Mindware

Log of discussions about
the philosophy of cognitive science

David Volquartz Lebech

January-March, 2009

Contents

1 Introduction **2**

2 Symbol systems **2**

3 Connectionism **4**

4 Complex interactions **5**

5 Cognitive Technology **7**

6 Conclusion **8**

1 Introduction

This log is intended to put together a summary of the thoughts that have arisen from the discussions with professor Arthur Farley (Art) at the University of Oregon about the book *Mindware* (2001), written by Andy Clark. The log is intended as a reference for later use to give an overview of these discussions and of the very complex philosophy behind cognitive science.

The major hypothesis throughout a great deal of the book is the question whether the brain and its inner workings has an analogy in computer hardware/software. Is the mind (our thoughts, beliefs, desires and so on) actually *mindware* running on a *meat machine* or (more eloquently) *wetware* (our brain with its neurons and synapses). In other words: Does mindware = software?

Mindware gives a very dense and good overview over several topics and opinions in cognitive science. Since the above question is still unknown there are many different points of view on the question and many of them are quite philosophical in their nature and not scientifically based. Throughout the chapters of the book, the complexity gradually heightens as more aspects are considered. In the end, nothing is really conclusive but at least it seems that the starting point with the question about mindware=software shows to be overly simplified.

The following sections follow the book in the way that I will first describe the basic ideas of a topic and then include some of the points from the discussions with Art. The sections will not be structured according to any specific chapters in the book though, but instead will be based on the ideas that I feel are most important. As mentioned, the book is very, very dense with information, especially because the topics are 100% new to me. I will try to convey my thoughts as best as I can.

2 Symbol systems

A lot of the philosophical proposals really boils down to whether or not the function of the brain can be seen as having some kind of inner symbol system to represent and cope with the world. Some philosophers, in particular Newell and Simon, advocate for this by formalizing what they call a *physical symbol system* (PSS) which consist of symbols and processes that can operate these symbols, located in a world that it can interact with (i.e. the real world). Newell and Simon claim, that a PSS is sufficient for intelligence (page 28). By saying this, they automatically say that they believe that mindware=software.

A famous argument against this is the *Chinese Room* thought experiment (page 34) which is proposed by one of the major opponents to the symbol system idea, John Searle. What the Chinese Room tries to show is that having a pure symbol crunching system does not lead to anything remotely intelligent. Another worry is raised concerning coping in everyday life. The question is how would a symbols system cope with certain situations. One (funny) example is: What if a Martian suddenly was in our kitchen (page 37)?

Other philosophers develop further on the idea of the mind as having *some kind of* inner symbol representations by describing *folk psychology* or commonsense psychology (page 45). Folk psychology tries to explain actual behavior in the real world rather than inner workings of the brain but uses this as explanation for why the idea of inner symbols is mainly true. For example, I might have the desire to drink water, because I am thirsty, and therefore I must have some kind of inner state transition "thirsty" $\rightarrow$ "drink water". Dennett dubs this the *intentional stance* (page 47) which is a concept that is used to explain daily behavior for well-designed beings where well-designed means intentional and purposeful. Ironically, the attribute can thus be ascribed to a quite large number of objects, including mechanical objects like a dishwasher. It should be noted, however, that Dennett does not claim that having the intentional stance necessarily requires an inner, organized symbol system (page 51).

The idea about folk psychology is believed to be true in the eyes of Fodor and partly Dennett but not Churchland who thinks that folk psychology is unreal.

Discussion

In our discussions, the above ideas have emerged again and again. Maybe this is because the thought of having a symbol system representing our mind seems very tangible for us. It is difficult to imagine, that our brain should be something "ghostly" and unexplainable (as hinted on page 8 and 43), especially because science have historically offered answers to almost all questions about our physical world.

Art mentioned in several of our discussions some recent research where fMRI brain scans showed that individual people's brain patterns were similar when presented with fairly abstract concepts as a hammer or a house. This is indeed intriguing and a good indicator that at least some kind of common framework is shared between humans.

As a scientific theory, it seems, symbol systems are good starting points for our discussions of the mind. But on the other hand, in my opinion, the idea seems to be overly restrictive and it does not take e.g. neurostientific research into account. Unless we insist that neurons and synapses are actually symbol carriers then the hardcore symbol cruncher theory, *strong AI* or *computationalism* (what ever we want to call it) do not seem to be very well grounded in the real world.

