From Low level to High Level Reasoning

Şule Yıldırım

Norwegian University of Science and Technology,
Institutt for datateknikk og informasjonsvitenskap Sem Saelandsvei 7-9 NO-7491 Trondheim, Norway, Phone: +47 73594567, Fax: +47 73594466, email: yildirim@idi.ntnu.no

Abstract

In this paper, we point out the long lasting debate between two main approaches of Artificial Intelligence: Symbolic AI and connectionist AI. The reasoning of the connectionist AI defenders depend on the indication of necessity of neuron modelling for knowledge representation and hence ignoring symbolic AI which fails in doing that. On the other hand, symbolic AI presents better models of representing knowledge, and higher reasoning capabilities. So we will investigate the issue of whether connectionists might be justified in their arguments since symbolic AI has no intent to consider neuronal level knowledge representation and manipulations and fails most of the time when agents it considers are embodied in a real world environment where as connectionist approaches mainly consider embodiment where there are yet no serious signs of how high level reasoning might have emerged from low level functioning of interconnected neurons.

Keywords: High level reasoning, symbolic AI, connectionist AI, integrated systems, brain, cogitive science.

1 Introduction

There's a long term debate that is still going on between symbolists and connectionists in artificial intelligence about representation and manipulation of knowledge in brain as far as the high level cognitive processes such as memory, thinking, problem solving, information retrieval etc. are concerned.

The symbolic approach assumes the presence of data structures in brain in analogy to the data structures used to store and manipulate knowledge in computer memory. Again algorithms in analogy to the computational algorithms are assumed to manipulate the knowledge in brain or to solve problems.

On the other hand, connectionist approaches reject symbolic approaches since symbolic approaches do not consider the neuronal level of knowledge representation in the brain. Furthermore they claim that with the use of neuronal level knowledge representations, the brain can no longer be considered to use algorithms to solve the problems as computers do. Their claim emerges from the point that neuronal level problem solving is totally different from algorithmic problem solutions and the brain functions with its neurons after all. Hence, connectionist theories model thinking using artificial neural networks.

The aim of this paper is to investigate the point whether connectionists can really be justified in their argument and whether brain research still has benefits to take from symbolic approaches.

2 Interesting Questions on Information Storage and Ordering of Actions

1956, George Miller summarized numerous studies which showed that the capacity of human thinking is limited, with short-term memory, for example, limited to around seven items. He proposed that memory limitations can be overcome by recoding information into chunks, mental representations that require mental procedures for encoding and decoding the information [1]. Cognitive theorists have proposed that the mind contains such mental representations as logical propositions, rules, concepts, images, and analogies, and that it uses mental procedures such as deduction, search, matching, rotating, and retrieval. The dominant mind-computer analogy in cognitive science has taken on a novel twist from the use of another analog, the brain. Connectionists have proposed novel ideas about representation and computation that use neurons and their connections as inspirations for data structures, and neuron firing and spreading activation as inspirations for algorithms [1].

Critics of cognitive science have offered such challenges as:

1. The emotion challenge: Cognitive science neglects the important role of emotions in human thinking.
2. The consciousness challenge: Cognitive science ignores the importance of consciousness in human thinking.
3. The world challenge: Cognitive science disregards the significant role of physical environments in human thinking.
4. The social challenge: Human thought is inherently social in ways that cognitive science ignores.
5. The dynamical systems challenge: The mind is a dynamical system, not a computational system.
6. The mathematics challenge: Mathematical results show that human thinking cannot be computational in the standard sense, so the brain must operate differently, perhaps as a quantum computer.

However, now we will discuss why these challenges against cognitive science or symbolic approaches might not be enough to abondon symbolic approaches.

So we start with a set of interesting questions to answer about the functioning or the high level capabilites of human mind:

- What are components of intelligence?
- Is hybrid intelligence (integration of symbolic and connectionist approaches to benefit from the advantageous sides of each) necessary/possible?
- How are symbols kept in memory physically?
- How are words (combined symbols) kept in memory?
- What's the relation between symbols, and the neural nets of the brain? Are they just inputs/outputs or do they appear in middle layers?
- What are the elements of mind?
- Is language learning able to say much more than other fields could do?

