UMINF 11.14 COMPUTING SCIENCE, UMEÅ UNIVERSITY

Achilles' heel of cognitive science[*]

Erik Billing[†]

November 23, 2011

Abstract

A broad discussion on the representational problems of cognitive science is presented. Two common classes of representations, symbol based propositional representations and input-output mappings referred to as weak representations, are described and their relation to the *physical symbol system (PSS)* and *connectionist approaches (CA)* is discussed. Four problems often used as critique of PSS, the ontological assumption, the frame problem, the symbol grounding problem and the frame of reference, is shortly presented and it is concluded that these problems to some extent also apply to many connectionist models. Embodied cognitive science is presented and it is argued that weak representations used in embodied agents avoid the four fundamental problems, but are instead subject to the problem of adaptation. It appears that, in order to solve the problem of adaptation, some bias for generalization has to be introduced, which also would reintroduce the problems with propositional representations. The conclusion is that the fruitful solution to the fundamental representational problems of cognitive science is to focus on these problems with a joint view of propositional and weak representations, rather than trying to rule out one in favor for the other.

Keywords: representation, physical symbol system, ontological assumption, frame problem, symbol grounding

1 Introduction

The question of how the human brain implement our mind is subject for intensive discussion within the field of cognitive science. One of the most heavily debated aspects of the mind is the notion of representations. The literature on models of cognition spans from classical symbol systems (Johnson-Laird, 1986) to anti-representalistic approaches such as Dreyfus' *Phenomenology* (Dreyfus, 1972, 1992). In between we find works such as Fodor's *Language of Thought*

[*] This report was written as a part of the course *Theoretical Perspectives on Cognitive Science*, Sewcog - Swedish National Graduate School of Cognitive Science, www.swecog.se

[†] Erik Billing, Department of Computing Science, Umeå University, Sweden
e-mail: billing@cs.umu.se

(Fodor, 1987b) and Dennet's *Intentional Stance* (Dennett, 1989). There works has numerous differences and similarities and could be categorized along a number of dimensions. One of the most emphasized categorizations found in literature may be the distinction between the body of work supporting the *Physical Symbol Systems (PSS)* and theories aligning with a *Connectionist Approach (CA)*. The notion of PSS was coined by Newell and Simon (Newell & Simon, 1976) and pushes the argument that cognition is essentially computation in terms of symbol manipulation, independent of the precise physical implementation (Pylyshyn, 1986). CA is less well defined (Ellis & Humphreys, 1999), but generally implies a refusal of the PSS hypothesis and emphasizes distributed and parallel models that resemble the physical architecture of the brain more closely than PSS models.

A number of other differences follows this division. Approaches relying heavily on symbols appear to be primarily concerned with higher level cognitive abilities, such as reasoning and long term planning, e.g., Johnson-Laird (1986). In contrast, CA are more often concerned with cognitive abilities on lower abstraction levels, such as swinging a tennis racket or recognizing an object in a scene (Dreyfus, 1998).

Large marts of the literature on the subject supports one of these two approaches, and explicitly or implicitly rules out the other. In the present work, I argue that both PSS and CA face the same fundamental problems and that better understanding of cognition is not reached by ruling out one of the approaches, but by uniting them.

2 Representations

The word *representation* entails a large variety of meanings and some elaboration of this term may therefore be in place. The term *propositional representation* is used to refer to the kind of representations most often present in classical cognitive science (Stillings et al., 1995), where mental representations are implemented by a symbol system. Symbol systems can be defined as 1) a set of arbitrary "physical tokens" scratches on paper, holes on a tape, events in a digital computer, etc. that are 2) manipulated on the basis of "explicit rules" that are 3) likewise physical tokens and strings of tokens. The rule-governed symbol-token manipulation is based 4) purely on the shape of the symbol tokens (not their "meaning"), i.e., it is purely syntactic, and consists of 5) "rulefully combining" and recombining symbol tokens. There are 6) primitive atomic symbol tokens and 7) composite symbol-token strings. The entire system and all its parts – the atomic tokens, the composite tokens, the syntactic manipulations both actual and possible and the rules – are all 8) "semantically interpretable:" The syntax can be systematically assigned a meaning e.g., as standing for objects, as describing states of affairs. (Harnad, 1990)

Symbol system rely on an *ontology*, i.e., a specification of what entities exist in the world, what attributes and relations they have and how they should be grouped and categorized. The ontology specifies what things a symbol could represent, without specifying how the relation between the symbol and the referant is established. Propositional representations can be implemented in a PSS (Newell & Simon, 1976).

