Is Mind-Brain Identity Theory a Valid Solution to the Mind-Body Problem?

The mind-body problem refers to that conundrum concerning how, and if, our physical bodies can give rise to and interact with our internal experience of a self. Dualism, the school of thought which suggests that the mind and body are strictly distinct, has been a dominant view throughout philosophical history. However, the success of empirical science has lead many philosophers to adopt a physicalist perspective; one which suggests that there is only one type of substance in the universe - the physical - upon which everything is said to supervene. Mind-brain identity theory (hereafter abbreviated to MBIT) is a physicalist theory which states that mental states simply are physical states, specifically of the brain. The theory, then, relies upon an ontological reduction or, in others words, the claim that one category of things is in fact one and the same with another. In this essay, I will be examining the validity of MBIT as a solution to the mind-body problem.

It appears to me that MBIT ought to be the default philosophical position in the absence of convincing counterarguments. The justification for this is given by Ockham’s razor, which suggests that when faced with competing hypotheses, all of which are consistent with the facts, the simplest one ought to be preferred. In the matter of MBIT, the claim that our subjective experiences simply are states of the brain is surely simpler than the claim of the dualist, that there exists a Cartesian ‘ghost in the machine’, or indeed the claims of other kinds of physicalists (logical behaviourists, for example). The theory is also persuasive in that it avoids the problem of casual interaction faced by many other theories of mind; if the mind and brain are not distinct but one and the same, it is quite clear that an explanation of how the mind interacts with the body is unnecessary. And, of course, the increasing ability of modern science to explain even the most seemingly mystical of phenomena in mechanistic terms suggests that consciousness may too, one day, be considered in this light. However, it is worth noting that science could only ever establish a correlation of mental states with physical states, not an identity between the two. This lies purely within the realm of the philosophical. And given that I believe that the burden of proof is on those who oppose MBIT, to justify the theory, one must evaluate the philosophical objections commonly cited against it.

Dualist philosopher Descartes raised a common objection to MBIT in his ‘Meditations on First Philosophy’, known as the divisibility argument (see Appendix A). Like many arguments against MBIT, this argument relies upon Leibniz’s indiscernibility of identicals, which states that if \(a\) is identical to \(b\), \(a\) and \(b\) must share all of their properties and differ in none. Descartes claims that the mind and the brain must be distinct as they differ in the property of divisibility; the body is divisible, and the mind is not. Therefore, according to Descartes, the mind cannot be the same as

\[\text{Gilbert Ryle, The Concept of Mind (1949)}\]
the body. This claim, however, is contestable. This could be said to beg the question.
To assume that the mind has the property ‘not divisible’ is to assume that the mind is a substance, as only substances can have properties, according to the philosophical definitions of ‘property’ and ‘substance’. But this is the very claim which Descartes is setting out to prove. Descartes, then, seems to be mistaken in his thinking about the properties of minds, but should instead think about mental properties, properties of physical objects. In this way, it becomes clear that the indivisibility of mental properties, such as thinking, is analogous to the similar indivisibility of physical properties, for example, being wet. These physical properties are also indivisible, but they can still be states of divisible physical objects, such as bodies. Therefore, there is no reason to believe that a divisible substance cannot possess properties which are not divisible, both mental and physical. It should be said that a similar objection to MBIT - that the brain and the mind cannot be identical as the former can be described spatially whilst the latter cannot - can be dealt with in the same way. Physical states also cannot be described spatially (it would be absurd to say that my being cold is three meters long), and yet we do not say that they are in some way distinct from our physical bodies. The same, I would argue, can be said for mental states. Thus, neither of these first two objections to MBIT appear to stand.