Roughly speaking a human is in interaction with its environment using its vision, sense, smell, taste sensors and he can observe both other objects and affects of its actions in the environment as well as he can observe himself (self awereness). He has high level reasoning capabilities which he uses language to communicate them to the others in his or distant environments.

On the other hand, following are the observations we have about how the mind might work:

- Symbols are kept in memory.
- They are manipulated in memory.
- Brain can make/think plans.
- Plans are a serious of action descriptions.
  An example plan:
  - Walk to the room.
  - Go in through the doorway.
  - Close the door.
  - Walk to the computer.
  - Start talking.
- Plans are kept in memory also.

The rest of the observations are as follows:
The steps in the above example plan have an ordering and the brain produced that ordering to solve the task of ”giving a talk while standing next to the computer”. When one actually wants to execute the steps in a plan, he can retrieve them and then execute them and this is the reason we think plans are kept in memory. The content of each step in a plan is a description of a behavior and is modeling that behavior however, the step itself resides in the brain and not just in the interface one has with other people, in the words one utters and on a paper that he puts them.

What about the times when there was no language and human beings caught animals, fed themselves with caught animals' meat and with plants, drank water from rivers, caught fish in the sea and felt pain when poisoned? Did they keep symbols and description of their actions in the brain then?

They should have. When one got hungry, he probably had the picture of a fish flashing in his mind suddenly.

If one was poisoned from water yesterday, he would flash appearance of his hand and arm movements in pain in his mind.

So if human beings were not able to form a verbal or a sign language to communicate with each other, would the brain ever evolve to hold symbols, plans etc.? We think the answer is “yes”. But the representation in the brain would probably be different.

How does the brain generate an ordering of the steps in a plan? Why doesn't ”go in” step come before the ”walk to the door” step in the above plan to solve the given task?

Should there be an algorithm in analogy to a computational algorithm in the brain that does the ordering? If there is, the algorithm certainly is embedded in the functioning of the physical network of neurons in the brain. Sometimes any two neuron connection in a brain neural net is intense and sometimes it is lose, depending on the task in the hand. The outputs of the algorithm (algorithm = the topology of the neurons in the brain and the strength of any two connected neurons) for given inputs are also kept in the brain in the area called memory and what is being kept are the names of people, names of objects, relation of an object to another one, one's impressions about people and objects, the plans one generates, etc. which are defined to be memory contents.

How does the memory store its contents? What kind of topology is the memory composed of to store a single symbol for example? Is there a single topology part of a bigger topology to represent a single symbol? When one talks about a red book for example, is there a topology for storing "red" and a topology for storing "book" separately? If so, what is the mechanism to relate these two topologies with each other or how do these separate topologies work with each other? Or is there a single topology to store everything? If not, what kind of topology lets one attribute color red to the book he is reading and at the same time color red to the sweater he is wearing and how does one store those concepts?

What kind of topology does the memory have for storing/keeping plans and language rules? The forehead of the brain is found out to be the part of the brain to do planning. Is it a single neural net topology in the forehead to do planning or is it a combination of them?

Could that algorithm to do planning be an implementation of a decision tree search (topic of symbolic approach) by the neural net(s) in the brain **in its (their) own way** and hence the outcome actions have an ordering as in a plan?

Does the algorithm itself have an orderly way of learning rules (i.e. language rules), learning algorithms and for planning? Does it make searches to solve a problem, does it use induction/deduction or does it decompose a problem and handle each decomposed part immediately or in time to solve the problem in an orderly manner?

Or does order come out of disorder? That is, do the brain processes (problem solving, planning, thinking, memory etc.) which we feel concretely to happen in our brains emerge as an order out of neurons interacting locally using simple rules ? [2]

After all these questions comes one of the main questions: How does the neuronal mapping from high level reasoning to low level neuronal activity occur? How do neurons map to/implement the algorithm(s) that do abstract planning, reasoning, thinking and form abstract memory? Do those seem to present abstract algorithms (higher level/complex order) emerge from random local interactions leading to global order such as in the world of ants or not? One can intuitively say yes or no also looking at what happens in other living systems, but does this answer the question?

