Dreyfus and the Dynamical Approach to Cognitive Science

Gert-Jan Lokhorst

Dreyfus Symposium,
Rotterdam, November 7, 1998

What I would have liked to talk about

Heidegger's brain

(talk by prof. dr. J. P. W. F. Lakke (Groningen), Dutch neurological association, November 6, 1998)

What I am going to talk about instead

Dreyfus and the dynamical approach to cognitive science

Three approaches in cognitive science

- Symbolism
- Connectionism
- Dynamicism

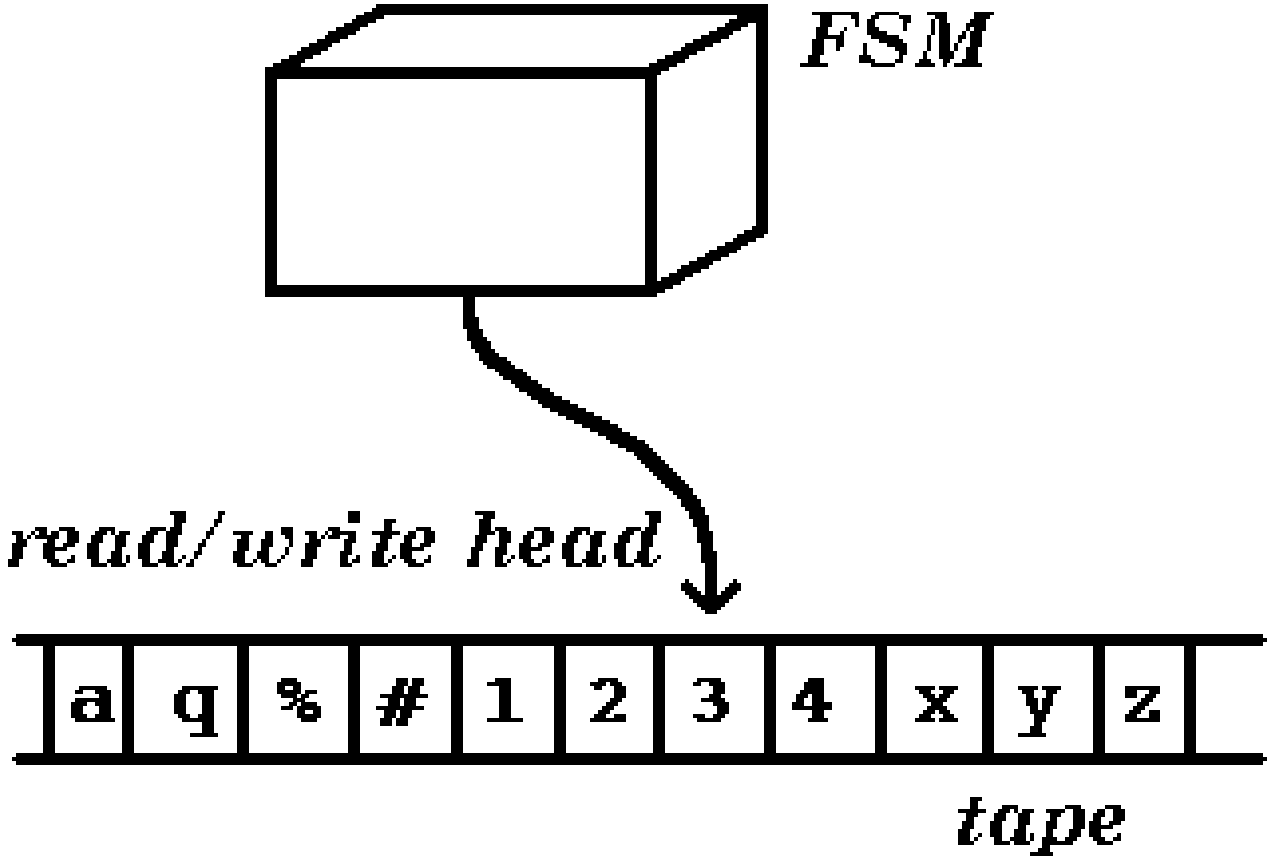
1 Symbolism

Physical symbol system hypothesis (Newell & Simon).

Paradigmatic example of a symbol processor: Turing machine (1936).

