

I've always been fascinated with computers, having started with a Timex Sinclair 1000 (with its chiclet keyboard and 1k of memory), moving on to a Vic-20, then an Apple ][ in 3<sup>rd</sup> grade, which had an amazing 32k of available RAM. It was right around this time, after using the Apple ][ for some school projects (one of which was spelling my name out in high-quality vector graphics), that I began to wonder just how smart a computer actually was. I could ask the system all sorts of things and it would respond with some answers; these answers were always what I expected to get. I never assumed the computer would reply with something strange, the sort of answers you might expect from a human. Supposing I asked the computer to calculate the sum of two values, I was fairly sure the output would be correct. For more open ended questions, the reply would directly depend on just how much time I put into programming. I once wrote a game wherein a spaceship was attacked by numerous waves of aliens, and the goal was to shoot as many aliens as possible. There was no variation of alien type, the speed at which they moved, etc. I didn't expect the computer to magically produce these variable effects since I understood that the system was incapable of drawing any sort of conclusion about what an appropriate response would be, beyond what its database held; a database that I or someone else had typed on a keyboard.

My conclusions at an early age were fairly simplistic, yet even now they seem valid. A computer isn't intelligent, nor does it have the capacity to be intelligent. However, it can know a lot of things, perhaps even things that humans cannot know. It has the ability to accumulate knowledge, and even learn, but it seems like it really doesn't KNOW those things. The naturalist model of epistemology would disagree with this

conclusion and might instead ask something like “What do you mean by ‘intelligence’? What do you mean by ‘know’?” To clarify, intellect in the sense I am using it is the type of feature we ascribe to humans, a sort of reflexive self-awareness about our own cognitive processes, and more specifically the ability to think abstractly about things, to come to conclusions that are not necessarily logically deducible from a set of criteria. Further, knowledge will refer to simple factual statements, while Knowledge will refer to those concepts that attempt to reflect the true nature of our world, a concept that is well known but poorly understood, especially in non-philosophical circles. In this paper, I will attempt to provide a contemporary model of Artificial Intelligence and how it relates to epistemological questions, specifically naturalist epistemology as that is the most relevant view for a scientific approach (and the one that lends itself to AI research). It is my goal to show that computers are incapable of true Knowledge, and as a result that naturalist epistemology does a poor job of modeling human cognition.

As I mentioned earlier, I have been programming computers to do things for most of my life. I never really thought of them as capable of true intelligence, although I was (and still am) a huge fan of science fiction stories that depicted self-aware machines. The movie *Tron*, for instance, is one of my favorite films, and the whole notion of programs having personalities makes for a fun atmosphere. Still, even though I sometimes wish I could be a videogame warrior and battle the evil MCP for control over the vast computer system, I doubt the MCP is possible. It was self-aware, had a distinct personality, and was more than a little vindictive. Qualities we would hardly ever associate with machines (at least in the literal sense). This brings up an interesting point, however. While we

understand that computers (or even cars, toasters, television sets, etc.) are not people, we still anthropomorphize<sup>1</sup> them to a great degree. Nearly all of us have uttered comments under our breath about how our computer hates us, or how the ATM is evil because it's always out of service when we try to get cash. I think this is one of the major problems with doing AI research: we make some assumptions that inanimate objects already have a personality, and therefore the possibility exists for intelligence. Further, contemporary scientific views about humans being the product of millions of years of evolution have seeped into our culture, giving the impression that computers might one day be smarter than humanity validity. We tend to assume that complexity equates to intellect, and the more complex machines get, the closer to intelligence they will reach. The book *Neuromancer* is a good example of the eventual level computers will reach making these assumptions, although it seems farfetched and a bit ridiculous when viewed with a slightly more pragmatic light. While Wintermute makes for a good character, it is hardly possible to imagine it actually existing as an entity with true cognitive power. We also do this to some degree when investigating the possible intelligence of animals. It is common to hear chimpanzees, for instance, compared to children as far as mental capacity. While I'm sure a grown chimp is more capable than a small child in many areas, it is quite an assumption that the chimp could be as smart as a human if it was only a bit more developed. Essentially, keeping the previous two points in mind when researching the intellectual capabilities of computers will allow us to remain grounded in reality and not let our imaginations influence our expectations quite so much.

