Extended Essay

Philosophy

A critical analysis of John Searle’s Chinese Room Argument

**Research question:** Is the Chinese Room Argument either sound or convincing?

**Candidate name:** Danish Dhamani

**Candidate number:** 002352-015

**School code:** 002352 (Aga Khan Academy, Mombasa)

**Session:** May 2013

**Supervisor:** Mr. Nicholas Alchin

**Word Count:** 3862
ABSTRACT

‘Can machines think?’ is a contentious question in the field of Artificial Intelligence (AI). One hypothesis, known as strong AI, postulates that intelligence can be created via a physical symbol system that has a mastery of syntax. However, since there is no empirical proof of the so-called strong AI yet, there are several a priori arguments that refute its possibility. One convincing argument, proposed 30 years ago is John Searle’s Chinese Room Argument (CRA). Searle presents a thought-experiment, which he uses to justify his argument against strong AI. At first glance, his argument is convincing, and the first part of the paper discusses what makes it so by analyzing the validity of the axioms and the logic of the argument. A brief description of syntactic and semantic understanding is presented in order to better understand the axioms. Nevertheless, a convincing argument does not make it self-sound, thus, the research question examined is, “Is the Chinese Room Argument either sound or convincing?”

The second part of the paper reveals a subtle flaw in the thought-experiment, and since Searle uses this thought-experiment to justify one of the axioms for the CRA, Searle’s argument is deemed not sound. Furthermore, Searle’s rebuttal to the so-called System’s reply is shown to be a non sequitur, thus, further strengthening the argument presented by this paper. After challenging the assumption Searle bases his argument upon, the original question about thinking machines is still left unanswered; however, the paper asserts that semantic understanding may be possible by a mastery of syntax by expounding upon Rappaport’s idea that semantics are merely causal relationships between words and reality. This may be possible through a superior computer program that has the potential to establish this link. Thus, we see a relationship between computer potential and degree of understanding.

Word Count: 297
DEFINITION OF TERMS:

The Physical Symbol System Hypothesis: A physical symbol system has the necessary and sufficient means for general intelligent action (Newell & Simon 1976, p. 116).

**Strong AI:** A physical symbol system can have a mind and mental states (Searle, 1999).

**Weak AI:** A physical symbol system can act intelligently (Searle, 1999).

**Syntax** for a set of symbol-types is to specify a set of rules, which lay down how tokens of those types are to be combined with, or related to, one another (Melnyk, 1996).

**Semantics** is to state what tokens of the symbol-types, whether alone or in combination, mean, or are about (Melnyk, 1996).

**Intelligence:** Power of understanding, ability to reason, ability to learn and relate knowledge, information (Hayward, 1969).
TABLE OF CONTENTS

Abstract........................................................................................................................................2

Definition of terms: ......................................................................................................................3

Introduction .................................................................................................................................5
  Can thinking machines be built? ...............................................................................................5
  The Chinese Room Argument .................................................................................................7
  Distinction between sound and convincing ...........................................................................9

The Chinese Room Argument: Convincing? .............................................................................11
  Difference between syntactic and semantic understanding ..................................................11
  Validity of the axioms and the logic of the argument ............................................................13

The Chinese Room Argument: Sound? ....................................................................................15
  Intuition pumps .........................................................................................................................15
  Flawed assumption in Axiom 3 ..............................................................................................18

Conclusion ..................................................................................................................................19
  What is it to understand Chinese? ...........................................................................................19

References: .................................................................................................................................21

Appendix .....................................................................................................................................22
  Appendix 1: SEARLE’S CHINESE-ROOM THOUGHT EXPERIMENT ....................................22

TABLE OF FIGURES:

Figure 1: The Chinese Room (Figure from, Franklin, S. (1995). The First AI Debate. Artificial minds (p. 105). Cambridge, Mass.: MIT Press.) ..........................................................................................7
INTRODUCTION

Can thinking machines be built?