An alternative use of the term representation is here called *weak representation* and refers to an non-symbolic input-output mapping. Weak representations can be implemented with a connectionist approach, e.g., artificial neural networks (Sejnowski & Rosenberg, 1987) or dynamic systems (Freeman, 1991). While these systems does not necessarily qualify as systems able to implement propositional representations they are able to construct complex input-putput mappings. Weak representations are representations in the sense that an output pattern can be reconstructed given an input pattern, but typically does not support deductive reasoning and has problems explaining declarative memory (Stillings et al., 1995, p. 115). For this reason, weak representations are often put forward by thaws who argue against representations, e.g., Dreyfus (1998); Brooks (1991b). Despite the lack of an explicit model on which the system can reason, systems implementing weak representations has been able to show many forms of intelligent behavior, e.g., Brooks (1991a); Pfeifer & Scheier (1997), and has therefore been put forward as a plausible model for cognition as a whole.

It should be pointed out that the conflict between the supporters of symbolic representations and thaws who favor weak representations are mostly in terms of how the system is analyzed. On the one hand, almost all CA found in literature, which typically implement weak representations, are themselves implemented in a computer and consequently are represented symbolically (Ellis & Humphreys, 1999). On the other hand, there are connectionist systems, e.g., cellular automata, able to implement a Turing Machine as a weak representation (Gardner, 1970).

Since CA are typically closer to the neurological implementation, it seams natural to think that it is a better description on a lower abstraction level, while PSS better describes brain function on a higher abstraction level. However, when using PSS and CA as descriptions of the function of a computer, their abstraction ordering appears to be the other way. PSS is clearly in closer resemblance with the low-level mechanisms of a digital computer, which may produce higher level computations better described in terms of CA.

3 Four problems of propositional representation

One of the most fundamental critiques against symbolic representations has been formulated as a refusal of the ontological assumption. The *Ontological Assumption*, as phrased by Dreyfus (1992, p. 206), states that “everything essential to intelligent behavior must in principle be understandable in terms of a set of determinate independent elements”. The challenge of the ontological assumption lies primarily in how it is formulated. The requirement for a set of determinate independent elements allowing knowledge to be stored as a set of facts has several problems. Dreyfus (1992, p. 38) argues that “there is no argument why we should expect to find elementary context-free features characterizing a chair type, nor any suggestion as to what these features might be". Further criticism has been put forward by Churchland (1981) where the domain dependence of propositional representations is pointed out, among many other problems. Within classical cognitive science, theories on concept forming such as *scripts* and *schemes* has faced similar criticism (Stillings et al., 1995, p. 88). Even though several alternative theories has been proposed, typically probabilistic versions of the original theories, a general accepted solution has not been found.

A neighboring problem with symbolic representations is the difficulty to reason about change. In order to allow an agent to reason about consequences of action the representation must in some way encode all possible consequences that may occur for each action. Even though one can explicitly encode all fluents, i.e. all changing conditions, that change as a consequence of action, it is much more difficult to formalize the fluents that does not change without explicitly naming them and by consequence introduce an exploding requirement for so called *frame axioms*. This problem was first identified by McCarthy & Hayes (1969) and have become known as the *Frame Problem*. In situation calculus, the frame problem has been solved using *Successor State Axioms* (Reiter, 1991), but the it remains unsolved in the general case.

