Sam and the Chinese Room

Roseli Gimenes

Universidade Paulista UNIP and PUC – SP TIDD

Abstract - The aim of this article is to describe the computer program SAM, introduced by Schank and Abelson, as well as its relationship with the Chinese room, presented by John Searle. Both programs, which belong to the field of Cognitive Science and Artificial Intelligence, are paths to approach the theory of knowledge and the ability of machines to tell a story.

Key words: Cognitive Sciences, Artificial Intelligence, Artificial Narratives.

1. Introduction

Cognitive Science has an interdisciplinary perspective with Philosophy, Anthropology, Linguistics, Cognitive Psychology, Neuroscience and Artificial Intelligence. This is fundamental for Psychology to understand how people think and act. Basically, Cognitive Science is the search for mind understanding and its goal is to explain how the mind works. Neuroscience studies the relations between mind and brain. Artificial Intelligence searches for model processes of human thinking in computer software and hardware. Linguistics investigates language use structures and what they can tell us about the mind. Anthropology observes mental thinking by analysing cultural features. Cognitive Psychology is the key to interdisciplinarity in these studies. [3]

What is Artificial Intelligence about? According to Boden [1] it is not about studying computers; it is about studying computer programming, that is, the use of computer programs and computing techniques to make a list of intelligence principles in general and human thinking in particular. Thus, computers would not be number triturators but symbol manipulators.

In this article, discussion will focus on John Searle's [6] inquiries on whether computers can think, the differences of the terminologies *strong Artificial Intelligence* and *weak Artificial Intelligence* Boden [1] and Schank and Adelson's [4] work commented by Searle as a simulation of human ability to understand stories.

If computers can think, can they narrate and understand stories?

2. Turing machine and the chinese room

In 1950, discussion on Artificial Intelligence was proposed by Alan Turing in one of his articles. [2]

Turing's Test aims at determining whether machines can show intelligent behavior. In Turing's original example, a judge talks to a man and to a machine which was created to perform just like a human being. The judge is not supposed to know when he is talking to the man and when he is talking to the machine. If the judge cannot tell one from the other it means the machine has passed the test. The conversation is limited to written texts (for example, a keyboard and a video monitor) so that the result does not depend on the ability of the machine to randomize words in audio.

Turing begins his article with some philosophic questions related to artificial intelligence, such as "Can machines think?". Considering "thinking" is difficult to define, Turing preferred to substitute his question for another one less ambiguous: "Is it possible to figure out digital computers which would be successful in the *game of imitation*? Turing believed there was an answer to that question. Then, he continues his article arguing against the great objections to "machines can think".

In 1980, John Searle wrote an article named *Minds, Brains and Programs* [7] proposing an argument known as *The Chinese Room* which was meant to prove that a computer program cannot give the computer a mind, comprehension or conscience, no matter how intelligent the machine may seem.

The *Chinese Room* is a thinking experience. It assumes there is a program which provides the computer with enough capacity to develop an intelligent conversation in Chinese writing. Searle supposes a man locked in a room. This man is given a paper pad with a text in Chinese. The man does not know written or spoken Chinese, so he does not recognize Chinese writing. A second paper pad, also in Chinese, is then given to that man with a syllabus of rules and he is supposed to relate the first information with the second one. The rules are in English and he knows this language. That allows him to relate a set of formal symbols with the other one. Formal is understood here as the ability to identify symbols by observing their shapes. Finally, the man is given a third paper pad which contains symbols in Chinese and he also receives some instructions in English to relate this one with the first and the second ones. The rules allow him to relate certain symbols in Chinese with certain configuration types. The man does not know it, but he is provided with texts with the following symbols: the first one is a syllabus; the second one is a story; the third one contains questions. The man in the room is then able to pass out the Chinese symbols which are correct answers to the questions. After some time, from the outside, Chinese people will say the man did really well for he answered the questions just like a Chinese speaker would have done. According to Searle, the answers are good enough, but the formal symbols in Chinese are meaningless. The man behaved as the computer: he executed computational operations; it is only a computer program instance.

