Can We Build a Match-Making Machine Out of Matches?

Marcin Chady
The University of Birmingham, UK
M.Chady@cs.bham.ac.uk

Abstract

This poster is an invitation to a free discussion on the limitations of the mind in its capability to understand itself and other minds. It poses some questions, presents examples and makes suggestions, in the hope to attract some comments and feedback from the readers. It does not attempt to solidly defend any theory or hypothesis. Readers are welcome to make challenging comments.
Note: Terms highlighted in CURLY FONT are considered highly ambiguous and prone to distorting the intended meaning of the sentence. They are used nevertheless, either precisely to highlight this effect, or for want of a more appropriate expression.

The Purposeful Machine

The human brain has evolved to model the world, so that we can avoid dangers and achieve our goals. It is not possible to simulate the universe exactly, so some, arguably significant, degree of approximation is required. In other words, in order to draw conclusions and predict future events, one must generalise. Consequently, any brain-endowed organism is a natural classifier, and we humans are a prime example of this: we categorise, generalise and organise things into hierarchies, so that we can make better sense of the world, and turn its rules into our advantage. We have developed languages, algebras and calculi. Our sophisticated system of symbols can describe abstract as well as physical things. However, as we face the enigma of our own MINDS, we must ask ourselves the question: Can we model the MIND with its own products?

The Ontologies We Use

When we talk about THINKING, we tend to use inherently ambiguous terms:

SYMBOLS	MENTAL STATES
EMOTIONS	INTENTIONS
EXPLORATION	CONTROL
MEMORY	Etc.

It is not until we actually try to realise them in a physical system, that it transpires how vague they are. Most of them have been discredited from scientific discourse, and confined to the realms of FOLK PSYCHOLOGY, but much of the confusion remains (cf. Smolensky (1988)).

SYMBOLS and MENTAL STATES are inherently *discrete* (Stich (1983)). They are convenient ABSTRACTIONS, like the idea of the centre of gravity (Dennett (1991a)), but they are, above all, and first of all, *conventions*. They exist only because they are useful to us.

> I expect that the actual internal states that cause behaviour will not be functionally individuated, even to an approximation, the way belief/desire psychology carves things up (Dennett (1987))

> Certain deeply intuitive principles of folk psychology, perhaps never before articulated, may have to be given up (Dennett (1991b))

> Mental states do not exist (eliminative materialism - Churchland)

The Architectures We Imagine

People like to view the MIND as a number of parallel processes: transforming information, exchanging signals, making decisions, each specialised in a different kind of operations. It is easy to CONCEPTUALISE such systems using the "divide and conquer" strategy and, perhaps most importantly, it is easy to *depict* them using boxes and arrows.

Although symbolic systems try to avoid brittleness by allowing a certain degree of flexibility and/or non-determinism in their architecture (cf. the ACT-R system), the component-based perspective still remains:

> As a result of asynchronous generation, motives may be in conflict, so motive comparators are also needed, for resolving conflicts. (Sloman (1999))

It is easy to forget that FUNCTIONALITY and MECHANISM are not necessarily in a one to one relationship. Such reasoning assumes that each FUNCTION identified in the designer's $\mathcal{MIND}$ has an architectural embodiment. Yet, it is easy to imagine competition without an arbiter, in which case the "comparator" has no counterpart.

The Systems We Build

Much of the non-symbolic AI, such as neural networks, inadvertently follows the same path by giving parts of their systems $\mathcal{EXPLICITLY}$ defined functionality, and combining them into interacting ensembles of modules. A prime example of this are implementationalist systems, where the logic is hand-wired into the connections between units (Shastri and Ajjanagadde (1993), Barnden (1991)).

Not only do they risk missing out a significant part of the solution domain, but parts of the problem domain too: they usually end up with a limited capability for learning.

Note that this is not an attack on modularity: it is merely pointing out that our way of segmenting a problem is heavily biased and generally incapable of producing a wide range of possible solutions.

All But Artefacts

It is hard not to structuralise, given that this is what our brains have been designed to do through millennia of evolution. However, one must not forget that everything our cognitive processes come up with is an *artefact*. It's a product of a process whose purpose is by no means general. On the contrary, its purpose is to filter out details that are irrelevant to survival, and produce a *convenient* model of the environment. There is no reason why our cognitive apparatus should be able to cope with its own workings, just like no one is predisposed to imagine 7-dimensional objects, which some physicists suggest our universe really consists of.

When we analyse a problem, we tend to be attracted to certain solutions, of which we say they are $\mathcal{NEATER}$, more $\mathcal{GENERAL}$, or just $\mathcal{NICER}$. However, let us not forget that beauty is always in the eye of the beholder: some solutions may just appeal to us simply because of the way our brains are built.

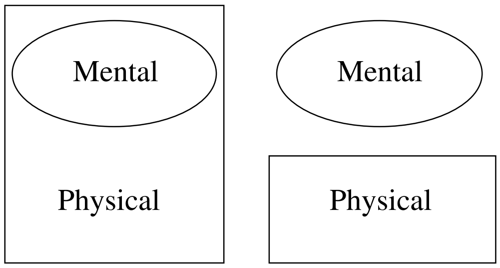
The dangers of taking symbols literally:

1. Mapping the $\mathcal{ENTITIES}$ of the real world directly onto $\mathcal{ENTITIES}$ inside a cognitive system. The pitfalls of this approach have been argued at great length by advocates of distributed representations and by antirepresentationalists in general. There is no need for internal representations to be $\mathcal{EXPLICIT}$. And if they aren't explicit, you might as well assume there aren't any.