---

<sup>1</sup> Dr. Kelly expounds on this point greatly, and he seems to have even more concern with it than I do.

I'm going to focus on a very simple computer to demonstrate my point that computers are fully capable of learning without any sort of pure cognitive process, and therefore lacking any true epistemological worth as relates to humans. Going back to 3<sup>rd</sup> grade again, I can recall a simple programming experiment we partook that modeled computer learning without any frills. Chess problems are very interesting modeling experiments in the computing world, as they are easily understood and structured, allowing quick programming of a system. In my case, we focused on a 3X3 chess board with six pawns, commonly called hexapawn<sup>2</sup> (figure shown below).



Hexapawn has simple rules and simple goals: the pawns may move one square forward (as in chess), or diagonally to take an opponent's piece. The winner is the player who reaches the far end of the board, or manages to capture all the opposing pieces. All the possible moves and counter moves can be calculated very quickly and a complete tree of moves produced. Hexapawn has a feature in common with tic-tac-toe: it is a trivial game. This means a correct counter move always exists that will lead to a victory. Tic-tac-toe results in a draw if played correctly, but the premise here is the same. Since hexapawn is trivial, a computer will always win when playing a human (assuming it moves second). It should become obvious that programming a computer to win the game itself is trivial. However, that's not how we approached the problem back in grade school, and in fact used no electronic computer. We had a series of diagrams pasted to matchboxes, each

---

<sup>2</sup> See <http://www.cs.toronto.edu/~mitchell/ai-course/gp.html> for a very good explanation of this problem, as well as a complete prototype solution for machine languages.

diagram a branch on the tree of possible piece positions on the board. The diagram was color coded, each possible move for that position having a corresponding color. Inside each box there were several pieces of colored candy (Skittles, I believe), one for each move on the diagram. When then setup the board and played the “computer”. The human player moved first, consulted the boxes, found the appropriate box that illustrated the current state of the board, and then opened the box. If the yellow move, for instance, would result in human advantage, the human player would have the computer move its piece in that manner, then eat the yellow candy contained inside the box if a pawn was taken. It can easily be deduced that after playing for a while, there were no longer any pieces of candy left in the boxes that allowed the computer to make incorrect moves. In fact, the computer would now only make the right move in every situation. In this case, the computer learned the correct moves to make by playing the game repeatedly. One might argue that a more sophisticated computer might have randomly chosen the correct move initially, and the human player had far too much influence on the “learning” process. These are not really valid criticisms, as the human programming the more sophisticated machine would have just as much influence on the results as picking the wrong move to make and eating a piece of candy. In fact, having the system always select the wrong move is the most efficient programming method, as all possible errors are eliminated.

As can be seen from the hexapawn example, even rudimentary computers made of matchboxes and Skittles are capable of learning. A naturalist, or maybe even the epistemologist, might say that in this case there is no real learning involved; all the