2.1 John's Propositions

In *Minds, Brains and Science* [6] and *Minds, Brains and Programs* [7]. Searle discusses strong Artificial Intelligence. He points to weak Intelligence in the sense of valorizing the computer for the study of the mind, as a powerful tool to formulate and test hypotheses precisely. On the other hand, strong Artificial Intelligence sees the computer not as a tool for the study of the mind, but as a computer program, and the brain as a digital computer. Thus, mind is to the brain as the program is to the computer hardware.

According to this conception, human mind would not have anything biological. Any physical system which had a correct program with the correct inputs and outputs would have a mind.

Searle [6] mentions that some researchers of strong Artificial Intelligence such as Simon, Newell, Dyson, Minsky and Mc Carthy state, among other things, that intelligence is a matter of physical symbol manipulation; there are no metaphors; as far as evolution is concerned, computers would have advantage over human beings and that even thermostats have beliefs. According to Searle [6]:

> *Mc Carthy says that even "machines as simple as thermostats have – one can say – beliefs". I admire Mc Carthy's courage. Once I asked him: 'Which beliefs does my thermostat have?' And he answered me: 'Mine has three beliefs: it is too hot here, it is too cold here and it is right here.' As a philosopher I appreciate these statements, they are reasonably clear and admit a simple and decisive refutation*. [6: 30]

According to the author, his refutations to the ideas of these researchers is that a digital computer has purely formal operations of abstract symbols – sequences of zeros and ones printed on a tape. These symbols have no meaning and no semantic content.

Back to the *Chinese Room*, [6] states that if the man cannot understand Chinese, a digital computer cannot as well because the computer has syntax, but no semantics:

> *And the digital computer can only have formal symbols because the operation of a computer occurs in terms of its capacity to run programs. And these programs are run in a purely formal way, that is, there is no semantic content*. [6:33]

Concerning the objections made by the researchers about the *Chinese Room*, Searle [7:72-80] points out:

> *System Objections (Berkeley): system cannot understand anything because system does not have anything that man does not have. If the man cannot understand, then system will not understand. System is merely a part of man.*

> *Robot Objection (Yale): a robot does not have intentional actions; it moves as a result of its electric systems and its programming. The creation of a program does not produce intentional states. All that it is done is to follow formal instructions about the manipulation of formal symbols.*

> *Brain Simulator Objection (Berkeley and MIT): Schank's works – which will be presented later - are mentioned here. According to Searle, the problem with the brain simulator is that it simulates wrong things about the brain. It simulates the formal structure of neural activities sequences, its causal properties and its ability to reproduce intentional states.*

> *Combination Objection (Berkley and Staford): Searle sees the robot as a mechanic puppet.*

> *Other minds Objection (Yale): Cognitive Science presupposes mental state can be accessed just like physical sciences presuppose physical objects can be accessed.*

> *Several groups Objection (Berkeley): Searle says that it trivializes strong Artificial Intelligence project by redefining it as anything that can produce and explain cognition artificially*.

His most important question comes out at the end of both his articles: Can a machine think? Can a digital computer think?

The author asserts that our brains are like digital computers because they carry out any number of computer programs. And our brains can think.

According to the author, the digital computer cannot think because it has only got syntax information. Thinking is more than meaningless symbol manipulation. What a digital computer can actually do, however, is to simulate human behavior. Even so, simulating is one thing and being real is another. When a storm is simulated, a shelter does not need to be reached because we know it is not for real, it is only a simulation.

Thus, Searle converges both articles like this: computers are not minds; a mind does not work by just activating a computer program. A computer does not make the causal connections a brain makes. Mental states are biological phenomena.

2.2 Schank's and Abelson's positions

Robert P. Abelson, a psychologist in Yale, and Roger Schank, Artificial Intelligence Researcher, wrote the book

Scripts, Plans, Goals and Understanding – An inquiry into human Knowledge structures [4].

