ARTIFICIAL INTELLIGENCE: AN ESSAY ON COMPUTERS AS LANGUAGE USERS

by
RICK WOODBURN

Submitted to the Department of Philosophy
of Saint Mary's University
in partial fulfillment of the requirements for the
Degree of Master of Arts

Halifax, Nova Scotia
1993

ISBN 0-315-90972-2

Members of The Examining Committee

Dr. June Blair

Wayne Grennan

Dr. Wayne Grennan

Dr. Sheldon Wein

Dedication

Although I would like to take full credit for this thesis, I cannot. There are those who have in one way or another put time and energy into the completion of this project. Academically, I would like to thank Dr. Wayne Grennan, who without his tireless rebuttals and rap sessions, this thesis would not have taken shape. Dr. June Blair for her sometimes maddening and relentless prodding turned a hapless research paper into a Master's level thesis. Sheldon Wein, who took the time to assess and correct the thesis. Finally the entire Philosophy Department for its help over the years. Non-academically, I would like to thank Lisa Paterson for her love and dedication over the years. And my mother, Barb, and my father, Eric, for their support in any of my endeavors over the years.

Abstract

The author has embarked on an investigation of Artificial Intelligence and Cognitivism. The focus is directed at AI's attempt to implement a program to endow a computer with intelligence. However, this endeavor may have been undermined by John Searle's Chinese room experiment. Searle, in Minds, Brains, and Science, rejects AI's fundamental claim that a properly programmed computer could ever be intelligent. His thesis relies on two main assumptions: (1) The formal structure of a computer is insufficient to produce understanding, and (2) the "hard wiring" of a computer, as opposed to the brain's "wet wiring," is insufficient to cause mind. These assumptions will be put to the test in rebuttals presented by several philosophers and AI researchers. However, each of these criticisms will be laid to rest, or at least questioned, by the author. The point of inquiry is now directed at the type of program needed to endow a computer with linguistic understanding. The quest begins with human language acquisition within a community of languages users and ends with a thought experiment. The experiment illuminates the nature of the program needed to produce linguistic understanding in a computer.

Contents

Introduction

Although this is mainly a philosophically oriented paper, the fields of artificial intelligence and cognitivism are ever-expanding and overlap into psychology, computer science and linguistics. It was the aid of modern neuroscience and even computer science that gave a new perspective in the philosophy of mind. However, with new insights there comes new problems.

It was with the discovery of how the brain operates (on a neurological level), and a new suggested model of mind, that AI experts set out to replicate mind according to a computational model. A replication of human thinking, understanding and intelligence was the hopeful outcome of this new discipline. The hopes of modelling human thought were boosted with a computational model in "the Turing Machine." Basically one could (in theory) model the neural nets of the human brain and this could be fed into Turing's computational model. These dreams died when it was found how difficult it was to design and implement these "nets." The project, but not the theory, was abandoned.

The AI experts believed if they could implement the proper program, a computer may be said to be intelligent. As computers became more complex and seemingly took on human-like characteristics there came a time to define (or redefine) "intelligent behavior." Alan Turing in "Computing Machinery and Intelligence," devised such a test for intelligent behavior. Based on the "imitation game" if something had the appearance of being intelligent, then it was deemed intelligent. A computer that could deceive a human interrogator into thinking the computer has intelligence (based on verbal responses) passed the Turing test for machine intelligence. Since most modern computers could pass the Turing test for machine intelligence, by this criterion the computers were intelligent.

However, a possibly fatal blow was struck at the heart of AI and cognitivism when John Searle presented "Minds, Brains and Programs" in 1980, and Minds, Brains, and Science in 1984. (Books that were later written by Searle are not considered in this thesis.) John Searle readily rejected any claims that a machine could either be intelligent or explain human cognition. Searle relies on two basic claims to reject AI: (1) The formal syntactical operations of a computer is insufficient to endow a computer with understanding, and (2) the "hard wiring" of the computer, as opposed to the "wet wiring" of the brain, is causally insufficient to produce mind.

The task of this thesis is threefold. First, I will review Searle's arguments against AI and cognitivism in Minds, Brains, and Science. This is in order to present a clear picture of Searle's detailed criticisms of the aforementioned. Second, I will present attempted rebuttals by several philosophers, cognitivists and AI researchers to Searle's main assumptions and thought experiments. However, I will show that each of these criticisms is either ineffective or problematic.

This process will show the difficulty in overcoming Searle's objections of AI.

The question at this point is: How can Searle be defeated, if at all? Has he dashed the hopes of AI researchers? The purpose is to find some middle ground between Turing and Searle. While Turing believes intelligence is a matter of exhibiting the proper behavior, Searle endorses the view that intelligence is a matter of having the correct biological make-up plus behavior. The argument is over the criteria for intelligence and what is deemed "intelligent behavior." Searle's criteria restricts intelligence to only those beings with the correct biological make-up exhibiting certain behaviors (language use, intentional actions, etc.), and hence he believes he has shown that the AI project is fundamentally flawed. However, I believe AI need not accede to all of Searle's claims: namely, his claim that only the biological wiring of the brain is sufficient to produce mind.

Part of what I believe to constitute "intelligent behavior" is linguistic understanding. If it can be shown in the third section how a computer may be able to come to understand a language much like a human language user does, then AI might have a chance. The third section of this thesis may ɔpen some otherwise closed doors. I inquire into human language acquisition in order to see how an individual comes to understand a language within a community of language users. This information may shed light on the type of program needed to (someday) endow a computer with linguistic understanding.

1.0

The first of three parts of this thesis is dedicated to exploring John Searle's Mind, Brains, and Science. It is also necessary from time to time to refer to Intentionality, which was also written by Searle. He dedicates the material in the former to refuting what is called strong AI and clearing the air surrounding some common misconceptions and mysteries in dealing with the mind/brain. The latter, while demystifying intentionality acts as a base to prove his thesis against opponents of AI.

Chapter one attempts to demystify the centuries-old mind/body problem and attempts to dispel the out-dated monist and dualist theories of mind/body. Searle explains that the problem for scientists and philosophers, alike has been the problem of explaining the connection between the mental and the world of the physical. That is, there are certain mental phenomena that are not easily explained in material world terms. Searle attempts to demystify four of these mental phenomena and explain their logical relationship to the physical world around us.

1.1

The first of these mental phenomena is consciousness. How is the gray, slimy mass we call a "brain" said to make us conscious? That is, how can the physical brain be said to support the intangible mental phenomena of consciousness? Searle begins to dispel the myth of consciousness by an investigation into the physical processes by which the brain operates. Since our knowledge of these processes has greatly improved since the days of the mechanists (the basic belief the body/brain works in much the same way a machine does), Searle states that neuroscience can lead us to the answers needed to clear up the problem. Searle believes that it is the neurological goings-on, plus other related features of the brain, that make it causally sufficient to produce a conscious being. It is because of the "wet wiring" of the brain that it is sufficient for consciousness.

1.2

The second unique feature of mind is intentionality. Intentionality, as defined by Searle, is a mental state realized in the biological structure of the brain, and is the product of the neurological workings of the human brain. But, how are these mental states connected to, or about, the physical world? It is important to note, when discussing intentions, Searle wants to make it clear that he is not talking about "an intention to do something." On his account, a state is intentional if it answers to such questions as "What is 'S' about?" "What is 'S' of?" "What is it a 'S' that?"[1] Intentions, for Searle, hold no special status in themselves, but fall under the category of intentionality just like beliefs, hopes, fears, etc. It is important to clarify that these forms of intentionality are not acts, as acts are what one does. These forms of intentionality are considered to be states or events, and hence will be referred to as intentional states.

The question posed by Searle is, "What is the relation between Intentional state or object and states of affairs that it is in some sense directed at?"[2] Searle states the "visual and auditory experiences, tactile sensations, hunger, thirst, and sexual desire, are all caused by brain processes and they are realised in the structure of the brain, and they are all intentional phenomena."[3] He uses the example of thirst to illustrate his point (however it must be modified for clarification). The neurological processes that occur when one is thirsty cause our desire (an intentional state) to drink. The connection between the intentional state and what it is about is tied in with the neurological processes of the brain. It is important to distinguish between two levels of causation, one on a physiological level, and one on the intentional level. The causal chain (as Searle sees it) begins at the neurological level and this leads to the intentional state which is realized in the structure of the brain. On the physiological level, one's thirst causes certain neurons to fire and this is causal in producing the intentional state "desire" to drink. It is this desire that is instrumental in one going to get a drink.

Searle's strategy here is to narrow the criteria for those that can have intentional states (mental states) to those that have the sufficient biological wiring, namely a human brain. If the brain's neurological workings produce intentional states (as Searle believes), then this narrows the kinds of things capable of thought to those possessing a brain. However, does Searle demystify intentionality? Is it enough to claim "the brain does it"?

