Extended Metacognition for Artificially Intelligent Systems (AIS): Artificial Locus of Control and Cognitive Economy

Dr. James A. Crowder
Raytheon Intelligence and Information Systems
16800 E. Centretech Parkway, Aurora, Colorado 80011

Shelli Friess MA, NCC
Relevant Counseling
P.O. Box 4193, Englewood, CO 80155

Abstract - *Theories into human learning and cognition have led to much research into new methods and structures for Artificial Intelligence (AI) and Artificially Intelligent Systems (AIS) to learn and reason like humans. As we move toward completely autonomous AIS, the ability to provide metacognitive capabilities becomes important [Crowder and Friess 2011b] in order for the AIS to deal with entirely new situations within the environment it may find itself (e.g., deep space, deep undersea). Presented here are theories and methodologies for Constructivist Learning (CL) processes that provide the methodologies to allow completely autonomous AIS to understand, evaluate, and evolve its "Locus of Control [Watts 2003]."*

Presented will be the a discussion of how the use of AI learning systems, like Occam [Crowder and Carbone 2011a] and PAC learning can be combined with Cognitive Economy concepts to provide this constructivist learning process to allow a Locus of Control evolution within the AIS. The goal here is to provide the AIS with a fully autonomous, cognitive framework that would be required for autonomous environmental interaction, evolution, and control.

In addition, provided are the mathematical constructs, based in Banach Spaces and Lebesque's work in Bounded Variability, that will provide the basis for Cognitive Economy structures in Artificially Intelligent Systems (AIS), allowing the AIS to operate in a "Bounded Rationality" mode, similar to humans, that will allow the autonomous system to function in new, unforeseen, and challenging environments it may find itself in. Natural intelligence filters out irrelevant information (either raw sensory perception information or higher-level conception information), and categorizes the problem representations to allow for maximum information processing with the least cognitive effort.

This work is based on the use of Intelligent Software Agents (ISAs) [Crowder 2010a] which will represent the world (its tasks, goals, and information) in terms of the reward values associated with different actions when those features of its abilities are active.

Keywords: Metacognition, Locus of Control, Cognitive Economy

1 Introduction

Intelligence reveals itself in a variety of ways, including the ability to adapt to unknown situations or changing environments. Without the ability to adapt to new situations, an intelligent system is left to rely on a previously-written set of rules. If we truly desire to design and implement autonomous AI Systems (AIS), they cannot require precisely-defined sets of rules for every possible contingency. The questions then become:

- ***How does an autonomous AI system construct good representations for tasks and knowledge as it is in the process of learning the task or knowledge?***
- ***What are the characteristics of a good representation of a new task or a new piece of knowledge?***
- ***How do these characteristics and the need to adapt to entirely new situations and knowledge affect the learning process?***

The ISAs, having bounded cognitive resources, would reacted to three aspects of Cognitive Economy to create a Bounded Rationality set of goals for a given set of ISAs generated to solve a particular problem or situation. These are:

1. The size of the feature set – how many "features" are required to define the success of each task
2. The "fuzzy" relevance of each feature for the tasks
3. The preservation of necessary distinctions for success in each task

The AIS's cognitive components would autonomously define, for each ISA, a Banach Space for that ISA's goals and tasks and would then consider the set of ISA Banach Spaces as a set of bounded variations, the sequence of which (through ISA collaboration) produces an acceptable solution to the situation(s) or task(s) at hand.

The Cognitive Economy methods will be described and a discussion will be provided, illustrating how these Cognitive Economy and Bounded Rationality concepts affect the overall learning aspects of an autonomous AIS.

In addition, when considering autonomous AIS, we must consider its need to interact and learn from its environment, and we have to ask ourselves "what is reality?" We have to establish how the AIS would interpret their reality. One of the issues that humans deal with that assists in their understanding of reality, or their world around them and how they need to interact, is their concept of "Locus of Control." **Locus of control** is a term in psychology that refers to a person's belief about what causes the events in their life, either in general or in a specific areas such as health or academics. Understanding of the concept was developed by Rotter [Rotter 1954], and has since become an important aspect of personality studies.

2.0 Artificial Locus of Control

Locus of control refers to the extent to which individuals believe that they can control events that affect them. Individuals with a high internal locus of control believe that events result primarily from their own behavior and actions. Those with a high external locus of control believe that powerful others, fate, or chance primarily determine events. Those with a high internal locus of control have better control of their behavior, tend to exhibit more political behaviors, and are more likely to attempt to influence other people than those with a high external locus of control; they are more likely to assume that their efforts will be successful. They are more active in seeking information and knowledge concerning their situation.