3 Connectionism

Chapter 4 is devoted to explaining *connectionism* which is the general theory and idea behind *Artificial Neural Network* (ANN) taught in many AI and Machine Learning classes. What makes the connectionist model appealing is the obvious and very direct analogy to the human neurons. The connectionist model describes a distributed and parallel system consisting of many units (or neurons) that can be connected to each other in various ways. As such, connectionism seems to be a good approximation to the parallel workings of the brain's neurons.

Connectionism is thus in great contrast with symbol systems that can only be in a single state at the time. Also, the connectionist system works not by having declarative logic statements but rather depends on weights and a certain architecture (page 67). This also makes the connectionist system lack a precise inner structure – a fact that is used to argue, that connectionism lacks the ability to represent human thought because human thought is supposedly systematic (page 76).

Discussion

The main focus of our discussions concerning connectionism was discussing whether or not connectionism is a theory entirely on its own or is in fact just a special kind of symbol system. It can be argued that every unit in the connectionist model actually corresponds to a self-contained symbol system, complete with input, computation and output. The connectionist system thus merely becomes a distributed set of symbol systems. However, I think that the systematicity argument in page 76-79 refutes this claim and connectionism and symbol systems thus seems related but different.

Accepting the argument of systematicity against connectionism seems to be (in my opinion and also hinted in page 79) an acceptance of the thought that the brain is indeed well-structured in the strong AI sense. Following the discussion from the previous section, this idea about symbol systems is not scientifically vindicated and I do not think that the argument works against connectionism. I think that a more striking argument is that the connectionist unit does not really follow the biological realities of neurons in the brain which was the original idea of the connectionism. Even if the connectionist model is only an approximation it does come with its problems (page 79-81).

4 Complex interactions

One of the drawbacks of the models described above is that they focus mainly on the mind and brain itself. Most of the latest research in cognitive science, however, seem to focus more on complex interactions with the world and environment. Summing it up like this is a definitely over-generalizing but nevertheless, I think it captures the overall picture described in chapters 5-7.

One example is the relationship between perception and action. In the symbol system world, perception is used only as input to an underlaying engine that computes the input and produces an output/action. However, perception and action goes hand in hand in the development of certain human processes (page 88 and 92). One example is distinguishing a certain creature from the ground which is helped by moving the head and body, i.e. action and perception are interactive elements.

The interaction idea is more notable in the research field of *artificial life*. Artificial life looks at whole systems instead of specific parts to study the complex interactions that take place in the simpler subparts. An example is the study of flocking of birds. It is shown that following very simple rules for each individual bird, the group of birds as a whole still are able to function as a whole. No centralized leader is required and the point is, that complex behavior does not necessarily need advanced reasoning to work well (page 107-109).

A whole field of cognitive science research called *Dynamic systems theory* or simply dynamics is also concerned with complex interactions as the name implies. One of the major new ideas of dynamics, compared to the ones described earlier, is that all and everything seem to be connected or “intermingled” across boundaries that seem unintuitive and also, the major proponents for dynamics totally deny the computationalism and symbol systems ideas. The dynamicists claim, that the idea of symbol systems limit our capability to see e.g. perception as a complicated system (page 130). To be even more bold, it is claimed that the Watt governor says more about the mind than both SOAR (a physical symbol system) and NETtalk (a connectionist system). This is the belief of e.g. Pollack and van Gelder.

However, later it is pointed out that some philosophers compromise and say that there might be a domain in between the two extremes and that they may not be mutually exclusive. Clark dubs this domain as *partial programs* (page 133) or *dynamic computationalism* (page 135) which are defined differently but has the same idea about dynamics and some kind of symbol system in unity.