In 1965 Herbert Simon – a pioneer in AI (Artificial Intelligence) - predicted that in 20 years, machines would be able to do any work a human can do (Crevier, 1993, p. 109). After 47 years, no such accomplishment has come about. Why did we ever think in the first place that a machine could do what a human can? Is the human mind just a biological machine, and if given the technology, can we come to understand how it works? Will we be able to create fellow human-being-like-beings or ‘robots’, who can think, act and behave exactly as we do now? These ‘robots’ would also learn from their surrounding and their mistakes, and maybe one day become superior to us, at least in limited domains.

It is the philosophical nature of the mind and its functioning that makes this quest a philosophical one. Maybe the cultural and political environment was never satisfactory, and the failure of AI is not due to a technological cap. Maybe this matter requires a ‘paradigm shift’ as coined by Kuhn, where we now need to look at AI in a complete overhauled manner, rather than in the orthodox ways. For example, emulating the behavior of individual neurons rather than the brain as a whole.

The Chinese Room Argument (CRA), by John Searle (1980) is a philosophical argument that declares ‘computer programs cannot think’. This is an a priori argument that does not require empirical evidence, rather just a series of logical statements combined to prove a point. Philosophy comes into play when the scientific technology is insufficient to quench the inquisitive minds of human beings. Since one of the best AI current technology can produce is the ‘Google Driverless Car’, we have a need to use theoretical means to answer this contentious question, ‘can thinking machines be built?’
The CRA argues against the ‘The Physical Symbol System Hypothesis’ and refutes strong AI, which claims that a physical symbol system (PSS) can have a mind and mental states. It is convincingly shown by the Chinese room thought-experiment that a PSS is not sufficient for intelligence. The first part of the paper explains what makes this argument convincing by looking at the validity of the axioms and the logic of the argument, as well as by differentiating between syntactic and semantic understanding. The second part of the paper, however, reveals a subtle flaw in the thought-experiment, and since Searle uses this thought-experiment to justify one of the axioms for the CRA, Searle’s argument is refuted and deemed not sound. Thus, the research question explored: *Is the Chinese Room Argument either sound or convincing?*
The Chinese Room Argument

The following is a persuasive reconstruction of Searle’s Chinese Room Argument for contextual understanding. See Appendix 1 for the argument in Searle’s own words.

Searle argues against strong AI (1980) by basing his proof on a non-empirical argument in his paper entitled “Minds, Brains, and Programs,” appearing in The Behavioral and Brain Sciences (Volume 3, 1980). He also describes in this paper the so-called strong AI; he states that, ‘computers given the right program can be literally said to understand and have other cognitive states.’ Thus, if he can show that a computer with the ‘right program’ shows no signs of understanding, then strong AI is false. This is exactly what he presents in the CRA, which he expresses by the Chinese room thought-experiment (see Figure 1):

![Diagram of the Chinese Room](image)

*Figure 1: The Chinese Room (Figure from, Franklin, S. (1995). The First AI Debate. Artificial minds (p. 105). Cambridge, Mass.: MIT Press.)*

Imagine a man locked in a room with only a comprehensive rulebook –written in English– for symbol manipulation. A Chinese interlocutor inserts a story and a question on it – written in Chinese– into the room. The man inside the room is monolingual and does not understand a single character of Chinese. His task inside the Chinese room is ‘symbol manipulation’; to manipulate the ‘input’ Chinese questions into an ‘output’ Chinese answer, by following the rulebook written in English instructions that he can understand. The only way he
can communicate outside the room is via written notes in Chinese. Although this process would take years to complete when done by hand, the end result is the same as the anticipated answer by the interlocutor. (The concept of time here is a practical issue that needs not concern this thought experiment.) Searle refers to this process of symbol conversion as mastery of syntax by the man with no mastery of semantics whatsoever. Searle illustrates in his thought-experiment how a man locked in a room can simulate a PSS and come up with ‘intelligent answers’, which he has no idea what they mean. Therefore, Searle concludes, “just manipulating the symbols is not by itself enough to guarantee cognition, perception, understanding, thinking and so forth” (Searle, 1990, p. 26); since the process of symbol manipulation does not allow the man to gain any understanding of his answers, a PSS is not sufficient for intelligence. Even though, the Chinese interlocutor has no doubts whether he is conversing with a person who truly understands Chinese, we, who know what is inside the room, can certainly agree to the fact that the man in the room does not understand a single Chinese character. Since strong AI states that a PSS can have a mind and mental states (Searle, 1999), and the above thought-experiment revealed no signs of understanding by the man (who is part of the PSS), strong AI can be effectively refuted non-empirically.