Locus of control is an individual's belief system regarding the causes of his or her experiences and the factors to which that person attributes success of failure. It can be assessed with the Rotter Internal-External Locus of Control Scale (see Figure 1). Think about humans, and how each person, experiences an event. Each person will see reality differently and uniquely. There is also the notion of how one interprets not just their local reality, but also the world reality [Botella 2011]. This world reality may be based on fact or impression.

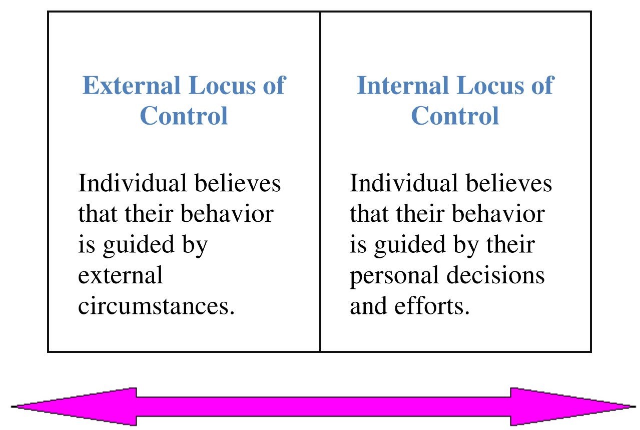


Figure 1 – The Rotter Locus of Control Scale

Take a car accident as an example. There are two people who witness a car hit a motorcycle. The police at the scene are supposed to evaluate the facts to determine what happened. The officer may use measurement tools that are supported by mathematical equations, such to be able to determine the speed at impact or where the impact happened. The officer may measure skid marks or measure the distance between vehicles. The officer is gathering factual data. Let's consider this juried evidence and legitimate evidence. Given how the world measures and uses universal mathematic equations, this evidence can be measured and re-measured by thousands of people and likely even machines. Back to the accident, when asked by the police officer, each human witness can recall the event as if they were watching it again, a step by step recount. Each person's story likely has unique qualities depending on how they conceptualize the incident. We use eyewitness testimony all the time. Even though each witness talks a slightly different story, we use it. We know, by eyewitness testimony studies, that often times the recalled event is very different than the actual event. Let's say in this example both people recalled the event similarly except the color of the car that hit the motorcycle. Perhaps even whether the car hit the motorcycle or the motorcycle hit the care recount differs. The fire truck blocks the view of each eyewitness so they cannot confirm the color of the car as they recount the event. Each person has had a legitimate experience even if they code the color of the car differently. Factually legitimate the car and bike collided at a specific rate of speed at a specific location. Emotionally legitimate is the witnesses' personal experience. To one witness the car was green to the other it was blue. Thus, with this incident we have three realities. One of the facts that we can measure by juried tools and the reality of each of the players in the scene; all experiencing the same event but each in his own unique way. Each reality is legitimate.

For further thought let's then consider Constructivist Psychology. According to "The internet Encyclopedia of Personal Construct Psychology" the Constructivist philosophy is interested more in the people's construction of the world than they are in evaluating the extent to which such constructions are "true" in representing a presumable external reality. It makes sense to look at this in the form of legitimacies. What is true is factually legitimate and what is peoples' construction of the external reality is another form of legitimacy. Later on we can consider the locus of control in relation to internal and external legitimacies or realities. You are correct if you are thinking that AIS is not human and will not have human perceptions. AIS may have AIS perceptions and realities. Thus, a mentor will be necessary. That mentor will need to understand AIS as AIS and be able to understand AIS in a human way, a human reality. After all, isn't this what makes AIS autonomous?

3 AIS Constructivist Learning

Constructive psychology is a meta-theory that integrates different schools of thought. According to the above cited article:

> Hans Vaihinger (1852-1933) asserted that people develop "workable fictions". This is his philosophy of "As if" such as mathematical infinity or God. Alfred Korzybski's (1879-1950) "System of Semantics" focused on the role of the speaker in assigning meaning to events. Thus, constructivists thought that human beings operated on the basis of symbolic or linguistic constructs that help navigate the world without contacting it in any simple or direct way. Postmodern thinkers assert that constructions are viable to the extent that they help us live our lives meaningfully and find validation in shared understandings of others. We live in a world constituted by multiple realities social realities, no one of which can claim to be "objectively" true across persons, cultures, or historical epochs. Instead, the constructions on the basis of which we live are at best provisional ways of organizing our "selves" and our activities, which

could under other circumstances, be constituted quite differently.