As it has been mentioned before, Schank is quoted by John Searle [7] for his project in Yale with machines which could understand stories. The book mentioned above became a classic and has been quoted by many social scientists:

> *I will analyze the work of Roger Schank et al. in Yale because I am more acquainted to it than to other similar works. Besides, it provides a clear example of the kind of work I wish to look into.* [7:65]

On his trajectory as a researcher, Schank showed how computers were able to process daily sentences in the English language as well as read newspaper articles. In 1976, he launched the first computer program which could read newspaper stories. With his projects, he realized computers had problems with memory – ability human beings have – but, on the other hand, computers could actually "remember" whole volumes, which is impossible for humans. What computers lacked was the ability to generalize. They could read a story, but they were not able to recognize aspects of a certain story in another one they had read before. Computers did not understand because they did not match similar occurrences. Schank realized the ability of generalizing and memory were interconnected.

Schank's connection to Abelson made him study the learning process. If he observed how people learn, he could apply this knowledge to computers and make them understand stories. Schank started to build up real world events for the computers. People remember things all the time. If things do not happen the same way, humans ask why. Hence, computers should have expectations.

Schank realized people store memory in packs. Man reconstructs several events when he needs to remember something. This was the basis, the dynamic memory, a theory of remembering and learning.

Schank's approach was more cautious than the approach of other researchers of strong Artificial Intelligence such as Simon and Newell. He defends computing as a means of testing cognition theory. According to Searle (1990, 2005), a weak Artificial Intelligence:

> *According to weak Artificial Intelligence, the main value of computer for the study of the mind resides in the fact that it provides us with an extremely powerful tool. For example, the computer allows us to formulate and test hypotheses in a more rigorous and precise way than before.* [7: 67]

The book mentioned is a theory of Cognitive Sciences about the understanding of stories. Basically, it suggests that meaning and cognition occurs by means of comprehension of concepts and sentences.

In this study, Schank and Abelson [1] point out that the work is not only about Psychology, Artificial Intelligence or Linguistics, but about the three areas together. Interdisciplinary. And that lead them to think about causal chain. Thus, interpretation is described as a filling in the blanks in a causal chain:

> *Psychology which studies knowledge systems wants to know how concepts are structured in the human mind, how such concepts develop and how they are used in comprehension and behavior. Artificial Intelligence researchers want to know how computer programs can understand and interact with the external world.* [4: 1]

The book is distinguished by chapters which discuss:

> *Scripts: composed of branches, roles, states, entry conditions and results. People act appropriately because they have world knowledge. A story is understood because people fill in missing information when reading.*
>
> *How do people organize all their knowledge in comprehensible sequences? How do people know which behavior is appropriate in a certain situation? (…) People know how to act appropriately because they know about the world they live in. What is the nature of this knowledge?* [4: 36]
>
> *Understanding is, then, a process through which people find what to see and listen in pre-stories of group actions they have already experienced. (…) Scripts intend to contain specific knowledge that people have. Most comprehensions have a basic script*. [4: 67]
>
> *Plans: they are means to reach for satisfactory objectives. There are USE plans, for example. They are made of general information so that "actors" can reach their targets. As the authors say* [4: 71]: *"A plan is a series of actions projected to reach a goal."*
>
> *Goals: There are seven forms of goals: satisfaction, enjoyment, achievement, preservation, crisis, instrumental and Delta. Instinct, necessity, values, way of living, beliefs. The authors classify the goals: a prime goal, a specific and substitute goal, a suspended goal and stylistic goals*. [4: 103]. *The authors ask the question: "Where do goals come from?" For them, the answer lies on what they define as 'themes'*. [4: 119].
>
> *Themes: the authors ask themselves: 'Where do goals and plans come from?', 'What are stories about?' For example, interpersonal relationships. In a theme list, people's goals are determined by social rules.* [4:132].

2.3 SAM computer program

In this book, intended to present computer programs, they try to construct intelligent machines taking into account human natural language processing and natural language processing in computers. The authors show several programs which were created in order to make machines understand stories. TALESPIN, PAM and SAM.

What is SAM? It is a computer program: Script Applier Mechanism. [4: 177]. This program was presented in Yale and it was designed so that stories could be understood as scripts. SAM links a series of causal concepts.

The track described, pointed out by Searle [7: 67], is the restaurant. A person goes to a restaurant, orders for something to eat, pays and leaves. They are short narratives, apparently uninteresting, but with great potential. The scripts have metadata, they describe the basis of the event and the system recognizes which script it is supposed to use: instrumental relations, places. These scripts are social rules, procedures, conventions. [4:36], scripts are powerful elements for cognition and comprehension of the world. If the scripts are known, we learn with the experience.