Searle (1990) presents the argument from the thought-experiment as following:

**Axiom 1.** Computer programs are formal (syntactic).

**Axiom 2.** Human minds have mental contents (semantics).

**Axiom 3.** Syntax by itself is neither constitutive of nor sufficient for minds.

**Conclusion:** Programs are neither constitutive of nor sufficient for minds.

∴ Strong AI is false.
Distinction between sound and convincing

Two conditions must be satisfied for an argument to be sound:

1. The premises must be true
2. The logic must be valid, i.e. the conclusion must follow from the premises

Observe the example below:

- I am a man.
- All men have cars.
- Therefore, I have a car.

In the above argument, the conclusion follows from the premises (the logic is valid), however the second premise is not true, therefore this argument is not sound.

In a purely logical world sound and convincing would be the same; if an argument is convincing, it is sound. However, invalid arguments can be made to sound convincing, and vice versa. Thus, a convincing argument does not require itself to be sound. Furthermore, the notion of convincing is subjective, and it varies for different people. An argument deemed sound by a logician might not be convincing for a person who might have limited knowledge on the subject matter, or vice versa. For example, it is widely accepted that mosquitoes are awful, and so, a logically valid argument based on this premise would be sound. However, if a person does not believe that mosquitoes are awful, then that premise is not justified for him to be true; thus, the argument (for him) is not convincing.
From the mosquito example, two conditions can be derived that must be satisfied for an argument to be *convincing*:

1. The person believes the premises
2. The logic of the argument must be valid

The CRA is a rather convincing argument with a highly plausible thought-experiment (as presented in the previous section). Nevertheless, are its premises true? If yes, is the logic of the argument valid? For the CRA to be a flawed argument, it has to be shown that the answers to either of these questions is no.
THE CHINESE ROOM ARGUMENT: CONVINCING?

Difference between syntactic and semantic understanding

It is crucial to establish the difference between syntactic and semantic understanding in order to better understand Searle’s argument.

The example below illustrates how the human mind can perform symbol manipulation that results in no semantic understanding.

\[3x + 1 = 7\]
\[3x = 7 - 1\]
\[x = \frac{7 - 1}{3}\]
\[x = 2\]

The value of \(x\) as 2 is deduced through syntactic steps by performing similar symbol manipulation like that in Searle’s Chinese room. This example shows that I have syntactic understanding of algebra and that there is no semantics involved here. Thus, when I solve the algebraic equation, I am not required to understand what it means to take ‘1’ on the other side.

Semantic understanding, on the other hand, would arise when algebra is performed on a finite quantity of real life objects (rather than on the symbols that represent nothing). For example, I buy three watermelons ($x$ each) and one apple ($1$ each) at a grocery store. If I pay $7$ for my grocery, what is the cost of one watermelon? Solving this problem semantically would involve saying to my self that, ‘remove $1$ from the $7$ since that paid for the apple. Thus, the remaining $6$ paid for the three watermelons.’ I then would divide the $6$ equally amongst the 3 watermelons giving me an answer of $2$ /watermelon.
These two examples show that the human brain has the capacity to perform mere syntactical manipulation, just like a computer program, as well as produce semantic understanding. How does it produce the latter? Cognitive science has not discovered this yet, though it appears to be at the root of our cognitive processes.

