Metaphysics of Consciousness

Ole Koksvik


Drinking a glass of cold water on a hot day feels a certain way. It is hard not to wonder why it feels that way, and indeed why it feels any way at all. In an influential and evocative way of speaking, a being is conscious just in case there is something it is like to be it (Nagel 1974). Similarly, a mental state is conscious just in case there is something it is like to be in that state. When there is something it is like to be a being, we say that that being has phenomenal experience, and when there is something it is like to be in a certain mental state, we say that that state has phenomenal properties, or phenomenology. (This notion of consciousness contrasts with access consciousness, see Block 1995, with which we shall not be concerned here.)

There are many interesting philosophical questions about phenomenal consciousness. For example: What is the relation between a creature’s being conscious and a mental state’s being conscious (Van Gulick 2009: section 2)? How is the overall phenomenology of a conscious being related to the more specific phenomenal experiences that being has (Bayne and Chalmers 2003)? What is the relationship between (phenomenally) conscious mental states and mental states that represent the world as being a certain way (Siewert 1998: ch. 8; Horgan and Tienson 2002; Chalmers 2004; Crane 2001: ch. 3; Pitt 2004)? Do all conscious mental states have content, and if so, of what kind (Siegol 2008)? Do we stand in a special relation to our own conscious states, a relation, e.g., of special authority, incorrigibility or infallibility (Macdonald 1995; Gertler 2008; Williamson 2000: ch. 4)?

Australasian philosophers have made important contributions to the philosophical understanding of consciousness in many areas, but there is insufficient room here to discuss them all in appropriate detail. Accordingly, a narrower focus is adopted: this entry shall be exclusively concerned with the philosophical treatment of certain metaphysical questions about consciousness, in the analytic tradition. (Even with this narrower focus, there are inevitably many regrettable omissions and simplifications.) What kind of thing is consciousness, and how does it fit in with the rest of the world? Australasian philosophers’ attempts to answer these questions have been enduringly influential.

The metaphysical questions about consciousness are questions to which one’s initial puzzlement about consciousness can quickly lead. They are also questions which have become absolutely central to the philosophy of mind, both in Australasia and worldwide. Why is that? An important part of the explanation is that apparent progress toward showing how mental states such as beliefs and desires fit into the rest of the world often seems incapable of being generalised to also show how consciousness fits in. Consciousness has thus come to occupy much the same role as was previously occupied by a more general concept of mind: it stubbornly resists explanation. One way to view the positions discussed below is as attempts to address this unsatisfactory situation.

The Identity Theory

The identity theory of the mind was developed at the University of Adelaide by Ullin Place, an English psychologist who was a lecturer there from 1951 to 1954. Place was
strongly influenced by discussions with J.J.C. Smart (on whom more below) and C.B. Martin. Martin was an emergentist (and not a materialist) who worked in Adelaide from 1954 to 1966. He then became a professor in Sydney, where Armstrong was. Although he did not at the time publish on the subject, his influence on the philosophers who met him is widely acknowledged.

Place (1956) argues that a reasonable scientific hypothesis is that the “intractable residue” of conscious experience is identical with processes in the brain. While the metaphysical independence of (kinds of) entities can often be inferred from the logical independence of statements about them, this is not always so, and conscious states and brain processes constitute one of the exceptions. (On this inference, see also Putnam 1973: 73-74.) In general, commonsense observations and scientific observations should be taken to be observations of the very same phenomenon whenever the latter, together with relevant theory, provide “an immediate explanation” of the former (Place 1956: 58). That is precisely what Place expects to see as our understanding of the brain advances: patterns emerging in the study of brain processes will eventually allow us to explain all our introspective observations.

