The Mind/Brain Identity Theory: A Critical Appraisal

Leslie Allan

Published online: 21 October 2016

The materialist version of the mind/brain identity theory has met with considerable challenges from philosophers of mind. The author first dispenses with a popular objection to the theory based on the law of indiscernibility of identicals. By means of discussing the vexatious problem of phenomenal qualities, he explores how the debate may be advanced by seeing each dualist and monist ontology through the lens of an evolutionary epistemology. The author suggests that by regarding each ontology as the core of a scientific research programme, each of these logically irrefutable hypotheses can be tested rationally.

To cite this essay:

Leslie Allan 2016. The Mind/Brain Identity Theory: A Critical Appraisal, URL = <http://www.RationalRealm.com/philosophy/metaphysics/mind-brain-identity-theory.html>

To link to this essay:

www.RationalRealm.com/philosophy/metaphysics/mind-brain-identity-theory.html

1. Introduction

Identity theories of mind can be categorized as those theories that contingently identify mental things, processes, states and events with physical things, processes, states and events. These theories can either be materialist, idealist (depending on the direction of reduction)[1] or neutral monist. For reasons that I will not explore here, I consider idealism and neutral monism to be untenable as ontological theories.[2] My interest in this essay is to consider what I think are the outstanding difficulties faced by a materialist identity theory.

I shall consider the most plausible type of materialist identity theory so far proposed; that which identifies the mind with the brain or, more accurately, the central nervous system, or part thereof. For the sake of convenience, I shall henceforth sacrifice precision and speak simply of the 'brain'. I shall begin by considering one common objection to a materialist identity theory. The discussion of the solution to this objection will serve to outline the direction that I think that a plausible identity theory should take and open the way for a critical survey of outstanding problems. I shall not consider problems that I believe have had a satisfactory solution expressed in the literature on this topic.

I want to say at the outset that I do not think that there are any logically sufficient reasons for rejecting any version of the identity theory. This is because I do not think that there are any such reasons for rejecting any internally consistent explanatory theory, including ontologies. (Ontologies are explanatory in the sense that they are attempts to explain the apparent diversity and the apparent unity of phenomena.) According to the well-known Duhem–Quine underdetermination thesis, explanatory theories do not entail observation statements in isolation. It is only when a theory is coupled with auxiliary hypotheses about initial conditions and other intervening mechanisms that an observation statement is entailed. Any apparent 'counterexample' to a theory can be dealt with in an *ad hoc* fashion, with the result that the apparent 'counterexample' turns out either to be a 'confirming' instance of the theory or an 'anomaly' to be solved at a later date. Any ontology that was logically incoherent can be *made* coherent without affecting its identity.

All of this is familiar to proponents of sophisticated conventionalist and evolutionary epistemologies. I want to suggest that competing ontologies be epistemically evaluated according to the criteria of Lakatos' evolutionary epistemology; the methodology of scientific research programmes. For Lakatos [1978a: ch. 1], an explanatory theory is to be evaluated via the history of the research programme within which it is embedded. A research programme is composed of a 'negative heuristic' and a 'positive heuristic'. The 'negative heuristic' specifies the 'hard core' of the programme; its metaphysical foundations or conceptual framework. In the context of this essay, the 'hard core' consists of some version of monism of dualism. According to Lakatos, this 'hard core' is deemed irrefutable by the methodological fiat of the programme's proponents. Every worthwhile research

[1]That the reduction is one way does not logically prevent the theory from being an identity theory. For example, tables, on present physical theory, can be reduced to cohering collections of molecules structured into a particular (tablelike) shape, yet it remains true that the expressions 'table' and 'cohering collections of molecules structured into a particular (tablelike) shape' corefer to the same entity. The reduced referring expression is not eliminated.

[2]For an argument against idealism, see my Allan [2016a].

programme develops in an ocean of anomalies. It is the function of the 'negative heuristic' of the programme to prevent such anomalies from refuting the 'hard core' by directing the scientists' attention to the revision of the 'protective belt' of auxiliary hypotheses and initial conditions. Just how the 'protective belt' is to be modified is specified by a partially articulated plan; the 'positive heuristic'.

