Artificial Psychology: The Psychology of AI

James A. Crowder
Raytheon Intelligence, Information, and Services
Aurora, CO 80112, USA

and

Shelli Friess
Relevant Counseling
Englewood, CO 80115, USA

ABSTRACT

Having artificially intelligent machines that think, learn, reason, experience, and can function autonomously, without supervision, is one of the most intriguing goals in all of Computer Science. As the types of problems we would like machines to solve get more complex, it is becoming a necessary goal as well. One of the many problems associated with this goal is that what learning and reasoning are have so many possible meanings that the solution can easily get lost in the sea of opinions and options. The goal of this paper is to establish some foundational principles, theory, and concepts that we feel are the backbone of real, autonomous Artificial Intelligence. With this fully autonomous, learning, reasoning, artificially intelligent system (an artificial brain), comes the need to possess constructs in its hardware and software that mimic processes and subsystems that exist within the human brain, including intuitive and emotional memory concepts. Presented here is a discussion of the psychological constructs of artificial intelligence and how they might play out in an artificial mind.

Keywords: Artificial Psychology, Artificial Cognition, Emotional Memory, Artificial Intelligence.

1. IMPORTANT INFORMATION

In order to not only design and implement these structures, but also understand how they must interact, cooperate, and come together to form a whole system, we must understand how these structures function within the human brain, and then translate these into how they must function within our "artificial brain." If our system is to possess an "artificial consciousness" then we must understand cognition, intuition, and other capabilities that humans possess [6, 10, 11].

In addition, if we are to create a complete artificial intelligent system, we need to understand how such a system would be received and perceived by people. The reverse is also true in that we must try to understand how the artificial intelligent system will react and perceive people [2, 3, 4, 5].

First, we will explore the concept of "Artificial Psychology" where we look at what it means to have Artificial Intelligence Systems (AIS) resemble human intelligence and when we need to start worrying about the "Psyche" of the Artificial Intelligence system

2. ARTIFICIAL PSYCHOLOGY

Psychology is the study of mental processes and behavior of individuals. Artificial Psychology is then the study of the mental processes of an Artificial Intelligence System (AIS) similar to humans [3, 4]. It is about the artificial cognitive processes required for an artificially intelligent entity to be intelligent, learning, autonomous and self-developing [4, 5]. In psychology there are several specialties or focuses of study. Take for example cognitive psychology that studies how the brain thinks and works. This includes learning, memory, perception, language, logic [5, 6, 13, 14]. There is also developmental psychology that considers how an individual adapts and changes during different developmental stages and what is appropriate to consider of a human based on development [17, 18, 19, 20, 21]. There is sports psychology that considers how to affect individual performance and how performance affects the individual. So Artificial Psychology for the purposes of this paper contains the artificial mental process considered necessary to create intelligent, autonomous, self-evolving, artificially cognitive systems. The AIS must mimic human processes in order to be intelligent. After all, isn't the human at the top of the intelligence spectrum?

Artificial Psychology is a theoretical discipline which was first proposed by Dan Curtis in 1963. This theory states that Artificial Intelligence will approach the complexity level of human intelligence when the artificially intelligent system meets three very important conditions:

- Condition 1: The artificially intelligent system makes all of its decisions autonomously (without supervision or human intervention) and is capable of making decisions based on information that is 1) New, 2) Abstract, and 3) Incomplete.

- Condition 2: The artificially intelligent system is capable of reprogramming itself (evolving), based on new information and is capable of resolving its own programming conflicts, even in the presence of incomplete information.[1]

[1] This means that the artificially intelligent system autonomously makes value-based decisions, referring to values that the artificially intelligent system has created for itself.

- Condition 3: Conditions 1 and 2 are met in situations that were not part of the original operational system (part of the original programming), i.e., novel situations that were not foreseen in the design and initial implementation of the system.

We believe that when all three conditions are met, then the possibility will exist that the artificially intelligent system will have the ability reach conclusions based on newly acquired and inferred information that has been learned and stored as memories. At this point, we believe the criteria exist, such that the new field of Artificial Psychology needs to be put into place for such systems [4, 5, 6, 7].

The ability of the artificially intelligent system to reprogram, or self-evolve, through a process of self-analysis and decision, based on information available to the system cannot provide the mechanisms for internal inconsistencies within the system to be resolved without adaptation of psychological constructs to AIS methodologies and strategies, and therefore, artificial psychology, by definition, is required.

Current theory of artificial psychology does not address the specifics of how complex the system must be to achieve the conditions presented above, but only that the system is sufficiently complex that the intelligence cannot simply be recorded by a software developer, and therefore this subject must be addressed through the same processes that humans go through to. Along the same lines, artificial psychology does not address the question of whether or not the intelligence is actually conscience or not.

