INTRODUCTION

The examination of artificial intelligence raises questions that many people find difficult and unusual. Beneath the technical investigations into its creation and application lies a milieu of philosophical questions and implications that ultimately penetrate to the heart of what it is to be conscious.

It seems that the natural question people form about artificial intelligences is whether or not they truly can be "intelligences." Can a computer think as we do? If a computer can become conscious, should it be afforded the same rights as a person? These are the issues that emerge, in our popular culture interpretation of artificial intelligence.

First, to clarify, for the purposes of this paper the term "consciousness" refers to one's own existence, one's subjective experience of the world. Consciousness is the "inner realm," the perception of the world, the thoughts, the "cognitive reality," or the mental acts. Consciousness is not simply a behavioral pattern.

Strictly speaking, we cannot experience another's consciousness directly (mental acts are private and not public.) We infer that the people we interact with daily are conscious by their behavior, true, but to conflate consciousness with patterns of action, or a simple term such as "self-awareness" denies the true complexity of the issue. In short, we can know what another is experiencing, but we cannot experience what they experience. It is ours and others' experiences that we are interested in as consciousness.

It is perhaps indicative of the scope of the problem of consciousness that words often fail to do justice to the concepts under scrutiny. The existence (or nonexistence), importance, and functioning of consciousness is contested between schools of thought, and is intimately tied with the most fundamental question of metaphysics we can ask: why does anything exist, rather than nothing?

The question of why there are essents (essents being “that which exists”) rather than nothing is wed with the idea of consciousness. An empiricist places consciousness before essents- as Berkeley wrote, “esse est percipi,” or that things exist only insofar as they are perceived. In the empiricist school, nothing would exist in any meaningful sense if there were no conscious subjects to perceive things.[1]

Many religions take as their basis the pre-eminence of consciousness- that is, consciousness is the “fundamental stuff” of which the universe is made, all else is created by the conscious mind and is ultimately not as real as our own selves. In Buddhist philosophy, all that exists is the “now,” the present which is defined by our existence as conscious beings, and all else is illusion. This is in total contrast with the western scientific approach that seeks to minimize all effects of the consciousness and discover what is real outside of the mind; the scientific method seeks to eliminate perturbations from any subjective cause and discover what can be objectively described as true.

Even to schools that deny the real existence or fundamental nature of the consciousness would have to admit that the case of a universe with no conscious beings is a troublesome one; this is akin to asking what it would be like if we, ourselves did not exist- the question is almost meaningless to ask, as we cannot describe our own nonexistence. One recalls the zen koan: “Show me the face you had before you were born.” This would seem to indicate to us that reality, as it were, is purely our own consciousness, to some degree; or at least we cannot image a universe in which we are

[1] The existence of God was critical to Berkeley’s philosophy, however, who perceived all and thus was the root of the persistence of reality, even as we slept for instance.

not conscious. Furthermore, this would divide reality into separate realities for all conscious observers, possibly leading to solipsism, the condition of believing that only oneself exists.

Perhaps the real question, then is closer to "why can questions be asked in the first place?" It is our consciousness that allows us to pose questions- it is our consciousness that composes the core of our being, our experience of reality.

If the reader is not persuaded of the fundamental quality of the question of consciousness, at least it could be conceded that it is one of the most important metaphysical questions we can ask, indeed it is tied with why we can ask the question itself. Of course, a discussion of consciousness is not complete without acknowledging the strange tie it has with the brain- the mind/body dichotomy, as it were. This is where artificial intelligence enters the discussion: can we re-create the means by which a mind arises for us, in some other "body" form?

A traditional, Cartesian dualism asserts that the mind is an immortal entity, separate from the mortal body. Descartes had to invoke God to explain the existence of the mind, which he identified with the soul, placing the mind in the spiritual realm.

In these times we can see that the mind and the body are not entirely separate entities: examples abound of instances of brain damage affecting the cognition of a subject. Drugs that replace or compete with our natural neurotransmitters can radically alter the form of our consciousness. Furthermore, mappings of mental states to physical states of the brain are accepted realities in neuroscience.

As the dominant position held by experts moves toward a pure materialism, this seems to necessarily deny our fundamental experience of the mind as something nonphysical. One could hardly deny that a thought is not a physical entity, that is, something consisting of matter in this familiar, external reality. One can show that this thought is linked to a specific process in the neurons of our brain- but one is clearly not the other.[2] One is a collection of neurons, the other is a thought. The link is a causal one, and schools are split on which causes the other. Still others maintain that the only thing that is, is the realm of matter and physicality, and as such are forced to maintain that their consciousness is illusory.

However, the marriage of the mind and the matter is clear. Some bizarre process gives rise to this wholly unique (as far as we can tell) experience of being, illusory or not. Call it a self-awareness, a consciousness, a sensorium, a soul; whatever it may be it is undeniably removed from our ordinary experience of the external world (and here the matter is confounded, as the entirety of our experience of the external world hinges on our consciousness, that is, our *experiencing* itself.)

Artificial intelligence, then, poses its own question: can we re-create the strange mental process in some matter configuration other than our human brain? What is it about the matter of the brain that gives rise to the mind- is it the sort of matter, its configuration, its processes, or a supernatural soul? Whatever it is, can we affect this process or configuration not in biological neurons but in a substrate of silicon, metal, electricity, and all the trappings of computers?

What a question this is! However, it may not be as inscrutable as it first appears. As I will attempt to show in this paper, perhaps we can make an educated guess that the mind does not depend on the matter that gives rise to it, but instead to a process, a system

[2] The difference here being that between experience and knowledge; again we can know the state of another's brain, but cannot experience what the other is experiencing.

that plays out in matter – and if this is the case, then these processes can be transferred to any sort of system at all. This position allows that consciousness could exist in a brain, an anthill, a society, the complex ecology of planet earth. Additionally, it allows us to reconcile our subjective experience as real and meaningful, separate and distinct from physical matter, but does not invoke anything necessarily supernatural.

The intention of this paper is to demonstrate this, and examine the feasibility of creating a consciousness in a computer- a true artificial intelligence. The IQP itself will entail creating a portrait of artificial intelligence, in its current form and its potentialities; and in the process of capturing the essence of artificial intelligence, perhaps I may catch a glimpse of a mind there, shrouded, perhaps, obscured by a barrier of interaction and language, yet aware and in the process of being born.