1.3

The third feature that needs to be demystified is subjective mental states, that is those that are not outwardly observable either through an individual's actions or utterances. Science has been mainly obsessed with the objective, physically observable, but each of us has our own, mostly unobservable, mental states. For example, we each have certain subjective pains that others cannot feel. One can only surmise that another is in pain from outside objective observable evidence (body language, screams, cries). Searle asks, ". . . how are we to accommodate the reality of subjective mental phenomena with the scientific conception of reality as totally objective."[4]

Searle argues that subjective mental states are facts that are overlooked by the objectivity of science. Just because it cannot be observed, then modern science feels the need to reject the notion. The e is, however, no denying that I feel pain, when I feel pain or I am consciously aware when I know I am consciously aware. We do have clues to these subjective states within our objective language-community framework. Each of us is taught language within a community-based framework that has built-in checks and balances which provide the community with an avenue to view our subjective states. As Wittgenstein states, an exhibition of pain behavior does not constitute pain.[5] However, the community can rely, within reason, on the "reliable" language user; that is, rely on a language user who usually plays within the bounds of the language game. However, this game has boundaries, namely, an individual must be operating within the said conceptual framework and be a reliable (as the community judges) speaker. With this, it would appear as if one's subjective states have a way to become outwardly knowable.

1.4

The final problem dealing with mind/body to be discussed is that of mental causation. How can a mental state cause something in the physical world to occur? For example, how can my desire, wish, etc. to raise my arm, initiate the raising of my arm?

Searle begins by allowing two separate levels of causation: one higher level of causality, of mental processes and one lower level of neuronal processes. For example, my conscious attempt to perform an action such as raising my arm initiates the movement of the arm. "At the higher level description, the intention to raise my arm causes the movement of the arm. But at the lower level of description, a series of neurons firing starts a chain of events that results in the contraction of muscles."[6]

Searle's description of how a mental state can cause something in the physical world to occur is at best vague. He explains how the mental causes the physical in terms of the "brain does it theory." "Intentional states are both caused by and realized in the structure of the brain. And the important thing . . . is to see both the fact that Intentional states stand in causal relations to the neurophysiological (as well as, of course, standing in causal relations to the other Intentional states), and the fact that Intentional states are realized in the neurophysiology of the brain."[7]

Searle attempts to explain away mental causality in much the same way that he did intentionality, consciousness and subjective mental states. He wishes to dispel the myths surrounding the mind and explain them in terms of the brain and body. In endorsing such a view, Searle is able to stress the importance of the brain to mental activities, hence excluding non-biological entities from these abilities. In particular he wants to exclude computers from having these abilities now or in the future.

2.0

In chapter two of Minds, Brains and Science, Searle discusses whether or not it is possible for computers to have (or can have) the ability to think in the same way that a normally functioning human does. It raises the question: could a computer with the right program be said to think in the same fashion a human can?

Searle begins this discussion with an explanation of strong AI. This view states that "the mind is to the brain, as the program is to the computer hardware."[8] In this view, the mind would not be essentially biological, and the brain would merely be a type of computer that contains programs which account for intelligence. So, if a computer had the right program it could be said to think. Strong AI supporters believe that while presently no such computer can think, in the future the trick to having emotions, intelligence and even mental states is merely a matter of implementing the proper programs.

Searle emphatically denies the claims made by supporters of AI. The basis of his position is that a computer does not understand or have the capability to understand the symbols it manipulates. On the other hand, humans have the ability to understand the language they use and come to have knowledge of the physical world that surrounds them. This is to say, while a computer cannot and does not attach any kind of meaning to the symbols it manipulates, humans have the ability to do so. According to Searle, a computer runs on a syntactical level and has no understanding of the symbols being manipulated. It is in this fashion (among others) that our mental processes are not identical with the functional capabilities of a computer. While our mental states have content, or are about things, a computer is limited to processing symbols according to a preset program.

2.1

Searle presents the Chinese room experiment to illustrate his point that computers are mere symbol manipulators and lack any semantic understanding.[9]

To outline the experiment briefly: Searle places an English speaker in a room (the Chinese room) full of tiny filing cabinets containing symbols (which mean nothing to the English speaker). Also in the room is a huge rule book which contains instructions for matching one symbol to another. Outside the enclosed room are native Chinese speakers who write down questions, etc., on a piece of paper and slide it into the room. The individual in the room takes the meaningless (to him) symbols and looks them up in the rule book. When she/he finds the match, she/he takes the symbol out of one of the cabinets, and when this process is completed for the entire inputted information, the individual slides the response back outside. It appears to the native Chinese speakers outside that the room (as a functional whole) knows Chinese. That is, the responses to the inputted information are good enough to convince the native Chinese speakers into thinking the room knows/understands native Chinese. All the while, the English speaker understands nothing of Chinese, she/he merely manipulates symbols.

Searle's point is that the Chinese room as a whole, or even the internal English speaker, cannot learn or come to know Chinese simply by manipulating Chinese symbols. This is because part of understanding Chinese involves understanding the meaning of those symbols manipulated. Merely taking inputted symbols and forming a response in accordance to a rule book (that tells one to match a squiggle to a squoggle) does not count as being sufficient to be said to understand the symbols manipulated. This analogy carries over to how a computer operates (on a basic level). Computers run on a syntactical level, their operations are defined syntactically. This means a computer could not have the ability to understand the symbols it manipulates, and mere symbol manipulation does not and will not add up to any kind of understanding. "It doesn't matter how good the technology is, or how rapid the calculations made by the computer are. If it really is a computer, its operations have to be defined syntactically, whereas consciousness, thoughts, feelings, emotions, and all the rest of it involve more than syntax."[10] While it appears "as if" it understands the information inputted, a computer merely manipulates the information inputted and outputs out the appropriate response in accordance with a preset program. It understands neither the information inputted nor its response. It cannot even be said to understand or come to know any of what it processes.

2.2

Searle presents an argument based on the Chinese room experiment that builds a stonewall that supporters of AI will have grave difficulty getting over. It consists of four premises and four conclusions.

(1) "Brains Cause Minds."[11] The "wet wiring" of the brain is causal to the mental processes which constitute mind. In other words, the neurological workings of the brain is sufficient for causing mind.

(2) "Syntax is not sufficient for semantics."[12] This, as explained earlier, means that the formal functions of computers are insufficient for them to have any understanding of the symbols they manipulate.

(3) "Computer programs are entirely defined by their formal, or syntactical, structure."[13] This is a key point Searle is trying to stress. Computers are information processors running on a syntactically based program, and in order to attribute "mind" to the computer it must understand the meaning of the symbols processed.

(4) "Minds have mental contents; specifically, they have semantic contents."[14]

These four premises lead to four conclusions:

(i) "No computer program by itself is sufficient to give a system a mind. Programs, in short, are not minds, and they are not by themselves sufficient for having minds."[15] This conclusion, if it holds, poses a fundamental criticism of the idea that any computer can possess a mind, or the idea that any program being implemented may ever duplicate the activity of thought.

(ii) "The way that brain functions cause minds cannot be solely in virtue of running a computer program."[16] The conjunction of premise (1) and conclusion (i) brings us to the conclusion that the "wet wiring" of the brain is an essential feature for mind. A mere duplication of the computational properties of brain is not enough to cause minds. The "wet wiring" according to Searle is a necessary condition to cause minds.

(iii) "Anything else that caused minds would have to have causal powers at least equivalent to those of the brain."[17] This particular conclusion allows for other conscious beings (e.g., alien beings). However, if we are to attribute a "mind" to these beings, they must have organs with at least the causal powers (to produce mental states) equal to or similar to that of our brains.

(iv) "For any artifact that we might build which had mental states equivalent to human mental states, the implementation of a computer program would not by itself be sufficient. Rather, the artifact would have to have powers equivalent to the powers of the human brain."[18] This conclusion is in conjunction with number (iii) to give a narrow conception of what it takes to have a mind or even be said to be thinking. Searle narrowed the scope down to humans in order to erase any hopes that AI had of succeeding in their mission. Searle regards mental states as necessarily caused by the biological workings of the brain. Without the "wet wiring" of the brain (Searle states) there can there be no mind. The necessary and sufficient conditions as laid out by Searle exclude computers from any hope now or in the future of possessing mind. In closing this section it is important to note that Searle may have narrowed the criteria down too much, and this, as we will later see, might get him into trouble. While it is true that at the present time no computers meet the criteria to be said to be thinking, I would not necessarily rule this out in the future.

3.0

Searle now poses some criticisms of the underlying assumptions made by the cognitivist. Cognitivism studies the brain on the information processing level, and therefore it ignores its mental capacity and the biological components, yet Searle believes these are the components that are necessary to make minds. However, the cognitivist states that in between the mental and biological components of the brain there is a gap; and this gap is filled with an information processor. By studying this gap between the mental and biological, the cognitivist hopes to replicate human thought in a computer.