Now from this difference between syntactic and semantic understanding, it can be understood what Searle implies in his Axiom 3, where he states that syntax is insufficient for semantics; by knowing how symbols can be combined with each other does not lead to the understanding of what they mean. Take for example learning a new language by memorizing only the syntax of the grammar; without a need to learn what the words mean -having no semantic understanding- one can read sentences in this new language and even conjugate verbs in different tenses solely with the knowledge of syntax. Nonetheless, Searle must assume for his CRA that syntax is insufficient for semantic for all types of programs, however superior they may be. It is this assumption that shall be argued against later on.
Validity of the axioms and the logic of the argument

Looking back at the CRA:

**Axiom 1.** Computer programs are formal (syntactic).

**Axiom 2.** Human minds have mental contents (semantics).

**Axiom 3.** Syntax by itself is neither constitutive of nor sufficient for minds.

**Conclusion:** Programs are neither constitutive of nor sufficient for minds. (Searle, 1990)

As discussed in one of the previous sections, for the above argument to be convincing (to a person):

1. The person believes in the axioms
2. The person believes the conclusion follows from the axioms

**Axiom 1:** The thought-experiment clearly shows that the person in the room (computer hardware) performs syntactic manipulations on the Chinese symbols (input) using the rulebook (program) to get the anticipated answer (output). Since computer programs function in a similar manner, they can be called *formal*. Searle (1990) further justifies this axiom by saying “symbols are manipulated [by the program] without reference to any meanings.” For example, if the Chinese story was on a restaurant, and the question asked was ‘did the waiter serve the food?’ The program will not need to know the meaning of the words ‘waiter’, ‘serve’ and ‘food’ to answer this question. Because it has a mastery of syntax, it can find the connections between the question and the story and correctly answer the question. Thus, this axiom is justified.

**Axiom 2:** This axiom appears to be self-evidently true from introspection and from ruling out the alternative; if this axiom is false, how else are humans able to understanding meanings in language? When we think of a word, say *horse*, our mind simultaneously assigns some form of meaning to that symbol of words, whether it is a picture or a description or even an anecdote. A
computer program, on the other hand, would merely represent the *horse* in symbols (h-o-r-s-e) and attach no meaning to it.

**Axiom 3:** From the thought-experiment, the process of symbol manipulation does not allow the man to gain any understanding of his answers. Thus, mastery of syntax appears insufficient to create mastery of semantics. However, it will be the validity of this axiom that shall be challenged in the next section.

**Conclusion:** Computer programs have mastery of syntax, and mastery of semantics is required for understanding to take place; furthermore, mastery of the former is not constitutive of the latter, and since programs have mastery of only the former are, therefore, unable to understand.

The above shows one-way in which Searle’s argument can be classified as convincing. However, as stated earlier, only in a purely logical world are *sound* and *convincing* the same, and since an invalid argument can be constructed to sound convincing, the CRA needs to be critically analyzed before deeming it sound.
THE CHINESE ROOM ARGUMENT: SOUND?

Intuition pumps

Before beginning the argument against Axiom 3 and the fallacious logic in the CRA, the concept of intuition pumps, as coined by Daniel Dennett, will be described.

A popular strategy in philosophy is to construct a certain sort of thought experiment I call an intuition pump [...]. Intuition pumps are cunningly designed to focus the reader's attention on "the important" features, and to deflect the reader from bogging down in hard-to-follow details. There is nothing wrong with this in principle. Indeed one of philosophy's highest callings is finding ways of helping people see the forest and not just the trees. But intuition pumps are often abused, though seldom deliberately.

(Dennett, 1984)

The CRA can be believed to be a ‘cunningly designed’ intuition pump that focuses the readers’ attention only to the man-in-the-room and what he understands. This not right for a philosophical argument as it biases and limits the reader into regarding the human as the only “locus of Chinese-understanding” (Copeland, 1993, p. 126). Searle’s entire argument lies on the fact that because this man does not understand, nothing in the Chinese room (the PSS) can understand. Copeland (1993, p. 126) further calls this misdoing as a “spurious pull towards Searle’s conclusion”. Let us tweak the Chinese room scenario where now there is no man at all to perform the symbol manipulation. In his place is Pebbles, the man’s best friend. This dog has been trained to flick switches representing the binary digits 1 and 0. This enables Pebbles to manipulate the Chinese symbols and come up with the Chinese answer in a similar manner. In this tweaked scenario, would the reader still be biased into thinking that since the dog does not understand, the system of which it is a part of, too does not understand? As long as the Chinese interlocutor outside the system gets an anticipated answer, he has no reason to believe that an understanding does not take place within the system. Note that the unintelligent dog has been
intentionally chosen, in contrast to the *intelligent* man; this is because in reality, the *thing* manipulating the symbols in a computer is an inanimate, unintentional, unthinking program. Therefore, it makes no sense to *lure* the reader into believing that in place of the non-understanding program performing the symbol manipulation, a man is used instead in this thought-experiment, who has the capacity to understand, but cannot due to the limits placed by Searle. This sort of argument is termed as the System’s reply by Searle (1980, p. 419) because it argues that the system as a whole needs to be looked at to validate the creation of understanding rather than pinpointing only at the *thing* carrying out symbol manipulation. One might wonder that if the man-in-the-room is not the only locus of understanding, what else in the room is capable of understanding? Is it the rulebook or the table in the room that understand? Although this rebuttal seems certainly valid, but to a person outside the Chinese room it seems that understanding has to have taken place for their question to be answered.