3 Possible Clues
It's important to get clues from the ongoing and previously done research in response to the question of which approach in AI is more likely to understand the functioning of human brain and whether an integrated system is possible/necessary to overcome the drawbacks of each approach.

One of the most investigated approaches is the integration of planning and reactivity for producing artificially intelligent agents [3]. There are different studies with this approach where a high level planner and a low level executer appear in different combinations together. In these schemes, elementary behaviors are implemented by a reactive controller [3] where as more complex behaviors can be produced by sets of low level control structures. The more complex behaviors are generated by a planner first and then they are compiled into simple behaviors (low level control structures). However, these integration studies are mainly for producing "artificially" intelligent agents. Their aim doesn't go beyond obtaining machines that can act intelligently although these machines are expected to mimic the functioning of brain and show brain like reasoning. However, in brain, low level is **not only** executing what is received from high level but **it generates/causes high level itself**.

So, in brain, planning algorithm itself (if it exits) is generated by what it is executed.

On the other hand in [4] Damasio proposes that human reasoning depends on several brain systems, working in concert across many levels of neuronal organization, rather on a single brain center. Both "high level" and "low-level" brain regions, from the prefrontal cortices to the hypothalamus and brain stem, cooperate in the making of reason. He also suggests that the lower levels in the neural edifice of reason are the same ones that regulate the processing of emotions and feelings, along with the body functions necessary for an organism's survival. In turn, these lower levels maintain direct and mutual relationships with virtually every bodily organ, thus placing body directly within the chain of operations that generate the highest reaches of reasoning, decision making, and, by extension, social behavior and creativity. Emotion, feeling, and biological regulation all play a role in human reason. The lowly orders of our organism are in the loop of high reason.

What [4] states gives a clear indication of there might be brain regions in brain busy with high level and low level functioning each of which might correspond to symbolic and connectionist approaches respectively.

Although some scientists believe that neuronal connections are in the form of every neuron having a connection with every other neuron, and that mind and behavior probably emerge from that messy connectivity in ways that no science will never reveal, [4] considers this thought to be wrong with the following statement:

> "On the average, every neuron forms about 1000 synapses, although some can have as many as 5000 or 6000. This may seem a high number, but when we consider that there are more than 10 billion neurons and more than 10 trillion synapses, we

realize that each neuron is nothing if not modestly connected. Pick a few neurons in the cortex or in nuclei, randomly or according to your anatomical preferences, and you will find that each neuron talks to a few others but never to most or all of the others."

Could this statement be a signal of brain processes having an order in execution in a sense of the order a computational algorithm has in execution – one step executes after the other? (Parallel computational algorithms also have an ordering in the sense that a group of steps executed in parallel are followed by a single or another group of steps to be executed in sequence or in parallel). Also [4] mentions about stages of reasoning which might easily correspond to step(s) of an algorithm.

In [5] Sobel defines the way the two disciplines (Symbolic vs. Connectionist AI) view knowledge or information with: "AI attempts to capture intelligent behavior without regard to the underlying mechanisms producing the behavior. This approach involves describing behaviors, usually with rules and symbols. In contrast, neural networks do not describe behaviors; they imitate them."

In [6] attention is drawn to the point that limitations of present-day machine intelligence stem largely from seeking "unified theories," or trying to repair the deficiencies of theoretically neat, but conceptually impoverished ideological positions. It is then mentioned that purely numerical connectionist networks are inherently deficient in abilities to reason well whereas purely symbolic logical systems are inherently deficient in abilities to represent the all-important "heuristic connections" between things - the uncertain, approximate, and analogical linkages that are needed for making new hypotheses.

Minsky also argues that yet symbolic AI has told us a little about how to solve problems by using methods that resemble reasoning in [6]. He goes on with saying: "**If we understood more about this, perhaps we could more easily work down toward finding out how brain cells do such things**." He points out that connectionist approaches have told us a little about the workings of brain cells and their connections. He states: "**More research on this might help us discover how the activities of brain-cell networks support our higher level processes**." He suggests the solution to be in doing more research on **how to combine** both types of ideas.

