion. But counterintuitive implications do not refute an idea. The interpretations required in Putnam's and Searle's examples are too complex for us to actually muster, putting the implied beings out of our interpretive reach, thus unable to affect our everyday experience. The last chapter of my recent book "Robot: Mere Machine to Transcendent Mind" explores further implications, and uncovers no self-contradictions nor contradictions with reality as we know it. Rather, the interpretive position sheds light on mysteries like the unexpected simplicity of basic physical law. It does predict many surprises beyond our immediate observational horizons, and offends common metaphysical assumptions. But today, when millions of 3D videogame players immerse themselves in increasingly expansive and populated worlds found in very special interpretations of the particle motions of a few unimpressive-looking silicon chips, is the idea of whole worlds hidden in unexpected places still beyond the pale?
Hans Moravec, January 7, 1999

May 26, 2000

A METHODOLOGY OF MEASURING THE PERFORMANCE OF INTELLIGENT SYSTEMS

Dear Advisory Board Member:

Everybody would benefit from the insights into the problem of Performance Measures gracefully submitted to us by Dr. Larry Reeker from NIST (a Member of our Advisory Board). Interestingly enough, the recommended techniques of measuring the performance could be applied to testing most of the Intelligent Systems with the elements of Autonomy.

This is what Larry wrote in his letter:

I thought you might be interested in Teasuro's discussion of evaluation of backgammon play. I remembered he had done some, and just ran into it as I reread his paper for an entirely different reason. If you are interested in the paper, you can read it at

http://www.research.ibm.com/massive/tdl.html#h1:temporal_difference_learning

It reflects three methods that have wider applicability: contests against other programs (particularly benchmark programs), contests against humans (coupled with subjective evaluation), simulation of the outcome of decisions made.

PERFORMANCE MEASURES

There is a number of methods available to assess the quality of play of a backgammon program; each of these methods has different strengths and weaknesses. One method is automated play against a benchmark computer opponent. If the two programs can be interfaced directly to each other, and if the programs play quickly enough, then many thousands of games can be played and accurate statistics can be obtained as to how often each side wins. A higher score against the benchmark opponent can be interpreted as an overall stronger level of play. While this method is accurate for computer programs, it is hard to translate into human terms.

A second method is game play against human masters. One can get an idea of the program's strength from both the outcome statistics of the games, and from the masters' play-by-play analysis of the computer's decisions. The main problem with this method is that game play against humans is much

slower, and usually only a few dozen games can be played. Also the expert's assessment is at least partly subjective, and may not be 100% accurate.

A third method of analysis, which is new but rapidly becoming the standard among human experts, is to analyze individual move decisions via computer rollouts. In other words, to check whether a player made the right move in a given situation, one sets up each candidate position and has a computer play out the position to completion several thousand times with different random dice sequences. The best play is assumed to be the one that produced the best outcome statistics in the rollout. Other plays giving lower equities are judged to be errors, and the seriousness of the error can be judged quantitatively by the measured loss of equity in the rollout.

In theory, there is a potential concern that the computer rollout results might not be accurate, since the program plays imperfectly. However, this apparently is not a major concern in practice. Over the last few years, many people have done extensive rollout work with a commercial program called "Expert Backgammon," a program that does not actually play at expert level but nevertheless seems to give reliable rollout results most of the time. The consensus of expert opinion is that, in most "normal" positions without too much contact, the rollout statistics of intermediate-level computer programs can be trusted for the analysis of move decisions. (They are less reliable, however, for analyzing doubling decisions.) Since TD-Gammon is such a strong program, experts are willing to trust its results virtually all the time, for both move decisions and doubling decisions. While computer rollouts are very compute-intensive (usually requiring several CPU hours to analyze one move decision), they provide a quantitative and unbiased way of measuring how well a human or computer played in a given situation.

Larry Reeker

May 27, 2000

Dear Advisory Board Members,

Attached, you will find, a document developed by a fellow Advisory Board Member, Dimitar Filev from Ford Corporation. Let me know if you have any comments.
Moderator.

AN IMPORTANT SUBSET OF INTELLIGENT SYSTEMS

This comment is focused on one special group of intelligent systems - the intelligent control systems. What makes a control system intelligent and is there a clearly defined border between intelligent and other (nonintelligent) control algorithms? The trivial answer to this question usually is

determined based on the control methodology used. Commonly, the soft computing based control algorithms (neural / fuzzy / genetic) are considered intelligent by default because of their knowledge-based content. Such a determination, however, is opposed by the control theorists who claim that modern (conventional) control methods with their strong mathematical foundations are not less intelligent than the above mentioned soft computing technologies.

Strictly speaking, all robust control algorithms (conventional and soft) fit the first part of the definition of Albus since they are targeted to work in uncertain environment and if properly designed they generate appropriate actions to "… increase the probability of success…" with respect to a given criterion. In a broad sense, however, there are very few, if any, control algorithms that satisfy the definition of system intelligence. While evaluating the level of intelligence based on this definition (to avoid the confusion of introducing a new one) we have to take into account:

- ***type of uncertain environment***
- ***strategy of achieving the goals***
- ***capability of the system to automatically create and update its subgoals.***

Most of the well-established methods for robust control design provide the capability to deal with small parametric and structural uncertainties and therefore include a basic level of intelligence in the control system according to the definition of Albus. Situational uncertainty, e.g. drastic changes in the environment that are due to completely different operating conditions, severe and unpredictable disturbances, etc., completely alter system dynamics, and therefore require control systems with much higher level of intelligence.

These strategies of achieving the goals that deal with analysis of the situation, selection of alternative control actions in accordance with the identified environment, and subsequent adaptation convey more than basic intelligence to the control system. In this scheme the gain scheduling, adaptive control and hierarchical control are only special cases of an intelligent control mechanism that brings in the elements of perception of situation and decision making.

The flexibility of the structures offered by the fuzzy and neural models and the natural granulation of the information that is associated with these models provide some of the basic building blocks for development of intelligent control systems. In my view, we have seen only some of the advantages of these methods over the conventional (equation based) control paradigm. So far, the gain of using these technologies comes mostly from using them as universal approximators and as tools for granulation (i.e. partitioning the space and natural decomposition of the system - typical example are the so called neuro-fuzzy systems where the fuzzy/neural model is used as a powerful tool for approximation of the plant model). Fuzzy/neural systems that are introduced in such environment are an alternative and powerful tool that enriches the control toolbox but does not automatically generate a higher level of intelligence. We are about to see more intelligent control strategies, e.g. task oriented control and

hierarchical control with dynamically created and updated subgoals when we start fully utilizing the knowledge-based content and decision making capabilities of the soft computing technologies.

Dimitar Filev

May 27, 2000

OUR EVALUATIONS OF MACHINE INTELLIGENCE SHOULD BE COMPATIBLE WITH OUR EVALUATIONS OF HUMAN INTELLIGENCE!

Dear Advisory Board Members,

The following Hans Moravec's thoughts will be interesting for you:

[THE PROPERTY OF INTELLIGENCE IS ASSIGNED BY US]

Perhaps the most unsettling implication of this train of thought is that anything can be interpreted as possessing any abstract property, including consciousness and intelligence. Given the right playbook, the thermal jostling of the atoms in a rock can be seen as the operation of a complex, self-aware mind. How strange. Common sense screams that people have minds and rocks don't. But interpretations are often ambiguous. One day's unintelligible sounds and squiggles may become another day's meaningful thoughts if one masters a foreign language in the interim. … Sometimes we exploit offbeat interpretations: an encrypted message is meaningless gibberish except when viewed through a deliberately obscure decoding. Humans have always used a modest multiplicity of interpretations, but computers widen the horizons. The first electronic computer was developed by Alan Turing to find ``interesting" interpretations of wartime messages radioed by Germany to its U-boats. As our thoughts become more powerful, our repertoire of useful interpretations will grow. We can see levers and springs in animal limbs, and beauty in the aurora: our ``mind children" may be able to spot fully functioning intelligences in the complex chemical goings on of plants, the dynamics of interstellar clouds, or the reverberations of cosmic radiation.

……

[THE CRUCIAL ROLE OF SELECTION]

There is no content or meaning without selection. The realm of all possible worlds, infinitely immense in one point of view, is vacuous in another. Imagine a book giving a detailed history of a world similar to ours. The book is written as compactly as possible: rote predictable details are left as homework for the reader. But even with maximal compression, it would be an astronomically immense tome, full of novelty and excitement. This interesting book, however, is found in ``the library of all

possible books written in the Roman alphabet, arranged alphabetically"---the whole library being adequately defined by this short, boring phrase in quotes. The library as a whole has so little content that getting a book from it takes as much effort as writing the book. The library might have stacks labeled A through Z, plus a few for punctuation, each forking into similarly labeled substacks, those forking into subsubstacks, and so on indefinitely. Each branchpoint holds a book whose content is the sequence of stack letters chosen to reach it. Any book can be found in the library, but to find it the user must choose its first letter, then its second, then its third, just as one types a book by keying each subsequent letter. The book's content results entirely from the user's selections; the library has no information of its own to contribute.

… The set of all possible interpretations of any process as simulations is exactly analogous to the content of all the books in the library. In total it contains no information, yet every interesting being and story can be found within it.

………

[WHO PERCEIVES AND WHO INTERPRETS?]

If our world distinguishes itself from the vast unexamined (and unexaminable) majority of possible worlds through the act of self-perception and self-appreciation, just who is doing all the perceiving and appreciating? The human mind may be up to interpreting its own functioning as conscious, so rescuing itself from meaningless zombiehood, but surely we few humans and other biota---trapped on a tiny, soggy dust speck in an obscure corner, only occasionally and dimly aware of the grossest features of our immediate surroundings and immediate past---are surely insufficient to bring meaning to the whole visible universe, full of unimagined surprises, 10^40 times as massive, 10^70 times as voluminous, and 10^10 times as long-lived as ourselves. Our present appreciative ability seems more a match for the simplicity of Saturday-morning cartoons.

……..

[NEW MODELS ENHANCE OUR ABILITY TO CREATE POSSIBLE WORLDS]

Although our eyes and arms effortlessly predict the liftability of a rock, the action of a lever, or the flight of an arrow, mechanics was deeply mysterious to those overly thoughtful ancients who pondered why stones fell, smoke rose, or the moon sailed by unperturbably. Newtonian mechanics revolutionized science by precisely formalizing the intelligence of eye and muscle, giving the Victorian era a viscerally satisfying mental grip on the physical world. In the twentieth century, this common-sense approach was gradually extended to biology and psychology. Meanwhile, physics moved beyond common sense. It had to be reworked because, it turned out, light did not fit the Newtonian framework.

In a one-two blow, intuitive notions of space, time, and reality were shattered, first by relativity, where space and time vary with perspective, then more seriously by quantum mechanics, where unobserved events dissolve into waves of alternatives. Although correctly describing everyday mechanics as well as such important features of the world as the stability of atoms and the finiteness of heat radiation, the new theories were so offensive to common sense, in concept and consequences, that

they inspire persistent misunderstandings and bitter attacks to this day. The insult will get worse. General relativity, superbly accurate at large scales and masses, has not yet been reconciled with quantum mechanics, itself superbly accurate at tiny scales and huge energy concentrations. Incomplete attempts to unite them in a single theory hint at possibilities that exceed even their individual strangeness.

……………………………………………………………………………………………………….

[COMMON SENSE OF MEASURING]

In principle, if not practice, the point of collapse can be pinpointed: before collapse, possibilities interfere like waves, creating interference patterns; after collapse, possibilities simply add in a common-sense way. Very small objects, like neutrons traveling through slits, make visible interference patterns.

Unfortunately, large, messy objects like particle detectors or observing physicists would produce interference patterns much, much finer than atoms, indistinguishable from common-sense probability distributions because they are so easily blurred by thermal jiggling.

Because, for humans, common sense is easier than quantum theory, workaday physicists take collapse to happen as soon as possible---for instance, when a particle first encounters its detector. But this ``early collapse" view can have peculiar implications. It implies that the wave function can be repeatedly collapsed and uncollapsed in subtle experiments that allow measurements to be undone through deliberate cancellation at the experimenter's whim.

… Einstein was troubled by the implications of quantum mechanics, and he devised thought experiments with outcomes so counterintuitive he felt they discredited the theory. Those counterintuitive outcomes are now observed in laboratories and utilized in experimental quantum computers and cryptographic signaling systems. Soon, more advanced quantum computers will allow the results of entire long computations to be undone.

Common sense screams that measurements are real when they register in the experimenter's consciousness. This thinking has led some philosophically inclined physicists to suggest that consciousness itself is the mysterious wave-collapsing process that quantum theory fails to identify. But even consciousness is insufficient to cause collapse in the thought experiment known as ``Wigner's Friend." Like the more famous ``Schrödinger's Cat," Wigner's friend is sealed in a perfectly isolating enclosure with a physics experiment that has two possible outcomes. The friend observes the experiment and notes the outcome mentally. Outside the leakproof enclosure, Wigner can only describe his friend's mental state as the superposition of the alternatives. In principle these alternatives should interfere, so that when the enclosure is opened one or another outcome may be favored, depending on the precise time of opening. Wigner might then conclude that his own consciousness triggered the collapse when the enclosure was opened, but his friend's earlier observation had left it uncollapsed.

……………………………………………………………………………………………………….

["MANY-WORLDS INTERPRETATIONS]

In a 1957 Ph.D. thesis, Hugh Everett gave a new answer to that question. Given a universally evolving wave function, where the configuration of a measuring apparatus, no less than of a particle, spreads wavelike through its space of possibilities, he showed that if two instruments recorded the same event, the overall wave function had maximum magnitude for situations where the records concurred and canceled where they disagreed. Thus, a peak in the combined wave represents a possibility where, for instance, an instrument, an experimenter's memory, and the marks in a notebook agree on where a particle alighted---eminent common sense. But the whole wave function contains many such peaks, each representing a consensus on a different outcome. Everett had shown that quantum mechanics, stripped of problematical collapsing wave functions, still predicts common-sense worlds---only many, many of them, all slightly different. The ``no-collapse" view became known as the ``many-worlds" interpretation of quantum mechanics.

………………………………………………………………………………………………………..