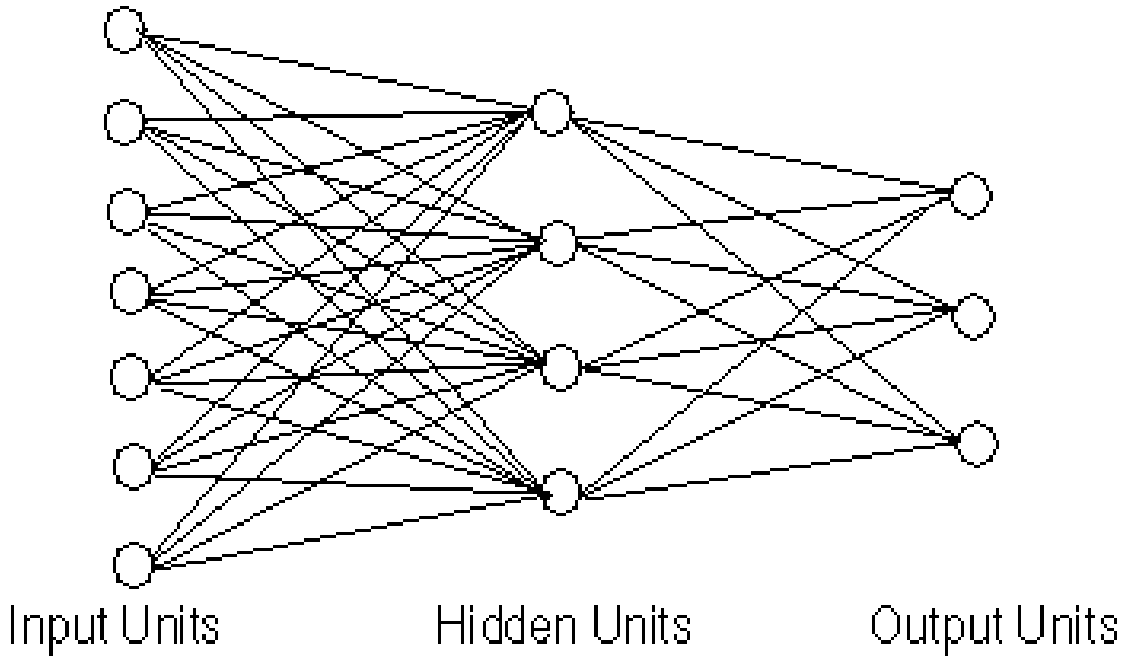
Paradigmatic example of a physical symbol processor: digital computer.

A Turing machine

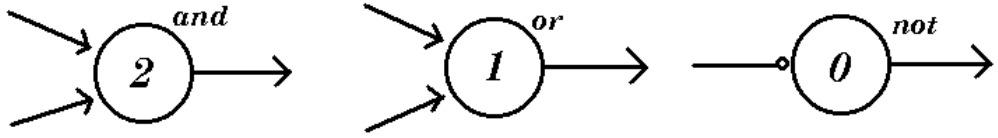


2 Connectionism

Theory of neural networks. A simple example:



- First neural nets: McCulloch & Pitts 1943. These consisted of simple "logical neurons":



- First one who attempted to make neural nets tolerant to noise: John von Neumann (50s).
- First one who tried to invent learning algorithms for neural nets: Alan Turing (around 1950). He failed, and went on to create the field of Artificial Life.
- First one who succeeded: Marvin Minsky, PhD thesis of 1954.
- Neural nets became very popular in the 1980s, after the invention of the backpropagation learning algorithm for multi-layered feedforward networks.

3 How are symbolism and connectionism related?

The "Bible of Connectionism", the famous PDP volumes by McClelland, Rumelhart and others, appeared in 1986. In the same year, the Dreyfus brothers published their book *Mind Over Machine*, in which they were highly critical of AI research as it had been practised until then. Hubert Dreyfus later wrote as follows about this coïncidence:

> At this point, like Charlie Chaplin in *Modern Times* emerging from a manhole with a red flag just as the revolutionaries came swarming by, we happily found ourselves surrounded by the rapidly growing ranks of neural-network modelers (*What Computers Still Can't Do* (1992), p. xiii).

Similar militant talk about "revolutions" and so on is to be found in the Dreyfus brothers' "Making a mind versus modeling the brain: AI at a crossroads" (Daedalus, 1988).