As a good example of extending connectionist approaches to have better high level reasoning capabilities, in [7], Meeden et. al. did experiments on robots which can exploit intelligent plan like behavior with a connectionist approach. In their experiments, there's a light source in a corner of a workspace as shown in Figure 1 and a robot in the workspace is given the task of approaching or moving away from a light source, one task following another task. Although the robot is not programmed in an explicit way as in a planning algorithm and uses an artificial neural net trained for the seek and avoid tasks, the observed behaviors of the robot to accomplish each of the task represent a behavior hierarchy which can be summarized as in Figure 2.

The behaviors in Figure 2 are not explicitly programmed but they emerge from the functioning of the neural net trained for seek/avoid behaviors. Although this work is a step towards increasing the representational and planning capabilities of artificial neural networks, it is still a very little portion of what humans can achieve with their high level reasoning capabilities.

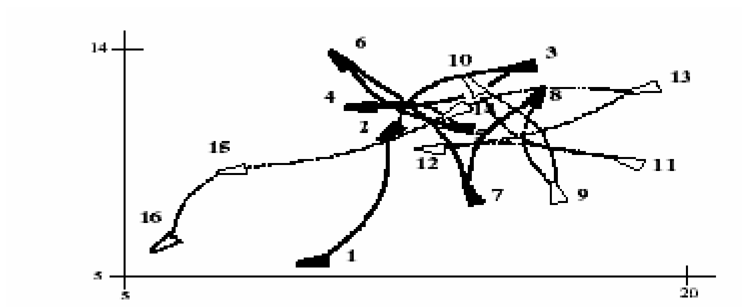

Figure 1. Path of a simulated robot through the playpen (units arc inches). The light is located at the origin. The direction of the arrows indicate robot's current heading. The numbers on the path refer to steps in a sequence of motion. 1-8 occurred during avoid-mode, 9-18 occurred during seek-mode. Note that it has satisfied its goals at steps 8 and 16.

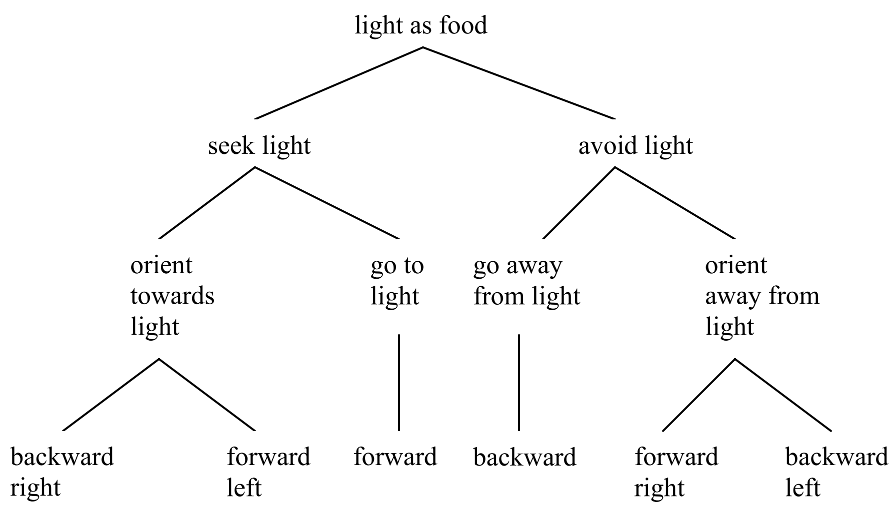

Figure 2. A hierarchical view of robot's behavior.

Finally, if one is given the task of summing 232 and 215 as in Figure 3 using the given simpler rules of summation on the left then does one use an algorithm as in analogy to a computational summation algorithm that uses the simpler rules of summation to do a more complex summation? Are these three rules used in an orderly manner as we would do in a computational algorithm? If complex global behavior can "emerge" naturally from collections of agents (here neurons) subject only to simple local interactions without the need for a high-level global controller, that is without the presence of a high level explicit summation algorithm in this example, then how are simpler rules of summation incorporated in the computation of the more complex summation operation?

```
2 + 2 = 4      232
3 + 1 = 4      215
5 + 2 = 7     +--------
               447
```

Figure 3. How is summation above done in the brain – by an algorithm as in analogy to a computational algorithm or by a network of neurons emerged to do the task?