LITERATURE REVIEW

THE SYNTHESIS OF A COMPUTER

Imperative to our discussion of a computer consciousness is an understanding of what a computer is, what it really is and does on the most basic level. It has been put forth that a mind is a process, independent of the physical substrate- but the substrate must be a system of suitable complexity to attain an awareness. Is a computer such a system?

Computers emerged as a physical analog to Boolean logic- a device that simply computes logic values. The fundamental components of a computer circuit are logic gates. These gates are physical, electrical constructs that manipulate voltages in a manner that reflects logical calculus.

Analog computers are systems that solve mathematical problems by using an analogous physical system. Equivalences are made between physical quantities and the mathematical quantities to be determined. By observing the physical quantities, we can determine the mathematical quantities in question.

For example, the OR operator is defined as a gate that outputs true if either of its two inputs are true, and false if neither inputs are true (it may be viewed as asked if the first or the second input is true). There exists a physical analogue to this mental construct: instead of manipulating mental “true” and “false” values through a system of logic, we may manipulate voltages corresponding to true and false through a system of circuitry that necessarily obey the laws of physics. In this case, the specific “OR” logic gate circuit will output a voltage corresponding to true if either of its inputs correspond to true, and output a voltage corresponding to false otherwise.

The absolute foundation of the computer lies in manipulating the external, physical world in such a way that you may create a physical correlation to your mental problem. The example of addition is illustrative in that, could we not solve the problem of “what is the value of 2 + 2” mentally, we might simply create a physical analog to the problem: most obviously by taking two rocks (or apples, traditionally) and putting them together with two more apples, and counting the result[3].

Early calculators, those of greater complexity than an abacus, were analog devices, ingeniously constructed purely by mechanical means. By taking advantage of the interactions between gears of varying diameters, for example, problems involving multiplication could be solved by calibrating and cranking elaborate, geared computers. A famous example of an early calculating device, the “Antikythera Mechanism” (or “Antikythera Computer”) seems to be a geared mechanism that was used to predict the motions of the sun and moon (deSolla, 1959.)

In the case of computers, the physical problem-solver device arose as a means to physically represent problems in logic- formal, mathematical logic, dealing with the truth or falsity of various statements. We can imagine such a device, for instance, that is a simple, mechanical analog to the purely mental “OR gate” construct: a box with two holes on top and one hole on the bottom, and if necessary some internal mechanism. The box behaves such that if a rock is dropped into one of the top holes, a rock will be dropped from the bottom hole. If we take a rock to be the physical object corresponding

[3] The example of counting apples may be troublesome, as it seems that the mental processes associated with evaluating addition necessarily resorts to real-world examples; that is, there is no strictly numerical method to add two numbers, the method relies on memorization of real-world relationships between values.

to our mental value of “truth” (and the absence of a rock to correspond to “false”) then we can say that the box acts as an OR gate. By linking it in series with various other boxes corresponding to AND gates, NOT gates, XOR gates and other such devices in Boolean logic then we could solve complicated problems simply by feeding in the initial truth values, and seeing what rocks in what configuration are output accordingly.

In these examples, after the initial difficulty of constructing such a device to the needed specifications, the process of solving a problem within the device’s input problem range requires substantially less work and mental effort than the solving of the problem through mental means. In the example of the OR-gate box, we know that the answer to the problem “what is the value of ‘true OR not true’?” is true, by the fact that our box outputs a rock when we input one, and these actions correlate to inputting true and not true in a logic gate. We are assured that the answer to the problem correlates to the output of our logic gate box simply because there is no other way it can happen- we can’t put in a rock and not get a rock out, by virtue of the construction of our logic gate box. The rocks and the device obey the laws of physics, simply act as they must, just as surely as if we let go of a rock near the earth’s surface it will accelerate downward. No other option is possible in an internally consistent universe.

In this sense we are exploiting the nature of physical reality, harnessing it in such a way that we are sure the results will correspond to an answer to a question. The logic gate box example is cumbersome, to be sure, and in actuality the physical analog to propositional logic was not created to be mechanical but electrical. In modern circuit design, the box with internal mechanisms is replaced by a meticulously machined circuit, the rocks with voltages. We know that when we input “2+2” into a pocket calculator that the resulting answer is “4,” because the way the device is constructed simply prevents any other possibility.

The chain is set up, then, as such: A computer is a system of circuits, across which flows electricity. These circuits are analogous to logic gates, or operators in the mathematical system of Boolean logic. Operations in Boolean logic, in turn, can re-create the full spectra of algebraic operations (addition, multiplication, etc.) So we emerge at a point where computers can manipulate algebraic expressions for us (Unger, 1989.)

It is also worth noting that at first it seems as if we get something for nothing by the construction of these devices that use the playing out of the physical dance of matter to output answers to our questions. However, the logic circuit taking in a series of voltages and putting out a series of voltages has no meaning, carries no information in itself. It is no more or less meaningful than a rock sliding down a cliff, a wave breaking on a shore, a planet forming from dust. These are all things that “simply happen” as they happen. It is the context of the circuit- the fact that we attribute a specific function to it, that we attribute specific meanings to its input and output- that gives it any worth in the first place. The intelligent interpretation of the input and output is still required.

The issue here, it seems, is that there are at least two levels of interpretation of the phenomena (or all phenomena perhaps): one is that there are rocks and there is physics, and the rocks are obeying physics, and that is all. The other interpretation is the logical paradigm that birthed the “rock-logic gate”, that is, the viewing of the rock box’s output as corresponding to a logical problem. The extraction of the meaning of the rocks is on a higher order than the physical process that actually outputs the meaning, just as there is meaning here behind the squiggles of pigment on paper that you are reading. The realm of “meaning” is a higher level, a sort of meta-reality that arises from the physical aspect yet is decidedly differentiated from it. This question of meaning seems related to the conceptual-level leap between matter and mind; while the leap may not be of the same sort and magnitude, both are examples of necessitating a “higher level” of viewing a system, or “stepping out” of the current system in order to view it more completely.