3. ARTIFICIAL COGNITION: WHAT DOES IT MEAN TO BE COGNITIVE?

Cognition is all about thinking. According to the book Ashcroft [22], "…cognition is the collection of mental processes and activities used in perceiving, remembering, thinking, and understanding, as well as the act of using those processes." Adding the term artificial identifies that the nonhuman system is a representation of a living intelligent system. Artificial Cognition refers to how the artificially intelligent machine learns, integrates, recalls, and uses the information that it receives [6, 7, 14, 15, 16]. It is also about how it receives the information. It is difficult at best to create an AIS as complex as human thinking. It is thought that a better understanding of human processes may come from being able to create a truly intelligent machine [22]. It seems that the reverse is also true. Thus, we have a whole new field, Artificial Cognitive Science.

4. ARTIFICIAL INTUITION: WHAT DOES IT MEAN TO BE INTUITIVE?

Saying what does it mean to be intuitive is basically asking the question: what does it mean to trust your gut? Another way to say it is to use your heart not your head. Intuition is another way of problem solving that is not the same as logic. According to Monica Anderson [24]:

> "Artificial intuition is not a high-level Logic model so there is no model to get confused by the illogical bizarreness of the world. Systems with intuition then can operate without getting confused with things such as constantly changing conditions, paradoxes, ambiguity, and misinformation."

In her article she also states that this does not mean that sufficient misinformation won't lead such a system to make incorrect predictions, but it means that the system does not require all information to be correct in order to operate. Intuition is fallible, and occasional misinformation makes failure slightly more likely. The system can keep multiple sets of information active in parallel (some more correct than others) and in the end, more often than not, the information that is 'most likely' to be correct wins. This happens in humans, and will happen in Artificial Intuition based systems. It greatly depends on how 'most likely' is defined. If it is only based on the experience of the system, then it can continually fall prey to anchoring and/or the availability heuristic. This implies the need to be supplied with initial data and the use of intuitive guides/rules (heuristics) to help during intuitive conceptual development.

The goal in our AIS is to provide the cognitive intuition required to deal with the world in a real-time, autonomous fashion. Included within the cognitive structure of our the AIS is a Dialectic Argument Structure, which is a methodology constructed for the AIS to deal with conflicting and ambiguous information and will allow the system the "cognitive gut" to deal with our paradoxical and ever changing world. In fact, according to Wired.com [25] IntuView, an Israeli high-tech firm has developed "artificial intuition" software that can scan large batches of documents in Arabic and other languages. According to the company's website, this tool "instantly assesses any Arabic-language document, determines whether it contains content of a terrorist nature or of intelligence value, provides a first-tier Intelligence Analysis Report of the main requirement-relevant elements in the document." So if we are going to provide the AIS with the ability to "follow its gut," do we then have to provide it with the emotions we use to make such decisions [9, 13]?

5. HUMAN VS. MACHINE EMOTIONS

In humans, emotions are still about thinking. According to Marvin Minsky [26]:

> "The main theory is that emotions are nothing special. Each emotional state is a different style of thinking. So it's not a general theory of emotions, because the main idea is that each of the major emotions is quite different. They have different management organizations for how you are thinking you will proceed."

Latest theories look at emotions as the way the brain consciously explains what has happened at a subconscious level. That is, we respond subconsciously (which is faster than subconscious thought) and the brain "explains" what happened with emotions, or arousal states (e.g., fear). So, for the AI system, the emotions produced are a reflection of the type of situation with which the system is dealing.

We can think of emotions in terms of arousal states. When a person is calm and quiet they are more likely to be able to take things in and listen, learn, or problem solve. Think about another emotional state, terror for example. When we are in a

state of terror we are not likely to be able to form complex problem solving. Typically with humans, that is why it is recommend to safety plan or practice evacuations. So at the time of crisis or terror the brain doesn't have to perform problem solving. Instead we can just follow the pre-thought out plan. Another example might be the instant you are in a car accident. The body is flushed with adrenaline, heart pounding, hands shaking, probably not a time to work out a calculus problem, for most of us anyway. Often times, emotional states also influence our perception. Take depression for example. It is not likely that a clinically depressed person will simply find the positives of a given situation. There is likely a more doom and gloom recognition. Take a rainy morning, a depressed person who has difficulty finding enjoyment, even if they like the rain may decided to stay in bed, whereas a non-depressed person, who may not even like the rain may, be able to determine that the rain offers opportunity to splash in the water or carry your favorite umbrella.