The final and for centuries asked question is: How does brain work?

4 Possible Ways of Answering Questions

Pick up a sub part(s) of the main problem to gain a better insight of how the solution should be shaped for the main problem. A possible sub part of the problem is memory and possibly answering questions of: How could memory be keeping and remembering stimuli(s) that have occurred in the past but need(s) to be remembered in future, possibly near future? How could such a memory be achieved?

Delayed response tasks are a standard way of investigating short-term memory (STM). The agent is typically assumed to **'remember' in some way** the necessary information about the stimulus (for example the side on which a stimulus appeared) during the delay period. Related experiments are carried on rats or with robots programmed to simulate the behaviors of rats in T-mazes as in Figure 4 where rat placed in a T-maze (or a robot) starts its motion from the bottom of a T-maze and it meets a light source on the left or on the right before it reaches the junction. When it reaches the junction, it takes a right or a left turn remembering which side of the corridor before the junction it saw a light source. The studies up to now are still far away from suggesting a specific model of how animals solve delayed response tasks [8].

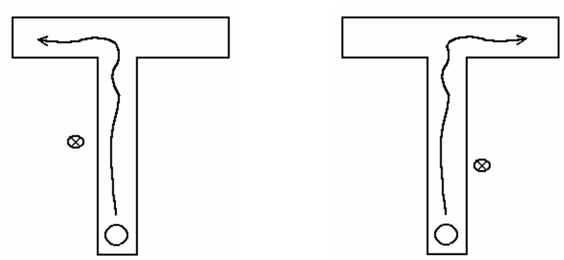

Figure 4. The two situations in the simple T-maze environment, adapted from [9].

From an observer's point of view it is relatively easy to attribute some kind of "representation" to an animal or a robot exhibiting the correct behavior as in [7], but the detailed analysis of studies can help to illuminate the actual mechanisms underlying that behavior [8].

Most of the studies used standard recurrent artificial neural nets in which certain neuron activation values are fed back and used as extra inputs to some of the neurons in a later time step (typically the next one). In this type of network, the synaptic connection weights are usually considered **long-term memory** since they are changed only by the training process, whereas the feedback activation values, which can change from moment to moment, is commonly considered to constitute **short-term memory**. [10] points out in review of computational models of working memory that memory-related activity has been observed in several brain areas, in particular the prefrontal cortex (PFC) which is “the brain structure most closely linked to working memory”. STM has been found in both single-neuron recordings in nonhuman primates and in human brain imaging studies [11, 12]. On the other hand, the role of neuromodulators such as dopamine is not understood yet. However, it has been observed that dopaminergic activity, which can affect synaptic currents increases during working memory tasks [13].

We have ongoing research in our department on short term memory and an extension to it: “counting”. If an agent in a corridor shown as in Figure 5 can be trained to make a turning to the correct corridor when given the corridor number then it can said to be able to count. As an example, when corridor number 3 is given, the agent should start its journey from a random position in the beginning of the main corridor and take a right turning to the third corridor. The entrance of each corridor is indicated by a zone sensor. The first results on this study and the artificial neural nets used for the implementation are given in [14].

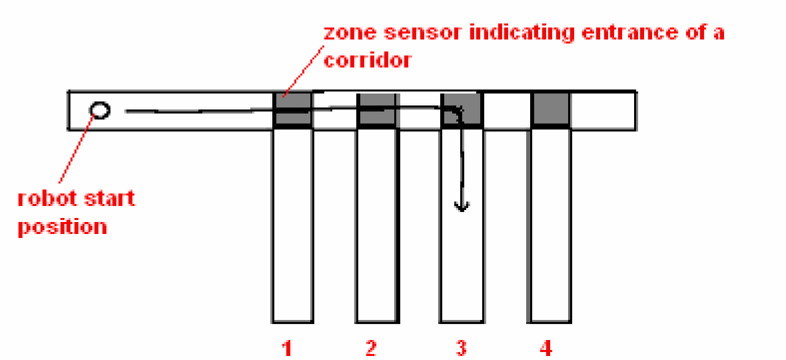

Figure 5. The robot can take the indicated path to make a turning to the 3rd corridor. Right after this task, the robot can be asked to make a turning to any other corridor.