Discussion

In our discussions about complex interactions, a major topic has been evolution. Evolution is an important piece of the whole puzzle, I believe, and a very difficult topic to discuss. One of the problems with symbol systems is that they seem to work too well for our purposes. Indeed, they are thought out and designed by humans. However, evolution is not restricted by its own mind. In the text, it is said that evolution works by "tinkering" (page 86) with the current "device" (i.e. the human). The lungs of humans are an example of this, as they are modified from fish swim bladders (page 86), an approach that a human designer probably would not have thought of. In contrast, human models of just about anything usually have very clear design and specifications. How then, can we understand evolution of the human mind when it is (1) constantly changing and (2) evolving in a fundamentally different way from the way we seem to handle problem solving?

Another question that arose when discussing evolution was whether or not human research in robotics (and gene manipulation and neuroscience an so on) is actually directly influencing evolution or just fighting a battle that will inevitably be won by the powers of evolution. To develop further on this thought: How is it that when evolution has created us, our mind and our thoughts that we can really think that we will ever be able to understand ourselves to the extent that we are able to reproduce our behavior in robots? It would, in my opinion, require a certain type of human made evolution, or at least tinkering with evolution itself, to use the same phrase. I do not think that this can happen with a snap of the fingers. In other words, I do not think that one day, an intelligent robot will emerge seemingly out of nowhere. I rather think that humans will start blending with robot technology, a thought that is further discussed in the last section of this small text.

5 Cognitive Technology

To round off the book, Clark presents his own ideas and current research (and he is self-aware of his bias). The main concept that is discussed here is what Clark calls *cognitive technology*, an idea that seem to situate itself in between the previous discussed theories of dynamics and symbol systems and borrowing from other ideas as well.

A cognitive technology is basically anything in our everyday life that we interact with. As a very mind-bugging example, Clark mentions the process of creative writing. Not only is this process dependent on our ability of abstract thinking and intelligence, it is also highly dependent on the props we use to carry out our task. That is: Pen, paper, computer, what ever. Clark points out, that the writing process is in fact an interaction between paper and person. The paper is thus a cognitive technology and, Clark states, is crucial for the creative writing process (page 142 and 154).

A very important point is made about language as being a cognitive technology. Clark argues that the reason we are able to have thoughts about thoughts (i.e. meta-thinking) is because we are able to treat a thought as an object. This is possible because we have developed a symbol language and that this is, according to Clark, probably our most important cognitive technology (page 145 and 147). Meta-thinking is actually something most theories have difficulty explaining but in the setting of language as a cognitive technology it seems to make sense.

Discussion

Cognitive technology is at one time an exciting and interesting explanation for human mindfulness but it also paints a rather disturbing image of our future. In the final discussion with Art, this was the main topic. As mentioned in the previous discussion section, I think that to construct intelligence (primarily in the form of intelligent agents), we have to tinker with evolution ourselves. Perhaps, however, if we believe in the theory about cognitive technology, evolution will automatically shape us, as it has done so far, alongside our cognitive technologies.

Today, we are becoming increasingly dependent on computers and the internet. At the same time, people get prosthetic eyes and limbs. These are all cognitive technologies and in the extreme case we can see this as a kind of symbiotic relationship (page 152). The question is then, if this increase in reliability of technology will actually turn us into – Cyborgs.

The above suggestion sounds like pure science fiction but I do not think that this development will happen overnight either. We discussed that it will happen in small increments. Consider, for example, the road sign. People are very reliant on the roadsigns to find their way around town. The next step might be, that everyone have GPS receivers and cannot figure out a way without. And then finally, what if everyone had built-in GPS receivers connected directly to their brains. Would evolution then render us incapable of simple path-finding to the extent that walking out of our house requires help from a mechanical unit hooked up to the brain? And will this evolution, if it happens like this, actually be the evolution that finally develops a kind of intelligent artificial life? I leave this questions open for concern.

6 Conclusion

One of the returning problems with correctly classifying the brain is the lack of very conclusive scientific research. Also, it might be the case that researchers are to focused in their specific field such as AI, neuroscience or cognitive psychology. As Clark points out in his own conclusion, the future of cognitive science has to rely on collaborative efforts in multiple fields. This is definitely on the rise e.g. in the field of robotics where large multidisciplinary research efforts are increasing.

Andy Clark's *Mindware* gives a very good overview about the basic philosophy of cognitive science. I see it as mainly a historic overview because the "recent" research in the book is now already 8 or more years old. But nevertheless, it is a good introduction to the field and it has definitely fueled my interest for further research into the topic.