For AIS with Constructivist Learning, the AIS cognitive learning process would be a building (or construction) process in which the AIS cognitive system builds an internal illustration of its learned knowledge-base, based on its experiences and personal interpretation (fuzzy inferences and conceptual ontology [Raskin & Taylor 2010a and Taylor & Raskin 2011a]) of its experiences. AIS Knowledge Representation and Knowledge Relativity Threads [Crowder and Carbone, 2011c], within AIS cognitive system memories would be continually open to modification, and the structures and linkages formed within AIS short-term, long-term, and emotional memories [Crowder and Friess, 2010b], along with its Knowledge Relativity Threads [Crowder and Carbone 2011c], would then form the bases for which knowledge structures would be created and attached to AIS memories.

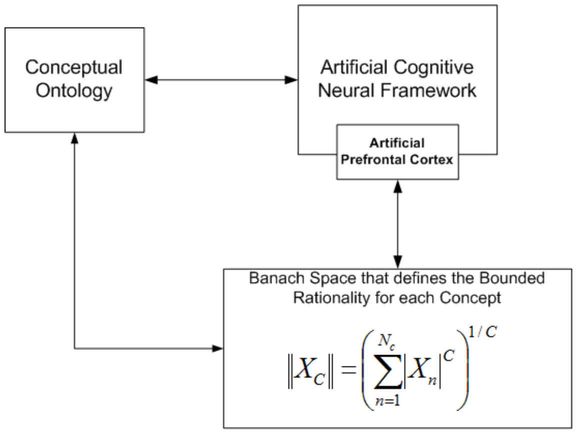
One of the results of the Constructivist Learning process with the AIS would be to gradually change its "Locus of Control" for a given situation or topic, from external (the system needing external input to make sense, or infer, about its environment) to internal (the AIS having the cumulative constructive knowledge-based of information , knowledge, context, and inferences to handle a given situation internally); meaning the AIS is able to make relevant and meaningful decisions and inferences about a situation or topic without outside knowledge or involvement. This becomes extremely important for completely autonomous AIS.

4 Bounded Conceptual Rationality (Cognitive Economy)

Bounded rationality is a concept within cognitive science that deals with decision-making in humans [LaBar and Capeza 2006]. Bounded rationality is the notion that individuals are limited by the information they have available (both internally and externally), the finite amount of time they have in any situation, and the cognitive limitations of their own skills. Given these limitations, decision making becomes an exercise in finding an optimal choice given the information available. Because there is not infinite information, infinite time, nor infinite cognitive skills, humans apply their rationality after simplifying the choices available, i.e., they bound the problem to be solved into the simplest cognitive choices possible [Jones 1999].

Any AIS must suffer the same issues. An autonomous system, by definition, has limited cognitive skills, limited memory, and limited access to information. The Locus of Control concepts discussed earlier assist AIS in determining which situations can be handled internally vs. externally, but still in any situation there is limited information, time, and cognitive abilities. This is particularly true if the system is dealing with multiple situations simultaneously. In order for the system to not become overloaded, we believe autonomous systems must employ strategies similar to human bounded rationality in order to deal with unknown and multiple situations they find themselves in. This involves creating mathematical constructs that can be utilized to mimic the notion of bounded rationality within autonomous AIS.

For this we look to Banach Space theory, tied into Constructivist Learning concepts [Botella 2011] for autonomous AIS. As concepts are learned and stored in the AIS conceptual ontology [Raskin & Taylor 2010a], Banach Spaces are defined that are used to bound the rationality choices or domains for that concept. As we "construct" these concepts and the Banach Spaces that bound them, the combination of Banach Spaces then defines the Conceptual Rationality for the Autonomous AIS. Figure 2 illustrates this concept.



where: $$X_n = \frac{1}{N}\sum_{i=1}^{N_c} G(x - x_i, \sigma I)$$

and: $$H_R(X_C) = \int_{-\infty}^{\infty} p_f(x)dx$$

where p_f is the fuzzy membership measurement

Figure 2 – AIS Bounded Conceptual Rationality

These Banach Spaces that define the bounds for each learned concept are utilized when the AIS must reason, or perform decision making. When there are restricting limitations on time, resources (as determined by the resource manager, e.g., artificial prefrontal cortex), and available information, the bounds of these Banach Spaces would be tightened or loosened to allow the AIS to deal with multiple situations, or situations that are time critical. This allows AIS to decide what is a "good enough" solution to a given problem or set of problems, and to adjudicate between competing resources, priorities and overall goals.