SAM [4: 176] is a kind of script "expert", just like some other programs such as FRUMP – the summary of a newspaper story based on scripts, TALESPIN – a story narrator which uses the program of plans and goals, and PAM – a program which "understands" stories by using plans, goals and themes.

Anyway, by the time the authors were working with SAM, there was not any other computer program enabled to understand stories.

Schank and Abelson [4] describe SAM by giving an example, as follows:

> *Input: John went to a restaurant. He sat down. He got mad. He left.*
>
> *John was hungry. He decided to go to a restaurant. He went to one. He sat down on a chair. A waiter did not go to the table. John became upset. He decided he has going to leave the restaurant. He left it*. [4: 189]

A script is a pre-organized inference chain related to a specific routine situation [4: 36]. It is a sequence of conceptualizations with some variables (*script* variables).

The restaurant *script* intends to capture a person's (actor's) knowledge about the sequence of events that occur when this person goes to a restaurant.

(1) Actor goes to the restaurant

(2) Actor sits

(3) Actor orders the waiter for a meal

(4) Waiter brings the meal to the actor

(5) Actor eats the meal

(6) Actor gives money to the restaurant

(7) Actor leaves the restaurant.

Schank and Abelson [4: 46] believe that people understand a story (an event) more easily when they have experienced it many times before. This experience is decoded in a script that, once constructed, is attached to the long-term memory and does not need to be recapitulated any longer. The script has a strong prediction power and failures in its structures can be recognized.

Actually, the detection of failures in a script is more related to the way it is organized than to the codified information. This organization, in turn, can be dynamically codified.

Initially, in the first versions of the theory, a script was seen as a structure which represented separate temporal sequences; one script did not relate to others. As the model was developed, the authors started to see scripts in a more modular way from which the interconnectivity of scripts starts to be investigated.

In this new conception, [5: 181-200] a model named Memory Organization Packets (MOPs) is developed. MOPs were meant to cut the script into small units called scenes. Then, the same scene could be shared by many MOPs because (a) it would not make any sense that the same information were represented in different "places" and (b) that would indeed facilitate learning. The example given can form the following scheme:

MOP 1: VISIT TO THE DOCTOR

MOP 2: VISIT TO A LAWYER

Shared scene: BEING IN A WAITING ROOM

A theory on memory organization "as a whole" is necessary when it comes to the dynamic modification of a MOP. Schank [5] then develops a theory named Dynamic Memory. His proposal is to connect MOPs the same way MOPs connect scenes. Hence:

MOPs would be connected by a set of abstraction hierarchies.

An example:

MOP: VISIT TO AN OFFICE (More abstract level).

MOP 1: VISIT TO A DOCTOR (Upper level MOP instance).

MOP 2: VISITI TO A LAWYER (Upper level MOP instance)

The MOPs would be connected by a set of packet links, connecting MOPs to other MOPs which frequently occur together in a broader context.

An example:

MOP: BUSINESS TRIP

MOP 1: TRAVELING BY PLANE

MOP 2: CHECKING INTO A HOTEL

MOP 3: BUSINESS LUNCH

2.4 SAM and the relation with a case study

Schank and Abelson [4: 177], still asking themselves ‘Where do scripts come from?’, point out that language acquisition is script acquisition.

As an example, Schank and Abelson [4: 228-237] test *script* with a child. The same script is narrated to the same child in three different moments: at the ages of two years and six months, three years and four months and four years and two months.

If the child has already been to a restaurant, he will answer the questions by using memory references: arriving, sitting, ordering, paying and leaving. This sequence may similarly repeat if the script is going to a pet shop: ordering, paying and leaving, for example.

The authors analyze that the sequence of actions is a crucial factor for the memory of the child and point out that the concept of memory was strongly activated in the first experience. Memory has made connections in context. Hence, scripts are learned to connect events and they are organized by goal structures which are used to meet the needs of these goals.