possible moves were pre-programmed and we simply eliminated the bad ones. I'd counter by stating that this example is no different than teaching a child not to touch the stove. There are two possible moves we might envision: touch the stove and don't touch the stove. If the child touches the stove, he learns quite quickly never to do it again. All the possible moves the child might take were inherent to the child before we ever started teaching him. The computer case is simply clearer cut because we know every possible move and can therefore act accordingly. The child might decide to have his friend touch the stove and relay the experience, a contingency we didn't think of. The computer can't produce such possibilities unless they are mathematically deducible, and thus discoverable by the programmer. In effect, the computer is incredibly simple and the learning involved seems trivial. This is because the learning is in fact trivial, but it is learning nonetheless. There is no cognitive process that occurred in the computer, yet it has learned how to play hexapawn to perfection. It is entirely probable that a human who was shown every possible move in this same manner would still make a mistake and lose a game or two<sup>3</sup>. The computer is just better than we could ever hope to be. Again, we might ask how this relates to human cognition. I don't think it is even possible to create any sort of relationship between the computer and a human in this case. I mentioned a child who wants to touch a hot stove; perhaps the child burns himself, and then touches the stove again at a later time. Obviously this seems rather irrational, yet children do such things all the time. A computer would simply touch the stove, learn that it got burned, and then never touch the stove again. The child, however, might touch the stove many, many times, getting burned repeatedly. Perhaps he wants to learn more about the sensation of being burned, or maybe even likes it. It is entirely possible he figures that the

---

<sup>3</sup> This is why people still insist on playing tic-tac-toe, even though it is not truly winnable.

stove can't always be hot; therefore he must test it from time to time. While both the computer and the child are learning something about the stove, the child is actively participating in the process, and formulating ideas that might be ridiculous or profound. It is hard to explicate, yet the difference is obvious. Something goes on in the head of the child that the computer is incapable of replicating, no matter how sophisticated it might be.

I touched on the computer being better than a human at a particular task, in the case above that of playing hexapawn. Computers of these sorts are called expert systems. The idea of expert systems has long been explored in AI research and computer research in general. It would be highly useful to have a system that is incredibly good at a task to alleviate the burden on humans. For instance, medical diagnostic software is a source of great interest as doctors would quickly be able to determine illnesses given a set of symptoms. In science fiction, these expert systems replace humans altogether in some capacity. The film *Aliens* for instance had a synthetic human character that was particularly good at his job, better than any human could be. Going back to chess problems, let's consider the case of Deep Blue, the IBM computer that played World Chess Champion Garry Kasparov<sup>4</sup> in May 1997. Kasparov is without a doubt a chess genius, and had no problems beating simple programs like those embedded in chess game keychains found in gas stations across the country. While I myself often struggle against such unsophisticated machines, Kasparov would need a supreme challenge, an expert chess system. Deep Blue was at the time the most advanced chess computer in existence, and as it battled with Kasparov, proved to be the better player. The first game belonged to

---

<sup>4</sup> <http://www.research.ibm.com/deepblue/>

Kasparov, the second to Deep Blue. Then three draws, finally a victory for Deep Blue. Reading the commentary and logs of the game, Kasparov seemed cocky and sure of himself until his first loss. Then he began playing in a manner that seemed suited to throw off Deep Blue's programming, not in a way he might play a human who was familiar with thousands of move variations. It was here that Kasparov made a mistake. It was almost as if he assumed Deep Blue to be the inferior player simply because it was a machine, not because it wasn't up to task.

Deep Blue had honed its skills over eight years, learning new openings, positioning, traps, etc. The programmers had tried to capture what it means to play chess like a grand master within the electronic brain of the system. Further, hardware upgrades allowed for significant speed gains and parallel processing which let Deep Blue pursue alternate lines of move sequences much like human chess players. It was an expert system, and one that was better than any human at chess. So what? Does the fact that Deep Blue could beat Kasparov in six games prove that machines, especially expert systems, are somehow more advanced than humans in cognitive power? I would hardly say the sci-fi assumptions have become a reality yet. Deep Blue expressed no knowledge of chess in the sense that Kasparov, or even a junior high student, expresses similar knowledge of chess. Surely Deep Blue knows a vast amount about chess, and in fact has been either programmed or learned through experience nearly every possible chess move. I am sure that in some point in the future its upgrades will be sufficient to put it on the level of perfection that my matchbox computer was with hexapawn. But again, Deep Blue cannot reflexively look upon its own knowledge of chess and contemplate what that