2. Using the $\mathcal{ENTITIES}$ and $\mathcal{OBJECTS}$ to describe the architecture and explain the operation of a cognitive system. This approach can be compared to building a match-making machine out of matches. All concepts, including $\mathcal{MEMORY}$, $\mathcal{CONTROL}$, and $\mathcal{PROCESS}$ are *products* of our $\mathcal{MINDS}$, not their *components*. Their usefulness is limited to the practical domain for which they are construed, and only to the extent to which the cogniser $\mathcal{UNDERSTANDS}$ the domain.

Illusions

The urge to $\mathcal{IDEALISE}$ is so strong that we much rather talk about $\mathcal{VIRTUAL}$ machines than their physical realisations (functionalism). Yet, we don't often realise that the underlying psychological motivation, if not the very principle of this approach is the same as in Cartesian dualism, which of course functionalism rejects.



Functionalism **Dualism**

Similarly, arguing that $\mathcal{SYSTEMATICITY}$ can only be achieved in a compositional system, i.e. such that is built on combinatorial syntax and $\mathcal{SEMANTICS}$, could turn out to be unjustified. After all, the concept of grammar is also an artefact.

It may well be that natural language, or even $\mathcal{THOUGHT}$ $\mathcal{PROCESSES}$ are best described to a human being using some kind of grammar (Fodor (1975)), but the operation of their generator doesn't need to be combinatorial, or even discrete in the $\mathcal{REAL}$ sense (see also the next section).

The existence of the *frame problem* is a strong indication that the symbolic view of the world is incomplete.

> I don't think the frame problem is unreal or insignificant: it seems to point up a basic error in the way we try to describe the everyday world. We aren't carving up nature at the right ontological joints, if you ask me. (Hayes (1987))

However, one can't help but ask, is there such a thing as the $\mathcal{RIGHT}$ ontological structure? Isn't it, after all, yet another artefact?

What is not an illusion?

Can we escape the kaleidoscope of our cognition? Most probably not. Even the machines we create will be biased.

However, we can reduce this bias by limiting our involvement in their design, putting emphasis on self-organisation (autopoiesis). Evolutionary strategies spring to mind: rather than designing complete architectural solutions, we can specify the constraints (using grammars in this case is fine) and let the system evolve to meet them.

However, one must not view evolution only as a means to arrive at a working system. An intelligent system is an evolving system. The brain constantly changes its architecture through creating new synapses and discarding old ones. It's not just $\mathcal{KNOWLEDGE}$ that changes, the $\mathcal{ENGINE}$ changes too. In fact, making a distinction in cognitive modelling between these two $\mathcal{CONCEPTS}$ is simply misleading.

Another approach is to study complex unconventional systems and look for new insights there. Of course, every $\mathcal{INSIGHT}$ will be subjective too, but there is nothing we can do about it. At least we may discover a new way of thinking (achieved through the usual brain rewiring) which will take us a further step back from the narrow vision of a "survival machine".

An example of such approach can be found in the work of Hanson and Crutchfield (1992). There, a complex behaviour of CA is analysed using classic paradigms from computational theory, such as Finite State Automata. The work demonstrates how complicated operation can result from simple interactions between simple units. Conversely, the use of classical computational models to describe the global behaviour shows that no explicit FSA machinery is necessary to produce FSA behaviour.

Roundup

Perhaps then, there are no parts responsible for $\mathcal{MOTIVATION}$, $\mathcal{EMOTIONS}$, $\mathcal{CONTROL}$, etc. They could be just facets of the same process, presenting themselves differently from different angles. Much like the facets of a hypercube intersected with a 3-dimensional hyperplane, or like ephemeral vortices of turbulent flow.

And last but not least, let us not forget that the $\mathcal{MIND}$ itself is an artefact: an attractive and convenient concept for picturing our $\mathcal{THOUGHTS}$ perhaps, but not necessarily the most useful one in tackling the "hard problem".

References

John A. Barnden. Encoding complex symbolic data structures with some unusual connectionist techniques. In John A. Barnden and J. B. Pollack, editors, *High Level Connectionist Models*, volume 1 of *Advances in Connectionist and Neural Computation Theory*. Ablex, Norwood, N.J., 1991.

D. C. Dennett. *The intentional stance*. MIT Press, Cambridge, 1987.

D. C. Dennett. Real patterns. *Journal of Philosophy*, 88: 27–51, 1991a.

D. C. Dennett. Two contrasts: folk craft versus folk science, and belief versus opinion. In J. D. Greenwood, editor, *The future of folk psychology*, volume 88, pages 135–148. Cambridge University Press, Cambridge, 1991b.

J. A. Fodor. *The language of thought*. Harvard University Press, Cambridge, 1975.

Jim Hanson and Jim Crutchfield. The attractor-basin portrait of a cellular automaton. *Journal of Statistical Physics*, 66:1415–1462, 1992.

Patrick Hayes. What the frame problem is and isn't. In Z. W. Pylyshyn, editor, *The robot's dilemma: the frame problem in artificial intelligence*, pages 123–136. Ablex, Norwood, 1987.

Lokendra Shastri and Venkat Ajjanagadde. From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16:417–494, 1993.

Aaron Sloman. What sort of architecture is required for a human-like agent? In Michael Wooldridge and Anand Rao, editors, *Foundations of Rational Agency*. Kluver Academic, Dordrecht, 1999.

Paul Smolensky. On the proper treatment of connectionism. *The Behavioral and Brain Sciences*, 11, 1988.

S. Stich. *From folk psychology to cognitive science*. MIT Press, Cambridge, 1983.