Perhaps the most famous objection to MBIT was devised by Putnam, in his paper ‘Psychological Predicates’ (see Appendix B). In this objection, known as the multiple realisability hypothesis, Putnam argues that it is unlikely that every animal - or indeed, possible extraterrestrial life - in every possible mental state has a brain state corresponding to ours when we experience this same mental state. It is much more likely, he suggests, that mental states are multiply realisable; they can be produced by many types of brain states, across many species, and so no identity can be maintained between the two. This argument, however, can be countered in two ways. The first is to make the distinction between two forms of MBIT. Type-identity theory is the kind which has been discussed thus far; the theory which states that particular mental states are always identical with physical states. A second kind is known as token-identity theory. This is based on the distinction between a type and a token of something. To use an example, Descartes’ ‘Meditations on First Philosophy’ is a type of book, but everyone who owns it still owns an individual book; a ‘token’ of the ‘type’. Token-identity theory, then, suggests that particular mental states are identical with particular physical states, in particular organism at a particular time, but that the same mental state - the type - may be created differently by individuals, or tokens. So in a person, pain may be caused by the firing of C-fibres, but in a reptile, pain may be caused by a different physiological event. It still, however, maintains an identity; in both cases, the pain still is the physiological event. Multiple realisability is built into the theory and so, if it is proven correct, it is no problem for the token-identity theorist. Opponents of token-brain identity theory, however, may suggest that this response has problems of its own. If there are many types of brain and, correspondingly, many ways in which pain can be realised, it might be asked how the token-identity theorist can say that all of these different brain states are “pain”? If the token-identity theorist responds that the identity can be deduced as creatures
undergoing this brain state *behave* as though in pain, then they might be said to have reverted to a behaviourist/functionalist definition of pain. To this, I would answer that, though the qualia may be deduced from pain behaviour, this is not to say that the experience *is* the behaviour, but rather that the behaviour is indicative of the experience (which can then be ontologically reduced without problem). It is, I would argue, analogous to a litmus test; the blue colour is not the alkaline, but rather is indicative of the alkaline, and the alkaline can then be ontologically reduced to a lack of hydrogen ions. Another way of countering the multiple realisability hypothesis, I would suggest, is to challenge the assumption that other species experience pain at all. Presumably, there are only two ways to determine if a creature is in pain: firstly, by examining its neural activity or secondly, by looking at its behaviour. If Putnam justifies his claim by the former, then he has surely conceded that neurophysiological events are identical with mental states. If by the latter, (or by a functionalist combination of neurophysiology and behaviour) then he begs the question; he assumes that animals experience pain by looking at their behaviour, but that pain behaviour is part of the definition of mental states is precisely the claim which the functionalist sets out to prove. Either way, it appears to be that, so long as token-identity theory is advocated over type, multiple realisability presents no problems for MBIT, at least as formulated by Putnam.

Another prominent criticism of MBIT was formulated by Jackson in “Epiphenomenal Qualia” (1982), in which he devised a thought experiment known as Mary’s Room (see Appendix C). Mary’s Room attempts to strengthen the notion that MBIT, and physicalism in general, leaves something crucial and undeniably real out of our picture of the world; our subjective experience, or qualia. Indeed, Jackson calls himself a ‘qualia freak’ in “Epiphenomenal Qualia”\(^2\). Many conclude from Jackson’s argument that physicalism is incapable of explaining this aspect of consciousness, and so it must be false. This, however, is not an argument I find compelling. Perhaps the strongest objection is to state that Jackson’s assumption that Mary learns something new is unjustified; he simply asserts that it is ‘just obvious’ that Mary will learn something new upon leaving the room. But intuitions such as this surely have no place in rational, philosophical debate. Mary’s room, then, proves nothing - it appears to me merely a convoluted statement of a particular viewpoint. The assumption is even less convincing when we realise that we do not, with our current understanding of neuroscience, understand what having ‘all the physical information’\(^3\) would be like. Jackson also seems to begin from a dualist’s perspective by the assumption that Mary learns something new. However, we surely ought to do the opposite; by Ockham’s Razor, we should start from the simplest position (physicalism) and only convert if proven otherwise. Our starting assumption ought to be that Mary does not learn anything new, is not surprised, and knew exactly what seeing colour would feel like. As Mary’s Room does nothing to dispel this claim, it is