However, research has also shown that minor stress can actually be good. This seems to point toward the notion that our brains are wired to pay attentions to certain things and that emotions (stress and fear in particular) are an indication that we should pay attention. In their work on Artificial Emotional Memories [3], Crowder and Friess investigated how to utilize these Emotional Memories in order to provide long-term implicit emotional triggers that provide artificial subconscious primers, based on situational awareness metrics.

Similarly, for an artificially intelligent entity, emotions are states of being. If the system is overloaded can it determine what resources to allocate to return to the homeostatic state or state of optimal performance? If for example there are enough indicators to arouse fear can the mediator, so to say, keep operations performing with the correct amount of urgency? Take terrorist threats for example. If an AIS is given enough information to conclude an attack on the country is imminent in the next 24 hours, could the system increase resources to determine the best plan of action? Just as the human level of arousal may contribute to what decisions we make, such as minor chest pain from strained muscle may result in taking an anti-inflammatory or severe chest pains may cause us to call the paramedic [27].

6. BASIC EMOTIONS

In his book on Emotion and Intuition [1], Bolte concluded:

> "We investigated effects of emotional states on the ability to make intuitive judgments about the semantic coherence of word triads… We conclude that positive mood potentiates spread of activation to weak or remote associates in memory, thereby improving intuitive coherence judgments. By contrast, negative mood appears to restrict spread of activation to close associates and dominant word meanings, thus impairing intuitive coherence judgments."

Bolte found a clear tie between emotions and the ability to have or exhibit intuition. This drives us to a model of basic emotions with the AIS that allow the system to channel resources and find solutions, based on emotional responses to its interaction with its environment. For the purposes of this paper basic emotions are emotions that are in simplest forms. Again they are states of arousal, states of being. For example, calm, alerted, stress, terror or trauma.

The jury is out whether AI will ever have emotions like humans. Consider though that human emotions are based on whether or not human needs are met. In nonviolent communication the author writes about how emotions are based on basic needs. One example is the human need for connection. When humans meet this need they feel valued and loved. As mentioned above, this appears to be a reaction to the mind processing at a subconscious level. It seems that this would be unnecessary for a machine. However, if the AIS is given constraints would those constraints then operate as needs? If the goal was to meet the constraint or satisfy the constraint would the AIS begin to feel. Would the machine reach a level of arousal based on a need or constraint? One possible implementation would be to introduce emotions in response to the system achieving, or not achieving, a goal or objective. This would be analogous to something happening subconsciously and the brain explaining it with an emotion.

Given the studies cited, can we give our AIS a sense of intuition without emotion? If we can, could it then exceed human performance on tasks that emotions influence? How separable is intuition and emotion? The question is: can the AIS perform predictions or problem solving without using states of arousal. We believe the answer is no, and we propose the concept of autonomic nervous system and arousal states within the AIS to provide the "emotion-like" features required to deal with the world around it [3].

Some of the questions that arise from this discussion involve how humans will perceive Artificial Intelligence, particularly with systems that display emotions. And the converse being how would an AI system that has emotional responses perceive humans and their emotional responses?

7. HUMAN PERCEPTION OF ARTIFICIAL INTELLIGENCE

According to Nass and Moon [23] humans mindlessly apply social rules and expectations to computers. They go on to say that humans respond to cues triggers various scripts, labels and expectations from the past rather than on all the relevant clues of the present, in a simplistic way. In the article, Nass and Moon illustrate three concepts to consider when thinking about human perceptions of AI. The first experiment they describe show that humans overuse social categories by applying gender stereotypes and ethnically identifying with computers. The second experiment they describe illustrates that people engage in over learned social behaviors such as politeness and reciprocity with computers. Thirdly they illustrate human's premature cognitive commitments by how humans respond to labeling. Nass and Moon conclude that individuals apply social scripts that are appropriate for human to human interaction not human computer interaction.

Sarah Harmon [28] shows that gender did not make a significant difference but that people paired characteristics that may have been affected by gender and embodiment. She showed significant correlation between things such as Passive and Likeable for the Male and Understandable and Pleasant for both male and female and Reliable and likeable for the male, thus showing that humans are willing to assign human characteristics to computers. Harmon does state however that

we need to consider confounding variables. Harmon also wrote that the degree of the entities embodiment influences how humans deem the characteristics with respect to each other such as the terminal and the robot had significant correlation for understanding/pleasant and friendly/optimistic. Yet the only the terminal showed significant correlation in regard to Understandable/Capable, Pleasant/ Reliable, and Helpful/Reliable.

Considering these authors work one would conclude that how AI is presented to humans will affect how AI is perceived. Even a navigation system in a car that one names seems to take on a whole different meaning once it has a human name. Clearly there are many variables influencing human perception of computers and any AI system. There is much research to be done on how AI could be presented that would make it best perceived by humans.

