

Modern Theories of Consciousness: Some Alternatives

The problem

Many philosophers have believed that it will forever prove impossible to explain mind in terms of matter. The famous philosopher, mathematician and physicist Leibniz (1646 – 1716) expressed this fundamental intuition thus: “Supposing that there were a machine whose structure produced thought, sensation and perception, we could conceive of it as increased in size with the same proportions until one was able to enter into its interior, as one would into a mill. Now, on going into it one would find only pieces working upon one another, but never would one find anything to explain perception” (Monadology, §17)

Nowadays the problem is often formulated in terms of brain and consciousness: Can consciousness be explained in terms of brain processes? We know a lot about how mental processes depend on neural processes. But what we may call “Leibniz’ problem” – that no matter how exhaustively you study the brain as a material mechanism, it seems that you will only find further material phenomena, not consciousness – is still with us.

Professor Magne Dybvig,
Department of Philosophy,
Norwegian University of Science
and Technology, NTNU.
CAS Fellow 2003/04.



The relationship between mind and brain: The main positions

In modern times, before the 20th century, the most popular interpretation of the mind-brain relationship was some version of *dualism*. It claims that mind is essentially non-physical. The brain is the place where this non-physical reality interacts with physical reality. The reason why you cannot “see” the mind when you inspect the brain is that the methods of inspection are adapted to the observation of material phenomena, and not to the observation of immaterial phenomena like e.g. thoughts. So what you can inspect using the methods of the natural sciences, is at most the *correlates* of consciousness, not the conscious itself.

In the 20th century, a series of *materialist*, or *physicalist*, alternatives to dualism have been developed. The main positions are (philosophical) *behaviourism*, the *identity theory*, *functionalism* and *eliminativism*.

Behaviourism: According to (philosophical) behaviourism the mind is simply the behaviour, or dispositions for behaviour, that an organism exhibits. The brain is not the mind, but the mechanism that enables mind – i.e. the underlying mechanism that enables the complex behaviour which *is* the mind. And the reason why you cannot observe mind by

simply observing the brain is not that mind is something immaterial. The reason is that you are so to speak looking in the wrong place – at the mechanism that makes mind possible, not at mind (the behaviour) itself.

Identity theory: A frequent objection to behaviourism is that we think of mind not as the behaviour itself but as what *causes* and *regulates* behaviour. And what causes and regulates behaviour are brain states; so mental states *are* brain states according to the (neural) identity theory. This mind-brain identity must be accepted as a kind of scientific truth, comparable to e.g. the identity of light and electromagnetic waves. So the states that you inspect when you inspect the brain *are* (some of them) mental states – it is only that you will not *recognise* them *as* mental states until you have developed the right ‘theoretical spectacles’.

Functionalism: An objection to the identity theory is that mental phenomena, e.g. pain, can be realised in the brain in many different ways, depending on what kind of organism we are talking about. According to functionalism, mind is not brain states, but something more abstract – namely the functional states the brain can be in. Anything (e.g. a complex robot, or an extraterrestrial being) with inner states that performed the right functions would have a mind, even if it did not have a biological brain. In functionalism the relationship between brain and mind is often compared to the relationship between hardware and software. And the reasons why you cannot observe mind by just observing brain processes, is that you are not focusing on a sufficiently abstract level – you are like an engineer who does not understand a computer because he only sees the electronic *hardware* and not the *software* (i.e. the set of programmed functions) that runs on this hardware.

Eliminativism/instrumentalism: What is common to behaviourism, identity-theory and functionalism is a belief that mental phenomena are real phenomena that can, in the end, be described in terms taken from the natural sciences (including biology) – either as behaviour, or neural states, or functional states. Eliminativism maintains that this is not the case – our common sense conception of mind is a *theory* of mind (“folk psychology”) that is basically wrong, so that nothing corresponds to mental phenomena “in the real world”. A correct theory will only refer to brain states and behaviour, not mind. Mind is at most a *useful fiction* (instrumentalism); and the reasons why you cannot observe the mind by observing the brain, are simply that the mind does not exist – there is no mind to observe.

Different phenomena may require different theories

None of the theories mentioned above have been generally accepted among philosophers working on the mind-brain relationship. Many look on themselves as some kind of materialists (or “physicalists”). Few are fully-fledged dualists, but elements of such a position can also be found in contemporary philosophy – notably the following two points:

Consciousness cannot be completely reduced to brain processes, and the study of it requires (in addition to methods found in the natural sciences) some special methods – a special kind of *self-observation* (introspection, or “phenomenological descriptions”) and perhaps some kind of *interpretation* of behaviour (similar to the interpretation of texts).