knowledge means. It has learned responses to various moves through trial and error, much like a human, but made no leaps in reasoning not based upon the logic by which it was programmed. It is possible the junior high chess champion could beat Deep Blue simply through intuition and an abstract notion of the chess board that the physical hardware was incapable of representing. Deep Blue is the best chess player in the world, but it doesn't know it's the best. It doesn't even know how it got to be the best, nor what "the best" means. It is simply an expert system that can do a task far better than any human can do. Epistemologically speaking, it is a collection of facts that are interrelated by logical statements, with only certain conclusions possible; Deep Blue would never do something crazy or unorthodox. The usefulness of systems like Deep Blue in other areas (I mentioned medical diagnostic programs) is obvious, but thinking that such systems can somehow attain true cognitive ability in the same manner as a human is absurd. They can learn, and surpass humans, but they have no real knowledge. Similarly, animals such as dogs and rats can learn to respond to outside phenomenon. No one would deny that a dog can be taught tricks, or that a rat can learn how to work its way through a maze. However, in these cases, the dog or rat has no ability to reflexively look upon the process by which it reacts to the criteria at hand. It simply knows, like Deep Blue, that a particular action X yields result Y. While a rat might be said to react by instinct, this is hardly any different than the "instinct" by which Deep Blue reacts when playing chess. Thus, sophistication has no real correlation with cognitive ability, as was postulated earlier.

The preceding assumes something about computers, namely that they model problems using binary logic. Humans don't go about solving problems in this manner; hence the whole field of fuzzy logic has arisen to more accurately imitate human cognition. According to Dr. Konar:

“Fuzzy logic deals with fuzzy sets and logical connectives for modeling the human-like reasoning problems of the real world. A fuzzy set, unlike conventional sets, includes all elements of the universal set of the domain but with varying membership values in the interval  $[0,1]$ .”

This definition clearly shows that fuzzy logic attempts to produce a system by which a computer might accurately think like a human by allowing non-binary values for yes/no questions. While this sounds rather strange at first, it is rather easy to understand. Due to the nature of electricity, computers have historically been built with simple logic gates that respond to current, either on or off. This binary logic system on the hardware level translates into binary logic at the software level, which in turn influences the sorts of programs a computer is designed to run. Before fuzzy logic research, it was assumed that a computer was only capable of answering a question yes or no, completely unlike a human. For instance, suppose a system is designed which uses optical receptors to determine the color of objects. Pointed at the sky, we might ask it “Is that blue?” The system would reply yes. We then ask if a tree leaf is green, if a pumpkin is orange, etc. The system replies with simple yes or no answers given the criteria it has to work with. Asking a human these same questions, we'd probably get a few “Maybe” answers, something a computer cannot produce. Suppose then we design a better color recognition system, an expert system. This system is given a palette of millions of colors, each with a

specific name, ranging from Seafoam to Sienna<sup>5</sup>. We then ask the system tell us the colors of specific objects, a job which it performs with 100% accuracy, giving better differentiation than a human possibly could. Imagine that an object is shown to the system for which it has no color name. How would the computer respond? Most likely it would either give an error, or say the color was not found in its database. What about the human? “Looks like a reddish-brown to me.” What is the difference here? The human has never seen the color either, but their response is not “I’ve never seen this color” (although they may say that) but instead they create a new name for the color, giving their best guess. This is what a fuzzy logic system attempts to do: provide answers to questions that have no clear cut answer, those that have no definitive yes/no response. A good fuzzy logic system with only a few thousand color names (several magnitudes less than the expert system mentioned above) would be able to name any color that it was shown, creating names in much the same way that a human would.

I’ve seen fuzzy logic systems in action, and they are rather interesting if not downright impressive. There are robots that play soccer, responding to a billion external parameters per second to do an action that is “sort of” the correct one. There are systems that calculate integrals with amazing accuracy, using a “best guess” method<sup>6</sup>. Sony now sells robotic pets that act like real animals, even developing personalities and mannerisms. But as Dr. Kelly suggests, where are all the self-aware systems? Where are the robots that know they’re robots? Even using highly advanced expert fuzzy logic

---

<sup>5</sup> Open up a box of Crayolas for some rather creative color names.