The System’s reply is based on the fact that there are some concepts that cannot be broken down to the basics. Science approaches stuff by reductionism - breaking down ideas to get to the basic point. However, if you keep going down till the lowest levels, you lose understanding of the system as a whole. Just as one water molecule would not be referred to as being wet, instead wetness arises from a collection of billions of water molecules together. Just because one molecule is not wet, does not mean that wetness does not exist. Similarly, a painting is made up of ink, and as a whole conveys meaning; one would not refer to one orange spec of ink as conveying the meaning of the painting. Consequently, in music not one note gives rise to a tune, rather, it as a whole system of notes in synergy. Thus, the System’s reply expounds upon the concept of a system that is much broader than just the man doing the symbol manipulation.
Therefore, just because at the basic level the man does not ‘understand’ does not mean that understanding does not take place in the system.

Searle (1980), in fact, saw this sort of reply forthcoming against his CRA and counters it by saying, “let the individual internalize all these elements of the system” (p. 419). His response asks the reader to suppose the man has all the syntax rules in his mind, resulting in no Chinese room anymore; all there is, is the man and the Chinese interlocutor. He concludes his rebuttal by asserting, “there is nothing in the ‘system’ that is not in me, and since I [the man] don’t [does not] understand Chinese, neither does the system” (Searle, 1990)

The above conclusion is unsubstantiated as shown by the system’s theory in the preceding paragraph; just because the man does not ‘understand’ does not lead to a conclusion that understanding does not take place in the system. Even if understanding, in fact, does not take place, the validity of the conclusion cannot make an argument sound as long as the logic of the argument is fallacious. Therefore, Searle’s rebuttal to the System’s reply is a non sequitur as it simply restates the CRA and draws no logical conclusion from it.
Flawed assumption in Axiom 3

Axiom 3. Syntax by itself is neither constitutive of nor sufficient for minds. (Searle, 1990)

As explained earlier, the justification for this axiom arrives from Searle’s thought-experiment. This, however, is a wrong assumption because of the empirical nature of the above premise, which cannot be justified by an a priori argument. Although there is no semantics in the program, only syntax, Searle cannot postulate from this that syntax is insufficient for semantics. Maybe, through progress in AI research, a superior program is created that can behave exactly as a human – think, reason, show emotions-, and until it is shown that more than mastery of syntax is required to create such a program, Searle’s Axiom 3 cannot be justified. Thus, the CRA is not sound as one of its axioms is not justified. One might argue as to why the CRA is not sound while the conclusion (the man does not understand) is true. As discussed earlier, even if the conclusion is true, it does not make an argument sound.

In addition to the above, imagine a situation where your best friend opened their chest to reveal an intricate system of wires and metal; most people would react to this by saying “oh robots can understand and talk after all.” However, it can be categorically assumed that Searle would react to this in a complete nonsensical way, “I do not care what he says to me because it is just complex symbol manipulation going on in his head. I know from my Chinese Room Argument that there is no understanding at all.” Therefore, for Searle, evidence is irrelevant. But what if one day we shall have machines that, only through symbol manipulations, will be able to have such conversations on abstract topics such as ‘the existence of God’ or ‘the meaning of life’? In these cases, it would seem perverse to deny the machine any semantic understanding, surely.
CONCLUSION

What is it to understand Chinese?