[INTELLIGENCE DETECTS SINGULARITIES WITHIN THE CHAOS]

No complete theory yet explains our existence and experiences, but there are hints. Tiny universes simulated in today's computers are often characterized by adjustable rules governing the interaction of neighboring regions. If the interactions are made very weak, the simulations quickly freeze to bland uniformity; if they are very strong, the simulated space may seethe intensely in a chaotic boil. Between the extremes is a narrow ``edge of chaos" with enough action to form interesting structures, and enough peace to let them persist and interact. Often such borderline universes can contain structures that use stored information to construct other things, including perfect or imperfect copies of themselves, thus supporting Darwinian evolution of complexity. If physics itself offers a spectrum of interaction intensities, it is no surprise that we find ourselves operating at the liquid boundary of chaos, for we could not function, nor have evolved, in motionless ice nor formless fire.

……..

[INTELLIGENCE DEVELOPS THE IMAGES OF "POSSIBLE WORLDS"]

The similarity between Everett's ``many worlds" and the philosophical ``possible worlds" may become stronger yet. In ``many worlds" quantum mechanics, physical constants, among other things, have fixed values. Gravity, in objects like black holes, loosens the rules, and a full quantum theory of gravity may predict possible worlds far exceeding Everett's range---and who knows what potent subtleties lie even further on? It may turn out, as we claw our way out through onion layers of interpretation, that physics places fewer and fewer constraints on the nature of things. The regularities we observe may be merely a self-reflection: we must perceive the world as compatible with our own existence---with a strong arrow of time, dependable probabilities, where complexity can evolve and persist, where experiences can accumulate in reliable memories, and the results of actions are predictable. Our mind children, able to manipulate their own substance and structure at the finest levels, will probably greatly transcend our narrow notions of what is.

..

[WE ARE SELF-INTERPRETING BOTH OURSELVES AND THE EXTERNAL WORLD]

Like organisms evolved in gentle tide pools, who migrate to freezing oceans or steaming jungles by developing metabolisms, mechanisms, and behaviors workable in those harsher and vaster environments, our descendants may develop means to venture far from the comfortable realms we consider reality into arbitrarily strange volumes of the all-possible library. Their techniques will be as meaningless to us as bicycles are to fish, but perhaps we can stretch our common-sense-hobbled imaginations enough to peer a short distance into this odd territory. Physical quantities like the speed of light, the attraction of electric charges, and the strength of gravity are, for us, the unchanging foundation on which everything is built. But if our existence is a product of self-interpretation in the space of all possible worlds, this stability may simply reflect the delicacy of our own construction---our biochemistry malfunctions in worlds where the physical constants vary, and we would cease to be there. Thus, we always find ourselves in a world where the constants are just what is needed to keep us functioning. For the same reason, we find the rules have held steady over a long period, so evolution could accumulate our many intricate, interlocking internal mechanisms.

Our engineered descendants will be more flexible. Perhaps mind-hosting bodies can be constructed that are adjustable for small changes in, say, the speed of light. An individual who installed itself in such a body, and then adjusted it for a slightly higher lightspeed, should then find itself in a physical universe appropriately altered, since it could then exist in no other. It would be a one-way trip. Acquaintances in old-style bodies would be seen to die---among fireworks everywhere, as formerly stable atoms and compounds disintegrated. Turning the tuning knob back would not restore the lost continuity of life and substance. Back in the old universe everything would be normal, only the acquaintances would witness an odd ``suicide by tuning knob." Such irreversible partings of the way occur elsewhere in physics. The many-worlds interpretation calls for them, subtly, at every recorded observation. General relativity offers dramatic ``event horizons": an observer falling into a black hole sees a previously inaccessible universe ahead at the instant she permanently loses the ability to signal friends left outside.

See more in URL:

http://www.frc.ri.cmu.edu/~hpm/project.archive/general.articles/1998/SimConEx.98.html

Be patient to get to the last quarter of THIS DOCUMENT. It raises important issues concerning world representation. As you know, H. Moravec was one of the first creators of vehicles with elements of autonomy.

Moderator

May 29, 2000

Dear Advisory Board Member,

Dr. Eric Horvitz (a member of our Board) has suggested to distribute the attached paper among the Advisory Board Members. This paper is related to dealing with uncertainties as a part of the process of decision making when the imprecisely computed Metrics are used.

Since any introduction of Metrics is linked with determining preferences under uncertainty, it would be meaningful to be prepared to the non-trivial situations of decision making using any proposed
Metrics.

It would be presumptuous to expect that our University education has prepared us to these situations exhaustively (partially? maybe).

I found this paper enlightening (to the extent I could understand it). All of you are expected to achieve some level of understanding of the peculiarity linked with introduction and use of Metrics for Decision Making.

Moderator

May 30, 2000

THOUGHTS AFTER READING THE WHITE PAPER

BY TOM WHALEN

The "white paper" draft seems to me to be a real gold mine, but like any mine it requires digging, sifting, and refining.

1. INTELLIGENCE AND AUTONOMY

I really like the idea of the autonomous climate control system being "motivated" to increase its autonomy by reducing the need for human intervention. (p.3) I think this could be the kernel for a better definition of what is meant by intelligence in autonomous constructed system.

Here's my stab at some global definitions:

Def.1

"A constructed system is autonomous if there is a likelihood that circumstances will arise in which no-one can predict in advance what it will do. This need not imply randomness, just complexity."

[This probably can imply both: complexity AND randomness. Moderator.]

Def.2

"An autonomous constructed system is intelligent if we can be reasonably confident that whatever unpredictable thing it does do will be something that tends toward success in the goals for which the system was constructed in the first place."

2. HOW TO FALSIFY THESE DEFINITIONS?

Thus, a claim that a system is autonomous and intelligent can be falsified in two ways: showing it is not autonomous by predicting all of its behavior in advance, or showing it is not intelligent by demonstrating that its behavior is stupid.

What an end user wants is a system that is trustworthy. If all behavior can be specified in advance, there is no need for autonomy; the intelligence and autonomy reside in the design team and not in the delivered system. If behavior can't be prespecified, then intelligence is necessary for trustworthiness; if it is lacking, the system needs to be monitored by a human operator and thus, again, lacks autonomy.

Note the statement on page of the White Paper that an intelligent system was "designed by humans (engineers and programmers)" is not true in machines that learn and self-organize except in a broadened sense of the word "designed." Even very large and complex programs that have no learning or self-organizing features need to be studied in much the same way we study social phenomena like economics or natural phenomena like weather, since no one person will ever know the program in its entirety. (For example, the Windows operating system.)

The paper has the beginnings of a structure for measuring the components of machine intelligence based on the six-box semiotic loop, but it's not very consistent.

3. MIND-BODY PROBLEM

The discussion of the "mind-body" problem crops up several times in the White Paper; I suggest making it specific by assigning perception, knowledge, and decision (behavior generation) as "mind" and assigning sensation and actuation as body. The sixth box, "world," is not part of the constructed autonomous system.

[Here, we should think twice: should the work-piece that we drill be considered a part of the drilling machine or not? The part of the world that immediately interacts with a system under consideration might be legitimately considered a part of this system. When we write the equations of the system, the "torque on the shaft" is a part of the equations of the system. Moderator.]

4. ABOUT CHINESE ROOM

Note: my personal view of the Chinese room is that performing the task without understanding Chinese is not in principle impossible, but the number of rules that would have to be written ahead of time and searched in real time by the occupant of the room is far beyond the trillions. Learning to understand Chinese is a much easier task, already mastered by over a billion people!

This is quite relevant to the issue of intelligent autonomous machines; there are tasks for which it may be within our grasp to produce a successful machine without autonomy, but it is actually easier to achieve the same level of success using an intelligent autonomous machine. A simple and very familiar example is the inverted pendulum, which is quite challenging to do with differential equations but a beginner's exercise to do with fuzzy control.

[I would like to emphasize this significant Tom's statement by capitalizing: THERE ARE TASKS FOR WHICH IT MAY BE WITHIN OUR GRASP TO PRODUCE A SUCCESSFUL MACHINE WITHOUT AUTONOMY, BUT IT IS ACTUALLY EASIER TO ACHIEVE THE SAME LEVEL OF SUCCESS USING AN INTELLIGENT AUTONOMOUS MACHINE. Moderator]

5. WHAT DOES IT MEAN TO BE "INTELLIGENT"

My initial impression is that while human intelligence testing does rely heavily on response time, my online thesaurus lists the following synonyms for "intelligent" (see the White Paper). Only three of the sixteen involve the idea of "quick-witted." Machines routinely do almost anything they can do at all more quickly than humans can do them, and they also are incapable of doing at all many things humans do quickly. There is only a small middle ground of things machines do roughly as fast as humans or that they do well but more slowly than humans.

I'm more interested in the prospects for machines that are canny. percipient, perspicacious, astute, and discerning.

Tom's insights are great!
Moderator.

May 30, 2000

THOUGHTS AFTER READING THE WHITE PAPER

by Sukhan Lee

I would like to congratulate (...) the effort to formalize the intelligent system research by establishing the measure of system intelligence. Establishing the measure (...) should not only be able to turn the intelligent system into a formal academic discipline but also provide a means of designing better and more powerful intelligent systems in practice.

I am generally impressed by the breadth as well as depth of [the proposed] measure of intelligence for a constructed system with autonomy. (...) various aspects of intelligence [are considered] including the need to learn as well as to generalize by an intelligent system. (...) a list of system specifications [are proposed] as well as the vector of intelligence as features representing intelligent functions of a system.

The list of features presented is very comprehensive. However, it is not clear how the measure of intelligence can be formulated out of such a list or a vector. Too many functional features may obscure the essence of how intelligence is generated, as they may not represent the engine but the expressions.

Having said that, I would like to pay attention to the following questions:

1) Should the intelligence measure be goal-dependent or goal-independent?

2) Should the intelligence measure be time varying or time-invariant?

3) Should the intelligence measure be resource-dependent or resource-independent?

1) (...) a question [emerges] whether there exists a universal measure of system intelligence such that the intelligence of a system can be compared independently of the given goals. A goal-independent measure may be more difficult to define, (if not impossible), and [it will be] more controversial.

A goal-dependent measure, however abstract the goal may be, can allow [for a] clear comparison among the systems of different architecture but with the same goal. For instance, for the latter case, an intelligence can be represented as how efficiently, and how optimally a system reaches the given goal by itself, i.e., the power of automatically solving problems defined as the discrepancy between the goal and the current state.

2) (...) [We should decide whether] the intelligence measure of a system should solely be based on problem-solving capability at time t or it should contain the potential increase of problem-solving capability in the future based on learning. My opinion is that we need both. But, it is better to define the two separately before integrating them together into one measure.

3) (...) [Finally, it is an important] issue whether the resources required for building systems and system operation should play a role for defining the measure of intelligence. As mentioned above, the efficiency in problem solving, I think, should be included in the measure: for instance, the time and energy required to reach a solution should be taken into consideration together with the optimality of the solution. But, I am not sure whether we should or should not include the cost of building a system.

[I WOULD APPRECIATE SENDING TO ME YOUR THOUGHTS ABOUT THE INTERESTING POINTS INDICATED BY S. LEE. Moderator]

May 31, 2000

THOUGHTS AFTER READING THE WHITE PAPER: C. WEISBIN'S QUESTIONS

Would it be appropriate for you to specify for the workshop a SMALL number (~5) questions which the workshop (perhaps within working groups, or abstracted from position papers) would try to answer with some degree of specificity? The field is so broad and the interests are so varied that I am puzzled how (whether?) tangible conclusions will emerge? The field is so broad and the interests are so varied that I am puzzled how (whether?) tangible conclusions will emerge?

I WOULD LIKE TO DISCUSS WITH ALL MEMBERS OF THE ADVISORY BOARD THE LIST OF QUESTIONS THAT I PROPOSED IN THE RESPONSE LETTER TO C. WEISBIN.
LET ME KNOW WHAT DO YOU THINK, THIS IS VERY IMPORTANT:

This is the list of questions that the Workshop will try to answer:

1. What is the vector of intelligence (VI) that should be measured and possibly used as a metric for systems comparison?

2. Should VI be measured in addition, or instead of measuring the vector of performance (VP) determined by the regular specifications?

3. If two systems have the same VP, what is implied by the difference in their VI values? Can this difference be represented in $ units?

4. Is it possible (and meaningful) to have different VI measures: a) goal-invariant, b) resource-invariant, c) time-invariant?

5. What should be recommended as a test of VI and how to normalize VP so that comparison be performed at the same normalized value of VP.

These are the five questions that you have asked about. As a reminder, I would like to formulate a set of other questions that are ingrained (directly, or indirectly) in the main five questions:

6. These are the less profound ("secondary") questions that should be addressed at the workshop and possibly unequivocally answered:

a) how to form VI for various architectures?

b) should the questions 1 through 5 be related to intelligent systems, or autonomous systems, or both?

c) what is the protocol of dealing with uncertainty when the uncertain metric is to be applied in the procedures of decision making? for example how the uncertainty of planning affects the cost of goal achievement?

d) what are the guidelines in constructing the world model and determining its scope in the variety of applications? how the scope of "world model" affects the sophistication of intelligent behavior?

e) how are the questions 1 through 5 related (and the answers applied to) the systems that are working under a hierarchy of goals.

f) should a competition between intelligent systems be considered a valid method of judging VI value?

Moderator

May 31, 200

SOME COMMENTS ON THE PROBLEM OF METRICS OF INTELLIGENCE

GEORGE BEKEY HAS PROPOSED THE FOLLOWING:

1. The selection of benchmark problems on which to measure the degree of autonomy and intelligence of a system, and

2. Something orthogonal to the previous discussions: A discussion of the moral and ethical implications of building increasingly intelligent systems.

For more details on these issues see his attachment. *Moderator*

Attachment: G. Bekey's comments:

1. Benchmarks

I am a strong believer in simplicity, so my definitions and metrics are very simple. I believe that the fundamental attributes of intelligence involve:

- ***Ability to perform tasks in unstructured environments***
- ***Ability to learn from experience***
- ***Ability to transfer knowledge from one domain to another***
- ***Ability to solve complex problems, requiring deductive and inductive reasoning***

(While stated differently, these issues are similar to Jim Albus's definitions). I suggest that the following simple measures can be used as metrics for such abilities in machines:

1. ***Size and complexity of programs required***
2. ***Memory requirement***
3. ***Solution time***

Clearly, such measures are useful only if (a) they are applied to benchmark problems, (b) all contestants use the same type and model of computer, and (c) all programs are written by comparably competent programmers, so that the programs are optimal in some sense.