J.J.C. (“Jack”) Smart, born in Cambridge and educated at the Universities of Glasgow and Oxford, held a chair at the University of Adelaide from 1950 to 1972. Originally a behaviourist view, he was convinced by Place (and also influenced by Feigl 1958) to adopt the identity theory. Smart (1959) argues that we must either understand conscious mental states as “nomological danglers” (the term is due to Feigl) or identify them with brain processes. We have, he says, good reasons to reject nomological danglers but no good reason to reject the identification, so the identification should be accepted. (Smart spends the best part of the paper replying to objections, objections from ordinary language considerations among others; with respect to the latter Martin, who was staunchly opposed to ordinary language philosophy, may have been influential.) Nomological danglers are epiphenomena, caused but not themselves causally efficacious. Smart’s objection to such entities seems to be based first on a denial of the very possibility of entities connected to the rest of the causal machinery of the world in this way, and second on the view that if such entities did exist, the laws relating them to the rest of the world would be strange, because they would relate microphysical objects with macroscopic phenomena.

The latter thought is presumably motivated by ‘unity of nature’ considerations: we find that the laws that do the real explanatory work elsewhere relate small entities to other small entities, so that is what we should expect where the mind is concerned as well. Whether this adds much independent weight to an already monistic outlook is perhaps doubtful: if, for independent reasons, consciousness is regarded as truly unique, the appearance of unusual laws relating it to other parts of the world should not be surprising.

Another aspect of Smart’s article is worth noting, because it may partially explain why it became so influential, even though it was largely concerned with defending a claim already made by another. Place had called the sense of identity he employed “the ‘is’ of composition” and had introduced it by means of analogies with cases such as someone’s table being an old packing case and someone’s hat being a bundle of straw (Place 1956: 56). This seems to leave at least some room for a distinction between the experience and the brain process: if four legs plus a tabletop compose a table, the result is usually taken to be six, and not five, distinct objects in
total. In contrast, Smart insists that sensations and brain processes are strictly identical (Smart 1959: 62).

The identity theory is often understood as claiming that when I experience pain, a certain (type of) brain process just is my experience. But other animals do, and extraterrestrial beings may, have brains that differ substantially from ours: they may not have brain processes of the same kind as mine (on a specification of kinds of brain processes narrow enough to yield sufficient variation in conscious states). Yet this does not seem to be a good reason to conclude that they do not feel pain, and it is compatible with having strong reasons to think that they do. If causing tissue damage to alien life forms causes them to retract the damaged body part, if they appear to be strongly opposed to having their tissue damaged and strongly motivated to bring about the cessation of the damaging process and to ensure that it is not repeated, etc., then it seems that we would have very good reason to think that they feel pain, and any knowledge of the aliens’ innards would not affect that. (Note that there is no reliance here on the claim that a certain repertoire of behaviour is all there is to pain.) Thus pain seems likely to be multiply realisable: it can be instantiated in brains quite different from ours, and perhaps even in systems we would not recognise as brains.

**Functionalism**

Functionalism is, roughly, the view that mental states are states whose identity or character is exhausted by the causal relations it stands in to (sensory) inputs, (behavioural) outputs and other mental states (Putnam 1973: esp. 76; Levin 2009; Lycan 1994; Block 1994; on the relationship between functionalism and the identity theory, see Jackson 1998: section 2 and Smart 2008: section 5).

Varieties of functionalism differ with respect to the information considered relevant for the individuation of mental states. On some views, the relevant information is available to everyone (at least all competent adults) in virtue of shared beliefs about how beliefs, desires, etc., respond to stimuli, interact with each other and result in behaviour. On other views, the pertinent information is that which has resulted or will result from scientific psychology (see Block 1994: 325). Australasian philosophy is associated especially with the former type of view, versions of which are defended by, e.g., David Armstrong, David Lewis, David Braddon-Mitchell and Frank Jackson. Here I concentrate on the first two.