3.1

Searle argues there are four reasons why one might adhere to the cognitivist point of view. The first has to do with language. When one speaks a language one conforms to certain formal rules of grammar or syntax, which is similar to the way a computer operates. Secondly, it seems as if humans follow rules when thinking which is similar to the way in which computers follow rules in information processing. Thirdly, there seems to be an internalized theory in the brain that grants humans the ability to learn language. A sort of innate "hard wiring" that allows humans to learn a language. The final reason to adhere to cognitivism is that there seems no other satisfactory way in which to understand the relationship between the mind and the brain other than that proposed by the cognitivist.

Searle now proceeds to discuss in turn. each assumption made by the cognitivist.

3.2

The first assumption deals with the notion that computers and humans conform to rules (grammar) in a similar fashion. However, Searle argues that there is a difference in the fashion in which computers and humans use rules. The human operates on a rule base that is guided primarily by the meaning of the rule in the backdrop of a syntactical structure. The meanings of characters lead to certain actions, thoughts, and mental states. The rules themselves do not have a direct causal effect but have a causal role in the production of behavior. On the other hand, computers do not act on the "meaning" or semantic content of the rule but merely act in accordance with certain "formal" procedures, whereas humans have an understanding (most of the time) of the rules which they follow.

The cognitivist argues that humans follow rules of syntax when they use language, and computers operate in a similar manner, therefore they are similar because computers, like humans run on a syntactical level. However, this is only true to a certain extent: a computer does not follow rules at all, it only acts in accordance with certain formal procedures. While humans conform to the rules of syntax for the use of language, computers run on a syntactical level.

Searle distinguishes between two distinct kinds of information processing. The first is "psychological information processing,"[19] and the second is the "as if information processing."[20] The first type exemplifies certain cognitive operations, whereby thinking is actually occurring. The second is similar to the first in every respect (behavioral) and it was "as if" there were cognitive operations actually going on. The point here is simple: it may appear "as if" computers follow rules the way humans do, and process information the way humans do; but they don't. It is "as if" they are following rules; but according to Searle, they merely act in accordance with certain formal procedures.

3.3

The third assumption of cognitivism is that an internal theory is necessary in order for humans to have any meaningful linguistic behavior, and this internal theory could be replicated and placed in a computer. Searle rejects Chomsky's idea of a universal grammar, which is based on the notion that we all have a complex set of rules in our brain allowing us to acquire grammar (or understand innately the grammatical structure of language). Searle claims instead that there is no need to have this innate ability in order to understand grammar; our abilities develop within the community-based framework. This framework presents us with the universal template of grammar backed by a way of attaching meaning to words. Rules certainly play a vital role in linguistic and other behavior; however, not all of our language behavior is rule-governed. As Searle states, we don't really have this innate internalized set of grammar rules but we "just do it" ". . . there may not be any theoretical mental level underlying those abilities [in reference to using language]; the brain just does them."[21]

3.4

Searle quickly concludes that there is no need for the intermediate level of a computational program that operates between the mind and the brain. There is no gap that needs to be filled as the mind and brain work synergistically together without the need for an intermediate program. At this point, the ability of a computer to explain or even replicate the workings of the brain which causes mind is out of the question.

4.0

Rejections of Searle's Thesis

I have compiled some of the better objections against Searle and his Chinese room experiment. At best they point to specific difficulties in Searle's thesis and at worst they are vague and misleading. Each objection will be followed by a rebuttal prepared by myself in defense of Searle. Using the Socratic method I build a strong case for the opposite view as best I can before questioning its validity. It may seem to be a fruitless journey to present views that are unsatisfactory; however, they need to be presented in order to paint a clear picture of the problems that need to be overcome for supporters of AI and cognitivism.

4.1

To begin, I will briefly review how it all started. The catalyst behind Searle's Chinese room experiment is what is called the "Turing test." By the tenets of the "Turing test," as devised by Alan Turing, one should take a computer (or similar artifact) to be intelligent if it could imitate certain aspects of human linguistic abilities. If a computer could reply to an interrogator in a way that appears to be indistinguishable from a human language user, then it could be deemed intelligent. The point of the test for Alan Turing was to "suggest a conceptual means of identifying intelligence.[22] By the behavioral criteria laid down by the test, under certain conditions one would have to attribute intelligence to a computer.

However, Searle rejects the Turing test on the basis that it is an insufficient measure of intelligence. Searle presents the Chinese room experiment and, like the computer, it passes the Turing test but does not understand language. Therefore, Searle concludes Turing's behavioral criteria are clearly insufficient if they allow something to be called intelligent when it is clearly not intelligent. According to Searle, a computer (like the Chinese room) runs on a formal program, and cannot be said to understand the symbols it manipulates. That is, the computer runs on a syntactical level and attaches no meaning to the symbols it possesses. Also, Searle argues, the computer does not have the right causal powers to produce mental states and hence no intelligence. A computer, while passing the Turing test for machine intelligence, fails the "Searle test" for intelligence.[23]

4.2

David Cole, in "Thought and Thought Experiments," attempts to undermine some of the claims made by Searle. As explained earlier, Searle's Chinese Room sets out to prove that the mere syntactical workings of a computer are insufficient to produce mind, and further Searle states that it is the biological nature of the brain that sets it apart from the machine. However, Cole raises two questions pertaining to the experiment. (1) "[D]oes Searle [or the homunculus] in fact fail to understand Chinese in experiment 3, as he [Searle]) claims?" (2) "[I]s the situation that Searle gives us (experiment 3) analogous to what goes on in the machine [computer]?"[24]

Cole questions the analogy that Searle draws between the Chinese room and a computer. Searle's point is that the Chinese room (as a whole functioning unit) is unable to learn or come to understand Chinese by only manipulating Chinese symbols. The Chinese room manipulates symbols in accordance with a rule book (that tells one to match a squiggle to a squoggle) and forms a response accordingly. However, Searle argues the display of behavior exhibited by the Chinese room is not enough to say it understands Chinese. This is because part of understanding Chinese involves understanding the meaning of the symbols manipulated and this cannot be achieved (according to Searle) by blind symbol manipulation. Searle believes the analogy carries over to how a computer operates (on a basic level). Computers, like the Chinese room, run on a syntactical level and their operations are defined syntactically. The point is that a computer could not have the ability to understand the symbols it manipulates, and hence Searle concludes that the blind symbol manipulation does not add up to any kind of understanding.

Taking this into account, Cole imagines a person (who can speak English) who has the ability to read and respond in Chinese (like the Chinese room on a behavioral level) but does not have the ability to translate. While odd but possible, Cole states that the person has the unique ability to understand both English and Chinese, yet has the odd disability of not being able to translate from one language to another. "I maintain . . . that he would understand Chinese, although with some odd disabilities — he also speaks English, yet can't translate."[25] Cole's main purpose here is to undermine the Chinese room's credibility. Cole states that the Chinese room exhibits native Chinese language user behavior and the homunculus inside has the ability to speak English. Contrary to Searle, Cole states that like his imaginary person, the Chinese room has the ability to understand English and Chinese although he/she does so in an odd way with some disabilities. Cole concludes that it has not been duly established by Searle that the Chinese room does not understand Chinese.

Cole now turns his attention to the second question. Is the Chinese room experiment analogous to what really goes on inside a real computer? Searle argues that the Chinese room operates on a syntactical level with no understanding of the Chinese characters it manipulates. It is in this way, Searle states, that computers are analogous to the workings of the Chinese room, as the computer also runs on a syntactical level with no understanding of the symbols it manipulates.

However, Cole argues that in the analogy that Searle draws between the computer and the Chinese room, the operations of a computer are more akin to the way in which a human language user behaves, and less like Searle's Chinese room. Let me explain. Just as language is a rule-governed activity, Cole believes the program in a computer governs the operations performed. Like a human language user the computer would exhibit the appropriate behavior while being "governed" by the program. So, just as rules govern our language activities, the program governs the computers operations that are performed. Cole explains "[t]hat program 'instructions' are not instructions understood or interpreted by the machine. Rather, they are operation determinators. [He further goes on to say that] a computer does not obey the 'instructions' or 'commands' in a computer program; rather, the lines of characters in the program just cause the computer to perform an operation. . . ."[26]The point is, as Cole argues, the operations of a computer are akin to a human language user on two levels. The first is the proper exhibition of language behavior; the second is the similarity between the way in which computers and human language users act in accordance with rules.

Cole further argues that if this actually were the way a computer works (as stated above), then there may be some sort of understanding: the type of understanding earlier specified, whereby the computer may understand in an odd way with some disabilities. Finally, Cole concludes that Searle's Chinese room experiment does not accurately depict the workings of the computer, as it is more akin to the natural language user (in the ways earlier specified).