Given these constraints, we could test intelligent systems A and B on the same benchmarks. The one that accomplishes the task more quickly, and does so with the least complex programs and least memory will be declared "more intelligent". What could be simpler than that?

If anyone agrees with me, I would be interested in leading a discussion on the selection of benchmarks on which we can all test our systems.

(There are several hidden questions here. One of them is the question of emergent behaviors. As a system learns more and more, and is able to transfer knowledge to other domains, and solve increasingly complex problems, it may begin to do so in totally unexpected ways. The emergence of such new behaviors will be appear in its ability to solve problems more rapidly, but will not be directly measurable).

2. On the ethics of building intelligent machines

Several recent books have dealt with this subject, such as the latest books from Hans Moravec or Ray Kurzweil ("The age of spiritual machines"). Kurzweil predicts that by 2025 computers will be more intelligent than people (but does nor provide metrics to measure this result!). Perhaps the most thoughtful analysis was published by Bill Joy (the chief scientist of Sun Microsystems), in the April 2000 issue of WIRED. In an article entitled "Why the future doesn't need us", he speculates that developments in robotics, nanotechnology and genetic engineering will inevitably lead to self-reproducing machines with increasing intelligence, whose behavior will be not only unpredictable but uncontrollable. Such machines may find human beings largely superfluous.

I would be interested in presenting a position paper summarizing current thinking on this matter and leading a discussion. It seems to me that if we are concerned only with measuring the intelligence of machines without any concern for the social implications of such intelligence, we are not fulfilling our responsibility to society.

George Bekey

June 3, 2000

WINTER'S CONJECTURE

(FROM MY CONVERSATION WITH VICTOR WINTER, MEMBER OF THE ADVISORY BOARD)

Victor Winter (V. W.):

The ability to learn is generally considered a "sophisticated" behavior. Given this ability, a machine can positively change its behavior over time thereby exceeding the sum of its initial input. By this metric, a more a sophisticated system can "do more" with less initial input than an unsophisticated system.

Alex Meystel (A. M.):

I believe that intelligent system should learn. By I am hesitant to say that anything that can learn is necessarily intelligent. Having an increased degree of sophistication? Probably!

V. W.

For example, in theory, a theorem prover can discover all of mathematics from its axiomatization. Of course there are some severe limitations to this in practice.

A. M.

I have already addressed this issue in the previous part of our conversation. No, a simple set of axioms is insufficient to prove these two theorems that I have mentioned above. (I worked this out with students).

V. W.

Nevertheless, we can think of an axiomatization of a closed system as a very compact model of that system. (We are talking about the need to supply knowledge of the initial general laws that pertain to a concrete environment; we call them "axioms").

A. M.

Compact - maybe. But sufficient - hardly.

V. W.

To the person writing the axioms the system need not be fully understood (e.g., knowledge of all theorems in mathematics is not required).

A. M.

This is an unwanted surprise: your definition of "understood" means "having known all theorems related." I think that as a label, you can use any word you want. But in reality "understood" means much more, for example: "having future behaviors anticipated".

(The term "theorems" should be understood as the provable rules that hold in the environment of the subsequent functioning).

V. W.

However, this compact representation of the system can be utilized by a sophisticated machine to solve a large collection of problems--problems that may not even have been initially foreseen.

A. M.

You are talking about "sophisticated machine" that has not been defined. However even if your "sophisticated machine" is equivalent "human intelligence", I doubt that having just compact representation (you have defined it as based only upon the axioms) we will be able to solve the unforeseen problems.

V. W.

An application of this idea is in a space system where there may be multiple ways to achieve a certain behavior (e.g., pitch or rotation of a spacecraft). If one possibility fails due to a malfunction of a component another possibility can be discovered -- provided a sufficiently complete system model has been given. In this example, a space bound system can be equipped with a computer that has

(1) a non-algorithmic specification of the behavior of the system, and

(2) an axiomatization of that system.

A sophisticated system would then be able to satisfy its specification whenever possible, even in the presence of unforeseen circumstances.

A. M.

We have to add the word ALL into your last sentence; then my refutation will be stronger: not "even in the presence of ALL unforeseen circumstances."

The reason I insist on this "ALL" is because you allowed for only limited variability in the space of searching the strategies of solution.

There is a threshold, a level of variability that is required to cover the space of future behaviors sufficient for, say, survival with a definite probability.

V. W.

The presence of such a specification would to some extent address the concerns/issues raised by Searle (as described in your White Paper). For example, a robotic system having a defective (e.g., bent) arm would be able to sense (through "sensors") that the behavior of the arm - in its current state -- is not satisfying the specification. Alternate means of satisfying the specification could then be generated and considered.

A. M.

Of course, many realistic cases of robots functioning in Space presently do not require and do not exercise any sophistication not to speak of intelligence.

V. W.

The notion of learning can be taken one step further and one can consider a system, which has the ability to dynamically generate axioms.

A. M.

I agree with you ecstatically.

V. W.

Given what I have mentioned above, I believe that an interesting metric would be the "size" of knowledge required or the amount of information that must initially be provided to a system in order for it to demonstrate a certain behavior -- the smaller is the required initial information, the more

"sophisticated" the system is. These kinds of systems (if practical) could effectively utilize the increases in computing power that will occur in the future.

A. M.

Victor, this is a great idea. I will call it Winter's Conjecture, and will circulate it among our Advisory Board Members. Prepare yourself for proving mathematically a number of theorems that entail your hypothesis. For example, the fundamental theorem that you should introduce is related to evaluation of the "amount" of this information based upon complexity of the system that is supposed to be equipped by "intelligence."

This theorem is on you. I would suggest to use a measure of variability of this minimal information and evaluate the probability of success in applying this minimum of information with a given measure of variability.

CLASSES OF INTELLIGENT SYSTEMS

June 3, 2000

Dear Advisory Board Member,

Attached, you will find, an important contribution by Steve Grossberg (see his paper in the next part of this book). This view focuses upon CONSTRAINTS that will undoubtedly make more difficult our road toward Metrics for Intelligent Systems in a particular class of intelligent systems.

The abstract also reminds about interesting problems that emerge in autonomous intelligent systems and make them different from the classical control systems. I would suggest to compare the class of systems that is described in S. Grossberg's paper and the class of systems that is described in D. Filev's statement (see my message on May 27).

Both classes belong to super-class of "intelligent systems."

But look how different they are!

I would not be surprised if their Metrics will have the same properties: belonging to the same super-class but very different.

Yours,

Alex Meystel

June 5, 2000

Dear Advisory Board Members,
I RECEIVED AN INTERESTING SET OF ANSWERS AND I WANT TO FAMILIARIZE YOU WITH THEM.

THIS LETTER I RECEIVED FROM WALTER FREEMAN, A MEMBER OF THE ADVISORY BOARD. IT CONTAINS SEVERAL FAR REACHING SUGGESTIONS, IMPORTANT FOR ALL OF US.

FIRST QUESTION:
1. What is the vector of intelligence (VI) that should be measured and possibly used as a metric for systems comparison?

W. FREEMAN'S ANSWER:
In my view, we don't measure intelligence; we infer it from measurements of performance.

[ACTUALLY, ANYTHING MEASURED IS CONCEPTUALLY INTRODUCED IN THE BEGINNING. IN OTHER WORDS, ALL THINGS THAT WE MEASURE WE INFER FROM MEASUREMENTS OF OTHER RELATED THINGS. WALTER IS MORE INTERESTED IN THE QUALITATIVE CHARACTERIZATION OF INTELLIGENCE AND HE ASKS:]

W. F. A better question is: What kinds of intelligence do we propose to emulate?

MY ANSWER TO THIS QUESTION IS:
I doubt that there are different kinds of intelligence. I know that multiple intelligences is a faith (very similar to the polytheism of ancient people). It is easy to declare any manifestation of perceptual and cognitive activity to be an intelligence. It is more difficult to find what do they have in common, maybe, absolutely the same.

We all can easily demonstrate that all known phenomena of intellect and intelligence are linked with a limited number of special computational algorithms including

--combining N-tuples

--searching

--focusing attention

--grouping (which includes "combining tuples")

--evaluating and ranking the results of grouping.

SECOND QUESTION: 2. Should VI be measured in addition, or instead of measuring the vector of performance (VP) determined by the regular specifications?

W. FREEMAN RESPONDS:

Hence, what kinds of performance do we choose as benchmarks for measurement?

THIRD QUESTION: 3. If two systems have the same VP, what is implied by the difference in their VI values? Can this difference be represented in $ units?

W. FREEMAN SUGGESTS:

Instead: How might we choose and construct benchmark tasks with graded difficulty, so that we can establish the competence of new systems and then challenge them with tasks of increasing complexity.

[I THINK, THIS IS AN EXCELLENT SUGGESTION]

FOURTH QUESTION: 4. Is it possible (and meaningful) to have different VI measures: a) goal-invariant, b) resource-invariant, c) time-invariant?

W. FREEMAN PROPOSES:

In that there are different kinds of intelligence, it follows that there are different VPs. For VI, in analogy to IQ, I would suggest using a ratio of performance VP to cost of construction of a system in time and money VC, giving a scalar value that will assign due place to building a machine gun to kill a fly:

VI[n] = VP[n]/VC[n], n = 1,N [number of classes of benchmark]

[I THINK THAT THIS PROPOSAL IS MEANINGFUL: NORMALIZING IS A PROPER

THING TO DO. BUT I DISAGREE TO CONSIDER THESE DIFFERENT MODES OF EVALUATION OF ONE INTELLIGENCE TO BE DIFFERENT INTELLIGENCES. WHY?]

FIFTH QUESTION: 5. What should be recommended as a test of VI and how to normalize VP so that comparison be performed at the same normalized value of VP.

WALTER FREEMAN EXPLAINS AND SUGGESTS THREE BENCHMARKS:

This question would be answered under #3 and #4. I suggest N = 3 benchmarks, relating to comprehension through perception, planning action, and dynamic reasoning through decision:

n = 1. Pattern classification - for example, detection of chemical explosives of increasing variety in increasingly complex background odorant environments. We have excellent chromatographs, but there is a need here for the artificial dog behind the artificial nose.

n = 2. Spatial navigation - for example, foraging for fuel in natural environments of graded complexity. Gray Walter's autonomous tortoises are still [in my opinion] the best of breed in this respect.

n = 3. Comprehension of instructions in a natural language – for example, accomplishing sequences of operations, each conditional on the steps preceding. Ross Ashby's 'homeostat' might offer a suitable early benchmark.

[I THINK THAT THIS GIVES AN ANSWER TO MANY QUESTIONS RELATED TO METRICS OF INTELLIGENCE]

WALTER BELIEVES THAT THIS MY SIXTH QUESTION IS PREMATURE, THAT WE ARE NOT AT THE STAGE THAT QUESTION No. 6 COULD BE APPROACHED RIGHT NOW. WELL, I DISAGREE.

Moderator

EMPHASES IN RESEARCH OF BUILDING INTELLIGENT MACHINES AND MEASURING THEIR INTELLIGENCE: THE ISSUE OF ETHICS

by J. Albus

June 6, 2000

I agree with George Bekey that we should spend some time discussing the ethics of building intelligent machines. I feel that the concerns voiced by Bill Joy should not go unanswered, and the predictions of Moravec and Kurzweil should not remain unchallenged.

I, for one, feel that the most important characteristic of intelligent machines is that they have the capacity to perform useful work, i.e., to create wealth. I also fervently believe that the biggest problem in the world today is poverty, i.e., the lack of wealth. (I see poverty is the fundamental cause of hunger, disease, ignorance, pollution, intolerance, oppression, lack of medical care, and lack of education.) I therefore would argue that we should focus on how to make it possible for intelligent machines to eliminate poverty world wide within 50 years. In my opinion, to do anything else is unethical.

As for the concerns of Joy, I believe they are largely unfounded and overblown, and the predictions of Kurtzweil, and Moravec are for the most part wildly exaggerated. However, these kinds of sensational concerns fan the flames of the Frankenstein myth in the popular imagination and thereby create a major distraction. They divert attention from real problems that intelligent machines could solve by inciting fears of scenarios that are highly improbable. And they divert attention from what realistically could be done to alleviate human suffering in the near future.

Jim Albus

June 6, 2000

MODERATOR'S RESPONSE:

I support Jim's letter emphasizing the utter importance of ethics in the area of Intelligent Systems. I also agree that one of biggest problems in the world is "lack of wealth."

However, there are two kinds of wealth: Material and Intellectual ones. Intellectual wealth is almost always an asset especially for those in poverty, while material wealth can't frequently help even those who is rich intellectually. Material wealth is not a universal remedy. While curing known problems it creates new. It is a mixed blessing.

I consider intelligent systems to be our helpers not in creation of material wealth as the first priority, but in elimination of Ignorance. Ignorance produces and maintains poverty both material and intellectual. Ignorance maintains unethical environment. Ignorance is the major adversary of Intelligence, so it becomes an adversary of the carriers of Intelligence.

Material wealth is frequently created by developing sophisticated and sometimes even elegant but unethical methods (including algorithms and even software packages) whose only goal is "to separate a client with his/her money." Intelligent Systems can become a powerful tool of this unethical process. It is especially terrible to be robbed by an Intelligent System tuned to create somebody's material wealth in an unethical way.

Sometimes, we give up on students that cannot figure out how things are associated with each other, and graduate them anyway -- this is when we make a step down even if we succeeded in helping them to receive a position with a major insurance company developing a huge material wealth. And this is also unethical, too.

We must think ethics before we construct anything intelligent. We should ask: hey, what is this for? whose material wealth will it increase?

But understanding how the intelligence works -- we must in all cases! This is why the analysis of Metrics for Intelligence should lead us in the right direction unmistakably.

But as far as applying this Intelligence in practice, we should ask firmly: hey, what is this for?