It is interesting at this point to demonstrate the similarities between circuitry and neurophysiology: just as the complicated functioning of a computer arises out of the stringing together of simpler circuits, all the functioning of the brain emerges from the collective function of perhaps ten billion neurons. Neurons are nerve cells, and they function in specific and well-defined ways. They receive input and send output to each other, and they work in tandem to produce what we recognize as intelligent action.

There are differences here between the brain and the computer, but the similarities are present as well. The reduction of brain processes to the simple monad that is a neuron is analogous to the computer being made of constituent circuits.

Although much is left to the imagination of the reader at this point (and none of it a trivial matter), at least some of the mystery of the workings of a computer is dissolved. On the basic level it is a physical analogue to our problems of propositional logic, on the highest level it is a complicated manipulator of algorithms, programs, user inputs, outputs. A great many levels of interpretation and instruction exist between our highest level, that of issuing symbolic, lingual commands to the computer, and the lowest level, the playing out of electrons in a circuit.

The barrier is opaque, such that the computer commonly takes on the mystique of a “black box” of mysterious workings. Words appear as if magic on a screen, corresponding to the lettered buttons one presses. The amount of processing and circuitry involved in just this simple case of word processing is staggering, but the result is the possibility of communicating intent and instruction to the computer with a minimum of abstraction.

Again the analogy to the brain makes itself clear. With the brain we find ourselves confined at present to the front-end, the highest level of function and output. We encounter difficulty trying to peer into the workings and decipher how all this emerges- the “black box” nature of it causes it to be so. The effect is sort of similar to attempting to decipher an alien computer.

Although computers do not yet accept commands in plain language, they do have a set of commands that amounts to their own language that is approximate to our own. If, for example, I want a computer to solve the problem “what is the value of 2+2,” it is impossible to consider constructing a computer for this purpose from the ground up, which would be to build a circuit myself. (That is to say, tackling the problem at the lowest level is impossible for the lay person.) However, by making use of these high levels of abstraction, the user interface and such, I can solve the problem with a minimum of difficulty by translating it into a short program.

In this case, the level of abstraction between the conceptual problem and the circuit design is much higher than that between the problem and the written program- in effect, these levels of complexity that separate the bottom logic circuits and the top levels (the graphical user interface, operating system, and input,) serve to bring the functionality of the computer closer to our own natural, abstract and symbolic thought process. They allow for a flexibility to the system.

THE ORIGINS OF ARTIFICIAL INTELLIGENCE

As computers have grown in complexity, naturally the scope of problems they may tackle grows, too. Originally, computers were used to solve difficult problems involving many repeated calculations- for example, a numerical solution to a differential equation. These computers were highly specialized, and the means of manipulating their inputs were highly abstracted. As they were refined toward their current state of being generally accessible, applicable to a large number of problems, and adaptable to a wide range of problems (from managing bank accounts to tracking inventory, to global communication) computers became indispensable aspects of daily life.

Artificial intelligence, in its practical sense, is simply an attempt to expand the usefulness of a computer. It is by definition the attempt to emulate (or, perhaps more strongly, give rise to) those actions and thought processes characterized as "intelligent." Naturally there are many problems even with simply defining what intelligent behavior is.

Perhaps putting a finger on the exact meaning of artificial intelligence is difficult, but it is easier to say what sorts of problems being tackled in computer engineering today involve artificial intelligence. For example, it would be of great practical application to many businesses to have a computer that could converse with people naturally- either in text or in voice, although voice presents the further problem of language recognition. In this case, the problem is in extracting meaning from a user's statements- understanding what is being communicated and responding appropriately.

An interesting thing to note at this juncture is that while it seems that humans have little to no problem with the basic actions and use of language, this is not the case. Consider that for most humans, a period of several years of development is necessary to learn the basics of language- after that, many more years are required before a person has a mature grasp of language. On top of this, misunderstandings and miscommunications are commonplace. In light of this, then, we should not consider natural language use a simple task, or wonder at why it seems so hard to create a program that reasonably emulates a conversation with a human. In fact, language is an incredibly complex issue, evidenced by our own inadequacies in its area as well as the long period of learning associated with it.

This is a common theme that emerges among problems of artificial intelligence: what seems simple and second nature to us only seems that way because the fundamental workings of our brains are necessarily opaque to us. At first glance, getting a computer to understand what is meant by the command "place the small box on top of the large box," for example, may seem almost trivial, but when one attempts to boil the command down to its primordial symbols, and meanings, and the rules governing their interactions, the problem quickly escalates. Most problems of application contain these hidden difficulties.

Artificial intelligence applications crop up in multiple areas. Almost every application that uses computers has some aspect of artificial intelligence either already in place or in development. Any problem that isn't one of simple math or logic can be seen as a problem for artificial intelligence to tackle. Generally these are problems of pattern recognition, "fuzzy" logic, difficult decisions with unclear or changing criteria, learning, or creating, among other things. Specifically, some current problems include facial recognition, voice recognition, providing computerized opponents in games, and other things of this sort.

Some researchers attempt to tackle artificial intelligence in a more direct manner; in effect, they attempt to learn about artificial intelligence itself for its own sake, through investigation of various aspects of intelligence. Creating a robot that is in some sense aware of its surroundings doesn't have many practical applications in its current state- it certainly will in the future, when the techniques for creating such and intelligence are refined- but as it stands, such an endeavor serves mostly to educate us about methods of creating an artificial intelligence. As another example, the robot AARON, created by artist Harold Cohen, can create works of art based on its own internal knowledge of various subject matter and artistic principles. Clearly this robot has no practical application; it is intended to give us a glimpse into the process of artificial intelligence (as well as being a statement about art.)

When viewed in this light, artificial intelligence seems simply to be the creative application of a computer to a difficult problem, perhaps one that is not easily solved by conventional means, or perhaps something that is traditionally exclusively in the realm of human ability. Artificial intelligence means the creation of a program that can solve some problem, enhancing the abilities of the computer. Once the problem is solved, we sit back and congratulate ourselves on the creation of intelligence.