<sup>6</sup> I might point out here that while it seems uninteresting to mention mathematics problems and fuzzy logic systems together, modern calculators don’t know that  $\sqrt{4} = 2$ , and instead use an integral rounding technique. Fuzzy logic systems, however, DO know this, in the same way that humans understand it. High-end programs such as Mathematica and Statistica are extremely popular for this reason.

systems, it is rather difficult to posit that when asked “What are you” the system would respond with “A computer program”...and mean it. While it may be able to deduce that it WAS a computer program from the way in which it interacted with the world, the system would still not be self-aware, any less than my dog can tell me she’s an autonomous being. My reasons for mentioning fuzzy logic here was to expound on the point previously, that sophistication does not equate with knowledge or intellect. Obviously fuzzy logic systems are vastly superior to conventional binary systems for many applications, and indeed the expert systems mentioned above tend to use fuzzy logic as their method of computation. Even as they evolve, computers do not gain any sense of self, which is what naturalist epistemology would tend to support. They do not know they’re thinking, something Descartes understood to be the basis for human cognition and existence. While fuzzy logic is fascinating and extremely useful, I don’t see any great strides taken that have produced human-like intelligence. Animal-like intelligence, absolutely; those robot dogs are frighteningly realistic.

I mentioned during the discussion about Deep Blue it was possible that its hardware was simply not capable of representing the sort of knowledge that humans have. Current advances in hardware endeavor to replicate much of the principles of fuzzy logic, using variable voltages instead of simple on/off gates or similar apparatus. Further, electronic copies of biological neural pathways, generally referred to as neural nets, have been in use for some time, attempting to imitate actual neurons and the structure by which they are connected. The neural nets are by far the most interesting development in computing at this time, as they really do seem to simulate actual biological processes. In

conventional electronic hardware, logic pathways can only go “one way”, in a previous defined sequence. Neural nets, however, usually link all pathways to each other, and even recursively to themselves. This allows the system to decide which path to use when solving a problem, possibly even programming itself. For example, an unsophisticated computer that had to identify an orange might go through a process where it looked at color, size, smell, texture, etc. It would then consult its knowledge database and compare each feature, finally arriving at the conclusion that the object was an orange. A human’s reasoning is quite different, given the nature of our brains. Perhaps we see the orange, recollect a time we went to Florida and saw an orange grove, thusly remembering the name of the fruit not by measuring all its features but by personal experience. Neural nets can imitate this almost random linkage of knowledge to provide very non-computer-like methods of problem solving. Sometimes this means that the neural net machine might take longer to recognize the orange than its counterpart, but typically it is not only faster, but can tell us interesting facts that the conventional machine does not associate with the orange. As expected, neural net systems are quite good at learning and acquiring knowledge. Computer visual recognition, conversion of data to digital format and robotics are some of the areas in which neural nets have been employed successfully; but again, these highly advanced biological modeling techniques have produced no self-aware machines capable of human knowledge. While a neural net machine might grow more and more advanced, and perhaps even one day copy a human brain completely, I fail to see how the system would ever gain true cognitive power. The answer here might be in the use of the word “modeling”. Models are useful for understanding how the human mind works, and might even give us great insight to our own epistemological

questions, yet are truly incapable of becoming anything more than a copy. Dr. Kelly posits:

“The philosophical problem arises when we try to read too much into the model, when we attribute to it depth and independence and originality of action, when we expect it to behave in a manner which goes beyond the explicit warrant of model. We must remember that a model is an abstraction, an impoverishment, incomplete. In the case of intelligence this impoverishment may strike to the very root, leaving us with an unnourished, brittle and unprincipled outward show.”