The first part of the paper showed that the CRA is a rather convincing argument, by looking at how the axioms could be true to a person and how the axioms follow to come up with the conclusion. The question raised here is why despite undermining the argument we actually, at first, found the argument rather convincing. Is our reasoning process not necessarily logical?

The second part of the paper conceded that in the thought-experiment there is only an understanding of syntax. However, since it has been shown that the argument is not sound, due the fallacious assumption upon which Axiom 3 is based, the CRA does not prove that syntax is insufficient for semantics. It was also shown that Searle’s reply to the System’s reply is a non sequitur, thus strengthening the argument against the CRA. Thus, looking back at the research question -Is the Chinese Room Argument either sound or convincing?- we find that the CRA is convincing but not sound.

The argument of this paper is based on the assertion that semantic understanding may be possible by symbol manipulation. Rapaport (1986) explains this possibility by suggesting that semantics is an external fact, which serves to link the representations of the symbols being manipulated to the objects it refers to. For example, when the man-in-the-room answers ‘table’ in Chinese, for there to be semantic understanding of the word ‘table’, there has to be a link between the simple symbols (t-a-b-l-e) and the object it refers to. It is only after this link would there be an understanding of the true meaning of the word. The computer program can then ponder what else can be classified as a table?; ‘Can it be any surface on which something can be placed upon or is it mandatory for it to have four legs and a flat surface?’ The idea that semantics are causal relationships between words and reality, seems highly plausible, and to achieve the
semantic link, Rapaport (1986) explains that this would require “just more syntactic symbol pushing”, which can plausibly be achieved by a superior program than that used in the Chinese room. Thus, it would be interesting to find out at what point on the continuum of understanding does a program move from merely manipulating symbols to truly understanding?

It was interesting to observe that Searle’s argument is a priori – argument comes before the evidence. This was especially the case when the validity of Axiom 3 was critically analyzed. Searle postulates that syntax is insufficient for semantics simply by means of his thought-experiment, which was shown to be a devious intuition pump. To say that I refuse to look at the evidence is terribly detrimental for innovation, especially in the case of AI. Thus, it is best not to classify this question as a priori or a posteriori, but rather to leave it open. The day science finally answers the question “can computer programs think?” this argument will cease to be philosophical and will become amendable to science. Aristotle used to argue that women were inferior to men because they have fewer teeth, until one day someone said “Hey there, hang on for a minute, why don’t we count them!” This argument then ceased to be philosophical. Does this mean that Science is the ultimate way of knowing, and will it ever encroach the idea of AI?

It must be noted that Searle (1980) does not go against the possibility that machines can think, because, he believes that the human brain itself is a biological machine that can think. He simply argues against strong AI by highlighting that only syntactical manipulation of symbols is insufficient for understanding, thus thinking may involve symbol manipulation to some extent.
REFERENCES:


APPENDIX

Appendix 1: SEARLE'S CHINESE-ROOM THOUGHT EXPERIMENT

Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore, that I know no Chinese. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes. Now suppose that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch "a script," they call the second batch a "story," and they call the third batch "questions." Furthermore, they call the symbols I give them back in response to the third batch "answers to the questions," and the set of rules in English that they give me, they call "the program." Imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view—that is, from the point of view of somebody outside the room in which I am locked—my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Let us also suppose that my answers to the English questions are indistinguishable from those of other native English
Danish Dhamani (002352-015)

speakers. From the external point of view—from the point of view of someone reading my "answers"—the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program.

Now the claims made by strong AI are that the programmed computer understands the stories and that the program in some sense explains human understanding.

It seems to me quite obvious in the example that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, a computer understands nothing of any stories.

We can see that the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and there is no understanding. But does it even provide a necessary condition? One of the claims made by the supporters of strong AI is that when I understand a story in English, what I am doing is exactly the same as what I was doing in manipulating the Chinese symbols. I have not demonstrated that this claim is false. As long as the program is defined in terms of computational operations on purely formally defined elements, what the example suggests is that these by themselves have no interesting connection with understanding. Whatever purely formal principles you put into the computer, they will not be sufficient for understanding, since a human will be able to follow the formal principles without understanding anything (Searle 1980, pp. 417-18).