8. HUMAN ACCEPTANCE OF ARTIFICIAL INTELLIGENCE

It seems that the no-intelligent robotics have had both positive and negative receptions from humans. On one hand the technology of AI could help humans to function better. For example, as stated earlier, AI could help to detect threats to national security. AI could also be used to train our forces and help solve complex problems. On the other hand AI could take over some human functions. Consider the effects of robots in the auto industry. The technology allowed for machines to do work that humans did. How much can AI out-perform humans? What will happen to human handled jobs and tasks? Thus AI could be well accepted or quickly rejected by humans.

It also seems, as with any technology, there is a usage and learning curve. AI may require humans to learn more about technology in order to be able to interface. As we can see with the internet and cell phone technology there is clearly a generational difference in use and acceptance, and there may be cultural differences in the willingness to accept AI. Thus, as with anything it may take time for humans to accept AI systems on a daily basis.

9. CONCLUSIONS

Clearly, there is some concern with how the future may go. There have been ethical guidelines for science to follow as they continue to create systems, although this is true in most fields of science. It makes sense to stop and consider ethics and human reactions to AI, after all this is heading to a superhuman technology.

10. REFERENCES

1. Bolte A, Goschke T, Kuhl J, "Emotion and Intuition." Institute of Psychology, Braunschweig University of Technology, Braunschweig, Germany. annette.bolte@tu-bs.de
2. Crowder, J.A., Friess, S., "Artificial Neural Diagnostics and Prognostics: Self-Soothing in Cognitive Systems." International Conference on Artificial Intelligence, ICAI'10 (July 2010).
3. Crowder, J. A., Friess, S., "Artificial Neural Emotions and Emotional Memory." International Conference on Artificial Intelligence, ICAI'10 (July 2010).
4. Crowder, J. A., "Flexible Object Architectures for Hybrid Neural Processing Systems." International Conference on Artificial Intelligence, ICAI'10 (July 2010).
5. Crowder, J. A., Carbone, J, "The Great Migration: Information to Knowledge using Cognition-Based Frameworks." Springer Science, New York (2011).
6. Crowder, J. A., "The Artificial Prefrontal Cortex: Artificial Consciousness." International Conference on Artificial Intelligence, ICAI'11 (July 2011).
7. Crowder, J. A., "Metacognition and Metamemory Concepts for AI Systems." International Conference on Artificial Intelligence, ICAI'11 (July 2011).
8. DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., and Gray, J. R. (2010). Testing predictions from personality neuroscience. Psychological Science, 21(6):820-828.
9. Marsella, S., and Gratch J., "*A Step Towards Irrationality: Using Emotion to Change Belief*." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems, Bologna, Italy (July 2002).
10. Miller EK, Freedman DJ, Wallis JD (August 2002). "The prefrontal cortex: categories, concepts and cognition". *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **357** (1424): 1123–36.
11. Newell, A., "*Unified Theories of Cognition*." Cambridge MA: Harvard University Press (2003).
12. Damasio A (1994) Descarte's error: Emotion, reason, and the human brain. New York: Gosset/Putnam.
13. Davis M, Whalen PJ (2001) The amygdala: vigilance and emotion. Mol Psychiatry 6:13-34.
14. Eichenbaum H (2002) The cognitive neuroscience of memory. New York: Oxford University Press.
15. Kosko, G., "*Fuzzy Cognitive Maps*," International Journal of Man-Machine Studies, 24: 65-75.
16. LaBar KS and Cabeza (2006) Cognitive neuroscience of emotional memory. Nat Rev Neurosci 7: 54-64.
17. LeDoux JE (1996) The Emotional Brain. New York: Simon and Schuster.
18. LeDoux JE (2000) Emotion circuits in the brain. Annu Rev Neurosci 23:155-184.
19. LeDoux JE (2002) Synaptic Self: How our brains become who we are. New York: Viking.
20. Levine, P., "Walking the Tiger: Healing Trauma." North Atlantic Books, Berkeley, CA (1997).
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.
22. Ashcroft, M., "Human Memory and Cognition." Prentice Hall Professional (1997).
23. Nass, C., and Moon, Y., "Machines and mindlessness: Social responses to computers." Journal of Social Issues, 56(1), 81-103.
24. http://artificial-intuition.com/intuition.html.
25. http://www.wired.com/dangerroom/2008/10/tech-firms-face/.

26. http://www.aaai.org/aitopics/pmwiki/pmwiki.php/AITopics/Emotion.
27. Psychol Sci. 2003 Sep;14(5):416-21.
28. www.cs.Colby.edu/srtaylor/SHarmon_GHC_Poster.PDF.