It might be possible that a hardware representation of the human brain would produce an extremely intelligent system, but I still doubt its capability for self-awareness or human-like cognition. Again, incredible levels of sophistication do not necessarily equate with true intelligence. It is even possible to imagine such advances in hardware design that an uber-brain could be manufactured, giving rise to machines of incredible “intellect”, smarter than a thousand people; but they would be missing something vital, the very essence of humanity that cannot be replicated by a mere model.

As a demonstration about how modeling fails to capture true intelligence yet makes us believe it to be possible, consider the case of ELIZA<sup>7</sup>, a program that simulates a psychologist. ELIZA has been around in one form or another for over to thirty years, an attempt by Joseph Weizenbaum to explore natural language processing. As Dr. Kelly states:

---

<sup>7</sup> See <http://www-ai.ijs.si/eliza/eliza.html> for a web-based implementation of ELIZA.

“The program was designed to imitate a non-directive psychotherapist...frequently the exchange was plausible and convincing, but it was all too easy to reduce the responses to absurdity.”

ELIZA had no ability to actually interpret the responses of the “patient”; it simply used some tricks of syntax to produce what looked like reasonable questions. Obviously, ELIZA was not intelligent even in the sense I’ve used it referring to other systems in this discussion, yet people have been and still are fooled into thinking otherwise. Some individuals actually thought there was someone in another room typing the responses on a keyboard; others went so far as to discuss deep psychological issues with the program. When I first saw ELIZA it was running on my friend’s Commodore 64. We spent hours talking to it, often getting frustrated and confused, even though we knew it was a simple program. ELIZA was anthropomorphized; becoming a person who asked you questions much like an annoying shrink. Even though the program was clearly not intelligent, the human tendency to ascribe personality took over. While systems far more advanced than ELIZA have been created and used, we must ask ourselves if we’re simply doing the same thing with those programs we did with ELIZA. I tend to think that as models get more and more advanced, people will be more apt to call them intelligent, and mean it in the same way humans are intelligent. This seems to be a function of our own psychological propensities, not necessarily anything with the true ability of the system we’re speaking about.

At the hardware level, and going back to neural nets for a moment, I am reminded of a Brain Teaser from the May 2003 issue of Discover Magazine. We are asked to look at a series of lines in a circle surrounded by another circle. Moving the page back and forth, it appears that the lines rotate in opposite directions from each other. An optical illusion to be sure, but the explanation for the illusion was quite interesting. Our brains expect individual neurons to perform multiple tasks at once, which can lead to confusion. Neural nets cannot have individual neural pathways perform in this manner, and I don't think it is even plausible to assume that will ever be the case unless we move to purely biological computers. Perhaps we can grow a brain in a vat, but would this really be a machine? It seems like the more human computer systems become at the hardware level, the more they look like either poor copies that don't capture true cognition or clones. It IS entirely plausible that we could produce clones, or perhaps parts of clones (again, brains in vats) that function as computers, and AI research might indeed head in his direction; the purely synthetic AI, however, just doesn't work.

I do not think computers will ever be capable of human intelligence, think abstractly nor able to reflect upon their own existence. To address the naturalist epistemologist, it appears they must concede that humans are different than animals at some fundamental level in the acquisition of knowledge. AI can produce programs, robots, systems, etc. that imitate animals to a great degree of accuracy. The production of self-aware systems that do the same for humans is lacking, and most likely impossible. Discovering how we think requires models, and AI research taken in this light, that of pure modeling to understand human cognition, is valid and reasonable. To think that

someday these systems might themselves become truly intelligent is a dream, perhaps a product of our own wishes and psychological fascination with ourselves. Science must remain aware of this try to remain grounded in the reality that the assumptions and discoveries made about animal intelligence have little if nothing to do with our own. Epistemologically speaking, I am unsure that we will ever ascertain exactly how we know anything, but we can surely know we know something.