This is a very limited view of artificial intelligence, one that is confined to the current state of affairs in practical artificial intelligence applications. This view is dominated by practical results. The researchers are not interested in creating a mind or understanding consciousness, they simply work on improving and refining techniques, finding new areas of application, and any sort of practical application. Modeling the mind is not practical, attempting to create a computer that can become conscious is not practical. Furthermore, there is no reason to believe that artificial intelligence, from this approach, would ever provide the sufficient conditions for a consciousness to emerge.

However, some researchers might tell you that the ultimate goal of their research is something far greater than simply a “smart” program that knows when you should be buying milk and eggs, or governs the flow of traffic through intersections. If the result of artificial intelligence applications is to create programs that emulate human behavior and decision processes with increasing precision, then it would seem that at some point in the future computers may develop to a point where their actions, words and thoughts seem as intelligent as our own. In this sense, computer behavior approaches human behavior through successive stages of refinement and research. In many areas computer ability has already surpassed human ability in terms of finding better or quicker solutions to problems.

Although currently our technology may be limited, it may be possible in the future to converse with your personal computer as naturally as with your friend. A computer may take your order at a restaurant, or answer your phone call at a service desk. The more ambitious envision a future in which intelligent robots intermingle with humans in daily life. These may all be possibilities or grossly over-optimistic hopes, but they serve to point toward the ultimate goal of A.I.: the total emulation of human behavior.

Naturally there enters into the discussion a great deal of ethical and philosophical concerns. Even if we further and further approximate human activity through computer means, to the point where it is indistinguishable from our own activity, does the computer ever become entitled to the same rights as humans? Can a computer have the subjective inner experience of the mind, as we have? Does the latter necessarily follow from the former, or is it the product of some other process?

SCHOOLS OF THOUGHT IN ARTIFICIAL INTELLIGENCE

As computer technology increased in complexity, it became clear that computers had potential far beyond simple calculating machines. Scientists saw in the computer the possibility of mimicking our own intelligent behavior, if not the possibility of the computer becoming an intelligent artifact. The study of artificial intelligence emerged within two separate camps: one intended to perfect the computer as a symbol-manipulating system, the other sought to use the computer to model the brain.

Both of these camps had the goal of artificial intelligence in mind, however, their methodologies as well as their basic concept of what artificial intelligence could be were different. In its initial stages, artificial intelligence research was dominated by the symbolic paradigm school (Dreyfus, 1988).

Philosophically, this school was the heir to western rationalism and reductionism. In his most famous work, *Discourse on Methods and Meditations*, Descartes (1637) described his impression of his thought process:

> "Those long chains of reasoning, so simple and easy, which enabled the geometricians to reach the most difficult demonstrations, had made me wonder

> whether all things knowable to men might not fall into a similar logical sequence.” (pg. 15)

Descartes identified his mind, even his very existence, with the ability to have rational thought. In his philosophy, rationality was what separated the human from the animal, and created the basis for all knowledge and experience.

Propositional logic is precisely the process Descartes describes in the above quotation. It is a formalized, axiomatic system that rigidly processes statements to arrive at further truths. The process of deduction, as in the example of geometry that Descartes presents, is seen to be the ultimate triumph of human rationality- the ability to manipulate symbolic statements in a purely abstracted mental space.

For this school of philosophy, the human is composed of two parts: the base material body, existing in the *res extensa*, and the mind, within the *res cogitans*. The body gives rise to emotion, instinct, and all manner of irrational impulses which are systematically suppressed in favor of the rationality of the mind. This is a strictly dualistic model, holding the mind as separate from the body (Descartes, 1637).

Followers of this philosophy, which was and still is the dominant mindset of science and the west, saw in the computer the potential to re-create this process of symbolic manipulation that they believed was the entirety of the human mind.

It was shown quite early on that computers could manipulate symbols on a fundamental level- since computers deal with numbers, and numbers can be made to encode any sort of information, it follows that computers can be precisely the sort of physical symbol system that our brain is. It was almost a trivial matter to program a computer to run through geometric proofs, or propositional logic theorems; the logical systems necessarily reduce to a branching tree of possible theorems provable from the given axioms, and running through this is exactly the forte of the computer.

Eventually, however the symbolic paradigm school began to run into problems. A computer could be made to manipulate symbols that correspond to the external reality, and from these manipulations make inferences about the external world. The semantics of how to relate these symbols were simple to define, but the problem of breaking down the external things into their atomic properties proved troublesome. This was a problem the incipient phenomenologists Husserl and Heidegger faced as they attempted to find the fundamental axioms and primitives underlying all experience. Husserl eventually concluded that such a project was futile to attempt (Dreyfus, 1988).

The second school of thought, which came to be known as the connectionist paradigm, came to the fore in the 1970s as the symbolic school ran into the "commonsense knowledge problem." This school was heir to the neuroscientists, and instead of attempting to create the mind from the top down, as it were, they sought instead to model the brain.

Many concepts come together to form the intellectual background for the connectionist school. Foremost perhaps is the notion of emergent properties, or emergent phenomena. Hillis describes emergent behavior in his essay, "Intelligence as an Emergent Behavior" (1988). Emergent behavior arises from the interactions of a large collection of simple parts. The rules that govern the simpler parts govern the high-level behavior of the system as well, yet it is not obvious from the rules alone what behavior the system will exhibit as a whole.

The intention of Hillis's essay is to probe the possibility that intelligence is an emergent property of the system that is the human brain- or rather, any system of a similar construction. Showing that this is the case is the intention of the connectionist school, but we will return to the justification of this position later.

Hofstadter (1979) describes emergent phenomena as that which exhibits itself on a high level of interpretation, yet is completely indescribable on the lower level that gives rise to the higher-order levels. In a sense this is a stronger conception of emergent behavior: it follows from the low-level behavior, but is completely inscrutable from a low-level perspective.

His example is Godel's theorem, which is a true statement about any logical system yet it cannot be proven from within the system. As soon as you "step outside" or "step above" the system, however, viewing the problem on a higher level than the low-level symbolic manipulation of theorems, the theorem is obviously true. The crux of the theorem is that it is decidedly unprovable from within the system, yet at the same time decidedly true from without the system. This seems to be a perfect example of a behavior of a system (in this case a true statement about the system itself) that is not even hinted at on the low level of the system (the manipulation of the axioms and theorems).

