

DigiPen Institute of Technology

# Artificial Consciousness

Research and Review on How Science Fiction Influences the Perception of  
Artificial Intelligence and Conscious Machines.

Dustin Knie  
4/19/2010

The science fiction genre has inspired advancements in real technology and created an intense fascination with artificial life and intelligence which has been used by some to gain increased understanding of how intelligence exists and what consciousness actually means. Western science fiction often takes the stance of artificial intelligence leading to destruction, harm, or is otherwise somehow not good. Other cultures interpret A.I. as a beneficial extension of humanity, or as an acceptable form of sentient life in which we communicate with at equal levels. Furthermore, there are other interpretations at every point within this scale, as well as extremes at either end. This paper intends to examine some common interpretations of artificial life and artificial intelligence and how fiction influences reality.

There are several views on when the concept of artificial intelligence began. Looking at history and the stories told it appears that A.I. is a concept that has existed in one way or another since mankind began telling stories. The Association for the Advancement of Artificial Intelligence<sup>1</sup> places the first history point of interest with A.I. at the 5<sup>th</sup> century B.C. with Aristotle. At this point in history, Aristotle (Smith, 2007) formulated his theory of inference which is a deductive reasoning system. At this point, defining intelligence is a necessity, but unfortunately is incredibly difficult to do so, as there are many definitions and theories behind what to classify as intelligence. Furthermore, the concept of what it means to be conscious is just as cloudy. The American Psychological Association published a report in 1995 called *Intelligence: Knowns and Unknowns* (Neisser, et al., 1995) which begins with:

*Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought. Although*

---

<sup>1</sup> Association for the Advancement of Artificial Intelligence (AAAI) [www.aaai.org](http://www.aaai.org)

*these individual differences can be substantial, they are never entirely consistent: a given person's intellectual performance will vary on different occasions, in different domains, as judged by different criteria. Concepts of "intelligence" are attempts to clarify and organize this complex set of phenomena. Although considerable clarity has been achieved in some areas, no such conceptualization has yet answered all the important questions, and none commands universal assent. Indeed, when two dozen prominent theorists were recently asked to define intelligence, they gave two dozen, somewhat different, definitions (Stenberg & Detterman, 1986).*

This disparity in definitions presents a problem. If an agreement on what intelligence is cannot be made, can artificial intelligence be adequately defined? Fortunately, since the most common goal of A.I. is to mimic intelligence, and not actually replicate it, liberties can be taken in order to provide a universal definition. With this in mind, a commonly accepted definition involves developing software that can adapt or at least provide reason to finding solutions to given problems. Though vague, this provides enough of a foundation to work with.

Science provides the theories, definitions and experiments that explain the world around us. Science fiction, as well as other forms of fiction, provides dreams, aspirations, imagination and a drive to move forward. In the last century, mankind has seen rapid advancements in technology, including realization of limited artificial intelligence. However, comparing current A.I. technology and progress with what is portrayed in works of science fiction, a divide in what is thought and what is real emerges. In most popular western science fiction, such as the works of William Gibson, Charles Stross, Phillip K. Dick, and others, artificial intelligence (and even

artificial life) is often looked at as sinister. The worst case scenario is often examined, though is perhaps a bit farfetched when compared to what is currently actually possible, and with how computing power is progressing. Though major advancements have happened and are coming, due to limitations in power, memory density, and the incomplete understanding of intelligence, as well as other unforeseen complications, creating something like Johnny 5 in Short Circuit, or Wintermute and Neuromancer in Neuromancer is unlikely and improbable. However, examining the progress made in the last sixty or so years since Alan Turing fathered the modern computer, it is unwise to say that something like that cannot or will not happen, however improbable.

On a lighter note, the animated television series, Futurama, took a different interpretation of artificial life. In several episodes, such as *Fear of a Bot Planet*, *I Second that Emotion*, and *I Dated a Robot*, a satirical approach to the genre of science fiction and the concept of artificial intelligence and life is portrayed. At many points throughout the series the concept that robots, as a form of artificial intelligence, lack emotion is played with many times with the robots reacting in a way that mimic emotion. In *Anthology of Interest II*, Bender says “I mean, being a robot’s great but we don’t have emotions and sometimes that makes me very sad. (Haaland, 2002)” In at least two episodes, the idea that an artificial intelligence has human-like aspirations and qualities is seen as well. For example in the second episode, *The Series has Landed*, a magnet is placed on Bender, which somehow causes him to begin singing folk songs. His response to this is “I guess a robot would have to be crazy to wanna be a folk singer... (Avanzino, Futurama - Season 1 Episode 2 - The Series has Landed, 1999)” In later episodes, other dreams are apparent, such as the desire to cook for his living friends and coworkers (Lynch & O'Brien, 1999)