However, I am unconvinced by Cole's argument against Searle's Chinese room and the conclusions drawn from that experiment. He responds to Searle on two levels: (1) Computers have the ability to understand in an "odd way" with certain disabilities, (2) the Chinese room is a disanalogy because it does not properly reflect the workings of a computer.

First of all, Cole fails to explain this "odd way" of understanding and further list the disabilities. It is difficult to comprehend how someone apparently fluent in two languages does not have the ability to translate. Merely inherent in the manner in which a second language is taught (in its crudest form) is relating objects one knows in their native language to words in the language being learned. Even on the crudest level of merely pointing to objects (ostensive definition) and saying what it was in the two languages one can translate.

Cole overlooks the fact that a bilingual (in the very meaning of the term) understands what they are saying in Chinese/English; that is their utterances have meaning to them. Whereas the Chinese room operator has no knowledge of what the symbols being manipulated refer to or what they are about; and this is a key element to understanding. Merely manipulating symbols is not enough.

Even if we are inclined to allow Cole the possibility of a person that is apparently fluent in two languages yet cannot translate, does it do him any good? Cole's purpose here is to undermine Searle's criteria for understanding. For Searle an exhibition of behavior (language behavior) is not enough to say that something understands (language). We can use the example of the idiot savant to show the point. Some idiot savants have the amazing ability (among other things) to calculate enormous figures in their heads in seconds, yet are considered mentally deficient in other ways. However, ask them what the numbers mean and they stare blankly into space or give a noncontextual answer (such as "grass is green"). Can we say the idiot savant "understands" what he/she calculates? I would have to say no. The exhibition of behavior in itself is not enough to say the idiot savant understands what he/she calculates. What is missing is the idiot savant's ability to attach meaning to his/her utterances. This point carries over to Searle's Chinese room experiment. The exhibition of behavior by the Chinese room (or computer) is not enough to say it understands, what is missing is an understanding of the symbols manipulated. This is the analogy that Searle wants to draw between the computer and Chinese room. Searle states that both the computer and Chinese room run on a syntactical level and by his criteria this is insufficient to produce understanding. One must have a way of attaching meaning to the symbols, characters or linguistic entities in order to attain an understanding of what is manipulated.

4.3

Richard Double takes a different approach to Searle in "Searle, programs and functionalism." It is slanted towards defeating Searle from a cognitivist point of view. He begins with an example of two Yale researchers who attempt to understand how humans acquire knowledge by studying computer processes. A brief story is given to a computer to which it can produce answers to questions about the intentional states of the characters in the story. The researchers argue the computers understand "and by understand he means: creates a linked causal chain of conceptualizations that represent what took place in each story."[27] The hopes of the researchers are to get the machine to perform a cognitive task (such as understanding language like human language users do) and through this we come to understand human cognition.

Double wishes to tackle Searle on one of his strongest points by questioning the validity of his theory of mental states. If the brain causes intentional states in the way Searle suggests, then only the "wet wiring" of the brain can be sufficient for causing mental states. However, Double questions how something that itself has no intentionality can produce intentional states. To this question he states there are no real answers. Following this line of thinking; Double asks what it is to understand. And even further, Double believes that by studying the processes of the computer we can come to an understanding of human cognition and what it is to say X understands Y.

Double is trying to establish a form of cognitivism whereby one can use the knowledge one has of how a computer works to illustrate how the human mind operates. The hope is, once one has the right program with at least the causal equivalency (for my purposes, language use) of the brain, then one can show how the mind works and perhaps replicate thought. He concentrates the attack on the fact Searle has no concrete evidence that only the neurological workings of the brain can cause the mental. Double states that more insight on how the mind works can be wrought from cognitive theory than the "brain does it theory" advocated by Searle. Searle can no more explain how the brain causes mental states than cognitive science can. Following from this, Double concludes that it is possible to implement a program with the causal powers sufficient to produce mental states.

While I am not satisfied with the core of Double's argument, it does raise a good point against Searle. The criteria that Searle sets down limits "mind" to only those beings that have the proper biological wiring. That is, only the brain's "wet wiring" is causally sufficient to produce mental states. I believe, as Double does, that something with at least the causal equivalency of the brain can cause mind. That is to say, besides the relevant behavioral criteria an artifact must have the correct causal mechanism to produce mind. I feel it does not necessarily have to be the biologically "correct" material that Searle requires. Does 't have to matter whether it is neuroprotein or silicon chips?

However, while Double make a good point about the criteria Searle sets down for mind, he fumbles with the notion of understanding. The two Yale researchers think they have stumbled onto a computer that appears to understand the story. It is not only simulating a human ability but also (1) ". . . the machine can literally be said to understand . . ." and (2) ". . . what the machine and its program do <u>explains</u> the human ability to understand. . . ."[28]

They (Yale researchers) define <u>understand</u> as "able to create a linked causal chain of conceptualizations that represent what took place in each story."[29] The machine answers questions about the story in a similar fashion that one would expect a human language user to respond. The researchers rely on the behavior of the machine to come to the conclusion it understands the story. The computer is questioned, and it responds correctly to the question based on information implicitly extracted from the story. Does this mean the computer understands the story?

The argument here is over the criteria for something to be said to "understand." If the criteria are set too low, then we allow things to understand that clearly do not understand. However, if the criteria are set too high, we eliminate things that might actually understand. To keep the discussion relevant to this thesis, I will discuss understanding as it relates to linguistic understanding. The problem is to define the criteria. Within the context of the story it may be possible that the computer understands the story (as understand is defined by the Yale researchers). However, can we say it understands the story in the same fashion an experienced language user does? I would have to say no.

Understanding a language involves more than an exhibition of the proper language behavior as shown by the refutation of the Turing test for machine intelligence. Behavior is necessary but insufficient to say X understands Y. What else is needed? In order to understand language, one must (among other things) understand the concepts involved, their relation to one another and the context of their use. We develop this ability within the community-based language framework. This is, among other things, what gives meaning to our utterances and an understanding of the language we speak. The computer exhibits these behaviors; however, why are we reluctant to say it does not understand?

This discussion brings to light some interesting issues. For example, how do we gain our own linguistic understanding, or what is it to understand? Also, is Searle's criterion for the production of mind too strict? What kind of program would need to be implemented in order to say the computer understands? Double attempts to give answers to these questions but never sheds any light on how to solve them. He states that we do not really understand how we understand and hence cannot know what it is to understand. The only way to discover how we come to understand (according to Double) is to study the workings of a computer (with this properly implemented program) and this will lead to insight on how the human mind works. However, he neither sheds light on how to implement such a program or even what kind of program it will be.

4.4

On the heels of that, Georges Rey redirects Searle's statement, "syntax is not sufficient for semantics,"[30] and asks, "how is the semantics of the internal code [causally] determined?" "What is the right sort of causal link to the outside world?"[31] This redirects the focus from a problem ("syntax is insufficient for semantics") to a solution to the problem of how to implement a system that could understand.

Rey believes the solution lies in the theory of natural meaning. Rey begins by giving an account of natural meaning similar to Grice and Fodor's, whereby one natural phenomenon can mean another. (Hence, Rey believes that one can get understanding in this way.) Using natural meaning, one can program a computer to get meaning by associating one natural kind with another, hence the computer would be running not only on a syntactical level but also a semantical one. That is, "the tree's having 50 rings means that the tree is 50 years old (if N [normal], then the tree displays 50 rings if the tree is 50 years old)."[32] By adhering to a theory of natural meaning, Rey argues he can implement a program that will enable the system to understand.

Rey believes, just as a thermometer is a good temperature detector, the Chinese room (or a computer) is an excellent proposition detector. Given a wide variety of propositions, the computer has the ability (via natural meaning) to come to understand what it processes. "Put some egg foo yung in front of its receptors and it will [respond] . . . 'Lo and behold, there's some egg foo yung in front of me now,' and it will behave similarly when confronted with Cashew Chicken, Peking Duck, a picture of Chairman Mao, and will put in anything you like."[33]

Rey attempts to provide a solution to the problem of what kind of program an individual has to implement in a system in order to make it understand natural language. However, convincing as it may sound, I do not think a computer can come to understand language using natural meaning. "For example, 'red' might be defined, prototypically, in this sense (natural meaning) with reference to blood or fire (as many dictionaries do indeed define it)."[34]

However, it would be difficult if not impossible to categorize the world around us into what is referred to as "natural kinds" for two main reasons. According to John Lyons, "(1) most lexemes in all languages do not denote natural kinds, (2) the denotation of those that do requires cultural support."[35] Each of these "natural kinds" may be denoted by certain lexemes in language, however the meaning of it still needs the support of the community-based conceptual framework. So to understand the meaning of a particular word involves (in a sense) knowing all the concepts involved. One cannot learn language in piecemeal in a building-block style, but must acquire an entire conceptual framework. Merely piecing together natural kinds is not enough to come to an understanding of natural language.