(Unless they offer us very good money...)

Moderator

WE DO NOT NEED JUST *ANY* KIND OF INTELLIGENT SYSTEM!

By John C. Cherniavsky

I'll comment on this though it's a bit far out of the "measuring machine intelligence" category.

First Joy is just raising questions that ethically should be raised by all scientists in the performance of their research. During a recent DARPA/NSF talk he fully acknowledged the benefits of NGR (Nanotechnology, Genetics, Robotics - I'm not sure I've got it in his order) in feeding the hungry, increasing llifespan, and generating general welfare for all peoples. Contrary to what Jim asserts, he is not ignoring the benefits of NGR. He is pointing out potential dangers. His main immediate concern, rightfully so in my opinion, is the possible creation of biological and/or nano-biological organisms that replicate and that have no natural controls on their replication. He is concerned that there are no regulatory bodies that oversee this sort of research which he sees as possible in the next 10 years or so. He is also very concerned about the possible low cost of entry for this research leading to possible inexpensive terrorism on an unprecedented scale (eg. the atomic bomb built in the garage scenario).

We certainly have regulatory bodies overseeing research on biological warfare agents, yet none on similar research that could accidently (or deliberately) release replicating organisms into the environment. How concerned you are depends upon your view of how likely this is to occur and how likely it is to be low cost and easy, but it should be thought about and not dismissed out of hand. This particular danger has very little to do with intelligence and controls on research could be very similar to controls on research for other potentially harmful biologicals (Ebola virus for example) with a twist that these controls won't work if the technology becomes too cheap, too easy, and too ubiquitous.

Joy's concern about intelligent robots is more long term and speculative and based a lot on Kurzweil and Moravec's writings. He again advocates serious study and perhaps safety controls on research by oversight scientific bodies. If you don't believe that a truly intelligent robot will be built, then of course you have no concerns.

If you do believe that such robots will be built, you should be concerned about rights for that robot - after all it's intelligent - and the possibility that such robots would pose dangers. Long term speculation true - but not too early to be discussing in an open forum which is just what Joy, Kurzweil, and Moravec are doing.

Will these sorts of concerns chill research? Quite possibly. Just look at the fairly benign research on genetically altered crops and the furor in the European Community. Is that necessarily bad? Again it's a question of risk perception. All of society should be involved in such debates, not just the knowledgeable scientists and the cost/benefits of such work be fully debated.

After all, there is another solution to world poverty and that is reducing the world's population. Maybe we don't need Moravec style robots.

John C. Cherniavsky, Ph.D.
Senior Advisor for Research, EHR
National Science Foundation *June 7, 2000*

June 8, 2000

WHAT MEASURES INTELLIGENCE: A COMPETITION OR A SPECIAL TEST?

A MEMBER OF ADVISORY BOARD R. GARNER SAYS:

> ***The question was: f) should a competition between intelligent systems be considered a valid method of judging VI value?"***
>
> ***I would want to argue that a competition measures performance, but a standardized test might measure intelligence.***

WHAT DO YOU THINK? Moderator

June 8, 2000

HOW TO MEASURE INTELLIGENCE?

A MEMBER OF THE ADVISORY BOARD MARVAN JABRI SAYS:

1. A list of tasks/conditions can be defined.

2. Learning (on-line or off-line) can be included.

3. The generalisation capability is obviously critical.

4. The resources utilised are important.

5. Speed of learning (lapsed time of a trial and number of trials) would be important.

In every community there are some benchmarks. Maybe the workshop can come up with a list of benchmarks that try to cater for various levels or dimensions of VI.

The difference between intelligent systems and autonomous systems is very vague. In some sense AI is like shooting on a moving target, what systems can aspire at doing today could be simple in the future. So ideally one would have a spectrum of tasks with various scales of difficulties. In other words something like a MIQ test.

WHAT DO YOU THINK ABOUT THESE STATEMENTS?

Moderator

June 13, 2000

ON THE UNIVERSALITY OF MECHANISMS OF INTELLIGENCE

By Thomas Whalen

I doubt that there are different kinds of intelligence. I know that multiple intelligences is a faith (very similar to the polytheism of ancient people).

It is easy to declare any manifestation of perceptual and cognitive activity to be an intelligence.

It is more difficult to find what do they have in common, maybe, absolutely the same.

We all can easily demonstrate that all known phenomena of intellect and intelligence are linked with a limited number of special computational algorithms including

--combining N-tuples

--searching

--focusing attention

--grouping (which includes "combining tuples")

--evaluating and ranking the results of grouping.

There are many humans who are superb at carrying out these five activities in one area but only average or occasionally even below average in applying the same activities in other areas. By 'areas" I mean things like math, music, human relations, mechanics, visual images, etcetera.

It may well be that to be considered "intelligent" in any field one has to be good at every one of these five within the context of that field, which would make the list a good universal definition of intelligence. But I suspect that the details of creating a constructed system that's good at these five things in one area will not be "plug and play" compatible with the details of doing so in another area.

It might be very instructive to look at just what is really measured by IQ tests and so on.

MODERATOR'S COMMENTS

Tom Whalen agrees that probably the skill of intelligence consists of these five intertwined components: combining N-tuples, searching, focusing attention, grouping (which includes "combining tuples"), evaluating and ranking the results of grouping.

But he is worried that in many bright people these five components work in one context (domain) and do not work in another. Therefore, he asks: "Why people having this mechanism OK in one area cannot apply it within other areas"?

Probably, we all agree that these five activities constitute the body of the "mechanism of intelligence". Then, we are all surprised that it does not make a genius in literature to be a simultaneous genius in discrete math.

I think that we all are mistaken about it. The literature genius maybe is closer than we think to the discrete math genius (and vice versa).

Tom gave a hidden answer to this question:

-- because this mechanism works "within the context of that field".

The keyword here is "context":

[1. CONTEXT - The part of a text or statement that surrounds a particular word or passage and determines its meaning.

2. CONTEXT - The circumstances in which an event occurs; a setting.

(from American heritage Dictionary)].

The mechanism of intelligence is here but it works only in the language of a particular domain (thus, can read only the context submitted in this particular language).

How can we escape this predicament? Either we should translate the problem and the context into the language that the mechanism of intelligence understands, or from the very beginning, we should be able to operate in a "metalanguage" and translate contexts from all languages into the "metalanguage" (was proposed by E. Messina).

It seems reasonable to expect that the mechanism of intelligence working with excellence in a particular language 1 of a domain 1 can be easily retrained into working in a "metalanguage". Some humans have problems with this because they are enslaved by their prejudices about multiple intelligences. Machines are more advanced creatures: they to not have software prejudices unless one put them into an operating system.

Then, translation of a problem from languages 2, 3, etc. into the metalanguage -- is just a technicality. This is why we can expect that the box of "intelligence" can be context and even domain independent: it will work in metalanguage. Just put at the input a translator from the domain language into the metalanguage.

Do I expect that this is simple? Taken in account a thick bark of prejudices that the problem is (and we are) covered, probably not.

Can we help this process? Probably yes.

June 14, 2000

INTELLIGENCE AS A GOAL-BUILDER FOR THE CONTROL SYSTEM AND THE PARAMETRIC EVALUATION OF INTELLIGENCE

Cliff Joslyn (C. J.):

A. Meystel (A. M.) :

A. M.

1. If the goal is somehow obtained (constructed), then we should build a model of the system and apply Hamilton/Jacoby (H/J) and Euler/Lagrange (E/L). Actually, this is a reference to Calculus of Variations that allows to derive the laws of motion, dynamics, physics without any need to refer to experimental data. (In textbooks, you can find derivation of Newton's Laws, for example, F=ma by applying E/L equations to the cost-function assigned as the expression for energy).

C. J.

OK, like a generalized least action principle? It also sounds similar to Jaynes' derivation of thermodynamic laws from an entropic maximization constraint.

A. M.

Then, we can introduce planning of the future motion as finding the minimum cost motion trajectories by assigning ANY form for the COST. This means that cost is the primary factor, and since assigning COST depend on the goal, then the goal becomes the primary factor.

C. J.

OK, I think I follow you here.

A. M.

Of course, the goal presumes that there exists a source of the goal, and in many cases, this source exists as the carrier of INTELLIGENCE. For example, for a single level in the hierarchy of intelligence the adjacent lower resolution level (level "above") can be considered a source of the "goal".

C. J.

This is what emerges from this line of reasoning. A definition of the amount of intelligence in a system might involve a quantitative measure

of:

****) the amount of phenomena under control;***

****) the number of environmental distinctions measured by the system;***

****) the complexity of modalities of measurement and control;***

****) the complexity of the environmental variety available to the measurement and control of the system.***

These are all related to each other in complex ways, but the nub of it is there.

A. M.

I appreciate your compliance with the option of considering intelligence as a player. However, I am not sure that I can accept your FOUR SUGGESTED PARAMETERS that you consider a set of quantitative measures for intelligence. To me, these four factors are rather characteristics of a system that is associated with the use of intelligence.

C. J.

What's the difference? The "intelligence" of systems (and I do NOT advocate the use of this term in this context) is based on their manifesting a semiotic relation which has been selected by evolution or by designers, allowing the system to "choose" to act counter to physical law.

A. M.

Semiotic or non-semiotic - it does not matter if you do not define the PLAYER who WANTS and the PLAYER who PAYS THE COSTS (they might be the same). The term "semiotic" might obscure the essence of the situation that can reveal the phenomenon of intelligence. It looks like the essence is in an existence of a source of INTENTION.

C. J.

In attempting to reconcile your usages of terms with mine, I would say the following prerequisites necessary to find an intelligence in the control system. First, a goal state is necessary, provided from an external source, call it "a want" (an intention) provided by a player. The action of the control system is to maintain the system aware from its natural equilibrium, and this requires action and work, which can be identified as costs. And yes, the goal (ends) constrains the possible actions (means), and vice versa.

A. M.

We are interested in understanding the phenomenon of INTELLIGENCE and thus decided to model the system with the factor of INTELLIGENCE taken in account. Therefore, we should determine what plays the role of COSTS, what is the source of GOALS (WANTS), and naturally, I call this source a PLAYER.

I state this again and again because you (in your statement above) said: The "intelligence" of such systems is based on their manifesting a semiotic relation, and this statement mutes the emergence of a player with his/her WANTS<=>GOALS=>PLAYER.

C. J.

Apologies: the (perhaps implicit) presence of a goal state is, of course, necessary.

A. M.

I thought that this is pretty obvious that Powers/Marken do not want to introduce the concept of intelligence. They were not interested in this, they have other goals. The strive toward minimalism is not a new phenomenon. But there is a limit to our possibility to minimize the number of factors to be taken in account. When I refer to your state of "have overgrown" I refer to the FACT that initially we all are trying to cut the number of factors involved. One should notice that at some point of system's complexity it becomes detrimental.

C. J.

OK, I understand the admonition. My problem is always that until we can agree on these fundamentals, I have little faith that we can or should move onto the complexities.

A. M.

OK, I would agree with addressing an example of the Inanimate World with similar complexities.

C. J.

Then, consider the flipping coin. In this context Representation plays the role of the causal forces acting on the coin, and Will (Intention) is an abstraction of whatever it is which resolves the uncertainty as to whether heads or tails will turn up (call it Chance, or Chaos, or Statistical Physics).

A. M.

I would not start with this example. It is very complicated because we have two Wills here: the Will of the Man who is flipping the coin, and the Will of the Physical Law that we are not equip to compute since we do not know well enough the point that the force has been applied, the value of this force, the angle under which it was applied, the air resistance and so on.

C. J.

In the problem as set up, we ignore the will of the flipper. Thus in a sense the coin "wants" to come up heads or tails.

A.M.

Cliff, this is not so. The want of the coin is determined by the physical laws that are not well determined

C. J.

Since the chaotic flipping process is unpredictable, we cannot resolve their will into a physical explanation, but rather must resort to a statistical description. The complexity of those chaotic physical processes we simply bundle into the "will" of the coin: that which resolves the uncertainty, chance.
In your usage you can extend control, and thus intelligence, to any physical process. You are free to do this, but I find it unparsimonious, extending the term beyond any useful boundary. Instead, we need a principled way to distinguish control from other processes.

A. M.

Still, some terminological issues will remain blinking on the screen and demanding for future clarification.

C. J.

My conclusion is that on strictly denotational grounds, every control system can be seen as a semiotically closed system (NOT that "any closure is a semiotic system"), but that this is not the sense that H. Pattee intended.

A. M.

It does not matter as soon as it is true, constructive and useful. I would say even more (and this is what H. Pattee probably intended to say) that ANY INTELLIGENT SYSTEM IS SEMIOTICALLY A CLOSURE.

C. J.

Rather, Pattee is referring to the situation where the selection of the semiotic (coding) relations present in the system is itself a referent of that very semiotic system. This is thus not "simple" closure between a system and its environment, but between a system and its own construction or creation.

A. M.

Cliff, I have no qualms about it. It is not related to our discussion of intelligence. I would like to focus upon the set of issues that leads us to discovery, clarification, and better understanding of the phenomenon of intelligence. Obviously, the area of CONTROL SYSTEMS has its inner issues.

C. J.

I agree, I was only trying to distinguish between a literal sense of semiotic closure and H. Pattee's sense.

June 14, 2000

IS THERE AN ALGORITHMIC INVARIANCE WITHIN ALL KINDS OF INTELLIGENCE?

Walter Freeman has doubts about it and he responds to my discussion with T. Whalen in the following way:

Tom Whalen agrees that probably the skill of intelligence consists of these five intertwined components:

--combining N-tuples

--searching

--focusing attention

--grouping (which includes "combining tuples")

--evaluating and ranking the results of grouping.

W. F. These are pretty simple-minded, things MLPs can do.

YES, THIS LOOKS PRETTY SIMPLE-MINDED. BUT WE SHOULD NOT FORGET (I REPEATEDLY STRESSED IT) THAT THIS IS THE SET OF ELEMENTARY ALGORITHMS THAT FORM INTELLIGENCE OF A SINGLE LEVEL. AFTER GROUPING HAPPENS WE RECEIVE ANOTHER LEVEL OF RESOLUTION, A LEVEL WITH GENERALIZED OBJECTS. HERE THE SAME SET OF ALGORITHMS WORKS IN A SIMILAR WAY. AS A RESULT, WE RECEIVE ANOTHER LOWER LEVEL OF RESOLUTION, AND SO ON.