Armstrong is a Melbourne-born philosopher, educated in Sydney and at Oxford (see Jackson 1998). He has lectured at London and Melbourne universities, and was the Challis professor of philosophy at University of Sydney from 1964 to 1991. “The concept of a mental state”, he argues, “is the concept of something that is, characteristically, the cause of certain effects and the effect of certain causes”, and this is all there is to our mental state concepts (1981: 21). Characteristic causes of mental states are other mental states as well as events and objects in a person’s environment; a characteristic effect is behaviour.

As is common with philosophical theories, functionalism is often put forward in sketch form: we are told in rough outline how the theory will analyse mental concepts and asked to trust (or to share the intuition) that the details can be filled in. One of Armstrong’s very significant contributions in *A Materialist Theory of the Mind* (1968) was his attempt to provide an analysis of a range of important mental concepts in considerable detail (see also his 1973 and 1981).

As one might expect, beliefs and desires play a central role. “[P]erception”, for example, “is nothing but the acquiring of true or false beliefs concerning the current state of the organism’s body and environment” (1968: 209). A challenge for
Armstrong’s position is that we are sometimes subject to known illusions, and in those cases we do not (generally) believe what we see. To account for this, Armstrong says that perception is either the acquisition of beliefs, the acquisition of degrees of belief (credences) which are ‘held in check’ by stronger credences (p.221), or acquisitions of dispositions to believe (pp.222-23). (See George Pitcher 1971: esp. 91-93 for a very similar account); Frank Jackson (1977: ch. 2) later argues that these manoeuvres fail to salvage the theory.

The key to see how the causal theory he advocates can encompass conscious experience is, Armstrong argues, to recognise that experience is transparent. (Armstrong is an early advocate of this thesis, which has received much attention of late.) We are not aware of properties of our experiences; what we are aware of are only the properties of the objects of those experiences. For example, in perception, the redness associated with certain perceptual experiences is to be understood as the redness of the perceived object (1981: 27-29). And this allows us, Armstrong thinks, to capture conscious mental states in the causal story: just as with all other mental states, a conscious state is that which has certain characteristic causes, like red objects in the environment (where ‘red’ is cashed out in terms of physical properties, such as surface reflectance profiles).

A similar account was developed independently (and published slightly earlier) by David Lewis. Lewis was American but had a strong association with Australia, due in large part to his close friendship with Smart. He visited Australia more than twenty times from 1971—when Smart had organised for him to give the Gavin David Young lectures at the University of Adelaide—to 2001, and is now considered an ‘honorary Australian’, at least by Australian philosophers (see Weatherson 2009; Nolan 2005). “The definitive characteristic of any (sort of) experience as such”, Lewis argued, “is its causal role, its syndrome of most typical causes and effects … which belong by analytic necessity to experiences” (1966: 17). One of Lewis’ distinctive contributions is his development of a general method for defining theoretical terms, which he then applied to mental state terms to yield an argument for the identification of mental states, including conscious states, with brain processes (see his 1970 for the development, his 1972 for a less technical presentation, and his 1995 for more details; the argument for the identification is in his 1966).

Start with a theory formulated by a long sentence (formed, perhaps, by conjoining the sentences that express the theory), ‘the postulate of T’: $T[t_1 \ldots t_n]$. Replace each of the $n$ terms which occur in that sentence with a new variable, $x_1$ to $x_n$, and then existentially quantify the result. This is the Ramsey sentence for the theory. It is silent on how many sets of entities stand to each other in the relations postulated by the theory; it claims only that at least one such set exists. Lewis argues, however, that theoretical terms are best understood as uniquely referring, or else as not referring at all. The introduction of a theory should be understood as claiming that there is exactly one set of entities which satisfy the theory, and so the modified Ramsey sentence, which states that the theory has a unique realisation, is what is of real interest: $\exists_1 x T[x]$ (or $\exists y \forall x (T[x] \equiv x=y)$).