A third fundamental problem with symbolic representations is known as the *Symbol Grounding Problem* (Harnad, 1990). The symbol grounding problem is essentially the problem of connecting the symbols in a PSS with their respective referents. Harnad demonstrates the symbol grounding problem both by the *Chinese Room Argument* (Searle, 1980) and his own Chinese learning example:

> *"Suppose you had to learn Chinese as a first language and the only source of information you had was a Chinese/Chinese dictionary! ... How can you ever get off the symbol/symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols?"* (Harnad, 1990)

The forth problem is known as the *Frame of Reference* (Pfeifer & Scheier, 2001, p. 112) and is different from the previous problems in that it criticizes propositional representations from a design perspective rather than a purely representative. With the Frame of Reference, Pfeifer and Scheier highlights the difference between the perspective of the agent and the perspective of an observer studying the agent. Even though the symbols may be suitable, maybe even necessary, for describing the behavior of an agent from the observer's perspective, it must not be taken as the internal mechanism of the agent. The behavior of an agent is always a combination of the internal mechanism and its environment, and can not be described on the basis of the internal mechanisms alone. The argument extends to include the complexity of behavior: "The complexity we observe in a particular behavior does not always indicate accurately the complexity of the underlying mechanism" (Pfeifer & Scheier, 2001, p. 112), implying that even if it is possible to describe cognition in terms of PSS, we may end up solving a much harder problem than evolution did when producing the cognitive system.

4 Defense of propositional representations

These problems have of course not gone unnoted by the supporters of the supporters of symbol systems. Fodor (1987a) presents an extensive discussion on the Frame Problem where he argues that it represents a deep fundamental problem in epistemological philosophy and links to the "Hamlet's problem: when to stop thinking". Fodor takes a pragmatical position and argues that the Frame Problem may partly be solved by identifying the *relevant* aspects of the situation. However, this response is far from satisfactory since the difficulty now is to determine what is and what isn't relevant (Zalta, 2004), which may be just as hard to do in the general case.

Along the same lines, Fodor argues that the Symbol Grounding essentially is about connecting the symbol system to the world in "the right way", e.g., (Fodor, 1994). I understand this position as admitting that Symbol Grounding may be difficult by never the less keeping the positing that it has to be done. Harnad (2002) criticizes Fodor on this point, since he has not provided a clear description of what the right way is. However, Harnad also admits that "if the symbolic approach ever succeeds in connecting its meaningless symbols to the world in the right way, this will amount to a kind of wide theory of meaning, encompassing the internal symbols and their external meanings via the yet-to-be-announced *causal connection*."

Other parts of Fodor's defense of propositional representations and specifically Language of Thought shoots in my opinion above the fundamental criticism. In his argumentation with Aunty, the voice of the Establishment, Fodor starts out from the following position:

> *"First, she concedes that there are beliefs and desires and that there is a matter of fact about their intentional contents; there's a matter of fact, that is to say, about which proposition the intentional object of a belief or a desire is. Second, Aunty accepts the coherence of physicalism. It may be that believing and desiring will prove to be states of the brain, and if they do that's OK with Aunty. Third, she is prepared to concede that beliefs and desires have causal roles and that overt behavior is typically the effect of complex interactions among these mental causes."* (Fodor, 1987b)

I find this unsatisfying since I believe that many supporters for weak representations would not even agree to these initial positions that Fodor obviously sees as a mater of fact. As I understand Fodor, he sees beliefs and desires as just as obviously existing as anything else in the world, and if we are to deny the existence of beliefs then we could just as well deny the existence of humans or trees. On this account, I think it is hard to argue against Fodor. However, he seems to directly draw the conclusion that since there are beliefs and desires, they have to be the mechanism of cognition. This is a much less obvious conclusion and is in my opinion a typical example of the Frame of Reference mentioned in Section 3.

Never the less, even if beliefs and desires may not produce cognition, it is reasonable to argue that they should reflect relevant aspects of cognition. From this point, Fodor (1987b) makes a strong argument for the Language of Thought. If a symbol processing system is aligned with the world such that the symbols are manipulated in a way that corresponds to the events of the world, I understand that Fodor would be happy to say that they have intentional content.

