UNDERSTANDING THE HUMAN MIND

This debate starts with Searle's argument on whether a Human mind is nothing more than a computer program and hence can a computer understand just like a human does by simply running a program which simulates human mind? [1] Searle argues that Computer programs are syntactic, they just do mere symbol manipulation based on some explicit rules and these symbols are some abstract notions, which are meaningless and have no causal properties. However, Human minds have mental contents which are semantically rich. What Searle tries to say is that is that programs just have a syntax which neither contains and nor is sufficient for semantics. He stresses on the point that symbols are formal elements which have no **intrinsic meaning**. He supports this claim by his Chinese room argument where he states that he can process Chinese characters and produce Chinese output by simply looking up the mapping of these symbols in a rulebook, when he doesn't understand even a word of it. Hence a machine that passes the Turing test (say, intelligently conversing with a Chinese person), cannot be claimed to 'understand' Chinese and hence cannot think like a mind does.

Searle strongly refutes strong AI which claims that intelligent machines can be produced which can understand in the same way people do. He also opposed Churchland's support of connectionists who believe that neural nets can simulate brain processes much more accurately due to the presence of parallel processing which is very similar to the working of neurons in our brain. According to him serial and parallel processing were **computationally** the same.

So why is it that the Chinese room argument works against computer systems and not our minds? To which Searle's response is that the computer programs are independent of their hardware implementation whereas our minds are **caused** by our brain's neurobiological processes. He claims that *any system would be capable of producing mind if and only if it had the causal properties equivalent to those of the brain*. Simulation is not the same as Duplication. Duplication of such causal powers is not possible by formal symbol manipulation.

There are some inconsistencies with Searle's arguments. He says that neurons in a neural net just do symbol manipulation but if that is true then how is it that these neurons learn through experience? Is it not akin to a changing rulebook? What if the feedback is incorporated in the rulebook? Consider the Kangaroo Pen example – When a child is shown the kangaroo Pen it identifies it as a kangaroo but then the cap is taken off and the kangaroo turns into a pen. So the child learns this and when the next time it's shown the same stimulus it thinks it to be either a kangaroo or a pen. Similarly, the rules in a rulebook can be updated. Thus in a way the rulebook can learn!

An important question is that what we expect of an intelligent system? Considering language, if a computer could give a reasonable response to every question that we ask it, wouldn't it be intelligent? This line of thinking was similar to Turing Test. Considering the basic question of the discussion, whether computers, when they attempt to understand natural language, should use symbols that have meanings or not.

What is understanding? Is it the ability to reply to certain questions. Example - Cleverbot , or the ability to make judgments, to think and provide creative ideas. The latter, we thought are higher functions that should not be brought into the discussion on the basic concept of 'understanding' a natural language.

Understanding just means giving reply to questions. You need a meaning associated to the symbols in a language, only to interpret what you hear/read. If understanding is the sole purpose, you can as well hard code everything into system!

We also considered that replying doesn't need to happen in the same language. Rather there just need to be a syntactically correct response. In case of humans, we use language to convey a certain sensory perception to another human.

To a computer, natural language would only be text symbols or audio symbols. But then isn't it just about the physical hardware? If a computer was mounted on a robot that had the ability to see, touch, smell, hear and taste, wouldn't in that case, combined with the ability to learn, the computer be able to reproduce a certain state that would correspond to understanding what is written in the text.

Here we were trying to use Harnad's model [2] on an advanced kind of robot. We have reached here an argument which Searle says people use against his theory (listed as f.). He refutes this line of thinking but has not given reasons for doing so.

However, Searle does say that being able to simulate hydrocarbon ignition or the human digestive system on a computer doesn't actually mean we can run a car or digest a pizza. Here we thought whether intelligence shouldn't be related to the hardware present in the system? Just like a system with different senses could understand better in our opinion, so could a system with the requisite hardware perform the tasks that the computer was simulating. A person with a hand cut-off still feels pain in it. So a computer does simulate tasks that have no physical significance until it is provided with that hardware.

In Nature of the Linguistic Sign[4], Perdinand De Saussure discusses the concept of sign. He says if a language is regarded as something which associates things with names then it makes a very big assumption that read-made ideas exist before words and symbols. Rather a linguistic sign is something which combines a sound-image (Signifier) with a concept (Signified). Sound image is the way a person perceives the stimuli. A Linguistic system is a collection of different such sounds (perceptions) associated with different ideas.

So, this means that only associations sanctioned by language exist in our reality and whatever else might be imagined, is ignored?? Discriminations help in categorization for some point of time but eventually you have to associate a linguistic label to the category!