This much follows, according to Lewis, from our concepts along with conventions for the introduction of new concepts (1972: 254, 1970: 439-40). Lewis argues that if empirical science discovers what actually uniquely realises the theory $T$, we are compelled, as a matter of logic, to identify the referents of our concepts with these realisers. So if, as he believes, a specification of conscious mental states in functional terms can be extracted from folk psychology, and if, as he thinks highly
likely, physical science eventually isolates the unique realisers (or near enough realisers) of those specifications as neural states, then the identification of conscious experience with neural states is forced upon us (Lewis 1966).

Functionalism can be seen as a response both to objections to the identity theory and to objections to behaviourism. Functionalism seems to allow for the multiple realisability of mental states, which presented a difficulty for the identity theory. For example, pain is the occupant of a certain functional role in the mental organisational structure (or, perhaps, the second-order property of having that role occupied; see Lewis 1995: 419-21). Provided that a functional characterisation can be given which is general enough to encompass all creatures who plausibly feel pain, the unpalatable implication that creatures with brains that differ from ours do not experience pain appears to be avoided. (An implication may be that there is ambiguity in the concept of pain; see Lewis 1983: esp. 128. However, for an argument which purports to show that functionalism is also vulnerable to the objection from multiple realisability see Block 1994: 330.)

For at least simple versions of behaviourism, a difficulty is that no single behavioural disposition is associated with a belief: which behaviour a belief will bring about depends on which other beliefs the person has, as well as on the person’s desires (Geach 1957: 8). But according to functionalism, the array of causal relations which individuates a mental state includes relations with sensory input, behavioural output and other mental states. So even if some creatures suppress all pain behaviour (Putnam 1968) and even if there are perfect actors, who imitate having an experience perfectly, functionalism would not force us to the mistaken conclusion that the creatures lack painful experiences or that the actor has them.

Functionalism seems to retain from behaviourism and the identity theory the virtue of offering a possible way of integrating the mind with the physical world (while also being compatible with their non-integration; see Block 1994: 326, 330): “if the concepts of the various sorts of mental state are concepts of that which is...apt for causing certain effects and apt for being the effects of certain causes, then it would be a quite unpuzzling thing if mental states should turn out to be physical states of the brain” (Armstrong 1981: 21). One might again ask, however, whether that promise really extends to consciousness. Are conscious states individuated exhaustively by their functional roles?

**Dualism**

Powerful arguments to the effect that they are not are presented by Australian dualists. In his seminal article “Epiphenomenal Qualia” (1982) Frank Jackson argues that facts about phenomenal experience are left out of all explanations restricted only to physical facts, even functional explanations. Through two thought-experiments Jackson presents his ‘knowledge argument’ for the view that physical information must leave something out. In one of them (for the other, see 1982: 130 and 1986) Jackson asks us to imagine that we encounter a person, Fred, who is capable of making a colour discrimination we cannot. For Fred, the things we classify as red fall into two groups, red₁ and red₂, as different from each other as yellow and blue are to us. Jackson argues that no amount of physical information, including functional information, will tell us what it is like to have the colour experience Fred has when he sees a colour that he, unlike us, can discriminate from the others. Therefore, there is more to know than what is encoded in physical information (1982: 128-30).

The knowledge argument had appeared, although not by that name, some 55 years earlier. John William Dunne argues that there is “a characteristic of red of
which...all seeing people are very strongly aware”, such that a blind person who has been told all the physical facts still “would have not the faintest shadow of an idea that [seeing people] experience anything of the kind” (1927: 15). However, Jackson’s vivid presentation of the argument generated a flurry of activity (see, e.g., Ludlow, Nagasawa and Stoljar 2004), and the fact that he himself no longer endorses the knowledge argument has not stopped the paper from remaining one of the most discussed and influential papers in recent philosophy of mind.

