Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 20 (2007) 947–954
Neural Networks
www.elsevier.com/locate/neunet

2007 Special Issue

Hierarchical approaches to understanding consciousness

L. Andrew Coward[a,*], Ron Sun[b]

[a] Department of Computer Science, Australian National University, Canberra, ACT 0200, Australia
[b] Cognitive Science Department, Rensselaer Polytechnic Institute, 110 Eighth Street, Carnegie 302A, Troy, NY 12180, USA

Abstract

There has been much discussion on what a scientific theory of consciousness would look like, and even whether such a theory is possible. Some common misunderstandings of the nature of theories (e.g., in the physical sciences) have confused the discussion of theories concerning consciousness. Theories in the physical sciences establish hierarchies of descriptions that relate high-level descriptions of macro-level phenomena to detailed-level descriptions at a micro level. Detailed descriptions are usually more accurate but information-dense and therefore often beyond human comprehensibility (unless limited to tiny segments of a macro-level phenomenon). High-level descriptions are usually much less information-dense but more approximate. The ability to map between levels of description, and in particular the understanding of when a shift from a higher-level to a more detailed description is needed to achieve a desired degree of accuracy, is fundamental to an effective theory in any field. The *form* of such a theory of consciousness is sketched, and the limitations of some alternative approaches described.

Keywords: Cognitive architectures; Complex systems; Consciousness theories; Modules; Scientific theories

1. Introduction

There are a range of possible approaches aimed at understanding consciousness. At one extreme, there has been debate over whether a scientific theory of consciousness is possible in principle (e.g. Chalmers (1995)). Another type of approach has been the creation of models that describe cognitive operations regarded as being conscious, but with limited consideration for how physiology could support the processes required by the model. Computer implementations of some of these models have been attempted (e.g. Franklin and Graesser (1999) and Sun (1999). Yet another approach has been to look for physiological activity that discriminates between conscious and unconscious states, the so-called "neural correlates of consciousness" (e.g. Crick and Koch (1998) and Rees, Krieman et al. (2002). Lamme (2006) claimed that neural activity occurred in different structures depending on the type of conscious or unconscious behaviour, and argued that the presence of neural activity should be part of the definition of whether consciousness was present, independent of cognitive measures. Yet another approach has been to claim that understanding of consciousness is only possible with reference to quantum mechanics (Penrose, 1994). Furthermore, there have been extensive debates over what phenomena should actually be labeled "conscious" (e.g. Block (1995)). The end result has been considerable meta-theoretical confusion.

For several centuries, the physical sciences have been regarded as the paradigm for valid scientific theories. We believe that much of the confusion over scientific understanding of consciousness derives from misunderstanding of what theories in the physical sciences actually deliver, and once these misunderstandings are cleared away, the form which a theory of consciousness must take becomes clearer. There are some good questions about whether a scientific theory of consciousness is possible, but these questions are in fact more empirical than philosophical, and relate to whether the brain is organized in such a way that understanding can occur within the limits of human information handling capabilities.

There are some significant similarities between theories in the physical sciences and the theoretical techniques by which the understanding of a complex computational system is achieved and maintained. Although there are minimal direct resemblances between brains and complex computational systems, natural selection pressures on brains have tended to have architectural effects which make these techniques relevant

* Corresponding author.
E-mail address: andrew.coward@anu.edu.au (L.A. Coward).

0893-6080/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2007.09.009

Table 1
A physical sciences hierarchy of descriptions relevant for understanding the refining process for crude oil

	Entities	Number of types of entities	Examples of entity types
Quantum mechanics	Individual elementary “particles”	<10	Electron, proton, photon
Atomic theory	Individual atoms	~ 100	Carbon atom, hydrogen atom, sulphur atom
Molecular theory	Individual molecules	All possible combinations of atoms	Benzene molecule, isobutene molecule
Chemical theory	Chemicals = large number of almost identical molecules	All possible collections of identical molecules in all possible temperatures and states	Different masses of benzene, at different temperatures, in solid, liquid or gas state etc.
Materials theory	Materials = mixtures of chemicals	All possible collections of chemicals	Crude oil