Before there was Futurama, there was Douglas Adams and his Hitchhikers Guide To The Galaxy series. In this series of books, a robot by the name of Marvin is introduced who is, essentially, clinically depressed (Adams, 1980). Again, traits that could be considered human traits are being applied to artificial beings. This may someday be possible, but first an understanding of why these traits exist and how they exist must be obtained. Some opinions on this kind of advancement has lead to fear and concern about how an artificial being would react to emotions that mankind struggles with. In an article published by The New York Times from July of 2009, John Markoff writes about the concern that machines could someday outsmart men<sup>2</sup>. Greg Fish<sup>3</sup> In May of 2009 self-published a short analysis on an artificial intelligence becoming self-aware and the idea that a machine could be created that was smarter than us. He states that creating an intelligent A.I. that was smarter than man is certainly possibly, but “[i]t’s a pretty lofty goal because designing something smarter than yourself requires that you build something you don’t fully understand.”<sup>4</sup> Then there is Nick Bostrom who has projected the creation of a superintelligence, which he defines as “an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.”<sup>5</sup> In essence, the field of artificial intelligence is surrounded with opinions from every direction. Artificial intelligence research is often done within the field of computer science, though neuroscience, philosophy and psychology are often included as well. This field is incredibly complex and requires a skill set that spans multiple areas of expertise. As a result, even with major advancements in the computing power, there is still an incredible amount to learn and comprehend with regards to intelligence and consciousness and how they exist.

---

<sup>2</sup> <http://www.nytimes.com/2009/07/26/science/26robot.html>

<sup>3</sup> <http://worldofweirdthings.com/about/>

<sup>4</sup> <http://worldofweirdthings.com/2009/05/29/looking-for-a-ghost-in-the-machine/>

<sup>5</sup> <http://www.nickbostrom.com/superintelligence.html> Originally published in *Int. Jour. of Future Studies*, 1998, vol. 2

If intelligence is what provides our ability to reason, communicate and survive, what is consciousness and how do they relate? Merriam Webster defines conscious as “perceiving, apprehending, or noticing with a degree of controlled thought or observation” or as “capable of or marked by thought, will, design or perception.” There are other definitions as well, but do not relate to the question at hand. Continuing, Merriam Webster defines consciousness as “the quality or state of being ware especially of something within oneself, or the state or fact of being conscious of an external object state or fact, or awareness.” Furthermore, for comparison and completeness, intelligence is defined as “the ability to learn or understand or to deal with new or trying situations, reason and the skilled use of reason, the ability to apply knowledge to manipulate one’s environment or to think abstractly as measured by objective criteria or tests.” Note that these definitions have been modified slightly for formatting, but content was not altered. What this means, from a dictionary definition point of view, is that consciousness implies awareness which implies the ability to perceive the world in which you exist. Intelligence, then, is the ability to act and react in a way that we consider reasonable. Looking at these definitions, and then returning to the A.I.s used in science fiction works like Neuromancer, Accelerando, Futurama, Hitchhikers Guide to the Galaxy, and others, they mostly fit.

With these definitions of intelligence and consciousness in mind, the next step is artificial life, as well as the adaption and preservation of human life in digital form. Looking beyond artificial intelligence into artificial life, many other possibilities of advancement emerge: genetic engineering, cloning, grafting robotics with biology and others. In Stross’s Accelerando the concept that a human mind can be captured in a way that can be digitized, stored and duplicated is explored. A similar idea was explored in Gibson’s Neuromancer with regards to being able to move ones consciousness around in a digitally connected environment, such as the internet. Cory

Doctorow also explored saved consciousness and clones in Down and Out In The Magical Kingdom. With Doctorow, clones are modified human bodies with circuitry that provides the ability to communicate naturally with something similar to the internet. These modified bodies allow for the backup of one's self, or ones consciousness. This backup allows them to be essentially reincarnated in the event of tragedy.