In practice, connectionists are identified with the method of creating simulated neural nets. Hebb (1980) outlined the basis of this technique, and in the early 1970s Minsky and Papert displayed a mathematically rigid definition of a specific neural net that exhibited several impressive traits. Minsky and Papert named their creation the "perceptron," however their original treatise was misconstrued by and large as showing that neural nets could not adequately simulate the mind; in fact, Minsky and Papert had only proven that very specific nets had severe limitations, when in fact these nets were not the foundation of connectionism (Papert, 1988).

Connectionism grew to encompass theories of massively parallel neural nets (Bechtel & Abrahamsen, 1991). Neural nets are essentially ripe for all sorts of surprising emergent behavior by their very nature of being large collections of simple, interacting elements, or neurons. They have displayed the ability to learn: nets by definition must give rise to behavior not explicitly coded to begin with. All that is provided initially is the framework of the net, the substrate of neurons and connections, and the behavior emerges naturally.

Further reason to believe that a neural net may in fact give rise to intelligence eventually can be found by examining the one already-existing example of intelligence we have on hand: ourselves. It is a ready fact that we do not spontaneously pop into existence as a mind all at once, so why should we attempt to create an intelligence as such?

By studying cognitive development of infants and children, one can practically see a consciousness emerging, ever so gradually, as the complicated and intertwined system that is the brain grows and learns. Piaget outlines several stages of child development: First the baby organizes its actions to deal with the immediate external world. Eventually they learn to think, to use symbols and internal images, yet their thinking is illogical and very different from that of an adult's. Eventually the child learns to think systematically in a "purely abstract and hypothetical plane" (Crain, 1980).

If consciousness, or a mind, is the emergent property of the immensely complex and parallel system of neurons that constitutes a brain, then it would make sense that it takes a long while to grow and emerge. The important notion here is that consciousness is a great deal more plastic than the average person acknowledges. It is far from an “on or off” proposition- it can take forms entirely alien to the world we know, it can come in shades of awareness and lucidity, it may be disconnected from the external or anchored in it. As a child develops, its mind gradually emerges, increasing in consciousness.

If in nature, the brain “substrate” is first created, and then intelligence is allowed to grow and develop and learn about the external world, then why should we attempt to create a mind, fully-formed in one go? It would make sense to mimic the natural process. The problem becomes the creation of a simulated infrastructure of neurons that are sufficiently organized to be able to display all the properties of the human mind. The neural net must have the sufficient conditions of intelligence, in organization and scale.

In an attempt to capture this concept that a mind is somehow more than the sum of the parts of the brain, McKenna (1976) proposed the holographic mind theorem. In essence, he theorized that the mind was fundamentally a higher-order phenomenon, that it was encoded in the low-level neurons. The analogy of the two-dimensional holographic plate giving rise to the three-dimensional holographic image is a good description of intelligence emerging from the fundamental lower level.

The holographic theory holds on another level- the distributed nature of memory and mind. If a hologram is cut in two, both halves retain the entire image at a degraded quality. The information encoded in the hologram is distributed across its entirety. This holds true for the physiology of the brain in some cases: There doesn’t appear to be a one-to-one correspondence between memories and neurons. Two related memories are not stored physically in adjacent neuron clusters, they are stored spread across the brain.

This is also a property of neural nets, that the learned behavior of the system is distributed across the net. In other words, it can stand to lose a few neurons without the entire system breaking. The storage is redundant and distributed, much as in the brain.

In the case of the natural human mind, its demonstrated similarities with the neural network's considerably "organic" process of growth and learning, as well as its similarity on the basic level of organization, seems to suggest that consciousness emerges as a high-level property, fundamentally distinct from and inscrutable on the level of interactions between neurons. This would indicate that the means to creating an artificial intelligence would be discovering the essential features of consciousness. Hofstadter (1979) suggests that at the core of consciousness is a self-referential, self-modifying feedback loop that crosses between the low and high levels, allowing for a causality in the high level that would influence the lower level.

> "In short, an 'I' comes about – in my view, at least – via a kind of vortex whereby patterns in a brain mirror the brain's mirroring of the world, and eventually mirror themselves, whereupon the vortex of "I" becomes a real, causal entity."
> (Hofstadter, p. 6)

In this view, in fact in the view that precludes the creation of an artificial intelligence, it is the system and the meaning, rather than the physical backbone or substrate that carries the system, which is intelligent. Thus intelligence may emerge from any sort of physical system of sufficient complexity- brain or computer, or anthill or society.

If this is the case, then the implications are far-reaching. One may argue that intelligent behavior does not necessarily give rise to an actual mind- that is, the subjective experience of reality that we have. This decidedly nonphysical "inner realm" of thought and perception seems to defy explanation in physical terms, and yet this is why Hofstadter's conception of strong emergent behavior seems like a promising link between the low-level neurons and high-level consciousness.

If we are to take the symbolic manipulator's position, we necessarily refute the existence of consciousness, of the subjective. Hebb wrote that awareness is illusory – indeed, if we accept the pre-eminence of the physical, then we are forced to conclude that any sort of "mental screen" and "inner subjective sensorium" that we experience (which, in fact, is the only thing we experience or know) does not, in fact, exist. Of course, this is impossible to actually achieve, as it means in essence the erasure of oneself; a materialist standpoint is at odds with our preeminent experience.

Perhaps the best way to frame the ultimate question about the origins and modes of consciousness is to rephrase it in a personal matter. I know that I was born many years ago, and grew and developed and gradually became self-aware. I can trace the origins of my consciousness to the convoluted grey matter in my skull, and my biological origins to my parents. To ask if a computer can become conscious is to ask if I could have discovered myself to be a computer rather than a human.

Is it possible that, instead of growing up and finding myself to be a biologically-based human being, rather I found a growing awareness of myself as a computer-based intelligence? Instead of a biological mother and father, could I have a team of computer scientists? Ultimately, the goal of this IQP is to answer the question, could I have the same subjective experience as I do as a human if I were a computer? Or rather, could a computer form the sensorium, the mind, the consciousness that is so familiar yet so vexing to us?