4.5

David Anderson, in "Is the Chinese Room the Real Thing?" discusses the criteria for distinguishing between understanding and simulated understanding. Anderson states that some forms of simulated understanding are so close to the real thing, it has to be called understanding. "[S]ome simulations go beyond merely copying the phenomenon simulated.

Some simulations are like 'clones' of the real thing. I have suggested a number of examples where it seems to be appropriate to draw no distinction (except of origin) between real things and exceptionally good simulations of them. If a simulation of understanding is indistinguishable from the real thing then I maintain that it is correct to think of it as understanding."[36]

Anderson presents some interesting examples to support this thesis. Imagine waking up in a hospital bed to find that your hand has been severed off in an auto accident. However, a new one has been surgically installed. The new hand is an exact copy in every aspect (moves the same, bleeds, etc.) except it did not grow on organically. While it is not numerically the same hand, the only difference between the old and the new is origin.

The second example depicts two piles of English pounds. They are identical in every aspect except one of the piles is fake. The counterfeit pounds can be spent, exchanged and passed off as well as the real thing. The only difference between the two is the fact that the real English pounds were issued by the appropriate authority. It is this institution that governs the use of, and worth of, real English pounds. Once again Anderson argues that the only difference here lies in origin.

These examples, among the others given, illustrates the point that Anderson is trying to get across. If a computer simulates human understanding that is indistinguishable from the real thing, can't this be called understanding? The criteria Anderson relies on to come to his conclusions about understanding are based on a computer's ability to behave verbally. That is, it appears to the computer user that it understands the inputted information because it responds appropriately in kind. As we have already seen when we criticized the Turing test his criterion was insufficient to prove that the computer understands anything. Anderson still stresses that a simulation (perhaps a far better simulation than is presently available) of computer understanding can add up to an understanding, and further the only thing that sets these two types of understanding apart is origin, which makes one only simulated.

Two questions must be asked at this juncture. (1) When do we have artificial understanding? and (2) How important is origin?

The answer to the first question depends on (a) how we define artificial understanding, and (b) the criterion for stating "X" understands. Searle believes that there is no need to define artificial understanding because either something understands or it does not. Searle illustrates his point when he distinguishes between two different kinds of information processing.[37] The first is "psychological information processing" and the second is "as if information processing." The first depicts certain cognitive operations whereby thinking that requires a mind is actually occurring. The second is similar to the first in almost every aspect and it is "as if" there was cognitive operations going on. The difference is clear to Searle, "People actually think, and this thinking goes on in their brains. Furthermore, there are all sorts of things going on in the brain at the neurophysiological level that actually cause our thought processes."[38] Searle's point is, while a computer can simulate understanding it is only "as if" it understands the symbols it manipulates. Searle concludes that a simulation of understanding "does not add up to understanding."

However, Anderson defines artificial understanding "as simulating human linguistic understanding so it is indistinguishable from the real thing (human understanding)." Taken within the context of Anderson's argument this appears to be a valid way to define artificial understanding. However, I believe Anderson is mistaken in believing that a simulation of a computer understanding can "add up" to an instance of understanding. Anderson relies solely on the linguistic behaviors of the computer for his criterion for understanding, which, on its own, is insufficient to endow "X" with understanding. Searle's criterion relies on the outward behavior plus the biological component which suggests that origin is important.

What about origin? Is the biological mass we call "brain" the only thing causally sufficient to produce mind, hence understanding? Or can there be something else causally sufficient to produce mind, hence understanding? If we are tempted to say origin is important, we are obliged to adopt the strict criterion presented by Searle ("brains cause minds").[39] However, if we are tempted to say origin (biological) is non-essential (as Anderson does in his example), then this may allow any number of things to possess mind.

To be careful not to fall on either side of this double-edged sword, I will contend that in order for something to possess mind it must be caused (mind) by something with at least the causal powers equivalent to a brain. This leaves the criterion strict enough to exclude non-thinking entities while opening the door for the future. The next chapter will attempt to peek into the future and hopefully give insight on how a computer might come to understand natural language the way a human language user does.

5.0

The question now boils down to: How are we to overcome the difficulties for AI that Searle has presented us with? Searle believes at bottom that AI in principle cannot work. He believes that a computer cannot possess mind now or in the future, for two reasons. First, a computer does not and cannot have the correct biological wiring that will enable it to possess mind. Also, according to Searle, a computer is defined syntactically and is unable to understand (or come to understand) the symbols it manipulates. Does AI research come to a grinding halt with no hope of ever reaching the elusive goal of incorporating a system into a computer that enables it to understand the linguistic symbols it manipulates. Should AI throw in the proverbial towel and concede to Searle? I think not!

Many philosophers and AI researchers such as David Cole, Patricia Hanna, Georges Rey suggest that there could be an implementation of a semantic element into a computer in order to enable a computer to understand linguistically. However, none of these people suggest either how this system is to be implemented or what the system should be. It is a seemingly good start but no one is willing to give a hint about how one should go about this complex procedure. The idea that a semantic element needs to be implemented suggests the problem is a linguistic one. If one clears up the fogginess surrounding language acquisition, perhaps this will lead to some answers of what kind of system actually needs to be implemented.

I want to clear the air surrounding this area by looking at how we (human language users) come to learn a language within a community of language users in hopes of stumbling across a way of implementing a similar system into a computer. I don't claim to have all the right answers, but by asking the right questions I may be able to make a positive contribution to AI research. I can't really tell how one is to implement such a system. However by coming to grips with how an individual acquires a language (initially) it might shed light on the type of system necessary for a computer to be said to have linguistic understanding.

Searle, as stated before, rejects the notion of providing a computer with a semantic element, as the computer is defined syntactically. It manipulates symbols without any understanding of what it manipulates. For this reason, Searle states that no amount of information processing (on any level) within a system of this kind will ever add up to understanding. Leaving this behind, Searle also rejects the possibility of computer understanding for a different reason. He also states that only the wet wiring of the brain (human) is causally sufficient to produce mind. Only the biological mass we call brain is causally sufficient to produce mind, and hence the hard wiring of a computer is causally insufficient to produce mind or any kind of understanding now or in the future.

However, I believe the latter to be incorrect, and the narrow view Searle espouses can be modified to suit our needs. The strict "only brains can produce minds" can be modified to a more intermediate claim that "only the brain or something with at least the causal equivalency of that of a brain can produce mind". I eliminate the stringent need for the biological wet wiring, so that the possibility for future systems having mind may be still in question. Is the idea of silicon chips causing mind any more absurd than neuroprotein having the ability to cause mind?

Now lets turn to Searle's initial complaint, that is, the computer is missing the semantic element necessary to be said to understand. Searle states: "syntax is not sufficient for semantics,[40] but what if a semantic element is implemented into the system? For my purposes it will be enough to say that this semantic element is something that allows the computer to understand the symbols it manipulates. Although it may not cause mind in the strictest sense, it may produce understanding. In order to explore this avenue, we must first come to see how we initially learn a language. This is not an inquiry into the origins of language, as this would take many more pages than I wish to devote to this topic, but an inquiry into how we came to learn language within a community already capable of communicating linguistically.

5.1

Ludwig Wittgenstein, when he wrote Philosophical Investigations, believed that language is uniquely tied to the community in such a way that the community is causally necessary for one to acquire a language (their first). He begins his investigations by refuting the notion that the meaning of a word is its reference. That is, one wouldn't necessarily come to understand the expression "blue" simply by ostensive definition (although he does stress the importance of ostensive definition in acquiring concepts, just not acquiring conceptual frameworks). A much wider context is needed to ensure that someone was actually describing a color expanse rather than, for example its shape. An individual, for example, builds up a battery of concepts; red, blue, green, etc. within the language activity that uses color words. But, how do we come to have the ability to identify and use them correctly?

An individual gains an understanding (linguistic) of their environment and language within community-based conceptual frameworks. The individual is brought into the framework and slowly inducted into the framework (a metaphor: the individual actually acquires the framework) and slowly brought into rule-governed language games. The guidance of the community (people who are masters of the language acivities) allows the individual to gain knowledge of concepts and the context in which they are used. The success of this enterprise depends on the individual's success in operating within (what

Wittgenstein called) "language games." That is, the community's approval of successful language acquisition, depends on the individual's ability tc communicate with others within rule-governed language games. Language has meaning within these frameworks, and stepping outside them can result in gibberish or non-communication with others.