Another very influential argument for dualism, often said to have revived the debate over dualism, is David Chalmers’ ‘zombie argument’ (1996: 94-99). Chalmers argues that it is logically possible that there be something—a phenomenal zombie—which replicates my physical makeup and (therefore) my functional organisation down to the most minute detail but which has no phenomenal experience. He claims that the possibility of a zombie twin shows that the facts about my functional organisation and physical makeup do not entail the facts about phenomenal experience. There is more to know than what the physical sciences can tell us. (For predecessors to the zombie argument, see Chalmers 1996: ch. 3, n.1. Chalmers in fact discusses five different arguments, which he regards as all pulling in the same direction.)

These arguments purport to show that what one can learn from one set of facts is not all there is to know. One might think that the metaphysical question about consciousness, whether or not consciousness itself is physical, is not immediately settled even if the knowledge and zombie arguments are successful. In particular, a popular thought has been that it is might be metaphysically necessary that consciousness is a brain process even if it is conceptually or logically possible that consciousness is something else. Both Jackson and Chalmers, however, take their arguments to have the strong metaphysical consequence that conscious experience is not identical to physical or functional states or processes. (See Chalmers 1996: ch. 4, esp. 131-40, and Jackson 1986: 291).

Important predecessors to both these arguments are found the work of Keith Campbell, who succeeded David Armstrong as Challis Professor at the University of Sydney in 1991. Behaviourist and causal (functional) analyses, he argues, leave out “the very thing which matters most about [conscious states]. Pains hurt: indeed that is their most salient feature” (1970: 71-72), and, in general, these theories cannot capture what it is like to be an experiencing subject (104). Campbell also briefly discusses an ‘imitation man’, a being who, like Chalmers’ zombie, lacks phenomenal experience, but is functionally similar (in Chalmers’s case, there is functional identity) to experiencing subjects (100).

The views on the metaphysics of consciousness discussed in this entry all have contemporary defenders, and are all controversial. A view that has not been discussed is that there really are no phenomenal experiences at all (a view often associated with Dennett 1988, 1991). Whether that view ought to be taken seriously is a question not discussed here. For those who remain convinced that there are conscious experiences, however, the continuing debate over their nature is profoundly influenced by the contributions of Australasian philosophers.

There are many other important Australasian contributions to research on consciousness (here again I apologise for inevitable omissions). Responses to Jackson’s knowledge argument have been put forth by John Bigelow and Robert Pargetter (1990, 2006), Lewis (1990), Cynthia Macdonald (2004), Philip Pettit (2004), Denis Robinson (1993), and Jackson himself (2003, 2004). The thesis that the zombie argument and similar arguments misleadingly seem plausible to us because
we lack knowledge of some physical truth has been given detailed defence (Stoljar 2006), and the view that the zombie intuition guides meaning while being metaphysically idle has been explored (Braddon-Mitchell 2003). Phenomenal concepts, the content of phenomenal and intentional states, and our knowledge of that content have been investigated (Chalmers 2004, 2005, 2003; MacDonald 1995). The lessons to be learned about consciousness from cognitive science is a focus of Australasian research (Hohwy and Frith 2004a, 2004b; Hohwy 2007; O’Brien and Opie 1999, 2000; Shea and Bayne forthcoming), as is the nature and phenomenology of perception (Bayne 2009; Fish 2008, 2009; Schellenberg forthcoming a; forthcoming b). Finally, at the Australian National University, the Centre for Consciousness continues to be a hub of research and discussion of the many intriguing philosophical questions about consciousness.

(I am grateful to John Bigelow, David Chalmers, Daniel Stoljar, Weng Hong Tang and the editors for very helpful discussion. I am, of course, solely responsible for any remaining shortcomings.)

References
Where two dates are given, the date in square brackets refers to the cited version, the other to the year of original publication. In the main text, the latter is used, to give the reader as good a sense of the chronology as possible.


Bayne, Tim; Chalmers, David J. ‘What is the Unity of Consciousness?’ In The Unity of Consciousness: Binding, Integration and Dissociation. Oxford: Oxford University Press.