WALTER FREEMAN SUGGESTS TO TEST THE CONCEPT RELATED TO A SINGLE LEVEL. THIS CONCEPT THAT SEEMS TO BE TOO SIMPLISTIC SHOULD BE CAPABLE OF RESOLVING SOME SERIOUS EXAMPLES:

Try the following:

- Abstracting figures from undefined backgrounds
- Creating adaptive images of what to search for
- Prioritizing conflicting demands for mental workspace
- Generalizing and classifying items that are not linearly separable in n-space
- Translating between natural languages

I WOULD MEET THIS CHALLENGE WITH AN OPEN VISOR. LET US TRY TO SOLVE:

Test No. 1

Abstracting figures from undefined backgrounds

--I scan the image with a sliding window [SEARCHING] and store properties of the image [e.g. average intensity, color, etc.] at regularly selected coordinates.

--I hypothesize clusters based upon both properties similarity and adjacency [e.g. FOCUSING ATTENTION and COMBINING N-TUPLES]

--I promote clusters that I have discovered into a rank of objects [GROUPING]

--I am browsing my memory looking for similar objects [SEARCHING]

…and so on.

Before I start browsing my images, I allow for some combinatorics upon created objects: the hypotheses of strings are considered together with their vicinities, and within the vicinity a local SEARCHING is executed (testing of combinations). This combinatorial freedom depends on the uncertainty of the results of clustering. When I perform browsing of my memory together with exploring combinatorial multiplicity of choices that comes from uncertainty.

(If the complexity of all this is too high, the problem distributes itself to other levels of resolution. This will reduce the complexity drastically).
What I have described in the previous two paragraphs is actually a solution of the second problem from W. Freeman's list:

Test No. 2
Creating adaptive images of what to search for

It would not be proper for me to go trough all examples of the list. But if one wants to do it, one will easily find that the solution for most of these problems can be represented by the five elementary algorithms that together are sufficient for modeling what some might call a "generic intelligence".

All cases of "gestalt" known from the literature allow for doing this.

I would suggest to all of you to make this and/or similar experiments. I am sure that if one have not resolved many similar problems earlier, it was only for the reason that one knew for sure that this is impossible. Sometimes, the expectation of futility of the possible effort is even more frightening than the complexity of problem.

It is really chilling to read something like: ...if you try to do this "you might wind up with is a collection of Turing Machines, that can talk to each other, but nobody else."

Sure, better even not to try...

In the meantime, if this collection will be a hierarchy of Turing Machines, the long term outlook might be very promising.
Moderator

June 14, 2000

WE CONTINUE TO DISCUSS THE ABILITY TO HAVE A CONTEXT-INDEPENDENT MODEL OF INTELLIGENCE

Paul Davis wrote:

(AND I WILL COMMENT AFTER EACH STATEMENT. A. M.)

P. D. Reactions to the set of five:

1. We probably need multiple levels and perspectives of intelligence's components. Quantum mechanics is beautiful, but it's not of much help to someone working at the levels of classical statistical mechanics, thermodynamics, engineering laws like Navier Stokes, or even cruder engineering scaling laws. The periodic table may be the essence of chemistry in some sense, but it doesn't take organic chemists very far. As the story goes in discussion of complex adaptive systems, different levels have their own laws.

NO DOUBT ABOUT IT. THE ELEMENTS OF THE GENERIC INTELLIGENCE (PAUL CALLS IT "THE SET OF FIVE" BUT IT MIGHT BE "SIX" OR "SEVEN") IS A SET THAT IS PRESUMED TO WORK AT A SINGLE LEVEL OF RESOLUTION. AS ONE CAN SEE, A PART OF ITS FUNCTIONING IS CREATION OF GROUPS, I. E. BUILDING UP A REPRESENTATION FOR THE NEXT LEVEL OF LOWER RESOLUTION.

SO, MULTIPLE LEVELS EMERGE AS A RESULT OF NORMAL FUNCTIONING IF THIS SET OF FIVE (OR SIX, OR SEVEN).

P. D. 2. Perhaps the set of five is a reasonable place to start discussion regarding ONE level/perspective. I suspect that it is incomplete, and I note the comments here of Walter Freeman. Beyond that, however, I wonder what empirical/theoretical basis exists for this or another set of underlying components or mechanisms. I would be very interested in a related discussion, because I'd learn a lot. But I don't believe that it would be nearly as useful as its proponents might hope (back to item 1, above).

PAUL IS CONFIDENT THAT THE PHENOMENON OF BEING MULTIRESOLUTIONAL IS MORE IMPORTANT THAN PROCESSES AT A SINGLE LEVEL. CERTAINLY! I AGREE WITH

YOU, PAUL. BUT THE MULTIRESOLUTIONAL SYSTEM OF REPRESENTATION EMERGES BECAUSE OF THIS "SET OF FIVE"!

P. D. 3. I would think that using Gardner's components of intelligence would not be a bad starting point from the other end, although others may have better suggestions.

AS YOU KNOW FROM MY PREVIOUS MESSAGES THE PHENOMENON OF MULTIPLE INTELLIGENCE IS EASILY TAKEN CARE OF BY INTRODUCING A TRANSLATION FROM THE DOMAIN OF APPLICATION INTO A NEUTRAL (META) LANGUAGE. IS THIS TRANSLATION IMPORTANT? OF COURSE! SHOULD WE DEVOTE ATTENTION TO THIS PHENOMENON? YES, OTHERWISE WE WON'T BE ABLE TO HANDLE IT.

P. D. 4 . While some may have the OPINION that multiple intelligences is a myth or an expression of "prejudice" (a rather inflammatory term), I have seen nothing in the e-mail to justify this opinion. The periodic table wasn't postulated or asserted; it was built up from empirical observations and minitheories.

PAUL, LET US GO TO AMERICAN HERITAGE DICTIONARY:

[Prejudice 1. a. An adverse judgment or opinion formed beforehand or without knowledge or examination of the facts. b. A preconceived preference or idea. 2. The act or state of holding unreasonable preconceived judgments or convictions.]

NO, I DON'T THINK THAT THE TERM "PREJUDICE" IS OR SHOULD BE TAKEN AS AN INFLAMMATORY ONE. THIS IS RATHER A TIMELY WARNING.

> SPEAKING ABOUT PERIODIC TABLE: I WANT TO REMIND YOU, PAL, THAT THE PHLOGISTON THEORY WAS BUILT UP ALSO FROM EMPIRICAL OBSERVATIONS AND MINITHEORIES...

P. D. There is an extensive body of pyschological literature supporting--at that level of description--the notion of multiple intelligences (and the failure of the single G-factor hypothesis).

YES, BECAUSE THE IDEA OF INVARIANCE OF THE INTELLIGENCE (WITH AN INPUT TRANSLATOR) MIGHT BE DIFFICULT TO BEAR FOR MANY. INDEED, ONE MUST BE VERY RESPECTFUL OF THESE TONS OF SWEAT, BLOOD, AND TEARS SHED TO GAIN HIS/HER DOMAIN KNOWLEDGE. IT IS HARD EVEN TENTATIVELY TO ASSUME THAT ALL THIS IS JUST A TRANSLATOR WHILE REAL GENIUS IS A SYMBOLIC ALGORITHM! WOULD I VOLUNTARILY ADMIT THAT ALL HIDDEN TRICKS OF MY DOMAIN OF NUCLEAR PHYSICS ARE REALLY RESOLVED IN THE SAME WAY LIKE THE PROBLEMS OF CULINARY OR PLUMBING DOMAINS? NO WAY!

AGAIN, THIS IS THE ESSENCE OF THE HYPOTHESIS AT HAND:
1) AN INTELLIGENCE AT A LEVEL IS THIS SET [SEARCH*FOCUSING ATTENTION*GROUPING*SELECTION* (MAYBE SOMETHING ELSE)]
2) TOGETHER ALL OF THESE PRODUCE THE NEXT LOWER LEVEL OF RESOLUTION WHERE THE SAME ACTIVITIES ARE INITIATED
3) [AND SO ON]
4) TOGETHER, THE HIERARCHY OF THESE LEVELS IS EASILY COPING WITH NP-COMPLETE PROBLEMS
5) ALL MECHANISMS MENTIONED ABOVE CAN WORK IN THE SPECIFIC LANGUAGE OF A PARTICULAR DOMAIN AND EQUIP THEMSELVES WITH VARIOUS AND THE NEAT CORNER-CUTTING TRICKS APPROPRIATE FOR THE DOMAIN LANGUAGE.
6) AS FAR AS MACHINE INTELLIGENCE IS CONCERNED, ALL IT COULD BE DONE SYMBOLICALLY (IN A METALANGUAGE) IN THE SAME WAY IN ALL DOMAINS; JUST AT THE INPUT AND OUTPUT WE HAVE TO HAVE CORRESPONDING LANGUAGE(i)-->LANGUAGE(meta) AND LANGUAGE(meta)-->LANGUAGE(i) TRANSLATORS.

Moderator

June 14, 2000

FROM THE RESPONSES TO C.WEISBIN'S QUESTIONS

Thomas Whalen wrote:

I WOULD LIKE TO DISCUSS WITH ALL MEMBERS OF THE ADVISORY BOARD THE LIST OF QUESTIONS THAT I PROPOSED IN THE RESPONSE LETTER TO C. WEISBIN. LET ME KNOW WHAT DO YOU THINK, THIS IS VERY IMPORTANT.

This is the list of C. Weisbin's questions that the Workshop will try to answer:

1. What is the vector of intelligence (VI) that should be measured and possibly used as a metric for systems comparison?

a) understanding instructions expressed in language convenient for the human giving them. *This is sometimes natural language, sometimes human-oriented technical language.)

b) understanding goal specifications and working independently to achieve goals presented to it in a language and level of detail convenient to the human whose goals they are.

c) generating (sub)goals in a useful but surprising way so as to improve the well-being of the humans using the system.

2. Should VI be measured in addition, or instead of measuring the vector of performance (VP) determined by the regular specifications?

I think it comes to the same thing, just with a different emphasis

3. If two systems have the same VP, what is implied by the difference in their VI values? Can this difference be represented in $ units?

If VP does not include cost, then a more intelligent system would sometimes be more costly, sometimes cheaper. If VP includes cost, benefit, and risk, including all externalizes, then nothing else is economically interesting.

Example: it might be possible to someday build two Chinese rooms, one that "really understands" Chinese and the other which just follows stimulus--response rules. If so, the intelligent one will probably be cheaper to produce.

4. Is it possible (and meaningful) to have different VI measures:
a) goal-invariant, b) resource-invariant, c) time-invariant?

I don't understand the question.

5. What should be recommended as a test of VI and how to normalize VP so that comparison be performed at the same normalized value of VP.

While I don't think that VP and VI are identical, I don't see a sharp enough distinction to be able to "normalize." If a human has VI>VP we attribute it to poor motivation or else to a specific disability. If a human has VP>VI we attribute it to a fault in testing or to extraordinary motivation.

June 14, 2000

Kirstie Bellman responds:

Those Weisbin questions are a reasonable start. It will be interesting t see how quickly discussions emerge on the behavioral correlates of "understanding" within different environments or artificial ecosystems. (Kirstie Bellman)

June 15, 2000

ARE THE CONTEXT AND DOMAIN INDEPENDENT MODELS OF INTELLIGENCE POSSIBLE?

W. F. What I want to say is that intelligence is to be found in the capacity for defining objects, which requires action by the agent (robot, animal, human) in respect to goals that the objects are to make achievable.

A. M. You admit that this is something we can understand, model, and simulate with the help of computer. To make it clear you refer to the fact that:

W. F. This is straight-forward theory psychology from the pragmatist and gestalt schools, and it has been incorporated by a number of avant-garde roboticists.

A. M. And, yes, you admit that this straight-forward theories can be fully understood and even simulated with the help of computer: this is what is actually available

W. F. "defining of 'objects' that are to be measured as 'n-tuples', sought, attended, grouped, and evaluated precedes these operations."

A. M. However, you firmly believe that DEFINING OBJECT is what we still cannot fully understand, and this is why computationally it cannot be done IN ALL CASES:

W. F. Once the objects are defined, Turing Machines will do fine, but Turing Machines can't do that ." [defining the objects. A. M.].[I would add "in all cases" A. M.]

A. M. Yes, we have a problem with defining the objects if this is linked with our "wants". You finger exactly in this direction:

W. F. My only contribution is to show by modeling brain dynamics that biological brains have this capacity,"

[defining the objects. A. M.] ***and its exercise is well described by the theory of intentionality."***

A. M. Yes, but in numerous domains, we start implementing "intelligent systems" that are solving more simple problems of defining the objects. In many cases this operation is within our reach, and we perform it successfully. The efforts continue, sophistication grows. Then, I hope, that modeling of "intentionality" will be in our reach soon, too.

All of this was just an introduction to the expression of your big doubts concerning the concept that "intelligence" in various domains might be modeled by the same computational structure

W. F. "I don't really object to proposing a common feature of 'generic intelligence', which may be in a class with other ideals such as truth, beauty and justice,"

A. M. Walter, the only thing that I propose is to have a multiresolutional model of knowledge representation that will have at each level of resolution a model-set [searching*grouping*focusing attention*... ...*evaluation*selection] that will do a definition of objects at this level from the objects of the higher resolution defined at the level beneath.

It is my conviction that this system can work both in the domain language and in metalanguage. In the latter case, it can be considered a context-independent algorithm (model) of intelligence.

I am far from a desire to talk about spiritual and other hot air producing issues that are, as you are saying, "in a class with other ideals". Some participants of the discussion, called this context-independent intelligence: "generic intelligence".

I have no objections against any relevant term. For me, the essence of this is the most important issue.

In conclusion, you said:

W. F. "…but I doubt that it could support judgments more compelling than to say "system A is smarter than system B". To make it stick, you have to say what each can do, at what cost."