This might make one think that symbolism and connectionism are somehow antagonistic to each other. However, they are not.

First, researchers such as Turing, Von Neumann and Minsky were—as we have seen—pioneers in *both* fields.

Secondly, leading researchers such as Minsky and Rumelhart have never seen their enterprises as conflicting. As Minsky wrote:

> Why is there so much excitement about Neural Networks today, and how is this related to research on Artificial Intelligence? Much has been said, in the popular press, as though these were conflicting activities. This seems exceedingly strange to me, because both are parts of the very

same enterprise. What caused this misconception?

Likewise, Rumelhart, the connectionist, still considers his work as part of the more general enterprise of AI. He regards the "AI is dead" talk which arose just after the publication of the PDP volumes as mistaken.

Thirdly, there are nice mathematical results which relate both fields to one another.

Computational powers of analog recurrent neural networks
(Hava Siegelmann, PhD thesis 1993)

connectionistic	symbolic
nets with integer weights	finite automata (Kleene 1956)
nets with rational weights	Turing machines
nets with real weights	Turing machines with oracles (*)

(*) First studied by Turing in his PhD thesis of 1939.

So for each model in one class there is an equivalent model in the other class. There is no antagonism at this level.

4 Dreyfus about the Turing machine

The neurophysiologically most adequate models of brain activity which exist today are the so-called "third generation" neural network models with **spiking neurons**. These networks are equivalent with analog recurrent networks with **real weights**, which are in turn equivalent with Turing machines with **oracles**. The latter machines are strictly more powerful than ordinary Turing machines. This suggests that man may conceivably be a **super-Turing machine**!

Now contrast this with the following claims made by Dreyfus.

1 *Any* process which can be formalised so that it can be represented as a series of instructions for the manipulation of discrete elements, can, at least in principle, be reproduced by [a universal Turing machine] (*What Computers Still Can't Do* (1992), p. 72).

- The machine table of any Turing machine with an oracle serves as a counterexample to this claim.

2 Human behavior "understood as motion" can "in principle be reproduced to any degree of accuracy" on a Turing machine (ibid., pp. 195–196).

- There is no basis for this claim. It would be false if man were a super-Turing machine.

3 It is "a fundamental truth that every form of information processing (even those which *in practice* can only be carried out on an "analogue computer") must *in principle* be simulable on a [Turing machine]" (ibid., p. 195).

- This is not a "fundamental truth". There is no mathematical theorem to this effect in the literature.

4 All physical processes can be described in a mathematical

formalism which can in turn be manipulated by a [Turing machine] (ibid., p. 195).

- This claim is just as unfounded. Even if a physical process is described in a mathematical formalism, there is no guarantee whatsoever that there is a Turing machine which is able to solve the equations.

In sum, we may say that Turing machines are more limited than Dreyfus thought, or alternatively, that Nature is richer than he thought.

In the meanwhile, it is rather ironic that, in a book entitled *What Computers Still Can't Do*, the author did not *underestimate* but *overestimate* their capacities.

5 Dynamicism

Let's now turn to the third player in the field, the dynamical approach.

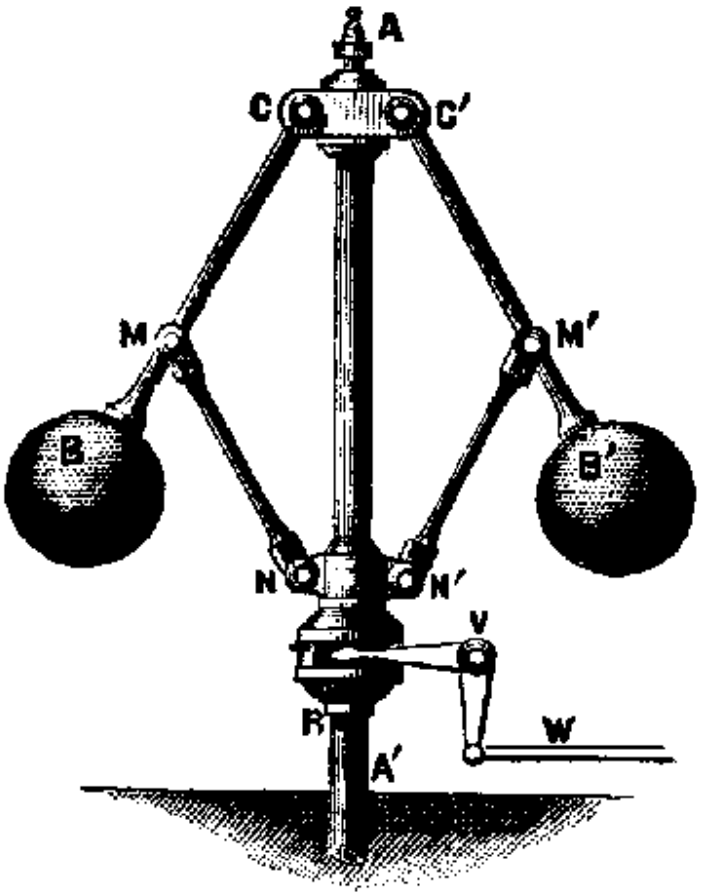
Main manifestos:

- Tim van Gelder: "What might cognition be, if not computation?" *The Journal of Philosophy* 91 (1995): 345-381.
- Tim van Gelder: "The dynamical hypothesis in cognitive science." *The Behavioral and Brain Sciences*, to appear.

Main thesis: Natural cognitive systems are to be viewed as *dynamical systems*.

Examples of dynamical systems: the solar system, the weather, James Watt's centrifugal governor (the dynamicists' favorite example).

The centrifugal governor (James Watt, 1788)



Like other dynamical systems, cognitive systems are to be described in terms such as:

- state space,
- trajectories through state space,
- attractors and repellers, limit cycles,
- stability, deterministic chaos,

and so on (see any book on chaos theory).

Objection This is not very new or controversial. Cf. the cybernetics movement of the 50s and 60s, e.g., Ashby's *Design for a Brain* (1952). Moreover, connectionists such as Sejnowski and Churchland have employed the same terminological framework.

Dynamicism has **two specific tenets**:

- Non-computationalism
- Non-representationalism

The first term seems a misnomer. The dynamicists do not deny that their systems are computational in a broad sense of the word (even though some of them are perhaps not Turing computable). The dynamicists should have used some term like "non-digitalism".

The second issue is more interesting.

Non-representationalists abhor talk of "internal representations". Instead, they say that each cognitive agent is "dynamically coupled" to the environment in which it is "embedded". There is a constant mutual interaction between the agent and its environment; they form one integrated system and cannot very well be isolated from each other. Cognition is most fruitfully seen as adaptive activity in an appropriate environment.

There are obvious points of similarity here with the philosophical positions adopted by Heidegger, Merleau-Ponty and the later Wittgenstein, a tradition which is nowadays being continued by thinkers such as Dreyfus & Freeman and Maturana & Varela. The dynamicists are well aware of this connection. They view their research programme as giving a scientific twist to the earlier, purely philosophical tradition.

On the other hand, there is a break with connectionism, which continually talks about distributed representations.

Personally, I think that the anti-representationalists are overstating their case.

- First, many cognitive capacities seem to have nothing to do with being in a tightly coupled relationship with the environment. I can reason about which retirement fund would be best in any non-distracting situation. Or think of mental imagery, doing mathematics and planning ahead: all these activities presuppose a considerable amount of *decoupling* from the actual environment. One might say that is precisely our ability to *break out* of the feedback loop with the environment that makes us human.

- Secondly, representationalism and dynamicism seem quite compatible. As Van Gelder himself put it, representations may be viewed as "trajectories or attractors of various kinds, or even such exotica as transformations of attractor arrangements as a system's control parameters change." As soon as representations are not naïvely identified with pictures or propositions but with some more sophisticated constructs, there is ample room for them in the dynamical framework.

Even if I am not particularly impressed with dynamicism, the recent arrival of this approach shows that there is still much vitality in cognitive science and AI. We are living in exciting times in which there is no scarcity of interesting ideas. There is no need to despair and to turn to Heidegger for illumination.

6 Literature about dynamicism

Go to David Chalmers, "Individuals with online papers in philosophy," section "Philosophy of mind (esp. artificial intelligence and cognitive science)," at
`http://ling.ucsc.edu/~chalmers/online.html`

7 Main comment by Dreyfus

"I am all for decoupling."