---

\(^2\) See Appendix C

\(^3\) See Appendix C
not a convincing argument against physicalism. Jackson himself accepted an objection similar to this, and accepted physicalism in ‘Postscript on Qualia’\(^4\). Others have argued that Mary doesn’t learn a new fact, but instead learns about the same fact in a different way. To give an analogy, one might say that one knows there is water in the reservoir. This is not the same, however, as saying one knows that there is H2O in the reservoir; one could know one fact and not the other and so, when told ‘there is H2O in the reservoir’, one might be said to learn something new. This, however, does not change the fact that water still is H2O. One has not learned about a new substance - H2O - which always co-inhabits the reservoir with water. Rather, one has learned a different way of thinking about the same fact; that there is something in the reservoir which can be called water or H2O. Mary also does not learn any new fact, or gain any new propositional knowledge, upon leaving the room. Rather, she merely learns a different way of thinking about colour; through qualia, rather than neurophysiology. This does not, however, challenge the identity of colour with physical events, any more than realising there are two ways to think about water and H2O changes that the two are really one and the same. Some philosophers have suggested that Mary does gain new knowledge, but it is acquaintance knowledge (knowledge gained by being directly acquainted with something) or ability knowledge (knowing of how to do something), not propositional knowledge (knowledge which has a truth value). If this is the case, Jackson’s argument breaks down, as no facts escape Mary’s physicalist perspective, and yet we can still concede that she learns something. Whichever approach is taken to counter Mary’s Room, it seems to me that it is not a convincing argument to disprove MBIT.

In conclusion, it appears to me that mind brain identity theory is a valid solution to the mind-body problem. By Ockham’s Razor, it is the theory which should considered most rational. It avoids many of the mystifying problems faced by dualists, such as that of causal interaction, and appears to be undefeated by many of the most prominent criticisms against it, such as the divisibility argument and other arguments from Leibniz’s Law, the multiple realisability hypothesis and Mary’s Room. However, this judgment has implications for many areas of philosophy. Surely we cannot have free will, for example, in a world where our consciousness is reducible to physical brain states, links in the deterministic chain? Surely, without personal autonomy, in a universe where we are just bio-chemical organisms, moral responsibility becomes unsustainable? And surely, if the mind is identical with the body, it must be inseparable from it, challenging the teachings of religions all over the world? With this conclusion, then, comes a barrage of new puzzles, quite as complex as the last.

Word count: 2, 161

\(^4\)https://faculty.unlv.edu/beisecker/Courses/PHIL-352/Dave%20-%20Consciousness%20PDFs/Jackson%20-%20Epiphenomenal%20Qualia/Jackson-Postscript_on_Qualia.PDF
Appendix A.

‘There is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself: I understand myself to be something quite single and complete... By contrast, there is no corporeal or extended thing that I can think of which in my thought I cannot easily divide into parts.’

Appendix B.

‘Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) the state in question must be a possible state of a mammalian brain, a reptile

---

brain, a mollusc’s brain (octopuses are molluscs, and certainly feel pain) etc. At the same time, it must not be a possible state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extraterrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may be pain. It is not altogether impossible that such a state will be found….But this is certainly an ambitious hypothesis. Finally, the hypothesis becomes still more ambitious when we realize (sic) that the brain state theorist is not just saying that pain is a brain state; he is, of course, concerned to maintain that every psychological state is a brain state. Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus (say ‘hungry’), but whose physical-chemical ‘correlate’ is different in the two cases, the brain-state theory has collapsed. It seems to me overwhelmingly probable that we can do this. ’

Appendix C.

“Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue.'

What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.”

---

6 http://www.phil.uu.nl/~joel/3027/3027PutnamPsychPredicates.pdf

7 https://www.sfu.ca/~jillmc/JacksonfromJStore.pdf