In a neighboring line of reasoning, Fodor (1987b) pushes the *systematicity* of thought, makes a strong argument why thoughts have to be systematic and concludes that their thereby has to be a language of thought. The alternative, as Fodor pushes it, would be "memorizing an enormous phrase book". His conclusion may be an oversimplification but is still effective since the initial proportions for weak representations were essentially a large set of *if X then Y*-statements, where X is a percept and Y is an action.

Even if not without problems, propositional representations may still be the best way to view mental representations. As argued by Pfeifer & Scheier (2001), much of the criticism presented in Section 3 applies just as much to connectionist models. Pfeifer and Scheier takes the famous NETtalk model as an example. NETtalk, by Sejnowski & Rosenberg (1987), is an artificial neural network that learns to convert English text to speech, by mapping sequences of characters to phonemes. The model is distributed, parallel and robust. It learns and generalizes in a way that has similarities with human language accusation and has plausible neural implementation. In the hidden layer of the network, an emergent categorization of vowels and consonants is produced. Still, it is in Pfeifer and Scheier's view essentially a symbolic system. NETtalk is not grounded in any environment, the meaning of both input and output of the network has to be interpreted by a human. It also implements an ontology, both in terms of it's seven character input window and it's set of phoneme output nodes. Furthermore, NETtalk is trained using supervised learning, where the correct input-output mappings are specified by the programmer, and the apparently emergent organization of categories is essentially nothing but data clustering.

The NETtalk example illustrates the necessity for another classification of representations. A very interesting attempt is made by Pfeifer & Scheier (2001) within the framework of Embodied Cognitive Science.

5 Embodied Cognitive Science

During the last decade, a branch of cognitive science known as *Embodied Cognitive Science* has grown more popular. In general terms, the trend pushes the role of the environment in cognitive processes and that cognition must not be seen as a process of the agent as much as a consequence of the interactions between the agent and its environment. Following the somewhat louse definition proposed by Pfeifer & Scheier (2001), a complete intelligent agent has to be adaptive, autonom, self-sufficient, embodied, and situated. An agent aligning to these terms solve all four problems outlined in Section 3. A simple example of such agents are the Braitenberg Vehicles (Braitenberg, 1986). These agents are not implementing symbols since the sensors and motors are not said to refer to something else, as was the case with the NETtalk model, and the control mechanism is nothing but non-linear connections between sensors and motors. Braitenberg Vehicles implement no world ontology and carry no model of the world that may be subject to the frame problem. They also align to the frame of reference by that they are designed in a strict agent centric manner. Even though their behavior may be seen as complex, involving decisions and intentions, these properties are merely constructs made by the observer.

A critical feature of the Braitenberg Vehicles, and all other agents that are complete in Pfeifer & Scheier (2001) terms, is that the actions of these agents affects the sensor input. These places the relation bedtween sensors and actuators in a new perspective since percepts can be seen as a function of action, involving the environment, just as much as an action is a response to a specific percept. With this argument in mind, Pfeifer & Scheier (1997) propose *Sensory-Motor Coordination (SMC)*, a control architecture where representations are created within the sensory-motor space. In contrast to the stimuli-response approaches proposed by Braitenberg (1986), Brooks (1986) and others, SMC implements control and perception are the same process, a coordination of sensors and actuators. It should be pointed out that Pfeifer & Scheier (2001) does not argue that weak representations used in SMC are ontology free. The sensory-motor space is also an ontology, referred to as a low-level specification, describing the agent-world interaction in terms of sensors and actuators. One important difference is however that the low-level specification is provided by the embodiment of the agent, and completely world and task invariant. The world may of course change, requiring the agent to develop new sensors and actuators in order to cope with its environment, but in terms of behavior it provides much more flexibility than the classical world ontology.

Since weak representations do not provide an explicit memory, learning is phrased in terms of adaptation (McFarland, 1991). Pfeifer & Scheier (2001) hereby makes a clear distinction between CA and weak representations by arguing that even though distributed, parallel and robust networks may have advantages over symbol systems, it is the embodiment and adaptation that does the trick.