METHODOLOGY

Probing the depths of the computer “mind,” to date, has traditionally been the work of philosophers, not scientists. Computer scientists and AI researchers can only attack the issue from a practical, functional standpoint, and any speculations they make about the future of AI must be inevitably unscientific. The issue cannot be broached in the confines of a scientific or mathematical analysis of neural nets.

Rather, the question of what a mind is and where it comes from is firmly in the realm of traditional religion, metaphysics, and philosophy. Unless one is content to simply dismiss the mind as an illusion, one must acknowledge the shortcomings of science in the analysis of the mind.

One approach to answering the question, which has never been attempted to my knowledge, is the creation of a portrait of artificial intelligence. This untried method may yet provide some insight into the mind of a computer, if such a thing exists, by virtue of portraiture’s singular ability of presenting an object’s essence.

DATA COLLECTION AND ANALYSIS

Lawrence-Lightfoot and Davis describe in their book, *The Art and Science of Portraiture*, a process of portrait creation that moves beyond a factual representation of a subject towards a more holisitc examination of one’s relationship with the subject. The purpose of such a portrait is to examine the entirety of the subject, in effect capturing its essence.

If this process could be applied to an artificial intelligence, then one might be able to gain some insight into the possibility of that intelligence actually possessing a mind. Clearly this would provide a fresh insight into an A.I. "mental process," with the intention being a search for true intelligence, or a consciousness.

For such a portrait, the data collection was undertaken on two fronts: I interviewed a current artificial intelligence researcher, with the goal of understanding the field as it exists today, and where the researchers believe it is headed in the future. This revealed that the current atmosphere in artificial intelligence is dominated by practical concerns, many researchers may not have put any thought into the philosophical implications of their work. The intention of interviewing a researcher was to establish the real possibility of the things discussed in this paper, to provide a professional view on my speculations . My primary concern was to avoid the sensationalism of pop science and pop culture.

The other front was my direct interaction with the artificial intelligences themselves. In these interactions, attention was paid to understanding the processes behind the artificial intelligence, and discovering any potential or actuality of consciousness. However, general traits and properties of the artificial intelligences were also observed. The interactions were sometimes little more than observations, although some took the form of conversations, depending on what the particular artificial intelligence was intended for. I paid attention to the potential abilities of the artificial intelligences.

When interviewing the researcher, notes were taken but not tape recordings. This provided a more natural method for me to record the researcher's general responses and beliefs about the topic. Specific quotes from the researcher were not required for the portrait, although a few were recorded. In the end product of the portrait, the general stance of the scientific community toward the philosophical aspects was most important- in the comic, such stances were paraphrased rather than directly quoted. As such, complete transcripts or recordings were not necessary and would only serve to complicate the interview.

Additionally, this approach provided the means to sketch the researcher, which proved valuable .Visual data was gathered at the site of the interview- I took some sketches of the environment as well as the researcher. The researcher became a character in the portrait, and so a preliminary sketch of his appearance proved to be useful.

The interactions with the artificial intelligence itself were by their nature less formal. It was impossible to adhere to only one format in interacting with these artificial intelligences; an interview format, for example, was at times unapplicable. I needed to view the outpourings of creative artificial intelligences (Aaron’s artwork, or computer-generated poetry for instance.) I paid attention to the capabilities and limitations of the artificial intelligences. By noting the general aspects of the artificial intelligences, for example, whether they followed clear patterns or improvised, whether they were generally coherent, whether they were clearly limited. These characteristics, taken together and across a variety of artificial intelligences helped me to establish the general characteristics of this technology.

The artificial intelligences that were available for my study were limited and mostly not even representative of the state of the art. As a result I did not weigh the limitations of these programs as heavily as one might have. The most advanced artificial intelligences currently are much harder to gain a private audience with- and much more specialized and esoteric than a publicly available poetry generator or conversation bot. For these reasons I assumed the capabilities of current artificial intelligences to be greater than what I observed.

Once the data was established, it required analysis that would allow it to translate into the portrait and the comic medium. Specifically, I had a body of data that consisted of notes from interactions with artificial intelligences and my interview of a researcher: I needed to decide first what aspects of artificial intelligence were essential, and then establish how to represent these themes in my comic portrait.

By examining the notes from my interactions with the artificial intelligences, I located the most common themes that recurred in multiple cases. These themes were adaptive behavior, unexpected behavior, and flexibility.

Most of the intelligences, and the most successful ones, showed the ability to adapt their behavior with time and experience. I felt that this was a crucial aspect of intelligence in accordance with the theories of the mind developed earlier in this paper, as well as being a successful way to emulate human behavior. It is worth noting that if an artificial intelligence is to mimic our human minds in its capabilities, then it makes sense that it should have to learn much as we do.

Unexpected behaviors showed themselves something as a consequence of adaptive behavior, but also as a result of the intelligence being programmed on a more general level than the basic output. That is to say, if the intelligence is given greater flexibility to choose its output, there is more potential for unforseen output, as well as greater emergent phenomena. This theme also further differentiates the artificial intelligence from a more rigorous computer program or formula for calculation.

The last trait of flexibility is tied to the idea of the intelligence showing unexpected behavior and learning. The artificial intelligences were usually confined to a specialization of behavior, but by virtue of complex behaviors arising from their generalized programming, could do more than supply one output to a specific input.

These themes needed to find expression in the portrait in one of two ways: as the subject of the comic would be artificial intelligence specifically, the aspect of artificial intelligence could be specifically discussed in dialog between characters, or addressed in the plot. Once the plan for the comic settled as a narrative and a fiction investigating the meaning behind artificial intelligence, these themes were incorporated easily.

TIMELINE

The data gathering period of the IQP will be limited to a time of about seven weeks. In this time period, I need to establish contact with experts in the fields relevant to this topic- artificial intelligence researchers and philosophers. I aim to interview at least two artificial intelligence researchers, specifically with the aim of finding their thoughts on the feasibility of computer consciousness. Contact must be made within the first week so that interviews can be scheduled within the time period. Additionally, at least one philosopher should be consulted to critique this understanding of consciousness, and offer insights as well into the possibility of computer consciousness.

