

# ARCHITECTURE-BASED CONCEPTIONS OF MIND

Aaron Sloman

*School of Computer Science, University of Birmingham*

*Birmingham, B15 2TT, UK*

<http://www.cs.bham.ac.uk/~axs/>

**Abstract** It is argued that our ordinary concepts of mind are both implicitly based on architectural presuppositions and also cluster concepts. By showing that different information processing architectures support different classes of possible concepts, and that cluster concepts have inherent indeterminacy that can be reduced in different ways for different purposes we point the way to a research programme that promises important conceptual clarification in disciplines concerned with what minds are, how they evolved, how they can go wrong, and how new types can be made, e.g. philosophy, neuroscience, psychology, biology and artificial intelligence.

## 1. Introduction

We seem to have direct access to mental phenomena, including thoughts, desires, emotions and, above all our own consciousness. This familiarity leads many people to believe they know exactly what they are talking about when they engage in debates about the nature of mind, and refer to consciousness, experience, awareness, the ‘first-person viewpoint’, and so on.

However, this conviction is at odds with the diversity of opinions expressed about the nature of the phenomena, and especially the widely differing definitions offered by various types of psychologists, cognitive scientists, brain scientists, AI theorists and philosophers, when they attempt to define concepts like ‘emotion’ and ‘consciousness’.

The confusion has several roots, one of which is the *hidden* complexity of both the phenomena and the architectural presuppositions we unwittingly make when we use such concepts.

Another is the common error of believing that we have a clear understanding of concepts just because they refer to phenomena that we experience directly. This is as mistaken as thinking we fully understand what simultaneity is simply

because we have direct experience of seeing a flash and hearing a bang simultaneously. Einstein taught us otherwise. That we can recognise some instances of a concept does not imply that we know what is meant *in general* by saying that something is or is not an instance. Endless debates about where to draw boundaries are a symptom that our concepts are confused, whether the debates are about which animals have consciousness, whether machines can be conscious, whether unborn infants have experiences, or whether certain seriously brain-damaged humans still have minds.

Such questions cannot be resolved by empirical research when there is so much disagreement about what sort of evidence is relevant. Does wincing behaviour in a foetus prove that it feels pain and is therefore conscious, or is it a mere physiological reaction? How can we decide? Does a particular type of neural structure prove that the foetus (or some other animal) is conscious, or is the link between physical mechanisms and consciousness too tenuous to prove anything?

This paper shows how the hidden complexity of our concepts and the phenomena they refer to explain why there is so much confusion and disagreement and indicates how we can begin to make progress beyond sterile debates.

Many of our concepts are implicitly architecture-based and different thinkers attend to different aspects of the architecture. They are also 'cluster concepts', referring to ill-defined clusters of capabilities supported by the architecture, and different views favouring different clusters. If we understand this we can see how to define different families of more precise concepts, on the basis of which answerable questions can be formulated. Which definitions are *correct* is a pointless question.

## 2. Architecture-based concepts

We can deepen our understanding of these concepts, and, where necessary, repair their deficiencies, by seeking an explanatory theory which accounts for as many phenomena as possible and then use it as a framework for systematically generating concepts. A common error is believing that we have to define our concepts before we seek explanatory theories. Typically it is only after we have a theory that we can understand the concepts describing the phenomena to be explained. So it is to be expected that we shall not be able to give good definitions of most of our mental concepts until we have good explanatory theories.

This does not imply that our pre-theoretical concepts are completely wrong. Our existing concepts of mind work well enough for ordinary conversational purposes (e.g. when we ask 'When did he regain consciousness?', 'Are you still angry with me?', etc.). So a good theory of the architecture underlying

mental states and processes should generate concepts which *extend* and *refine* our previous concepts, rather than replacing or eliminating them.

New theories of the sub-atomic architecture of matter extended and revised our concepts of kinds of elements, kinds of chemical compounds, and kinds of physical and chemical processes. We still talk about iron, carbon, water, etc., though we also now know about isotopes and new sorts of elements and compounds, and many new kinds of processes involving previously known kinds of physical stuff. We still talk about solids, liquids and gases though we also know about other states of matter supported by the architecture.

### 2.1. Architecture-based cluster-concepts

Muddles in our pre-theoretical concepts of mind surface when we try to ask philosophical or scientific questions, e.g. 'How did consciousness evolve?' 'What are its neural correlates?' 'Which animals have it?' What we normally refer to as consciousness involves the exercise of a large, diverse, ill-defined cluster of capabilities (many of them unconscious!) supported by our information processing architectures. If there is no well-defined subset of capabilities which are necessary or sufficient for consciousness, then some of our apparently meaningful questions, like many questions involving cluster concepts, may be ill-defined. Many mental concepts share this semantic indeterminacy, e.g. 'emotion', 'intelligence', 'understanding', 'pleasure', etc.

The idea that there are cluster concepts, that various kinds of indeterminacy or, what has been called *open texture*, pervades ordinary language is very old, e.g. in the writings of Wittgenstein (1953), Waismann (1965) and many others. I shall attempt to explain how it comes about that ordinary mental concepts have that feature, and what to do about it.

### 2.2. Multiple architectures generate multiple families of concepts

The analogy suggested above between the way theories of the architecture of matter extend and refine ordinary concepts of kinds of stuff and the way a new theory of the architecture of mind could illuminate concepts of mentality, is only partial, because there is only one physical reality and one architecture for physical matter (although it may have many levels of abstraction), whereas there are many kinds of minds with different architectures.

Figures 1 (a) and (b) illustrate two typical architectural decompositions of an intelligent organism, software system, or robot. Figures 2 combines the two views and add further detail. Figures 3(a) and (b) elaborate further. Organisms with simpler architectures have fewer architectural layers, and simpler perceptual or motor subsystems. They would then support simpler collections of processes, and different concepts would be applicable to them. If insects

In the Scope of Logic, Methodology and Philosophy of  
Science

Volume Two of the 11th International Congress of  
Logic, Methodology and Philosophy of Science, Cracow,  
August 1999

Gärdenfors, P.; Wolenski, J.; Kijania-Placek, K. (Eds.)

2002, VIII, 362 p., Hardcover

ISBN: 978-1-4020-0930-3