From this argument it should be clear that the four problems of propositional representations are essentially not problems of symbol systems as such, but rather the problem of representations that are not grounded in behavior. Even though discussions like the one above has been taken as arguments against the possibility to implement truly intelligent machines, e.g. Dreyfus (1992), I see no reason why the input-output mapping of an embodied agent can not be established by a symbol manipulating system. However, embodied cognitive science can still be seen as an argument against the PSS hypothesis in the sense that cognition, from an embodied perspective, takes place in the interaction between the agent and its environment, and it is consequently hard to see where a computational level should appear.

6 Adaptation

In the previous section, it was concluded that the four problems of propositional representations is solved by an embodied agent that learns by adapting its input-output connections, i.e., changing its weak representation. But there is one fundamental problem left: How should the agent adapt?

Weak representations merely map input to output, and if some new knowledge is to be stored, the mapping must change. However, it is far from obvious how this mapping is to be changed. In contrast to propositional representation where the facts are explicit and may be updated as necessary, a weak representation does not make facts explicit and it is therefore very hard to pick out the relations that is to be modified.

Adaptivity is often said to have two components, one conservative and one innovative (Pfeifer & Scheier, 2001). The tradeoff between these two aspects of adaptation can be found in many forms through large parts of the literature. Piaget (1952) distinguished between assimilation and accommodation as two aspects of child development, Carpenter & Grossberg (1988) described a trade-off between stability-flexibility in adaptive pattern recognition, and the notion of exploitation versus exploration is well known through the literature on reinforcement learning, e.g. Sutton & Barto (1998). The two components can even be found in evolutionary theory as the tradeoff between inheritance and mutation.

In a general sense, the difficulty to choose between "doing what you know" and "doing something new" is captured by the "no free lunch" theorems (Ho & Pepyne, 2002; Wolpert & Macready, 1997). The argument illustrates the need for bias in learning. In order to know how to adapt to a certain situation, some pre-judge is required, specifying how the information that we have can be generalized. This is essentially what the ontology of the propositional representation provides. The reason reactive robots are so good doing insect-like behavior is that their ontology, the low-level specification, corresponds to the selected tasks. Braitenberg Vehicles are just as unable to play chess as Deep blue is unable to avoid obstacles, the difference is essentially in terms of what kind of ontology that is provided.

7 Conclusion

Two distinct categories of representations has been identified, *propositional representations* and *weak representations*. The distinction between these two classes reflects an debate within the field of artificial intelligence that has been present for the last thirty years. Supporters of weak representations argue that the propositional approach is doomed since symbol based representations are based on the ontological assumption, subject to the frame problem and does not ground symbols in the world. On the contrary, reactive agents employing weak representations are unable to abstract and create representations outside their own ontology, the low-level specification. Both kinds of representations can in principle solve the problem: The symbol system merely has to connect to the world in *the right way*, and the weak representations have to adapt to the new situation, in *the right way*.

I believe that one important point put forward in the criticism against symbolic representations are that all knowledge are context dependent to some extent. The problem with world onthologies is not the requirement for explicit facts per say, but the requirement for context-free facts. Heidegger introduces the notion of *Being-in-the-world* in his argument that the world itself provides the fundamental context (Guignon, 1983). Dreyfus (1992, p. 207) continues: "in order to understand an utterance, structure a problem, or recognize a pattern, a computer must select and interpret its data in terms of a context. But how are we to impart this context itself to the computer?"

As I understand Dreyfus' requirement for *context* it is essentially a situation-dependent ontology, i.e., a specification of all aspects of the world that are relevant for the *present situation*. This may of course not be what Dreyfus argues since he clearly can not see how such context dependent knowledge could be represented in a computer, but I still believe it pinpoints a critical problem of PSS and more importantly, provides a possible solution.

I believe that, in order to propose a coherent theory of representation, it is time to stop focusing on the debate between symbolic and non-symbolic approaches to representation, and in stead focus on their common issues. Even though a discussion of possible solutions is outside the scope of this report, I believe that we have come a long way if we have identified the fundamental problems common to both classes of representations, namely to balance the bias provided by an ontology with the proclivity of changing the ontology in order to connect it with the world. An introduction to several approaches with the ambition to solve this common representational problem, inherit from information theory, computational neuroscience, and robotics, can be found in Billing (2009).