Now, if language use is a game of sorts, then there must be some guidelines; some rules as it were. Wittgenstein rejects the notion of an individual acting because of knowledge of the rule. That is, the peιson follows Rule (X) because of knowledge of Rule (X). Wittgenstein seems to say an individual is able to follow Rule (X), not because of explicit knowledge of the Rule (X), but because the rule is in place, and the person is somehow trained to use it. In a sense he advocates a notion of rule conforming behavior. However, Wittgenstein is still unclear on how rules are to be incorporated into his language game. All that is said is that "the rule may aid in teaching the game, the learner is told it and given practice in applying it — or it is an instrument of the game itself — or a rule is employed neither in the teaching nor the game itself; nor is it set down in a list of rules. One learns the game by watching how others play. But we say that it is played according to such-and-such rules because an observer can read these rules off from the practice of the game — like a law of nature governing the play."[41]

Wittgenstein argues that it is not necessary to claim that the rules of language games are in the person's mind thus avoiding the pitfall that the meaning of words are found in the head. However, while avoiding this problem, I feel he creates another problem. How can we come to read these rules off the "practice of the game"? The sentence, "can read these rules off the practice of the game," seems to presuppose that one somehow already has the ability to acquire language rules. For example, watching a chess game is not necessarily sufficient to say I understand the "game" of chess. It assumes we already have the ability whether acquired or innate to learn these rules. If it (the ability) is acquired, then how do we acquire it? If it is innate then obviously our biological makeup is important to language use. What is needed is an explanation of how we "read the rules off the play of the game."

It is agreed that language is a rule-governed activity; however, it is how we use or follow these rules that is important to this thesis. Chomsky assumed that we possessed an innate generative grammar that guided our language acquisition and use. However, this incurs regress when one asks how we learn to use the rules associated with this innate grammar underlying our language use. Chomsky's innate underlying grammar suggests that language acquisition is species specific. That is, only those with the correct underlying structure can have the ability to acquire and use language. John Searle also suggests that our ability to acquire and use language is essentially biological. Without the biological component, states Searle, language use is impossible. His conclusion is based on the premise that only the biological wet wiring of the brain is causally sufficient to produce mind, hence the ability to acquire and use language is essentially biological. Searle believes: ". . . mental phenomena are biologically based: they are both caused by the operations of the brain and realized in the structure of the brain. On this view, consciousness and Intentionality are as much a part of human biology as digestion, or the circulation of blood."[42]

I agree that the correct "wiring" or structure must be in place in order to support language acquisition; however, I do not believe it has to be "biologically correct" wiring. Chomsky narrows language acquisition to those with the correct underlying structure, and Searle requires the "causally sufficient wet wiring." What I want to show is that language acquisition does not necessarily require the correct biological structure but just the correct structure. In order to get a lead on what this "structure" would have to look like in order to support language use, I am trying to get a handle on how a human language user acquires and uses his/her native language. One of the more mysterious areas of language acquisition and use is how we use or follow rules. In the next section I will attempt to clarify the types of rule following that guide language use.

5.2

The need for clarification of the types of rule following is essential to the project at hand, for if we rely merely on an innate structure for language acquisition then the consequence would be that AI is not a possible model of mind.

That would mean origin (for language acquisition) is definitely essential for acquiring a language. I'll attempt to sort this out by presenting the two present views on rule following.

(A) Mere conforming to the Rule: unconcious conforming to the rule (e.g., I do X in conformity to the rule but not because of the rule; much like the illiterate language user. Although she/he does not know the rules of grammar and can neither read nor write, she/he can speak correctly and fluently).

(B) Conscious Rule Following: Obeying Rules (e.g., "I do X <u>Because of</u> my <u>knowledge</u> of the Rule", this implies a knowledge of the rule and compliance is because of the rule).

If these are the only two choices available, then it seems that we are at an impass. Both of these accounts of rule following have difficulties associated with them. I believe that "A" is too weak to support language use (and acquisition). On its own it is too weak, and requires a supplementary account of how we come to conform to these rules. It also leaves the account mysterious as to how we come to get new uses of language. That is, if we are conforming to the rules, how do we come to label new instances of language use that do not fit the old (uses); for example, generating new sentence types, or naming new objects. On these grounds, I believe a stronger account is needed.

However, there are also problems with "B". According to "B," Learning to use language (L) is learning to obey the rules of (L). This account suggests that following the rule requires one to know the rule prior to learning language. Part of the problem with this is rules do not carry their own application, so even if one knew the rule a supplementary account is needed in order to explain their application. This leads to a vicious regress "[s]o that learning to use a language (L) presupposes having learned to use a meta-language (ML). And by the same token, having learned to use (ML) presupposes having learned to use a meta-meta-language (MML) and so on."[43]

I must point out that sorting out the types of rule following had a reason; that is to see how we acquire language. "A" and "B" are sufficient guides to rule following if an individual has already acquired language but the question is: How do we come to either ("A") comply or conform, or ("B") have knowledge of these rules? "A" is too moderate to support the acquisition of language use and "B" presupposes too much. There must be another alternative.

Wilfred Sellars in "Some Reflections on Language Games"[44] suggests a third alternative to following rules as a way of explaining language acquisition. Sellars suggests an intermediate position between the moderate "Rule conforming" and stringent "Rule obeying" which is "pattern governed".

> (C) Pattern governed: "I do X because of the rule" to which one conforms, there is no "knowledge" (the supposition here is that no prior knowledge of the rule is needed) per se of the rule.

So, because there is no (supposedly prior) knowledge of the rules it avoids the regress "B" fell into and it has the ability to avoid difficulties associated with "A". Let me explain. The leading idea is that the rule must somehow be causal in the behavior of the language user, that is our brain must somehow encode the rule in some neurophysiological pattern. However, pattern-governed behavior need not fall into the problems of this alternative. It is not enough to say the brain must encode the rule (case B); it also needs to be said that the neurological pattern itself is caused by the rule. It is in this way that we can say the "pattern" is an encoding of the rule.

However, the problem that lies ahead is two-fold: (1) how does the rule itself cause a neurological pattern, (2) If these rules are not innate, where do they come from? The solution lies in two main facts" (1) that the rule exists as a concrete pattern of rewards and sanctions in the community (2) that this concrete pattern has causal efficacy in producing the neurophysiological patterning which is the proximate cause of our rule conforming behavior."[45]

This removes the mysteries of an innate meta-language (or rule base) for governing language use and places the responsibility of this chore in the existing community of language users. The community, through a system of rewards and punishments, "patterns" the individual language user. That is, reward is given when a pattern is properly encoded and exhibited by the language learner and the individual is sanctioned when it is not (encoded and exhibited properly).

This system of rewards and punishments is not to be confused with Skinner's "pigeon tricks". It should be taken much broade: than that. The reward for the proper encoding of the rule is the learning of a language, and the punishment is a mere correcting of your linguistic errors. "It's not Green; It's Blue!"

Through this process the community acts as the rule base which patterns the individuals' language acquisition. For example, "coming to see something as red is the culmination of a conceptual process which is the slow building up of a multi-dimensional pattern of linguistic responses (by verbal expressions to things, by verbal expressions to verbal expressions, by meta-linguistic expressions to object language expressions, etc.), the fruiton of which is, as [conceptualization] occurs, all these dimensions come into play in such direct perceptions as that this physical object (not that one) over here (not over there) is (rather than was) red (not orange, yellow, etc.)."[46] The community guides the individual to learn "red" and acquire a complex neurological pattern, whereby one does not just learn a concept but a whole battery of concepts. This is because to be in possession of any given concept presupposes a whole battery of concepts tied into a complex neurological pattern as dictated by the Rule base (community). It is "[o]nly after a long period of training not only with regard to this [Red] and other color words but also with regard to a whole network of concepts involved in color recognition can the "child" finally be said in a given instance to be truly saying (as opposed to uttering) "This is red."[47]

The causal chain goes something like this:

(A) Rule Base (This is the community: in the practice of rewards and punishment -- the community is causal)

(B) The rule base causes certain (neurological) patterns to arise in the individual (The Rule base is causal in the encoding of certain neurological patterns in the individual)

(C) Individuals who exhibit behavior that is governed by the pattern as caused by the community are rewarded, those who do not are sanctioned

(D) The pattern is encoded (the individual eventually, through this process, becomes a language user).

At first the picture painted may appear a little strange. However it is no more difficult to see that we learn a language in this fashion than that our ability to learn language comes to us innately. The argument is over how we acquire this structure, and instead of leaving the explanation blank or filling it with innate underlying abilities, Sellars endorses the plausible account that the community is responsible for it. It is very plausible to suppose the community is causal in development of our linguistic behaviors. This has never been questioned, but, Sellars goes a step farther to make the community the rule base which patterns an individuals' language acquisition.

The question now arises: How do we come to develop these patterns or acquire a conceptual framework? An individual (states Sellars) certainly doesn't acquire concepts in a piecemeal fashion merely by observing certain regularities in the surrounding environment and labelling them. This would be to buy into the empirical notion of ostensive definition. However the individual does acquire entire conceptual frameworks (patterns) and inferentially links them to other overlapping frameworks, until the individual can be called a competent language user within the community. These patterns are necessarily caused by the community through a system of rewards and punishments. To use an example: One cannot be said to learn to tell time merely by observing a clock and labelling certain pin-pointed times such as "four o'clock", "six-thirty", or "nine forty-five", etc., until finally one totals the separate times and comes to know how to tell time. It isn't until one is trained or a concrete pattern is formed by this training that can one come to tell time. That is, one must have the concepts and context dealing with time before one can be said to "know how to tell time". Only knowing a few scattered positions of the hands is insufficient for telling time, because to know how to tell time is a combination of knowing all the concepts involved, or in other words, acquiring the proper conceptual framework.