In this experience, the child is asked to tell stories. In the beginning of the experiment, when the child is younger, many details are told and these details are different from those which really matter in the scripts for adults. As the child grows older, he changes his system to tell stories: from a model based on scripts to a model based on plans. In that case, the program SAM – based on scripts – gives way to the program TALESPIN – based on plans.

The child is told two stories. In one of them a man gets on a train, sits, is robbed and leaves. In the other story, the man who left the train goes to a restaurant. He gets in, sits, orders, eats and when he is going to pay he realizes he does not have any money so he will have to wash the dishes.

At first, the child does not understand the man was robbed because that child has not experienced a robbery before. This is also the reason the child does not understand why the man has to wash the dishes in the restaurant, as a way of paying for his debt. The child’s script is: the man gets into the restaurant, sits, orders, is served, pays and leaves.

The use of scripts depends on the perfect comprehension and the conditions under which someone decides the use a given script.

Schank and Abelson [4: 237] conclude the experiment with the child by asserting that ‘the limit of comprehension is the limit of the world knowledge comprehension’: *The child understands to the limit of his world knowledge (…) You know what you can understand. This is true for children and for adults.*

3. Conclusions

John Searle [7: 67] pointed out that Schank and Abelson [4] created SAM – the computer program – to simulate human ability to understand stories. According to Searle [7], human beings do understand a story when they can answers questions about it, even if some information necessary to answer these questions are not explicit in the text.

About Schank and Abelson’s [4: 178] restaurant script, Searle [7: 68] comments that a man went to a restaurant and ordered for a hamburger. When his order came he realized the hamburger was toasted so he left the restaurant feeling furious. He did not pay the bill and did not tip the waiter.

The authors observe that if the following question was asked: ‘Did the man eat the hamburger?’ anyone would probably answer: ‘No, he did not.’ This would be the answer, given the circumstances.

A man went to a restaurant and ordered for a hamburger. When his order came he was quite satisfied with the meal and before paying for the bill he tipped the waitress generously.

If the following question was asked: ‘Did the man eat the hamburger?’, anyone would certainly say: ‘Yes, he ate the hamburger.’

According to Searle [7: 68], Schank and Abelson’s machine [4] is enabled to answer questions like those about restaurants. To do so, the machine needs to be programed with the information human beings have about restaurants.

When the machine is provided with the story and a question is asked, the machine will print out the same answers humans were expected to give.

To Searle [7: 68], people who back up strong Artificial Intelligence assert that this question-answer sequence indicates not only that the machine is simulating a human ability but also that it understands the story. And what the machine and its program do explains human ability to understand stories and answer questions about it.

To throw aside the statements above, that is, to discard the idea that SAM – the computer program created by Schank and Abelson [4] – understands the story, Searle launches the argument of the Chinese Room.

And they conclude by stating that if the man cannot understand Chinese, the machine cannot as well.

Schank and Abelson [4: 237] say that 'the limit of comprehension is the limit of the world knowledge'. Taking that into consideration it is possible to say that, just like the child observed during the experiment, anyone can answer questions to the limit of his knowledge or to the limits of the script, plans, goals… which, in a certain moment, we have.

4. References

[1] Margaret A. Boden. "What is artificial intelligence?" In: Artificial intelligence and natural man. New York: The harvest press limited, 1977.

[2] William S. Robinsom. "The Turing Test". In: Computers, minds & robots. Philadelphia: Temple University Press, 1992.

[3] Gregory Robinson and Bridget Riegler. "Cognitive Psychology". Applying the science of the mind. Boston: Pearson, 2004 (2009).

[4] Roger C. Schank; Robert P . Abelson "Scripts, plans, goals and understanding. An inquiry human knowledge structures". New Jersey: Lawrence Eribaum Associates, Inc. Publishers, 1977.

[5] Roger Schank; Alex Kass. "Knowledge representation in people and machines". In: Umberto Eco, Marco Santambrogio, Patuzia Violli (Eds). Meaning and mental representations. Indianpolis: Indiana University Press, 1988.

[6] John Searle. "Can computers think?" In: Minds, brains and science. Penguin book, 1984 (1991).

[7] John Searle. "Minds, brains, and program". In: Margaret Boden. A .The philosophy of artificial intelligence. Oxford: Univ.Press, 1990 (2005).