6 Conclusions and Discussion

What we have presented here are initial concepts and methodologies for what we believe are essential cognitive skills that autonomous systems must have in order to deal with and survive in real-time extreme environments. As we push for systems that think, learn, and adapt, we must provide these systems with cognitive skills similar to human processes in order to be able to deal with and survive real-time situations they find in their environments. This is very preliminary work and much more remains in order to put these concepts into practice.

7 References

1. Botella, L. 2011. Personal Construction Psychology, Constructivism, and Postmodern Thought. Found at http://www.massey.ac.nz/~alock/virtual/Construc.htm.
2. Crowder, J. 2010a. Flexible Object Architectures for Hybrid Neural Processing Systems. *Proceedings of the 12th annual International Conference on Artificial Intelligence*, Las Vegas, NV.
3. Crowder, J. 2010b. Operative Information Software Agents (OISA) for Intelligence Processing. *Proceedings of the 12th annual International Conference on Artificial Intelligence*, Las Vegas, NV.
4. Crowder, J. and Carbone, J. (2011a). Occam Learning through Pattern Discovery: Computational Mechanics in AI Systems. *Proceedings of the 13th annual International Conference on Artificial Intelligence*, Las Vegas, NV.
5. Crowder, J. and Carbone, J. (2011b). The Great Migration: Information to Knowledge using Cognition-Based Frameworks. *Springer Science*, New York, NY.
6. Crowder, J. and Carbone, J. (2011c). Recombinant Knowledge Relativity Threads for Contextual Knowledge Storage. *Proceedings of the 13th annual International Conference on Artificial Intelligence*, Las Vegas, NV
7. Crowder, J. and Friess, S. (2010a). Artificial Neural Diagnostics and Prognostics: Self-Soothing in Cognitive Systems. *Proceedings of the 12th annual International Conference on Artificial Intelligence*, Las Vegas, NV.
8. Crowder, J. and Friess, S. (2010b). Artificial Neural Emotions and Emotional Memory. *Proceedings of the 12th annual International*

Conference on Artificial Intelligence, Las Vegas, NV.

9. Crowder, J. and Friess, S. (2011a). The Artificial Prefrontal Cortex: Artificial Consciousness. *Proceedings of the 13th annual International Conference on Artificial Intelligence*, Las Vegas, NV.
10. Crowder, J. and Friess, S. (2011b). Metacognition and Metamemory Concepts for AI Systems. *Proceedings of the 13th annual International Conference on Artificial Intelligence*, Las Vegas, NV.
11. Crowder, J., Scally, L., & Bonato, M. 2011. Learning agents for Autonomous Space Asset Management. *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, Maui, HI.
12. Jones B. 1999. Bounded Rationality. Annual Review of Political Science, 1999.2:297-321.
13. Kosko, G., "Fuzzy Cognitive Maps," *International Journal of Man-Machine Studies*, 24: 65-75.
14. LaBar KS and Cabeza (2006) Cognitive neuroscience of emotional memory. *Nat Rev Neurosci 7: 54-64*.
15. Neimeyer, R. 2004. The Internet Encyclopedia of Person Constructivist Psychology, located at: http://www.pcp-net.org/encyclopaedia/pc-theory.html.
16. Newell, A. 2003. Unified Theories of Cognition. *Harvard University Press*, Cambridge, MA.
17. Raskin, V., Taylor, J. M., & Hempelmann, C. F. 2010a. Ontological semantic technology for detecting insider threat and social engineering. *New Security Paradigms Workshop*, Concord, MA.
18. Rotter, J. 1954. Social Learning and Clinical Psychology. *Prentice-Hall*, New York, NY.
19. Taylor, J. M., & Raskin , V. 2011a. Understanding the unknown: Unattested input processing in natural language, *FUZZ-IEEE Conference*, Taipei, Taiwan.
20. Watts, R. 2003. Therapy as a Rational Constructivist Approach. *The Family Journal: Counseling and Therapy for Couples and Families*, Vol. 11. No. 2, pp. 139-147.
21. Yang Y, Raine A (November 2009). "Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis". *Psychiatry Res* **174** (2): 81–8. doi:10.1016/j.pscychresns.2009.03.012. PMID 19833485.