It may be that different theories fit different types of mental phenomena. In philosophy it is for instance usual to distinguish between states that have a kind of semantic content, similar to the content of

words (e.g., both a thought and an assertion can have the content ‘it is raining’), and states that lack such a content (sensations). The first are sometimes called propositional attitudes (or intentional states), while the latter are called qualia. It may be that the understanding of *propositional* attitudes requires some kind of interpretation, and that a full understanding of qualia is impossible without some kind of introspection. It has also been claimed that while propositional attitudes can be understood in functional terms, qualia require a dualist theory – they are irreducibly mental. And some materialists (who reject dualism) have claimed that a functional theory may be true for propositional attitudes, while some kind of identity theory is true for qualia.

The answer may also depend on how we conceive the relationship between mind and brain

Traditionally philosophers have thought of the relationship between mind and matter either in terms of identity (‘the mind is nothing but brain states and/or behaviour’) or in terms of causality (‘mind is different from brain states, but somehow caused by brain states’). Lately it has been proposed that it would be better to think of the relationship as a kind of *supervenience*-relationship. Mental states supervene on brain states if it is impossible to have a change of mental states without *some* change in brain states. Or conversely: Complete similarity in brain states entails complete similarity in mental states. Such a relationship implies that the mental is a kind of function of the brain even if it should prove impossible to formulate exact causal laws for how mind depends on the brain.

It has also been pointed out that *individual* mental events (e.g. the pain that I feel just now) can be identical with individual brain events (e.g. the firing in C-fibres going on just now) without *the properties* of mental events necessarily being identical with neurological properties. The first type of identity is called “token identity” while the latter is called “type identity”. If this view is accepted one can for example say that the pain I feel is in fact token-identical with some brain event, while it has properties (e.g. ‘being a throbbing pain’) which *cannot* be identified with neurological properties (though they probably *supervene on* such properties). Such a view is often called *non-reductive* physicalism, and may be considered a kind of compromise between a physicalist and a dualist position.

Understanding mental phenomena may require a 3-stage functionalist strategy

Among the physicalist theories mentioned above, it is probably functionalism that has had most adherents in recent years. If you accept such a theory, it is natural to claim that the study of mind has three stages or dimensions:

Stage 1: Formulate a functional analysis of the mental phenomenon in question

Stage 2: Describe a psychological mechanism that implies (“implements”) the function

Stage 3: Describe a physiological mechanism that implies (“implements”) the psychological mechanism

Let us take ‘consciousness’ as an example. What does it mean that a state is conscious rather than unconscious? If you are a functionalist, you will try to describe the characteristic way in which a conscious (in contrast

to an unconscious) state functions. One proposal is that I can immediately make use of what I am conscious of, and that I can use it for many different purposes. So a functional analysis of consciousness could be: Consciousness = immediate & global availability.

Next you might ask what kind of cognitive mechanism implements (makes possible) this immediate and global availability. A possible proposal is e.g. that mind is organised around a ‘global workspace’ (a kind of ‘working memory’; Baars 1988) and that a piece of information is conscious at a certain moment in so far it forms part of the global workspace at that moment.

The third question will then be what kind of *neurological* mechanisms implement the *psychological* mechanism (e.g. ‘global workspace’ or ‘working memory’) which implements *consciousness*, functionally defined. A proposal is that some kind of synchronized firings of neurons engaged in related tasks may play a key role in integrating (“binding”) and storing information in working memory – or Baars’ ‘global workspace’ – so that they become immediately and globally available in action and speech (for a brief survey of some cognitive and physiological mechanisms that might implement consciousness, see Chalmers 1996).

Is there an ambiguity in the concept of consciousness?

Though the above proposal for a 3-stage functionalist analysis of consciousness is highly speculative, one can at least see how consciousness ‘in principle’ could be a necessary consequence of the way in which the brain is organised – *if* a functionalist analysis of consciousness is possible. But is such a functionalist analysis really possible? It has been claimed (e.g. Block 2002) that we must distinguish between two concepts of consciousness. The first is consciousness as what one has *access to* (“access-consciousness”); the other is consciousness as *experience* (“phenomenal consciousness”). The first may be analysed in functional terms (more or less as suggested above); the other cannot be so analysed.

If one accepts this distinction, one may also claim that “consciousness” covers different phenomena that require different theories, as suggested above. One possibility is to combine a functionalist theory of access-consciousness with some kind of dualist theory (or non-reductionist materialist theory) of phenomenal consciousness. Another possibility is to combine a functionalist theory of access-consciousness with an identity-theory of phenomenal consciousness, where the latter is taken to consist in the occurrence of sensations (“qualia”).

References

- Baars, B. J. 1997. *In the Theatre of Consciousness*. Oxford: Oxford University Press.
Block, N. 2002. Concepts of Consciousness. In Chalmers 2002.
Chalmers, D.J. (ed.) 2002. *Philosophy of Mind*. Oxford: Oxford University Press
Chalmers, D.J. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
Leibniz, G.W. 1961. *The Monadology*. In Leibniz: Philosophical Writings. London: J.M. Dent & Sons.