References

Billing, E. A. (2009). *Cognition Reversed - Robot Learning from Demonstration.* Lic. thesis, Umeï¿œ University, Department of Computing Science, Umeï¿œ, Sweden.

Braitenberg, V. (1986). *Vehicles - Experiments in Synthetic Psychology.* The MIT Press.

Brooks, R. A. (1986). A Robust Layered Control System For A Mobile Robot. *IEEE Journal of Robotics and Automation*, 2(1), 14–23.

Brooks, R. A. (1991a). Intelligence without reason. *Proceedings, 1991 Int. Joint Conf. on Artificial Intelligence*, (pp. 569–595).

Brooks, R. A. (1991b). Intelligence without Representation. *Artificial Intelligence*, 47, 139 – 159.

Carpenter, G. A. & Grossberg, S. (1988). The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. *Computer*, 21(3), 77–88.

Churchland, P. M. (1981). Eliminative Materialism and Propositional Attitudes. *The Journal of Philosophy*, 78, 67–90.

Dennett, D. C. (1989). *The Intentional Stance*. The MIT Press.

Dreyfus, H. L. (1972). *What Computers Can't Do: A critique of artificial reason*. Harper & Row.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. The MIT Press.

Dreyfus, H. L. (1998). A Phenomenology of Skill Acquisition as the basis for a Merleau-Pontian Non-representationalist Cognitive Science.

Ellis, R. & Humphreys, G. W. (1999). *Connectionist Psychology: A Textbook with Readings*. Psychology Press, 1 edition.

Fodor, J. A. (1987a). *Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres*. Norwood, NJ.

Fodor, J. A. (1987b). *Psychosemantics*. Bradford Books/MIT Press.

Fodor, J. A. (1994). *The elm and the expert: mentalese and its semantics*. MIT Press.

Freeman, W. J. (1991). The Physiology of Perception. *Scientific American*, 264(2), 78–85.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223, 120–123.

Guignon, C. B. (1983). *Heidegger and the problem of knowledge*. Hackett Publishing.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica*, D(42), 335–346.

Harnad, S. (2002). *Symbol grounding and the origin of language*, (pp. 143–158). MIT Press.

Ho, Y. C. & Pepyne, D. L. (2002). Simple Explanation of the No-Free-Lunch Theorem and Its Implications. *Journal of Optimization Theory and Applications*, 115(3), 570, 549.

Johnson-Laird, P. N. (1986). *Mental Models*. Harvard University Press.

McCarthy, J. & Hayes, P. J. (1969). *Some Philosophical Problems from the Standpoint of Artificial Intelligence*, (pp. 463–502). Edinburgh University Press.

McFarland, D. (1991). What it means for robot behaviour to be adaptive. In *Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats* (pp. 22–28).: MIT Press. Cambrage, Massachusetts.

Newell, A. & Simon, H. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communication of the ACM*, 19(3), 113–126.

Pfeifer, R. & Scheier, C. (1997). Sensory-motor coordination: the metaphor and beyond. *Robotics and Autonomous Systems*, 20(2), 157–178.

Pfeifer, R. & Scheier, C. (2001). *Understanding Intelligence*. MIT Press. Cambrage, Massachusetts.

Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.

Pylyshyn, Z. W. (1986). *Computation and Cognition: Toward a Foundation for Cognitive Science*. The MIT Press.

Reiter, R. (1991). The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. *Artificial intelligence and mathematical theory of computation: papers in honor of John McCarthy*, (pp. 359–380).

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417–424.

Sejnowski, T. J. & Rosenberg, C. R. (1987). Parallel Networks that Learn to Pronounce English Text. *Complex Systems*, 1, 145–168.

Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L., & Rissland, E. L. (1995). *Cognitive Science*. Cambridge, Massachusetts: MIT Press.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press. Cambrage, Massachusetts.

Wolpert, D. H. & Macready, W. G. (1997). No free lunch theorems for optimization. volume 1 (pp. 67–82).

Zalta, E. N. (2004). The Frame Problem.