A. M. We will be able to say: "system A is smarter than system B" ? Not bad! We "have to say what each can do, at what cost" -- no doubt about it!
Moderator

June 16, 2000

TURING TEST, SUCCESS VS. LUCK, SUPERVISION AND AUTONOMY

Dan Repperger wrote:

"Some comments after reading the white paper:

(1) Your efforts to quantify intelligence, especially from the perspective of a machine present a difficult problem. Your example of a Chinese room negates the Turing test as a possible definition of machine intelligence."

A. M. DON'T YOU THINK THAT IT IS A RIGHT TIME TO STOP JUDGING INTELLIGENCE

BY A SIMPLE SKILL TO PRETEND BEING "INTELLIGENT" ? (A. M.)

D. R. "(2) The definition of J. S. Albus seems comprehensive enough and you transfer the responsibility to defining success, rather than being due to pure luck."

A. M. I HOPE THAT ONE COMPONENT OF THE VECTOR OF SUCCESS OF OUR MEETING WILL BE A CONSENSUS ON MEASURING THE SUCCESS OF IS FUNCTIONING

D. R. "(3) Your vector of intelligence on page 5, I thought, would have a component of the speed at which it accesses information. It did not have this component but you address it later on."

A. M. YOU ARE RIGHT: THE SPEED OF ACCESSING INFORMATION IS A MAJOR ISSUE

D. R. "(4) I agree that supervisory control and defining autonomy in subordinate systems is a key problem to be addressed in the next 10 years or so. The Air Force is very interested in this problem in the design of unmanned air vehicles."

A. M. SUPERVISORY CONTROL-->A DEGREE OF AUTONOMY-->AUTONOMY THAT MAXIMIZES EFFICIENCY -- PROBABLY, THIS WILL BE OUR PROGRESSION IN TIME.

Moderator

June 16, 2000

A RESPONSE TO THE DRAFT OF THE WHITE PAPER: FROM COMPUTING WITH WORDS ←TO CHINESE ROOM IN REAL CHINA

I. B. Turksen wrote:

B. T. "I have read your white paper with interest and enthusiasm. I agree with you that metrics of intelligence need to be developed. However, I would like to suggest that such metrics should be developed not just with "Computing with Numbers" paradigm in your reference to Lord Calvin, but as a synthesis of "Computing with Numbers" and "Computing with Words" paradigm of Lotfi Zadeh."

A. M. CERTAINLY, IT WOULD BE SUPERFICIAL TO UNDERSTAND THE NEED IN METRICS AS THE NEED IN A SOLELY QUANTITATIVE FORM, OR A FORMULA. ULTIMATELY, THE NEED IN A METRIC IS DETERMINED BY THE NEED TO COMPARE ALTERNATIVES, IN OUR CASE, TO COMPARE INTELLIGENT SYSTEMS: "WHICH ONE IS MORE INTELLIGENT," OR "WHICH ONE IS PREFERABLE FOR OUR NEEDS."

IF THIS PREFERENCE CAN BE FOUND WITHOUT NUMBERS AND RANKING CAN BE DONE AS A RESULT OF SOME LOGICAL INFERENCE -- SO BE IT!

B. T. "Which you indirectly say but not clarify as one should. For example, you write: '...by living creatures, and especially by humans: ability to work under a hierarchy of goals, ability to perceive the external world and organize objects, actions and situations,...' (quoted from the first page in the last paragraph). Note that humans at least do this with the use of their natural languages. Thus my point about the 'Computing with Words' of Lotfi. Note that he recently began to talk about 'Computing with Perceptions'."

A. M. ...WHICH REMINDS US THAT DETERMINING PREFERENCES MIGHT BE DONE NOT ONLY WITHOUT NUMBERS, BUT EVEN WITHOUT EXPLICIT LOGICAL INFERENCE: "AH! I LOVE THIS LANDSCAPE (OR THIS BEAUTIFUL FACE, OR THIS POWERFUL PAINTING). WELL, IN ALL THESE EXAMPLES THERE IS SOME INTERPRETATION OF PREFERENCE, AND THIS INTERPRETATION MIGHT BE DONE ON A PRE-LOGICAL LEVEL (IF IT EXISTS).

B. T. "Let me put it in a different way. Recently I participated in a teleconferencing with some Italian Colleagues. They have developed an artificial 'Nose'. They want to use their electro-mechanical device with novel sensor which provide lots of information. They have used 'principal component' analysis, neural networks, etc. all numerical based analysis. But they are aware that they cant represent humans' ability, e.g., sense of smell, to detect variations in food, e.g. , cheese, and drinks, e.g., wine. These are however expressed with linguistic variable that humans use. Clearly 'fuzzy set and logic' approach is a preliminary but effective way to begin and conceptualize such complex metrics of intelligence."

A. M. THIS IS ONE MORE ARGUMENT IN FAVOR OF NON NUMERICAL EVALUATION OF THE DEGREE OF INTELLIGENCE

B. T. "In page 6, item (g) of the White Paper, you talk about (CIRCLE) why not also include (ellipse) and other more complex shapes?"

A. M. YES, IN EVALUATION OF THE UNCERTAINTY FOR EACH COMPONENT OF THE VECTOR OF INTELLIGENCE ANY CONFIGURATION OF THE UNCERTAINTY ZONE CAN BE EXPECTED. I AM TALKING ABOUT "CIRCLE" BECAUSE IT SHOULD MEAN EQUALLY LARGE UNCERTAINTY FOR EACH COMPONENT OF THE VECTOR OF INTELLIGENCE.

B. T. "In page 7 of the White Paper, you talk about "gestalt" concept. Is there a relationship between gestalt representation and its word representation and the potential semiotic representation and its interpretation?"

A. M. WHEN WE ARE TALKING ABOUT "GESTALT" TODAY, WE ALL AGREE THAT THIS IS THE TERM THAT HAS BEEN INTRODUCED TO ACCOUNT FOR RECOGNIZING ENTITY FROM THE MULTIPLICITY OF ITS SEEMINGLY UNORGANIZED COMPONENTS. IT SEEMS REASONABLE TO EXPECT LINGUISTIC GESTALT SIMILAR TO THE GESTALT IN VISUAL PERCEPTION, AND GESTALT WORKING IN ANY SYSTEM OF SYMBOLIC REPRESENTATION, I. E. SEMIOTIC GESTALT.

B. T. In page 8 of the White Paper, J Searle and Chinese room experiment are mentioned. Let me tell you my personal experience in 1982 in Taipei Taiwan. A friend and I tried to locate a Bank with a map with Chinese characters. We were able to locate the Bank by matching the street labels on the map to the street labels on the street name plates. Even though we didn't know what the street names meant or how they were pronounced, we were able to find the Bank. Hence mission was accomplished and the goal was achieved without knowing the Chinese language."

A. M. THIS IS A FASCINATING STORY! BUT IT DOES NOT SAY ANYTHING GOOD ABOUT THE NOTORIOUS TURING TEST. INDEED, YOU DEMONSTRATED MULTIDIMENSIONAL, MULTIFUNCTIONAL INTELLIGENCE. FIRST, YOU WERE CAPABLE OF PUTTING IN CORRESPONDENCE THE NOISY 3D-REALITY OF THE CITY WITH THE NOT VERYCONGRUENT SYMBOLICS OF THE MAP. THEN, YOU DEMONSTRATED INTELLIGENCE BY UNDERSTANDING THAT THE CAPTIONS ON THE STREETS SHOWN IN THE MAP ARE THE SAME AS CAPTIONS ON THE STREET POSTS IN THE INTERSECTIONS. I APPRECIATE THE FACT THAT ALL OF THIS WAS IN A HIEROGLYPHIC SIGNS AND FINDING SIMILARITY BETWEEN HIEROGLYPHS IN DIFFERENT SOURCES REQUIRES INTELLIGENT OF SIMPLE SEARCH FOR SIMILARITY.

YES, YOU'VE DEMONSTRATED YOUR INTELLIGENCE!!!

BUT THE MAN SITTING IN THE CHINESE ROOM AND COMPARING WRITTEN HIEROGLYPHS BY SIMILARITY DEMONSTRATES ONLY A LITTLE BIT OF IT (AT LEAST HE UNDERSTOOD THE ALGORITHMS OF COMPARISON OF SIGNS AND SEARCHING IN A TABLE).

Moderator

June 16, 2000

DEFINITION OF INTELLIGENCE AND SEPARABILITY HYPOTHESIS

by B. Chandrasekaran

Consciousness is usually treated as an intrinsic property. I experience my being, but I don't experience your being. However, I usually hypothesize that you have the same property, consciousness. I'll never know for sure about you, just as you'll never know about me for sure.

Except for this theoretical caveat, we pretty much attribute to each other the property of consciousness and get on with our lives.

On the other hand, the term "intelligence" has both an extrinsic and an intrinsic connotation, depending on context. I watch an agent's behavior, and based on certain characteristics of the behavior, I may conclude that the agent's behavior is intelligent. In this sense, it is an extrinsic characterization. On the other hand, sometimes the term has an intrinsic connotation, that of having a mind, an entity that experiences being, experiences having thoughts, and so on. Thus, calling a thermostat intelligent is OK as long as it is intended as an extrinsic characterization (and as long as you agree with the criteria that were used for judging the presence of intelligence in thermostats).

Claiming intelligence in the intrinsic sense for them is much more problematic. In AI and cognitive science, people often slide from one sense of the term to the other without being aware that they are doing so. That is because, until very recently, there was no reason to separate the extrinsic and intrinsic senses of the term. The only entities that we called intelligent -- biological agents of various sorts, including humans -- showed intelligence extrinsically, and we were reasonably confident of attributing to them intelligence in the intrinsic sense.

But technology has made it necessary to separate the two senses. It now seems theoretically possible to conceive of entities that *behave* intelligently -- have intelligence in the extrinsic sense. But it is much harder to be certain about their having intelligence in the intrinsic sense. We are missing the full panoply of the evidential basis that allowed us to abduce intrinsic intelligence from evidence of extrinsic intelligence in biological agents.

The situation is not unique to the term "intelligence." There are other biologically based concepts that seemed pretty clear until recently, but now suddenly seem problematic. Consider the concept, "mother." One normally thinks that whether A is a mother of B is a matter of fact, not point of view. However, consider the case where woman A's fertilized egg is implanted in woman B's womb, and the infant that is born is immediately given to woman C, who adopts and raises the child.

*There is no self-evident answer to the question, "Who is the *real* mother of the child?". That is because contributing the egg, carrying the fetus in the womb and raising the newborn for several years are all typically done by one woman, and thus we normally do not separate these three properties associated with the concept of "mother." Depending on the purpose behind the question, however, we can answer the question. Thus, if the question is asked from the viewpoint of finding a donor for kidney, woman A is the mother. From the viewpoint of finding a woman who can suckle the infant, woman B is the mother. From the viewpoint of finding someone to solace the child when crying, woman C is the mother.*

I think that similarly, because of its natural orgin, at least two properties, perhaps more, come packed in one word "intelligence." If we don't recognize this and argue about what "really" is intelligence, and whether the thermostat is "really" intelligent, we will be like the people who argue about who "really" is the mother of the child in my story above.

It is possible to argue that the criteria used by Albus to attribute intelligence in the extrinsic sense to thermostats were too weak. Someone making this argument would hold that a meaningful characterization of extrinsic intelligence would seek to capture a much larger range of adaptation and behaviors than thermostats possess. Such an argument would identify higher mammals perhaps as a reasonable place to start, if not just focus on humans. But this is not a debate that has a clear correct answer either. One can choose one characterization as more interesting, more productive, and so on, but not as the one that is truly correct.

And, carrying this argument further, one might claim that the more complex forms of extrinsic intelligence can only be generated by systems that also have intelligence in the intrinsic sense. One way to interpret Penrose is that he is saying that the extrinsically intelligent behavior of a top-flight mathematician is not possible without certain essential characteristics of intrinsic intelligence. According to him, the mathematician directly "experiences" the truth of certain mathematical propositions. This capacity of intrinsic intelligence is essential for his behavior of finding a proof of the theorem.

While it may turn out to be true as a matter of empirical fact that we will only solve the problem of making artifacts that have a significant extrinsic intelligence only by making them have intelligence in the intrinsic sense, the latter is not logically a prerequisite for the former. At least, no one has shown it is. I have proposed what I call the "Separability Hypothesis" as a good working hypothesis for AI, namely, that it is not necessary to solve the problem of consciousness or intelligence in the intrinsic sense, to produce artifacts that show intelligence in the extrinsic sense. Those who are curious, see:

http://www.cis.ohio-state.edu/~chandra/separability.pdf

Chandra

B. CHANDRASEKARAN, OHIO STATE UNIVERSITY

June 17, 2000

THE IMPORTANT POINTS OF W. FREEMAN'S MESSAGE

Dear Advisory Board Members,

All of you received W. Freeman's message. I would like to emphasize some of the further developments that his message triggers.

1. ABOUT THE CONSCIOUSNESS

W. Freeman wrote:

I agree with Chandrasekaran, that "consciousness" need not be considered as a goal in machine intelligence, nor for that matter in biological intelligence and intentionality.

A. M.:

In other words, we have an additional support for the view that consciousness is a "GUI" for monitoring functioning of the system and its Umwelt.

2. INTELLIGENCE WITH AND WITHOUT LEARNING

W. Freeman wrote:

[An] intelligent system learns through practice. This rules out ordinary thermostats.

A. M.:

Is decision-making without learning "intelligent"? We can agree and accept that we will call "decision making without learning" a lesser degree of intelligence than "decision making with learning." It is possible even to postulate that a system with learning is supposed to be better performance, reliability, and so on. Of course, a cockroach learns ***not*** within a single generation. It learns at a lower resolution, at a specie level.

W. Freeman continues:

A child learns to recognize a spoon when it sees one, because it has practiced eating with it. Similarly, a machine can learn to recognize an electrical connector, if it can practice plugging itself in to recharge its batteries. You might say that a smart machine knows what it is doing, and a dumb one does not, but that invokes "knowing" of knowledge (or information), which is irrelevant to the design.

A. M.:

Walter, your last sentence would not be controversial if you rephrase it like this:

"...a smart machine learns what it is doing, and a dumb one does not, and that invokes "learning" of knowledge (or information), which is relevant to the design since the devices for learning should be designed."