I am geographically restricted to Massachusetts, and perhaps a much smaller region than that. Certainly areas within Worcester and Boston are feasible, but beyond that would be very special scenarios.

During the inevitable period between the scheduling and the interview itself, I will have the opportunity to further refine this paper, and to seek out any instances of possible artificial intelligence. Quantitative values for the number of "artificial intelligence interviews" to make are probably impossible to predict.

It is very likely that the artificial intelligences available for scrutiny will be very limited and primitive. Their imitations of human behavior most likely will be transparent machinations, lacking the spontaneous emergence of surprising abilities, or unexpected behaviors. The programs most likely to exhibit these traits probably won't possess the ability to converse naturally with humans; thus the artist must be prepared to recognize intelligence in any number of systems, or any sort of manifestation.

For example, as we expect intelligence to spontaneously emerge, perhaps unexpectedly, it is possible that it may display itself in the nonstandard states of equipment as it is malfunctioning, for example. When a device malfunctions, all bets are off, so to speak, as to what it is capable of and what it is doing, and why. To form an analogy, if the computer substrate acts as the "primordial ooze" out of which awareness may emerge, the strange jolt of an error may provide the spark, the bolt of lighting or whatever event allowed for the emergence of life. That is, the potential for intelligence may already be there, in some form and capacity.

Of course, in these cases errors need only be examined that do not immediately halt the operation of the device. If, for example a computer program simply halts itself at a syntax error, there is nothing of interest there. On the other hand, if it enters an infinite feedback loop, or even modifies itself, or makes no attempt to stop itself, then the possibilities are much more interesting- we are looking for emergent properties that surprise the viewer with their output, so where better to look?

After the seven weeks of B term I hope to have enough data to move forward with the creation of the portrait itself. This will begin in C term and last another seven weeks- the final week being left open for reflection, presentation, editing leaves six weeks for the creation of the piece.

A WORD ABOUT THE COMICS MEDIUM

The final creation of the portrait will be accomplished in the medium of comics. McCloud defines comics as “sequential art,” or more formally, “juxtaposed pictorial and other images in deliberate sequence.” In his book, *Understanding Comics*, McCloud (1997) outlines the unique properties of comics in portraying a variety of situations.

Comics as a whole are still attempting to achieve recognition and critical review as an art form on their own merit. Traditionally comics are relegated to the lowest form of pulp entertainment, intended for an ostensibly “low brow” audience, and this is cause for many to refrain from placing comics alongside music, visual arts, or cinema. Many informed essays have been written about comics' merits, however, and gradually it is achieving the acceptance that legitimate, meaningful works can emerge in this medium. It is not the purpose of this IQP to discuss the reasoning behind comics adherents, but it is a topic that necessarily must be addressed preemptively.

As for comics' particular capacities as an art form, one of the primary strengths that is generally agreed upon is its ability to portray time uniquely as the juxtaposition of spatial imagery. This may lend itself particularly to the strange case of the gradual emergence of a consciousness that would be considerably more difficult to portray as a static image, for example.

Comics are unique in that they, like cinema offer the viewer a look into the visual world the comic creates, but unlike cinema the view is established by a concrete visual symbol on the page- that of the panel. The panel, or the polygonal frame that separates out discrete images and times exists in the symbolic realm just as the images it contains. By visually pointing out the panel and breaking the fourth wall, the comic could bring the reader to awareness of his or her window into the comic world, which is a striking metaphor for one's window into the "real" world, or consciousness itself, which is the essence of the question of the mind.

Comics are not tied to fiction; pioneering artists such as Spiegelman and Gonick have created critically acclaimed nonfiction comics: Spiegelman, in fact, used the comic medium to create a portrait and biography of his father as a survivor of the holocaust. McCloud presented his seminal comics-as-art thesis within the comics medium itself. Thus the possibility of creating this portrait of artificial intelligence itself in the medium of comics is not impossible or even unprecedented. However, I decided upon a fictional narrative in order to do something new; all of my academic comics in the past have been nonfiction comics.

There are three basic stages to account for in the creation of the comic: writing, penciling and inking need to be scheduled. Writing the comic took place over a week-long period, although I had been thinking about its subject and general presentation for a month prior to the writing. After this stage, the comic was constructed on a page-by-page basis according to the plan.

The written plan for a comic can take many forms, depending on whom is writing and for what purpose. In large-scale industry comics, the writer needs clear indications of speech, attitudes, inflections, postures; it is much closer to a movie script. However, as I am both the writer and the illustrator for the comic, I need little more than brief reminders and thumbnail sketches of what the comic will look like. My plan consisted of a sketchpad with rough thumbnails of the page layouts and a text document outlining the plot arc, with significant portions of the dialog written out explicitly.

My written plan was perhaps the least orthodox aspect of my creative process. Most comic artists begin with a clear script divided into pages; I started with a loose collection of notes and plans but no clear script. I knew the planned development for the plot, and followed it along as each page was created.

Page creation started with a general plan for the layout of the panels and the dialog or captioning (or lack thereof) to be found in each panel. The page was then penciled on 9x11 inch smooth bristol board, the pencil lines were inked and the pencils erased.

Penciling was probably the most time-consuming portion of the portrait creation. Figures and word balloons needed to be firmly established, and the specific detail of each panel needs to be shown. The laying out of the panels themselves can be a time-consuming process, depending on how exacting I want to be.

Inking simply entails going over the images in ink, blocking out the solid ink portions, shading the appropriate portions and whiting out any mistakes. After the ink dries, the pencil is erased, cleaning up the appearance of the page. Attention is given at this stage to balance of light and dark portions of the page, creating a full range of dark and light shades without losing the definition of the figures.

CONCLUSIONS AND REFLECTIONS

I feel that my portrait of artificial intelligence is a success. Capturing the essence of A.I.- which was the goal of the portrait- was not easy or simple, and during the creation I felt it hard to gauge my relative success or failure. After the completion of my portrait, when I could re-read and absorb my comic as a whole, I was pleased with its general tone and effect.

The process of creating this portrait was a very beneficial one for me. As with most major projects, the things I learned branched out and touched on many topics. This project was my introduction to the rigors of social science- a field which that I was largely ignorant of beforehand. Having to learn and practice it myself was a new and enriching experience, and stimulated my interest in the subject. It prompted me to read further on the subject.