Also, single neuron recordings and brain imaging techniques, focus on the measurement of neuronal activity, whereas synaptic changes are more difficult to monitor. The role of studying plasticity in STM might contribute to neuro-scientific and/or cognitive scientific theories and models of the corresponding biological mechanisms. In addition to the short term memory studies [15], studies on internal simulation of perception where controls are presented an environment and are asked to walk around the environment by remembering what they have seen to avoid obstacles are other approaches for studying memory and hence getting a step ahead on how the mind works [16].

5 Conclusion

In contrary to the debate put forward by connectionists against the necessity of symbolic methods in brain research, we present the points in this paper why this argument might not be valid. Since none of the approaches has a valid model that explains functioning of brain yet, it’s also not possible to prefer an approach over the other although connectionist approaches put effort in modeling brain functions in a way the brain itself does: with neurons. However, there is not yet any evidence that any of these models actually represent brain functions. Although brain is thought to emerge its functioning, there is again no evidence that brain does not function in a manner a computational algorithm does.

References

[1] Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/cognitive-science).
[2] S. Johnson, Emergence: The Connected Lives Of Ants, Brains, Cities And Software, Simon & Schuster, New York. (2001)
[3] A. Saffiotti. Some Notes on the Integration of Planning and Reactivity in Autonomous Mobile Robots., AAAI Spring Symposium, Foundations of Planning, pp. 122-126, Stanford, CA, USA, March. (1993)
[4] A. R. Damasio, Descartes Error: Emotion, Reason and the Human Brain. (1994)
[5] C. P. Sobel , The Cognitive Sciences, An interdisciplinary Approach. (2001)
[6] M. Minsky, Logical vs. Analogical or Symbolic vs. Connectionist or Neat vs. Scruffy, MIT press. (1990)
[7] L. Meeden, G. McGraw and D. Blank, Emergent Control And Planning In An Autonomous Vehicle, Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 735-740. (1993)
[8] T. Ziemke and M. Thieme, Neuromodulation of Reactive Sensorimotor Mappings as a Short-Term Memory Mechanism in Delayed Response Tasks, Adaptive Behavior, 10 (3/4), pp. 185-199. (2002)
[9] C. Ulbricht, Handling Time-Warped Sequences with Neural Networks, Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, SAB96, September 9th-13th, Cape Cod, Massachusetts. (1996)
[10] D. Durstewitz, J. K. Seamans and T. J. Sejnowski, Neurocomputational Models of Working Memory **,** Nature Neuroscience**,** 3(Supp), pp. 1184 – 1191, November 2000.
[11] A. Baddeley, Recent Developments In Working Memory, Current Opinion in Neurobiology, 8(2), pp. 234-238. (1998)
[12] Synaptic reverberation underlying mnemonic persistent activity. X. Wang. Trends in Neurosciences, 24(8), pp. 455 – 463.
[13] T. Ljunberg, P. Apicella and W. Schultz, Responses Of Monkey Dopamine Neurons During Learning Of Behavioral Reactions, Journal of Neurophysiology, 67(1), pp. 145-163. (1992)
[14] E. A. Buskum, S. Gervasi, R. Gross, B. De Vylder, S. Yildirim, Evolving a Robot Neurocontroller For Memorization Tasks, EVONET Summer School, Parma, ITALY. (2003)
[15] D. Jirenhed, G. Hesslow and T. Ziemke, Exploring Internal Simulation of Perception in Mobile Robots., Fourth European Workshop on Advanced Mobile Robots - Proceedings, Lund University Cognitive Studies, 86, Lund, Sweden. (2001)
[16] Hesslow G., Conscious Thought as Simulation of Behavior and Perception, Trends in Cognitive Science, 1(6/6), pp.242-247. (2002)