3. USING CHAOS AS A TOOL OF RANDOMIZATION

Then, W. Freeman said:

The learning in biological systems depends on chaos for hypothesis formation. This process is more closely related to statistics than to logic,

A. M.:

Walter, the statistics does not exclude logic, neither logic is fully complete without the logic of statistics. The logic of class formation, the logic of cause-->effect derivations, the associated deductions, inductions, and abductions neither disappear not lose their strength.

Randomization for hypotheses formation is a legitimate tool of reducing the complexity of computations. Chaos is a tool of randomization to collect more or less persuasive statistics. If chaos is generated that does not help to randomize properly, the statistical results may happen to be deceptive even if they look meaningful.

Therefore, biological system are not unique in using the tools of randomization for complexity reduction. Yes, they learned about these fascinating tool before Neanderthals, and before Cro-Magnons, and even before Haken's Research Institute in Europe and Santa Fe in US started exploring these things.

But in the engineering, these methods are utilized without too much associating these tools with mechanisms of intelligence.

W. Freeman:

[As] Johnny von Neumann wrote, brains "lack the arithmetic and logical depth" that we expect in machines, so he concluded that whatever the language of the brain might be, if it has any, it is "not mathematics, or at least not what we consciously and explicitly call mathematics" (1958). In other words, brains don't have numbers, but they do have a "way" of functioning which is highly successful in certain domains.

A. M.:

All musings of great people sooner or later become interpretable. In this particular case, we should not overestimate this "number-versus-another way" dilemma. In the previous letter, I. Turksen commented on Computing With Words paradigm that opens room for any symbolic system to be a language of the brain. I am sure that you would not reject the hypothesis that the language of the brain is symbolic (proof: based on the definition of "language").

W. Freeman:

That "way" is simulated in my KIII model (Freeman 2000). I look on it as a "machine-in-embryo", but it can already do useful work, such as reliable pattern classifications, that no other existing system can do, at all. Of course, it is simulated in software using numbers, so it is 100-fold slower than the sensory system it models, but it is realizable in hardware that could do the tasks in real time. The use of stochastic chaos instead of deterministic dynamics is what I perceive to be missing in your suggested approach. Am I wrong?

A. M.:

Walter, from my previous comments you could already deduce that randomization (and "stochastic chaos" is just a useful tools of randomization) is a regular and legitimate computational measure of complexity reduction. We probably are not aggressively propagating the word Chaos, we limit ourselves with a milder term "randomization" but - we have plenty of examples.

Today, NIST uses randomization at all levels of resolution of the Autonomous Intelligent Vehicle control. My first use of randomization can be dated to 1982-1984. In the system of Computer Aided Conceptual Design of robotic structures, I used it for hypothesizing assemblies. If one don't use it, the amount of computations grows unbearably[1]. But, I would agree, that "randomization" as a term would lose the contest with the term "Chaos".

Moderator

June 17, 2000

PAUL DAVIS ON THE DISCUSSION WITH B. CHANDRASEKARAN

Paul responds to my discussion with B. Chandrasekaran. Please, take in account that some thoughts concerning consciousness belong originally to B. Chandrasekaran (the letter is attached in the end).

Paul Davis wrote:
"The term "performance" is associated with externally observable actions such as the maintaining a temperature task."

A. M.

[The specification of each target variable like "maintaining a temperature task" is actually more complicated. The specification sounds rather like: "maintain the temperature in the room within a particular interval of temperatures [from t1 - to t2]"] and in a more realistic scenario:

"maintain the expectation of the temperature within a particular interval [from t1 - to t2] while maintain the variance [sigma-t] within a particular interval [from sigma-t1 to sigma-t2]."

Then, each variable will be supplemented by the list of constraints like:

> [as you maintain what I asked above, please, do not exceed
> the total number of "on-off switching operations" higher
> than N; do not exceed the total consumed power (energy per time)
> higher than Pk, and so on].

---

[1] Paper by A. Meystel and M. Thomas on this subject was published in Proceedings of the IEEE Conference on Robotics and Automation, Atlanta, GA, 1984, pp. 220-229.

Paul, you are talking about the temperature. In the room, we always have a non-stationary arbitrarily shaped field of temperature. Usually, the temperature sensor is a part of the thermostat. It measures the value of temperature in the vicinity of the thermostat, and the temperature in other parts of the room is just assumed.

I would presume that a thoughtful engineer will install 5-7 sensors in different locations in the room and will require from this single thermostat to provide for, say, an average temperature to be within some interval. Donald Trump or a governmental diligent lab might be willing to distribute several thermostats in the room so that the particular temperature field be provided.

Paul Davis continues:

"when the white paper talks about intelligence, it includes things like memory, processing speed, etc. In one frame of mind, these "sound like" other measures of performance. "Boy, that hummer really performs: it's a gigaflop machine!"

A. M. This sounds like a measure of performance only if your real problem is not (or cannot be) specified properly. There are many such problems: they are UNDERSPECIFIED, and we are interested to have some measure of the system universality and/or smartness so that the unexpected factors would not caught us unprepared!

BUT WHAT SHOULD BE CONSIDERED AN INDICATOR OF SMARTNESS: THIS IS THE SUBJECT OF OUR INQUIRY!

P. D.

"Thus, one might think that we are going down the path of saying that both performance and intelligence are about doing tasks (e.g., crunching a billion arithmetic operations)."

A. M. I hope that all we have already understood that

a) in a well specified case this is plain WRONG

b) in the underspecified case to rely on some buzzwords from commercials (for laymen as well as for the scientists) is plain silly,

or as Paul is saying UNSATISFYING.

P. D. "That is correct - unsatisfying. It would hardly help with the Chinese Room problem. Even adding items such as number of objects discerned, or number of levels of resolution used, still don't sound like intelligence. Hmm."

A. M. The greatness of the moment is in the fact that we have realized already that the characteristics of the system can and should be looked at carefully. In our white paper we have divided them into two groups:

VECTOR OF PERFORMANCE (that characterizes the output variables) and

VECTOR OF INTELLIGENCE (that characterizes the properties and features of the system of control)

Paul whose language is a little bit different prefers to talk about

EXTERNAL (equivalent to OUTPUT, or PERFORMANCE set of variables), and INTERNAL (equivalent to INTELLIGENCE set of variables).

Then, the picture looks for him more peaceful:

P. D. There are two parts to the solution, I think. First, we need to distinguish between internal and external. If a machine has Gigaflop capability, that is an internal capability related to potential intelligence. To be sure, we might measure that capabilitiy by having the machine do a task. However, it might also be possible to study the "anatomy" of a machine and infer that it has parallel processing capability; if so, that would be another way to infer internal capability--one that doesn't require having the machine "perform."

A. M. But, of course! To study the ANATOMY of our intelligence is probably the only way to judge upon the future functioning especially if we are uncertain about it!

P. D. If it were possible to read some of the machine's programming, then we might infer that it has the capability to "detect" objects of certain classes, or even to give names to patterns that it "discovers." We might not know how the machine would perform at tasks, because we might not understand the totality of the programming, but we could at least see potentialities.

A. M. Hurray! This is one of the best descriptions of the Vector of Intelligence ONE could ever dream. This is how COMPUTER VISION people are talking about their systems: in the terms of classes available for distinguishing and interpreting. This is how PLANNING/CONTROL people are talking about their systems: in the terms of classes of terrain they can handle, and classes of obstacles that they can efficiently avoid. In both cases: vision and planning/control the description of classes can be general enough, yet adequately presenting the properties of the future problems.

P. D. ***It seems to me that, so far, we are on good ground distinguishing between performance and components of intelligence, or, perhaps, enablers of intelligence VI. Moreover, it seems to me that none of this is mystical.***

A. M. In other words, we are capable of specifying for all cases the relevant VP and VI.

Paul Davis goes on:

The other part to the solution requires doing something about the "emergent properties" business. Some aspects of intelligence seem to demand this. I don't think that we can claim to have tackled intelligence without at least building place holders in for capabilities such as world modeling, including world modeling that adds inferred features that were not already lurking in the machine's data base.

A. M. These are formidable observations! We, the people, all our life are doing one thing: (as Paul describes it) searching for "inferred features that were not already lurking" in our memory. Obviously, the contemporary audience has learned about the "emergent properties" first and only after this it has noticed that it actually infers new features.

[The next paragraph is more related to positions from the B. Chandrasekaran's letter (see a couple of letters back)]

P. D. ***And what about emotion and its machine analogues if there are some? Perhaps Crick's work is relevant here (e.g., his book using vision as something we more nearly understand that may be related to consciousness). Perhaps it would be possible to determine whether machines have and adapt internal models of the external world by giving it certain tests. We wouldn't necessarily "see" the machine's model (unless it was a simple program that we inserted in the first place), but we might be able to have a strong basis for inferring the existence of a model. Going back to our furry friends for examples, some of us believe that certain of their actions go beyond something explainable by straightforward "behavior:" we would argue that the animals are "thinking," although we haven't a clue what their mental "picture" or reasoning is like in any detail. We could be wrong (and, certainly, some scientists are vociferous in insisting that other animals don't think), but our inferences have some basis.***

[With this background, here are some P. Davis' responses to the questions earlier presented to Advisory Board Members. See Attachment 1]

1. Yes, I think there is a difference between VI and VP. One has to do with internal capabilities; the other has to do with accomplishment of externally observable tasks. The first may be inferentially measured by having the machine do tasks.

2. VI should me measured in addition to VP.

3. If two machines have the same VP, it might be because we only had a meager set of tasks and, as a result, didn't make the distinctions we might have. It certainly seems unlikely to me that we shall soon be able to infer VI from VP.

4. Some aspects of VI (the enablers of intelligence, if not intelligence itself) might be goal invariant, such as processing speed, memory, etc. I'm not sure, however, what is meant by the several invariances.

5. This is a really tough question and I don't understand yet how to do it, except in some very simple respects. Getting at the existence and richness of internal models seems important here.

(the question was: 5. What should be recommended as a test of VI and how to normalize VP so that comparison be performed at the same normalized value of VP. A. M.)

6. I would think that we could construct broad problem spaces, measure performances that give us hints about intelligence components, etc., without focusing on any particular problem area. If we did, however, the result would not be context independent, but rather information about how intelligence varied with context! Unless, of course, we did some gross averaging. I am very skeptical about simple measures in this business.

7. I think that resources are relevant, but shouldn't be allowed to dominate.

[This is the end of P. Davis' commented message]

Attachment 1

This is the list of questions that the Workshop is to answer:

1. What is the vector of intelligence (VI) that should be measured and possibly used as a metric for systems comparison?

2. Should VI be measured in addition, or instead of measuring the vector of performance (VP) determined by the regular specifications?

3. If two systems have the same VP, what is implied by the difference in their VI values? Can this difference be represented in $ units?

4. Is it possible (and meaningful) to have different VI measures: a) goal-invariant, b) resource-invariant, c) time-invariant?

5. What should be recommended as a test of VI and how to normalize VP so that comparison be performed at the same normalized value of VP.

The subsequent supplementary questions are ingrained (directly, or indirectly) in the main five questions:

a) how to form VI for various architectures?
b) should the questions 1 through 5 be related to intelligent systems, or autonomous systems, or both?
c) what is the protocol of dealing with uncertainty when the uncertain metric is to be applied in the procedures of decision making? for example how the uncertainty of planning affects the cost of goal achievement?
d) what are the guidelines in constructing the world model and determining its scope in the variety of applications? how the scope of "world model" affects the sophistication of intelligent behavior?
e) how are the questions 1 through 5 related (and the answers applied to) the systems that are working under a hierarchy of goals.
f) should a competition between intelligent systems be considered a valid method of judging VI value?

June 17, 2000

LEARNING, GOALS, INTELLIGENCE: COMPARATIVE AND DEVELOPMENTAL PSYCHOLOGY

Thomas Whalen wrote:

I think another good way to get perspective on the questions we have been wrestling with is to look at biological systems other than fully developed humans. There has been some discussion of animals already, but I think it could be made more systematic.

All biological organisms manifest learning by their genome, and what the socio-biologists call the "goals" of the genome. Whether this is per se a manifestation of "intelligence" is a metaphysical question.

An insect's behavior shows only this kind of learning, like a thermostat's behavior manifests learning and intelligence but not learning and intelligence of the thermostat's own, just the learning and intelligence of the thermostat's designers.

Higher animals such as mice show learning of their own. The trained behavior of a simple neural net or even regression equation also manifests learning of the system's own, but the goals and intelligence of the designers.

Carnivores like dogs and cats, and even more so primates, seem to have goals of their own beyond the goals given by instinct. Do our current autonomous constructed systems have goals of their own in this sense?

Does a gorilla or chimpanzee, especially one who has learned to use language or at least a language-like systems, have intelligence of its own?

More to the point, what does the question mean? Is it the same question as when we ask it about a constructed system?

Coming from another direction, a newborn infant's brain is physically immature as well as having vast amounts of learning ahead of it. Very young babies already manifest rapid learning of their own, overshadowing billions of years of genomic learning. But a newborn baby does not manifest intelligence of his or her own, while a five year old certainly does. If we watch the emergence of intelligence in children and collectively introspect about when we want to use the word "intelligent" we amy learn things useful in answering the same questions about constructed systems.

As an example, does the intelligence of a little child emerge as a unified phenomenon or do some "kinds" of intelligence emerge before others? Does the "vector of Intelligence" emerge in concert, or one element at a time?

I hope to break loose some time to review current comparative and developmental psychology looking for slues we can carry over into understanding constructed systems.

Tom Whalen
July 3, 2000

Answer by the Moderator:

Tom,

It is my deep conviction that your questions, both legitimate and interesting, CAN be addressed and answered within the formalisms that we use in the multiresolutional nested hierarchical planning/ control systems, when we introduce learning phenomena.

(see URLs: http://www.ee.umd.edu/medlab/papers/Final/Final.html

which serves to enable a structure similar to

http://www.ee.umd.edu/medlab/papers/trep/trep.html

or ask me about published references)

It is true, presently we do not build intelligent systems that have instincts enabling them to get involved in efficient learning. Living creatures have an interesting distinction from intelligent constructed creatures. Our robots are concerned with survival of

themselves

their team

their master

an assigned object.