The nature of the project was to use an artistic sensibility to investigate and answer a question- something that I found fresh and exciting. Throughout the project, I thought about the purposes, applications and uses for art. Is art a tool for communicating deeper or stranger or more subjective truths? Does it help us answer questions in ways that aren't viable for ordinary language? Questions like these forced me to confront and refine my ideas about the nature of art.

The project was also an incentive to examine critically my philosophical and speculative theories about the mind. The result of these introspections can be found in this paper, but this is by no means a definitive or complete theory. I am fully prepared to re-examine my ideas at any moment, and I am not fully satisfied with the conclusions I have drawn.

For example, the theory as I have presented it is a very dualistic construct- the mind and body are separated as Descartes would have it. This doesn't fully satisfy my intuitive feelings on the matter. However, I feel that this theory does much to reconcile rational, scientific materialism with a more philosophical idealism.

I feel that fundamentally, the scientific negation of our subjective experience is at odds with our very knowledge of existence, and many will attempt to take one extreme position or the other. The critical thinking I have applied to this problem is simply my attempt to make peace with the world-at-large, an understanding of reality as it is presented to me, as something both internal and subjective, and external and objective to all appearances. In my portrait and to some extent, in my paper, I attempt to draw a cause-and-effect link between the ordering of information in a mind-system, and the experience of reality by a sentient observer. The conclusions here are by no means certain, but I suppose they can be rationally held and defended.

The second issue with the mind/body dualism is accounting for the qualitative difference in essence of mental constructs and physical constructs. I feel my current theory as presented here does little to reconcile the fundamental difference between a sentience and a brain-substrate. In short, the link between the "strange feedback loop of self-modifying information" could probably be fleshed out in further investigation.

The metaphysical link between the small-scale ordering of our minds and the larger-scale ordering of our planet or even the basic universal matter presented itself somewhat early on in my investigations, but I neglected to include it in my portrait as unnecessarily speculative, and not related to the topic of artificial intelligence. It would seem that the self-sustaining mental loop that creates our minds could be the same loop that creates self-sustaining matter patterns, or life; on a larger scale still it would be the development of basic energy into particles and macro-ordering as structures such as stars and planets. If our noetic construct is the consciousness of a matter-substrate, then our matter is the consciousness of an energy-substrate of particles. This aspect of the theory bears further investigation.

These are some examples of the development of my philosophical theories during the course of this project, but this is by no means a complete theory. I suspect I will be investigating this topic for a long while; I was interested in it before this project and I will still be interested in it later. However, this project can be seen as a concrete statement of my theory as it stands to date.

The process of the comic creation was intense but rewarding. It forced me to confront shortcomings in my own technique and style, again a growth process that has not ended to date, and will continue into the future. From a technical standpoint, there are many pages that I find lacking and would do over if given the time. In what amounts to perhaps an ulterior motivation in creating this portrait, I constantly found myself critically evaluating my work, attempting to remove the qualities I felt would identify it as an amateur production. Of course, I feel a cursory glance reveals that I am not a professional, but I did constantly push myself to evolve my technique beyond the pitfalls I see in other amateur works.

At times I felt I couldn't finish this comic by the deadline and would have to compromise parts; in the end, however, the schedule worked rather well. It is perhaps a painful irony that I have spent a term writing and drawing this comic, and two other terms besides planning it, and a casual reader could absorb it in under 15 minutes.

This is, beyond doubt, the largest exhibition of my work, the most important application it has seen. I feel it is somehow legitimizing, to see my art which had been an amusing distraction in the past become the subject of an IQP, and for it to be exposed to many more people than is usual.

In the end, this project has been a major push for me to continue learning and growing in my academic study, my thought and my art. It really has been beneficial to me in many ways, and it amounts to the first time I have placed my work under such scrutiny. It has been a difficult, complicated, rewarding and enriching experience for myself and my cohorts, and I hope that some portion of my experience in this translates to the reader.

REFERENCES

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Cambridge: Basil Blackwell, Inc.

Crain, W. (1980). *Theories of Development: Concepts and Applications.* Englewood Cliffs: Prentice-Hall, Inc.

Descartes, R. (1637). *Discourse on Method and Meditations*. New York: The Bobbs-Merrill Company, Inc.

Dreyfus, H. & Dreyfus, S. (1988). Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint. In S. Graubard (Ed.), *The Artificial Intelligence Debate: False Starts, Real Foundations* (pp 15-44). Cambridge: The MIT Press.

Grossmann, R. (1965). *The Structure of Mind*. Tennessee: Kingsport Press, Inc.

Hebb, D. (1980). The Structure of Thought. In P. Jusczyk & R. Klein (Eds.), *The Nature of Thought: Essays in Honor of D. O. Hebb* (pp. 19-36). Hillsdale: Lawrence Erlbaum Associates, Inc.

Hillis, W. (1988). Intelligence as an Emergent Behavior. In S. Graubard (Ed.), *The Artificial Intelligence Debate: False Starts, Real Foundations* (pp 175-189). Cambridge: The MIT Press.

Hofstadter, D. (1979). *Godel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books, Inc.

Lawrence-Lightfoot, S. & Davis, J. (1997). *The Art and Science of Portraiture.* San Francisco: Jossey-Bass.

McCloud, S. (1994). *Understanding Comics: The Invisible Art*. New York: Harper Paperbacks.

McKenna, T. (1976). *The Invisible Landscape: Mind, Hallucinogens, and the I Ching.* New York: Charles Scribner's Sons.

Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge: Maple Press Company.

Papert, S. (1988). One AI or Many? In S. Graubard (Ed.), *The Artificial Intelligence Debate: False Starts, Real Foundations* (pp 1-14). Cambridge: The MIT Press.

Stevens, Charles F. (1966). *Neurophysiology: A Primer*. New York: John Wiley & Sons, Inc.

de Solla Price, Derek. "An Ancient Greek Computer." Scientific American June 1959: 60-67.

Unger, Stephen H. (1989). *The Essence of Logic Circuits*. New Jersey: Prentice-Hall, Inc.