Descriptions at a detailed level (such as quantum mechanics) use fewer different types of entities with relatively simple interactions between them, are more mathematically exact, but have an information density that makes them incomprehensible if applied to a complete macro phenomenon. Descriptions at a higher level have more types of entities with complex interactions between them that can be approximated by a simple description, have a much more comprehensible (lower) information density, but are more approximate. However, their accuracy is adequate for many practical purposes, and there are clear rules indicating when a more detailed level is required for adequate descriptive accuracy.

and applicable (Coward, 2005), thus providing a basis for a scientific theory of consciousness. The general form which such a theory would take can therefore be sketched.

2. Theories in the physical sciences

The ability of theories in the physical sciences to accurately, quantitatively predict the outcome of experiments has made physics the paradigm for a good scientific theory. In quantum mechanics, the extreme accuracy of such predictions has contrasted with the abstract and sometimes counterintuitive nature of the concepts used. This contrast has sharpened the philosophical debate between those who believe that the predictive power of a scientific theory provides insight into the underlying causal structure of reality (or “natural laws”), and those who would argue that such theories simply provide effective but ultimately approximate descriptions of reality (i.e., realists versus empiricists). In the neurosciences, the counterintuitive nature of quantum mechanics has perhaps influenced the development of the view that “folk psychology”, or everyday understanding of human psychology, will have no place in a genuine science of the brain (Churchland, 1989).

There is an aspect of the practice of the physical sciences that is often missing both from the philosophy of science and from thinking about the ultimate form of a genuine brain science. This aspect is the existence and use of hierarchies of description, which make it possible to describe causal (and other key) relationships within a phenomenon on many different levels of detail. The higher levels are more approximate than lower (more detailed) levels, but all the different levels of detail are essential to create a comprehensible scientific understanding of the phenomenon. An indispensable part of a scientific theory is the ability to map between different levels of description, including rules to indicate when a transition to a deeper (lower) level is required to achieve a desired degree of accuracy.

In order to illustrate this aspect of the physical sciences, consider how the science of matter can be applied to designing and adjusting the refining process for crude oil. Some key elements of the five-level description hierarchy are summarized in Table 1.

At the materials level, it can be understood that crude oils vary from light to heavy (referring to density and viscosity), and from sour or sweet (reflecting lower temperatures at which corrosion occurs with sour crudes). Even at this level, such understanding can be applied to adjusting the refining process. At the chemical level, the knowledge that crude oils are a mixture of chemicals such as butane (a paraffin), benzene (an aromatic), cyclohexane (a napthene), butene (an alkene), butadiene (an alkyne), and many other chemicals of these types provides deeper understanding of the differences between light and heavy crudes which may be required to design the process. At the molecular level, such knowledge as that paraffins have the general structure C_nH_{2n+2} while aromatics have the general structure C_6H_6-hydrocarbon branch(es) is necessary for some process design aspects. At the atomic level, the knowledge that the presence of sulphur atoms causes reactions with iron alloys at lower temperatures is also necessary to address adjustments for sour oils. At the quantum mechanical level, the knowledge that some electrons in the benzene ring are delocalized to the point that they are shared across all the carbon atoms in the ring is also necessary to understand the relative stability of such rings.

Note, however, that any attempt to describe a complete oil refining process in quantum mechanical terms would need to follow the behaviour of perhaps 10^{30} electrons, protons, photons etc. and such a description would be completely incomprehensible to human intellect. In practice, most of the thinking about the process occurs at the materials and chemicals level, but with the awareness of when the descriptions at these levels will become inaccurate and more detailed levels will be required, and of how to shift to more detailed levels. More detailed descriptions can only cover very tiny segments of the entire process (the interactions of a few molecules, for example), but there are also rules (such as statistical mechanics) for how to scale up.

Consider some key properties of this hierarchy. At the most detailed level, there are often relatively few possible types of entities (electron, proton, photon etc.) and relatively few types