Living creatures are concerned in addition with survival of their species, and we even don't know how exactly this thing is implanted within their architectures. My robot will defend, or bring information to his own team, to me and to an assigned subject because this is what I have implanted explicitly into its architecture, or its knowledge base.

Some people are saying: We can implant the ability for evolutionary development of the specie of my robot. We just cannot wait a couple of billions years to see how it will develop. They are right. We will have a problem of funding if we propose a type of research that should be even 100 years short.

Clearly, analyzing goals implanted into genome within the constructed unmanned robots, seems to be impractical. It is much easier to manipulate with its software at the stage of design. Thus, I cannot wait until he develops a feature, or a goal: I must predict what I want and prescribe both the goal and the feature.

Saying this, I do not feel sadness. I feel joy.

Maybe, because I was constructed to be an engineer.

Moderator

ARCHITECTURES OF INFORMATION EMPLOYED BY INTELLIGENT SYSTEMS

C. Landauer responds to A. Wild's Questions

C. L. We have been studying exactly the kinds of questions that Andreas Wild suggested, and we have answers for some of them and approaches for others.

A.Wild's questions:

1- How could an information system be architected such that heterogeneous elements, like different types of computation (reasoning ?), may coexist, interact and add value to each other ? What would be the interfaces between sub-domains looking like?

2- Can such a system evolve by including domains that became relevant after the system was built, or by modifying or eliminating some of the domains implemented at its "birth" ?

3- Is there a way for a system to control interfaces among its own sub-systems, e.g. define new ones, eliminate or modify existing ones ? Can a system re-architect itself ?

4- Can this happen across hierarchical boundaries without generating unbearable chaos ?

5- Is a non-hierarchical, self-configuring, heterogeneous system at all possible ? If yes, are there any rules to follow, are there impossible situations to avoid, or, alternatively, anything goes, and the solutions will be selected by trial and error ?

C. L. These are our[2] answers / approaches / expectations / hopes:

1-We have been writing about Constructed Complex Systems for some time now, providing them with a Knowledge-Based infrastructure that supports exactly this kind of heterogeneity, and a further property called Computational Reflection, which means that the

---

[2] K. Bellman and C. Landauer

system has a complete model of its own behavior (internal and external), to some level of detail, so it can examine and change its own functions and behavior[3].

2-We have added domains that only became relevant after the fact, since the system can defer until run time its search for relevant resources to apply to a problem (including the problem of interpretation of the problem statements in a language not defined until later in the system lifetime) - the system has no privileged resources at all, so anything can be changed, on the fly – this is partly achieved by explicit and uniform separation of the problems posed during system operation and the resources used to address those problems (the Problem Posing Programming Paradigm), and flexible mappings from problems to resources (Knowledge-Based Polymorphism), that together lead to a new approach to Generic Programming[4].

3-We have an approach to the creation and management of internal interfaces in a system, based on a new knowledge representation structure we called a "conceptual category", which separates the purpose of an interface from its appearence in code (we have found the term used much earlier for something different, but we intend to keep it for this data structure anyway, since it is the right term for what we are trying to model)[5]. We are even trying to arrange that a system can change its own basic symbol systems, since we have shown that it must, if it is to persist for an extended time in a complex environment[6].

4-We think so, but have not proven it

5-We think so and are trying to prove it - the key here is that not all of the computations can be in the application domain - many of them have to be in the ``organization of the computing system'' domain, that is, much more internal infrastructure needs to be available than

---

[3] How this works with autonomous computing systems is discussed in the following paper: C. Landauer, K. L. Bellman, ``Computational Embodiment: Constructing Autonomous Software Systems'', pp. 131-168 in Cybernetics and Systems: An International Journal, Volume 30, No. 2 (1999)

[4] This architecture is described in C. Landauer, K. L. Bellman, ``Problem Posing Interpretation of Programming Languages'', Paper etecc07 in Proceedings of HICSS'99: the 32nd Hawaii Conference on System Sciences, Track III: Emerging Technologies, Engineering Complex Computing Systems Mini-Track, 5-8 January 1999, Maui, Hawaii (1999); C. Landauer, K. L. Bellman, ``Generic Programming, Partial Evaluation, and a New Programming Paradigm'', ibid., Track III: Emerging Technologies, Software Process Improvement Mini-Track ibid; revised and extended as C. Landauer, K. L. Bellman, ``Generic Programming, Partial Evaluation, and a New Programming Paradigm'', Chapter 8, pp. 108-154 in G. McGuire (ed.), Software Process Improvement, Idea Group Publishing (1999)

[5] The approach is partly described in C. Landauer, ``Conceptual Categories as Knowledge Structures'', pp. 44-49 in A. M. Meystel (ed.), Proceedings of ISAS'97: The 1997 International Conference on Intelligent Systems and Semiotics: A Learning Perspective, 22-25 September 1997, NIST, Gaithersburg, Maryland (1997)

[6] The proof and discussion are in C. Landauer, K. L. Bellman, ``Situation Assessment via Computational Semiotics'', pp. 712-717 in Proceedings of ISAS'98: the 1998 International MultiDisciplinary Conference on Intelligent Systems and Semiotics, 14-17 September 1998, NIST, Gaithersburg, Maryland (1998)

in most systems, and it needs to be much more capable and knowledgable than in most systems (it is clear from AW's description that this was at least part of the problem for the example)[7].

We think we are making progress on all of these fronts, though more slowly than we would like.

Christopher Landauer

July 4, 2000

LARRY REEKER SHARES HIS THOUGHTS:

1. We Measure In More Than Numbers Alone

Though numbers are very useful, other entities are always involved in measurement, in various ways. The most obvious entities that are involved are linguistic ones. Words are used to indicate the dimensions of the measurement (mass, time, etc.) and the units used (kg., lb., sec., etc.). But there are other linguistic means by which caveats on the measurements are made, and these may be both numerical and non-numerical.

Numbers were invented for purposes of measurement, either of size (cardinal numbers) or of sequence (ordinal numbers). They were invented because they can provide a succinct, precise means of expressing size or sequence, and that is why people like them and have faith in them. But that can be a disadvantage, and qualifications are therefore necessary. One type of qualification has to do with precision or possible error, and other numbers can be used to measure such a qualification.

Additionally, there are descriptions of the data on which measurements are made. In information storage and retrieval, Recall and Precision are measures made numerically, but the text corpus over which the measurements are taken is important in interpreting these numbers. Similarly, the terrain over which a robot's navigation abilities are measured has a lot to do with the relevance of the evaluation to particular applications.

---

[7] An overview of the approach can be found in C. Landauer, K. L. Bellman, ``Architectures for Embodied Intelligence'', pp. 215-220 in Proceedings of ANNIE'99: 1999 Artificial Neural Nets and Industrial Engineering, Special Track on Bizarre Systems, 7-10 November 1999, St. Louis, Missouri (1999)

Despite qualifications, there is often a tendency to misuse numbers just because they are so handy and seemingly clear, whereas the qualifications are tedious and boring, often a little like reading an insurance policy or a legal contract. Thus one can be tempted to think that there are "lies, damned lies, and measurements" (to rephrase a popular adage about statistics).

I guess all of that is pretty obvious. It is also obvious that there are in principle ways to measure that are not numerical, as long as they involve lattice relations. One can define such a relation on a vector of numbers, and thereby rank the vectors according to some criterion. One is just ordinary normalization, but one can use weightings, too. For a particular application, a system evaluation vector could be multiplied by a vector that characterizes the needs of the system.

2. Lord Kelvin (A Historical Digression)

I have always had some reservations about Kelvin's famous statement on measurement, since really numerical measurements are neither necessary or sufficient for a scientific theory. Even as he wrote or uttered his famous statement,

"I often say when you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge of it is of a meagre and unsatisfactory kind...", he had to know that numbers are not a sine qua non for measurement.

He knew, of course, that line congruence was the equivalent of numerical comparison in Euclid's geometry, but not expressed in numbers, so that one could clearly find ways to express significant facts about significant domains without numbers. (Despite the breadth across which his brilliance was spread, he was first a mathematician, and the son of a mathematician.) His aphoristic claim that is sometimes cited, "To measure is to know", is probably a better statement of his real feelings, for that reason. But Kelvin knew how to say things that would be remembered (no wonder that he eventually became a dean!).

Kelvin also said "If you can not measure it, you can not improve it", which is certainly true in the wider sense that the Workshop on Performance Metrics in Intelligent Systems is about, even if we take as the measurement device contests, or even human judgments of gradation. It would be very difficult to define "improve" in a sense that is not circular, without expressing some measurement concept. "To enhance in value or quality" is a common definition, and value and quality refer to measures, even if subjective.

It is unlikely, however, that Kelvin would have accepted any measures that were overly subjective. He might well have asked for more than Turing's "imitation game", test if he had been alive later. If we look at the context of the statement above, we see some expression of qualifications that are also revealing:

"In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. [Here follows, "I often say that", as cited above.] but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be."

Here too, I would guess (for the reasons mentioned above) that "numerical" means "mathematical" in a wider sense than just numbers.

I would argue that in the last part of the quotation, Kelvin is wrong, but that he was making a valid point within the prevailing Anglo-American view of science in his time. Though Kelvin had been born almost 200 years after the death of Francis Bacon, he still lived in an era where the predominant view of science was due to Bacon, with contributions of Hume and other philosophers. Charles Darwin, working in roughly the same time frame that Kelvin was growing up, tells us that he (Darwin) "worked on true Baconian principles, and without any theory collected facts on a wholesale scale"

This collection of facts "on a wholesale scale" often lacked any more than subjective observation, and Kelvin's notion that measurement was important was potentially, in that framework, an important way of improving the inference of causal laws and disentangling phenomena that are only superficially related. But (to return to his contemporary Darwin) Lord Kelvin should certainly have recognized the contribution of the leap from collections (no matter how large) of phenomena to models (of which Darwin's "theory of evolution" was a model that advanced science materially). Had he done so, he would not have declared that such contributions (others of which he himself had made, as well as Darwin and others of the period) "scarcely advanced to the state of Science..."

3. Back to Measurement (After A Short Philosophical Digression)

Today, our view of scientific theory has changed from that held in the 19th Century. The bare-bones version of a scientific theory is that it consists of a model composed of abstract theoretical constructs and a calculus that manipulates these constructs in a way that can account for observations and accurately predict the value of experiments. The theoretical constructs have a relation with observed entities, properties and processes that may be quite abstract, not necessarily readily available to human senses, but following directly from calculations based on the theory. There are a number of principles applied to a model that give us increased confidence in the theory, but the one most relevant here is that we can measure the observed entities to confirm the predictions of the theories.

It is relevant to observe that the "calculus" mentioned above is used in the dictionary sense "a method of computation or calculation in a special notation (as of logic or symbolic logic)". That means that it may be numerical or non-numerical. In fact, as Herb Simon and Allen Newell pointed out so strongly almost a half century ago, that calculus might be expressed in the notation of a computer program, the better to speed its manipulation of the theoretical constructs.

With respect to the field of Artificial Intelligence, the point has sometimes been made that the value of individual research results is difficult to confirm, and this has been used to cast aspersions on the entire AI enterprise. It is not uncommon for this criticism to be made, at least implicitly, by people whose own knowledge of AI is "of a meagre sort", and then one suspects that it is motivated by feelings of the sort that Herb Simon was describing when he wrote:

"I continue to marvel at the fact that, after 45 years, the naysayers can still be taken seriously, when they deny that computers (sometimes) think, or place that happy possibility in the distant future. I am afraid that at the outset of our adventure I greatly underestimated the emotional need many members of our species have to believe in its uniqueness… Patience! All that will pass. (Herbert Simon, E-mail, 26 Jun 1999).

On the other hand, there are well-informed AI critics whose views often reflect the fact that systems have been described qualitatively in ways that cannot be backed up by objective evaluation. There have indeed been system developers with simplistic ideas about scientific theory, suggesting that their computational models were theories of actual human cognition merely on the basis of surface resemblance. These suggestions have also been of concern to the majority of scientists in the AI and Cognitive Science Communities, who have been in search of solid theoretical concepts to underpin the field. The issue here is not necessarily the fact that the measures are qualitative, however. It is the fact that they are not meaningful as part of a scientific theory. They are therefore vague in a way that Kelvin would not have admired; but more importantly, they cannot as easily lead to improvements, as Kelvin so aptly observed.

4. Some Implications for Intelligent Systems Metrics

Perhaps (and Lord Kelvin would like this idea, I hope) our emphasis on finding metrics can solidify the theoretical constructs of the field, as well as providing a means of measuring progress.

The key to doing this is not to think of evaluation only as measurement of some benchmarks or physical parameters (which I will call "behaviors") that are manifested in the operation of the systems being evaluated. We need to think in terms of the inner workings of the systems and how the parameters within them relate to the measured externally manifested behaviors.

Consider the measures of Recall and Precision as an example. Given a particular text corpus, one can consider various weighting schemes, use of a thesaurus, use of grammatical parsing that seeks to label the corpus as to parts of speech, etc., within a system and see how these items (I realize they are more resources than theoretical constructs) relate to precision and recall in the context of a particular corpus, or of a corpus with particular characteristics (these might be theoretical constructs). I believe this sort of thing has been done, but it is not a field that I have followed recently.

It may be more interesting in the Workshop to consider component systems for things like reinforcement learning (RL). There are a number of different techniques within the RL, all of which have many possible applications. The concepts include the states chosen, the reinforcement function, and the policy. The area is becoming quite sophisticated, and there are known facts about the relation of these to outcomes in particular cases.

Suppose that a reinforcement learning system constitutes a part of the intelligence of an intelligent system. There should be some way of predicting how that system would do upon encountering problems of a certain nature. By knowing how it chooses the concepts in its system and how they react on problems of that type, one can provide a partial evaluation of how effective the learning system would be. By obtaining such figures for all such subsystems, one could relate them to the performance of the full intelligent system.

I am working on a general framework of this sort, and hope to discuss it further at the Workshop. I hope that we will eventually be able use such a framework systematically relate measures, whether numerical, non-numerical, or a combination thereof, at all levels of the system, from internal capabilities to external performance, to, as Lord Kelvin might say, "advance to the state of Science".