Unlike the previous examples, in Star Trek: The Next Generation, the introduction of androids was done using an artificial intelligence that largely does not contain emotion. Data, the first android that is introduced in the series, is, a machine built to follow commands and serve as a starship crewman. However, Data is a curious creature and attempts to learn more about humanity and proceeds to learn about we call humanity. The Terminator movie series created by James Cameron uses a similar concept at a base, but here, the robots are attempting to exterminate humanity, not learn more about it. It is due to these fictional interpretations on what A.I. can do and cause, that influence and present fears regarding what could happen. Certainly, anything is possible, but there is much that is improbable based on the points explored above.

Where does that leave us now, though? Science fiction tells to fear artificial intelligence and approach the field with extreme caution. This is likely a view that should be used in any area of uncertainty. As Markoff mentioned in his self-published article referenced above, creating an intelligence that is better than us is possible, however it requires the design of something smarter than ourselves, as mankind. How accurate is this view, though? In the medical field, understanding the cause of many forms of cancer is still limited, though we are able to treat it in outstanding ways. Many of the developments in the history of man were not due directly to understanding, but experimentation. We are a curious creature that likes to

experiment and see what happens. It is perhaps this attribute that will be the most difficult to replicate in an artificial being. However, that is simply an opinion of the author and nothing in the research done for this paper provided any usable insight into such a statement.

Science fiction has provided fuel to inspire those with the ability to create. In this genre of imaginary worlds and imaginary societies, most works that include artificial intelligence use it as something that has existed for a sufficiently long time to be common place. Some works do look at the creation in a more detailed look, such as Isaac Asimov. We have seen where things could go given the right set of circumstances and environment. With this in mind, let us examine some current technology that is approaching, or has entered the uncanny valley. Boston Dynamics<sup>6</sup> was founded in 1992 by former MIT students and has developed multiple robots for various uses. Sometime in the last couple years, they announced BigDog<sup>7</sup> which has possessed some incredible movement abilities. Unfortunately a date on when this robot was made public could not be found, however the demonstrations which the company has released have provided a glimpse at something amazing. BigDog shows how far we have come in providing legged mobility in a robot, however they have not shown any other signs of recent advancements in artificial technology beyond movement. The Stanford Artificial Intelligence Laboratory<sup>8</sup>, on the other hand, has developed learning algorithms that allow an untrained robot to learn how to pick up objects and handle them properly. This is seen in the STAIR<sup>9</sup> project.

Outside of academia and government funded research, a growing hobbyist electronics market, could lead to more independent and personal ventures into artificial intelligence.

---

<sup>6</sup> [http://www.bostondynamics.com/bd\\_about.html](http://www.bostondynamics.com/bd_about.html)

<sup>7</sup> [http://www.bostondynamics.com/robot\\_bigdog.html](http://www.bostondynamics.com/robot_bigdog.html)

<sup>8</sup> <http://ai.stanford.edu/>

<sup>9</sup> <http://stair.stanford.edu/>

Companies like Make<sup>10</sup>, SparkFun<sup>11</sup>, Adafruit Industries<sup>12</sup> and others are dedicated to the do-it-yourself type people and hobbyists; however it is yet to be seen how this will affect artificial intelligence research, if at all. Funded research is increasingly being driven by the ability to capitalize on an end product. Research for the sake of curiosity is becoming harder and harder to justify. The rise of the hobbyist could influence this, however, and provide another means of research and development by way of the curious researching, designing and developing in their homes.

Advancements in artificial intelligence and the pursuit of sentient machines are loosely tied to the authors of the many works of science fiction. Without impossible worlds to explore in our minds with impossible societies, impossible cultures and impossible technology, it is uncertain what our current state of technology would be. These works of fiction have inspired people to dream big and attempt the impossible. This same drive can be seen in nearly every other field as well. It is also clear that within the realm of scientific research and progression, there are mixed assumptions and opinions on what should happen and what could happen. The implementable capabilities of artificial intelligences will likely increase as computing power increases. Due to the limited understanding of the reasoning behind concepts such as intelligence, consciousness and awareness, it is not clear if or when these concepts will appear complete within an artificial intelligence. However, based on advancements made in other fields, and even the amount of progress made within the field of artificial intelligence, a sentient or at least pseudo-sentient A.I. may occur sooner rather than later thanks to experimentation, even though adequate understanding may not exist when it happens.