> To learn pattern governed behavior is to become conditioned to arrange perceptible elements into patterns and to form these, in turn, into more complex patterns and sequences of patterns. Presumably, such learning is capable of explanation in S-R [Stimulus-Response] — reinforcement terms, the organism coming to respond to patterns as wholes through being (among other things) rewarded

> when it completes gappy instances of the Patterns.[48]

One would have to be blind to not see the similarities to Wittgenstein that Sellars invokes. However, the important difference (among others) lies in the fact that Sellars believes language to be pattern governed behavior and not rule governed. This is not a sly side-stepping routine on Sellar's part; it is an important distinction. The community, according to Sellars, is causal in developing the neurophysiological patterns which guide the use of language in the individual. The community operates with the rule base that causes the encoding of patterns in the individual that enables the individual to acquire language. On the very top of the heap, rules cause an encoding of a pattern, however, the individual's language use is pattern-governed. The key here is how rules are followed.

> Now it is obvious that acquiring the concept of red cannot be equated with coming to obey a semantic rule...the application of the concept red to an object in the process of observing that something is red, cannot be construed as obeying a semantic rule, for a rule is always a rule for doing something in some circumstances, and obeying a rule presupposes the recognition that the circumstances are of a kind to which the rule applies.[49]

So, if an individual was to obey a semantic rule as illustrated above, the rule, for example "call red objects red" would necessarily presuppose our understanding that red is a color word. It is important to remember that Sellars refers to language acquisition and our linguistic abilities when referring to rules. He doesn't want to state that without language we can't perceive differences in objects of perception, i.e., different shades of red. However, it is when we wish to label these objects and color patches within a community-owned conceptual framework that the notion of rule-following becomes sticky.

It appears that pattern governed behavior is a solution that may suggest a way out of this problem. Instead of "rules" governing our behavior (linguistic), the community is causal (the rule base) in encoding certain patterns that guide our linguistic behavior. In this manner, we eliminate the need for an innate underlying meta-language, or a complex explanation of how we manage to conform to rules. The individual merely follows a community-encoded neurological pattern. Some might say we "can't follow neurological states", but why not? It is not a following of a "neurological state" but a pattern; a pattern that is encoded through repetitious community involvement.

6.0

Why is it necessary to clarify whether we follow rules or follow patterns caused by rules (the community)? The reason why a seemingly long-winded clarification was necessary was two-fold: (1) in order to come to understand how we are to implement a system into a computer that will enable it to have linguistic understanding , we must first be clear on how we acquire language; (2) If it is true that we actually follow neurological patterning caused by rules, then we may be able to clear a path for AI.

Computers per se, don't follow rules. They act because of a program that operates on a syntactical level, and for this reason (among others) they are guided by a program (programs are sets of rules) and not rules. The program is guided by a rule base however, a rule base interjected by the programmer. So the programmer creates the program following a set of rules, then puts it into a computer. The computer then follows the pattern as laid out by the programmer. My point here is crucial. If it can be proven that our language use is not directly rule governed, then there may be a way to implement a program that "learns" language in a similar way in which humans seem to: that is, by following patterns. These patterns would be (could be) formed by a complex relation between the programmer and the computer. The programmer would act in much the same way as the community does in language acquisition for humans. Let me illustrate my point.

There is a chess program that "learns" from its mistakes. That is, within a complex matrix the computer plays chess with an individual (usually human) and when it loses a pawn, bishop, what have you, it "learns" that if it is in the same position again (in another game) it should not (I don't want to use this word to sound like it is "thinking about move X" or even following any kind of semantic rule base) make the same move again. This is not to be confused with choices, as the computer learns not to do the same move again if in the same position again. In a sense, the more games the computer plays the better it gets. It learns, on a very rudimentary level, that doing X is incorrect and doing Y is correct. This learning is a result of sanctions from the community (by being "beaten" by the other player) for making a bad move and rewarded (by "winning") for making a good move. It has the basic capacity, within the context of chess, to distinguish a good move or a bad move and hence acts on that capacity by becoming, in the end, a better chess player. The learning is not self-initiated, but comes from a process that requires interaction with an outside community. Just as putting a baby alone on a deserted island is insufficient for that baby to develop language, so is it for the computer, mainly because of the missing community interaction.

Now I want to apply this capacity in a thought experiment; one that is going to take some open mindedness. What if we had a computer that had the ability to perceive its environment? There are presently systems available that could allow a computer to do this.

> As work in "computer vision" has shown, metal and silicon are undoubtedly able to support some of the functions necessary for the 2D-to-3D mapping involved in vision. Moreover, they can embody specific mathematical functions for recognizing intensity-gradients (namely "DOG-detectors", which compute the difference of Gaussians) which seems to be involved in many biological visual systems.[50]

This "vision" would not be exactly like human sight, however, it would have the ability to pick out objects in a visual field. To properly visualize the environment, a computer would have to have mobility, simply sitting on a desk staring blankly ahead will not do. With all the advances in robotics it would be relatively easy to build a robot that could move around, pick up objects, even perform complex tasks. However, all we have is a machine wandering around as dictated by a preset program; it knows nothing and understands nothing.

However, what if a program similar to the chess example is implemented into our robot? A computer operates by a system of patterns as designated by its programmer. It does only those things that the programmer intends for it to do. The computer processes information and behaves in accordance with a pre-programmed itinerary with no understanding of what it processes. If something is inputted or appears within its visual field that is strange to it, it does not have the capacity to adjust, to overcome the difficulties it is having. It lacks the power of recognition. However, the chess program has the ability to "learn" or overcome difficulties it may have with a certain series of moves that persistently give it certain negative results. It adjusts to overcome its difficulties accordingly in order to become a "better player" within the framework of the game "chess". It has the ability through community sanctions (the human player beating it or being defeated by it) to learn to produce different outcomes. Once again, by the community (individual playing with the computer) sanctioning the computer for an incorrect move (by taking the pawn, bishop, or by beating it) the computer encodes a new pattern to be used in a later games. Within the context of a "chess game" the computer "learns" from its mistakes so they will not occur again.

6.1

How does this help in our search for a computer that may be able to have linguistic understanding? Well, just as humans acquire a language (their first at any rate) through the guidance of the community, so may a computer. Let me continue the thought experiment from what was already said. Lets start with the "robot" that has the ability to perceive its environment in a rudimentary fashion. If we implement a program similar to the chess program, only take it out of the context of "chess game" and instead place the context within "language game", whereby the computer "learns" certain conceptual frameworks as patterned by the community (interactive programmers) we may be able to break ground. That is, the rule base would be the community which would be causal (through constant interaction with the computer in its learning stages) in patterning certain frameworks. These patterns would grow and become linked in overlapping patterns until the computer can be said to be interacting and correctly identifying its environment. What is important in this is the notion of the community (or programmer) sanctioning the computer for incorrect behavior, hence allowing it to overcome each mistake in its effort to acquire language in a rudimentary fashion. This is not to be mistaken for mere ostensive training, as the computer would actually learn a "pattern", not by piecemeal fashion, but as complex "patterns" as represented by the rule base (communities causal interaction). This is not a quick process however; it takes the average language learner up to 10 years to acquire their first language in full.

The computer would not just learn "red", but as guided by the community, it develops complex patterns whereby it not only learns "red" but other concepts associated with colors. These, in turn, fit neatly into the overlapping patterns, such as colored objects, and then go on to name objects such as Apple (which is red) which is involved with learning the pattern involved with fruit. Just as with language learning for humans, the computer cannot be said to grasp the individual concept until other concepts associated with it are grasped.

Can a computer with these capabilities be said to know, or even understand what it patterns? Both of these words bring with them a host of related problems that are not yet sorted out. If knowledge in the classical sense is "Justified True

Belief", then it seems our computer would not have knowledge. The computer does not have the ability to produce mental states, hence belief states, therefore it doesn't have knowledge. However, understanding may be another thing. The criteria for whether someone (or something) understands lies in the community. The community is the judge of whether someone understands what is being said to them (asking them questions for instance), or able to correctly interact within a community of language users. Does this mean the computer understands? Merely meeting the criteria as laid out by the community does not necessarily mean it will, or can, understand the linguistic entities it patterns. A test would be to see if it refers and predicates adequately. For instance:

> <u>Community</u>: "Hey look at the blue chair"
>
> <u>Robot</u>: "You mean the red chair, don't you?"
> (referring to the only chair [red] in the room)
>
> <u>Community</u>: "Yes, I'm sorry, the red chair"
>
> <u>Robot</u>: "Well, what about it?"