To address the question of lack of intrinsic meaning in symbols Harnad in his paper on 'The Symbol Grounding Problem' [2], [3] tries to model a system that is the hybrid of a symbolic system and a connectionist system. The symbol grounding problem concerns with how the semantic interpretation of a formal symbolic system can be made intrinsic to a system rather than being parasitic or something in our heads. Harnad talks about the different kinds of mental representations – iconic, categorical and symbolic, the first two being intrinsic and non-symbolic. Iconic representation is the sensory projection

of any object while the categorical representation ("warped" transformation) is the selection of certain invariant features from this sensory projection of the object which is learnt over time, which helps us discriminate between objects and place them into different categories. Symbols are the names given to these objects. Now consider the scenario where a person tries to learn Chinese and all he has is a Chinese to Chinese dictionary: This will result in an infinite regression problem! There has to be certain Chinese symbols whose meanings are known to that person. Thus few basic symbols should be given a meaning – **The Symbol Grounding Problem.** According to Harnad's Model these symbols are grounded bottom-up in their non-symbolic cognitive representations: icons and categories.

Deacon(1997) said that only humans can acquire symbols and languages. They have these extra processing abilities due to their enlarged prefrontal cortex. In human languages, symbols are associated with each other by syntactic rules. These associations form words. Only a smaller set of elementary symbols need to be grounded, higher level symbols can inherit semantics from them through the syntactic rules. For example, horse and stripes are two basic categories. Now, a linguistic definition of zebra = horse + stripes can be easily understood and given a categorical representation of a horse with stripes, even when the person has not actually seen a zebra (symbolic theft [2]).

In connectionism, Artificial Neural networks simulate this cognitive model which possesses the ability to discriminate and categorize. Thus these networks can replicate 'warped' transformations (the "Categorical Perception' effect) where members of the same category look more alike and members of different categories look more different. This CP process provides for **discrimination** (iconic, difference in sensory perceptions) and **identification** (to assign a stable category) of objects. The categories can be attained by either direct sensorimotor interaction which is termed as **sensorimotor toil** or indirectly by combining grounded symbols which is termed as **symbolic theft**, for example zebra**.** A hypothesis has been made that symbolic theft is the basis of the adaptive advantages in a language (Harnad, 1996). However, some basic categories must still be learned by toil to avoid an infinite loop in the symbol grounding problem. The experiments done with the computational model of mushroom world supports this claim.

An interesting question is how does an infant ground the symbols. Unlike the experiments conducted on neural nets, the scenario is more complex. Can it be simulated?

We also looked into the paper on "Language is spatial" [6] where the authors claim that a "major part of our language in mapped to certain spatial representations in our minds which are grounded in action and perception". They conducted several experiments which showed that indeed certain image schemas exist for concrete (run, eat etc.) and abstract (love, respect etc.) verbs. The experiments involved two kinds of tasks: 1: forced choice task where they had to choose a suitable representation from given options (with varying primary axes – vertical, horizontal) to represent several verbs, 2: Free form task where the subjects had to make their own representations of sentences using simple structures. Both the tasks had comparable results which showed that there exists a great deal of coherence among the subjects in their underlying imagery representation of the linguistic terms, when presented with the same inputs. Also the orientations of these schemas depend highly on the native language of the

subject. For example, English speakers use horizontal spatial metaphors while describing time while Mandarin speakers use both horizontal and vertical.

REFERENCES:

[1] John R. Searle; Is the Brain's Mind a Computer Program?; In Scientific American, January 1990, pp. 26--31

[2] Angelo Cangelosi, Alberto Greco and Stevan Harnad; Symbol Grounding and the Symbolic Theft Hypothesis; InCangelosi A & Parisi D (Eds) (2002). Simulating the Evolution of Language. London: Springer

[3] Harnad, S.; The Symbol Grounding Problem; (1990); Physica D 42: 335-346

[4] Saussure, Ferdinand de; Nature of the Linguistics Sign; In Charles Bally & Albert Sechehaye (Ed.)(1916), Cours de linguistique générale, McGraw Hill Education. ISBN 0-07-016524-6

[5] Charles Sanders Peirce, 1932: The sign: icon, index, and symbol (excerpt)

[6] Daniel C. Richardson, Michael J. Spivey, Shimon Edelman, Adam J. Naples ; “Language is Spatial”: Experimental Evidence for Image Schemas of Concrete and Abstract Verbs; Department of Psychology, Cornell University, Ithaca, NY 14853 USA