---

<sup>10</sup> <http://www.makezine.com/>

<sup>11</sup> <http://www.sparkfun.com/>

<sup>12</sup> <http://www.adafruit.com/>

Approaching a topic like artificial intelligence can be tricky. There are a number of ways in which this idea can be approached, and every single one of them will likely lead in a different direction with a different analysis at the end. As with actual intelligence, artificial intelligence is seen differently by nearly every single person that approaches it. This can be seen in the multiple interpretations in science fiction, as well as the varied progress in academia. What this all leads to is simply that curiosity is what got us to where we are today, and curiosity will carry us forward. It may even lead the first near sentient A.I. to evolve itself. In addition, thanks to our understanding of electricity and power and everything else that is required for a machine to operate, many of the worst-case-scenarios we read about in science fiction are just that: fiction. Individual curiosity is likely going to bring about the little steps that are needed to progress in this field, though major breakthroughs and advancements are just as likely going to be the result of DARPA or otherwise funded projects. Caution, as always, should be exercised, but we should not be afraid of what will happen. An A.I. is a block of code that we can alter to our hearts content.

## Bibliography

- Adafruit Industries. (2010). *About us* << *adafruit industries blog*. Retrieved April 18, 2010, from Adafruit Industries Web site: <http://www.adafruit.com/blog/about/>
- Adams, D. (1980). *The Hitchhikers Guide to the Galaxy*. New York: Hamony Books.
- Avanzino, P. (Director). (1999). *Futurama - Season 1 Episode 2 - The Series has Landed* [Motion Picture].
- Avanzino, P., & Baeza, C. (Directors). (1999). *Futurama - Season 1 Episode 5 - Fear of a Bot Planet* [Motion Picture].
- Badham, J. (Director). (1986). *Short Circuit* [Motion Picture].
- Boston Dynamics. (2009). *Boston Dynamics: Dedicated to the Science and Art of How Things Move*. Retrieved April 18, 2010, from Boston Dynamics Web Site: [http://bostondynamics.com/robot\\_bigdog.html](http://bostondynamics.com/robot_bigdog.html)
- Bostrom, N. (1998). How Long Before Superintelligence? *International Journal of Future Studies* .
- Cameron, J. (Director). (1984). *Terminator* [Motion Picture].
- Cameron, J. (Director). (1991). *Terminator 2: Judgement Day* [Motion Picture].
- Doctorow, C. (2003). *Down and Out in the Magic Kingdom*. New York: Tor Books.
- Ervin, M. (Director). (1999). *Futurama - Season 2 Episode 1 - I Second That Emotion* [Motion Picture].
- Fish, G. (2009, May 29). *looking for a ghost in the machine*. Retrieved April 18, 2010, from Weird Things: <http://worldofweirdthings.com/2009/05/29/looking-for-a-ghost-in-the-machine/>
- Gibson, W. (1984). *Neuromancer*. New York: Ace Books.
- Haaland, B. (Director). (2002). *Futurama - Season 3 Episode 18 - Anthology of Interest II* [Motion Picture].
- Lynch, J., & O'Brien, K. (Directors). (1999). *Futurama - Season 1 Episode 7 - My Three Suns* [Motion Picture].
- Markoff, J. (2009, July 25). *Scientists Worry Machines May Outsmart Man*. Retrieved April 18, 2010, from The New York Times: <http://www.nytimes.com/2009/07/26/science/26robot.html>
- Neisser, U., Boodoo, G., Bouchard, J. T., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1995). *Intelligence: Knowns and Unknowns*. Board of Scientific Affairs of the American Psychological Association.

O'Reilly Media, Inc. (2010). *makezine.com: MAKE: technology on your time*. Retrieved April 18, 2010, from Make Magazine: <http://www.makezine.com/>

Purdum, J. (Director). (2001). *Futurama - Season 3 Episode 15 - I Dated a Robot* [Motion Picture].

Smith, R. (2007, December 14). *Aristotle's Logic* (*Stanford Encyclopedia of Philosophy*). Retrieved April 18, 2010, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/aristotle-logic/>

SparkFun Electronics. (2010). *SparkFun Electronics - About Us*. Retrieved April 18, 2010, from SparkFun Electronics: <http://www.sparkfun.com/commerce/aboutus.php>

Stross, C. (2006). *Accelerando*. New York: Ace Books.

Various (Director). (1987 to 1994). *Star Trek: The Next Generation* [Motion Picture].