There is a difference between merely uttering noises and actually saying "the red chair". I believe a computer of this kind can predicate, and does come to an understanding (in a rudimentary fashion). At this juncture, some may state that a computer may not need the intermediate learning process, hence a program worked on by a team of experts is enough to provide a computer with linguistic understanding. However, I believe that there are certain abilities one can't program into a computer with great success: one of these is linguistic understanding. This is because, I believe just inserting a program is not enough; one needs the interaction of the community in order to learn a language. AI seems to miss this important point, as the general belief among the experts is that the implementation of the "proper program" is in itself sufficient to produce a competent language user. However, it is my belief that the community (in the case of a computer, an "interactive programmer") is essential to the language learning process. Just as humans rely on the community for language acquisition, I believe that interaction between the community (interactive programmer) and a computer is necessary for a computer coming to understand the linguistic entities it possesses. AI was correct when it required the "proper program"; however, this requirement is insufficient on its own to produce linguistic understanding in a computer, as I believe the community is a necessary requirement. Part of understanding, as I see it, is the learning process involved with coming to understand. Just as humans are not born with a language and have to be guided by the community in order to acquire a language, I believe that one cannot simply "endow" a computer with linguistic understanding by just implementing a program. A computer, in my opinion, must go through a similar process that the human language learner does in order to acquire linguistic understanding. That is, the community must be actively involved in the language acquisition process.

7.0
Conclusion

I think I can go no further in my quest; there are questions for which I do not have answers, such as: How is this program to be implemented? How might a computer or even a human come to have these "patterns" on a neurological level? The first question I leave to the computer design experts, and the latter I leave to the neurophysiologists. It is enough for me to suggest theoretically the kind of system needed to replicate learning of language, hence language. Such a system, I suggest, is capable of replicating an understanding (linguistically) in a rudimentary fashion. By rudimentary, I mean we would basically end up with an entity with no feelings, no concept of past or present, only an understanding of the now. It would be a logical being much the way Star Trek's "Spock" is.

This quest has placed me somewhere between the extreme behaviorism of Turing and the narrow biological criteria laid out by Searle. To find an answer is to find common ground between these two men.

Endnotes

[1] John Searle, Intentionality: An Essay in the Philosophy of Mind (Cambridge/New York: Cambridge University Press, 1983), p. 2.

[2] Ibid., p. 4.

[3] John Searle, Minds, Brains, and Science (Cambridge/New York: Cambridge University Press, 1983), p. 24.

[4] Ibid., p. 16.

[5] Anthony Kenny, Wittgenstein (Harmondsworth, Middlesex, England: Penguin Books, 1973), p. 184.

[6] John Searle, Minds, Brains, and Science (Cambridge/New York: Cambridge University Press, 1983), p. 26.

[7] John Searle, Intentionality, An Essay in the Philosophy of Mind (Cambridge/New York: Cambridge University Press, 1983), p. 15.

[8] Searle, Minds, Brains, and Science, p. 28.

[9] Ibid., pp. 31-34.

[10] Ibid., p. 37.

[11] Ibid., p. 39.

[12] Ibid., p. 39.

[13] Ibid., p. 40.

[14] Ibid., p. 40.

[15] Ibid., p. 40.

[16] Ibid., p. 40.

[17] Ibid., p. 40.

[18] Ibid., p. 41.

[19] Ibid., p. 49.

[20] Ibid., p. 49.

[21] Ibid., p. 53.

[22] A. M. Turing, "Computing Machinery and Intelligence," Mind, Vol. LIX, No. 2236 (Oct. 1950), pp. 433-60.

[23] It is later open to debate whether the criteria Searle presents are too tough, just as Turing's criteria were too soft.

[24] David Cole, "Thought and Thought Experiment," Philosophical Studies, 45 (May 1984), p. 434.

[25] Ibid., p. 437.

[26] Ibid., p. 440.

[27] Richard Double, "Searle, Programs, and Functionalism," Nature and Systems (1983), p. 107.

[28] Searle, "Minds, Brains, and Programs," p. 418.

[29] Double, p. 107.

[30] Searle, Minds, Brains, and Science, p. 39.

[31] Georges Rey, "What's Really Going on in Searle's 'Chinese Room,'" Philosophical Studies, 50 (1986), p. 177.

[32] Ibid., p. 178.

[33] Ibid., p. 179.

[34] John Lyons, Language and Linguistics (Cambridge: Cambridge University Press, 1981), p. 316.

[35] Ibid., pp. 316-17.

[36] David Anderson, "Is the Chinese Room the Real Thing?" Philosophy 62 (1987), p. 393.

[37] Searle, Minds, Brains, and Science, p. 49.

[38] Ibid., p. 50.

[39] Ibid., p. 39.

[40] Ibid., p. 39.

[41] Anthony Kenny, Wittgenstein (Harmondsworth, Middlesex, England: Penguin Books, 1973), p. 171.

[42] Searle, Intentionality, An Essay in the Philosophy of Mind, p. ix.

[43] Wilfred Sellars, Science,Perception and Reality (Routledge and Kegan Paul, 1963), p. 321.

[44] Wilfred Sellars, Science Perception and Reality (Routledge and Kegan Paul, 1963), pp. 321-358.

[45] Tom Vinci, 1992 (in conversation).

[46] C. F. Delaney, The Synoptic Vision: Essays on the Philosophy of Wilfred Sellars (Notre Dame: University of Notre Dame Press, 1977), p. 7.

[47] Ibid., p. 24.

[48] Ibid., p. 7.

[49] Wilfred Sellars, Science, Perception, and Reality (Routledge and Kegan Paul, 193), p. 333.

[50] Margaret A. Boden, "Escaping from the Chinese Room," Computer Models of Mind, chapter 8 (Cambridge: Cambridge University Press, 1980), p. 93.

Bibliography

Anderson, David. "Is the Chinese Room the Real Thing?" Philosophy, 62 (1987), pp. 389-393.

Boden, Margaret A. (ed.). The Philosophy of Artificial Intelligence. Oxford: Oxford University Press, 1990.

Boden, Margaret A. "Escaping from the Chinese Room." Computer Models of Mind, chapter 8. Cambridge: Cambridge University Press, 1980.

Cam, Philip. "Searle on Strong AI." Australasian Journal of Philosophy, 68 (1) (March 1990), pp. 103-108.

Cole, David. "Thought and Thought Experiment." Philosophical Studies, 45 (May 1984), pp. 431-444.

Delaney, C. F. The Synoptic Vision: Essays on the Philosophy of Wilfred Sellars. Notre Dame: University of Notre Dame Press, 1977.

Double, Richard. "Searle, Programs, and Functionalism." Nature and Systems (1983), pp. 107-114.

Hamlyn, D. W. The Theory of Knowledge. Macmillan Publishers, 1970.

Hanna, Patricia. "Translation, Indeterminancy and Triviality." Philosophia (Israel), 14 (December 1984), pp. 85-98.

Intention and Intentionality. Essays in Honor of G. E. M. Anscombe. Ithaca, N.Y.: Cornell University Press, 1979.

Jacquette, Dale. "Adventures in the Chinese Room." Philosophy and Phenomenological Research, 49 (1989), pp. 605-623.

Kenny, Anthony. Wittgenstein. Harmondsworth, Middlesex, England: Penguin Books, 1973.

Lyons, John. Language and linguistics. Cambridge: Cambridge University Press, 1981.

Maloney, J. Christopher. "The Right Stuff." Synthesis 70 (1987), pp. 349-372.

Martin, Robert M. The Meaning of Language. Massachusetts: M.I.T. Press, 1987.

Pitt, Joseph C. The Philosophy of Wilfred Sellars. Durdrecht, Holland: D. Reidel Pub. Co., 1978.

Rey, Georges. "What's Really Going on in Searle's 'Chinese Room.'" Philosophical Studies 50 (1986), pp. 169-185.

Searle, John. Minds, Brains, and Science. Cambridge, Mass.: Cambridge Havard University Press, 1984.

Searle, John R. "Minds, Brains, and Programs." The Behavior and Brain Sciences, 3 (1980): 417-24.

Searle, John. Intentionalty, An Essay in the Philosophy of Mind. Cambridge/New York: Cambridge University Press, 1983.

Sellars, Wilfrid. Science, Perception and Reality. Routledge and Kegan Paul, 1963.

Turing, A. M. "Computing Machinery and Intelligence." Mind, Vol. LIX, No. 2236 (Oct. 1950), pp. 433-60.

Vinci, Thomas. In conversation.

Wagman, Morton. Cognitive Science and Concepts of Mind. New York: Praeger Publishers, 1991.

Wittgenstein, Ludwig. Philosophical Investigations. New York: MacMillan Press, 1958.