A research programme was regarded by Lakatos as 'progressive' if the successive modifications to its protective belt satisfy the following two conditions. Firstly, each successive modification must be 'theoretically progressive', or have 'excess empirical content' in the sense that the new theory, which consists of laws of nature, auxiliary hypotheses and initial conditions, must predict some hitherto unexpected, novel fact. Secondly, the modifications must be 'empirically progressive' in the sense that the predicted novel facts must be at least occasionally corroborated. Conversely, a programme that is not 'progressive' is deemed 'degenerating'. Lakatos had considered that for a research programme to be 'scientific' it must be at least theoretically progressive. And for one research programme to supersede a rival, it must be progressive while its rival is degenerating and must satisfactorily explain the previous predictive successes of the rival.[3]

Seeing ontological theories in this historical context as competing research programmes bestows significant advantages. Judging them on the basis of their evolutionary development helps clarify their present epistemic status and indicates the directions in which the competing theories should develop. In this vein, the difficulties that I will present here for the identity theory are not logically sufficient reasons for rejecting the theory, but outstanding empirical problems that stand in the way of currently accepting the theory.

[3]Elsewhere, I have modified and further developed Lakatos' methodology of scientific research programmes. I have attempted to give reasons for accepting a slightly modified version of Lakatos' epistemology by characterizing the demands of a general objectivist epistemology and demonstrating how Lakatos' criteria for theory appraisal satisfy these demands. See my Allan [2016b].

2. The Problem of Phenomenal Properties

The problem that I wish to begin with derives from the application of the law of indiscernibility of identicals (Leibniz's Law). This law states that if two items are numerically identical, then for any property, it is a property of one if and only if it is a property of the other. Formally,

$$x{=}y \rightarrow (\mathrm{F})(\mathrm{F}x \leftrightarrow \mathrm{F}y)$$

The identity theorist, in identifying mental items with physical items, means by 'identity' this notion of strict numerical identity.

The objection to making this strict identity is that there is a class of properties, the phenomenal properties, that do not satisfy the law of indiscernibility of identicals. Phenomenal properties can be roughly categorized as those that can be ascribed directly to sensations, but only indirectly to items that are not sensations. These phenomenal properties, the objector claims, are easily applicable to mental items. But, they argue, it is either meaningless or false to attribute these same properties to the physical items, such as brain states, that identity theorists wish to identify with mental items.[4] It makes sense, they continue, and is sometimes true to say that we have red, round, loud or sweet sensations, tingling, burning, sharp or severe pains, and happy, angry or nostalgic feelings.[5] But it is either meaningless or false to say that we have red, round, etc., brain states.[6]

The reply that I want to make in defence of the identity theorist is that phenomenal properties cannot strictly be construed as properties of sensations. This response obviates the application of the law of indiscernibility of identicals in this instance. The surface structure of the common idiom, 'I have a red sensation', can fool us into thinking that there are sensations that are red. But sensations are not coloured. There is not another set of colour receptors lying behind our sensations that can discriminate their colour. We do not *see* sensations. The idiom, 'red sensation', is simply a colloquial locution of the more accurate phrase, 'sensation of redness'.

These same comments apply to the locutions 'round sensation', 'loud sensation' and 'sweet sensation'. Sensations do not make noises, nor can they taste like sugar. Nor do they come in various shapes and sizes.[7] Once again, the same can be said for 'happy', 'angry' and 'nostalgic feelings'. Literally speaking, it is not our feelings that are happy or angry or

[4] Direct realists and phenomenalists will strongly disagree with this rough definition, but this will not matter for my purposes here. For those who disagree, the phenomenal properties can be extensionally indicated as colour, pitch, taste, smell and feel.

[5] I am aware that it has been seriously disputed whether this last class of items, namely feelings, is a class of sensations. Nonetheless, I have included it here for the sake of completeness.

[6] For an example of the 'phenomenal properties' objection, see Rollins [1971: 27]. Here, he says, 'Prima facie, no language would be complete for describing experience if it did not allow us to speak of intensity, dreariness, nostalgia, a sense of familiarity, awe or overwhelmingness, placidness, or, in general, of something occurring to us which is of such a nature as to neither have nor lack a locus. In our understood language, descriptions like these could not be eradicated without loss of descriptive power necessary for the language of experiences; yet none of them, of course, can intelligibly be predicated of a brain process itself.'

[7] It should be noted that 'round' is not strictly a phenomenal property, for it can be directly ascribed to items that are not sensations. I include it here simply for the sake of completeness.

nostalgic, but ourselves. And our feeling happy or angry or nostalgic is constituted by the fact that we have feelings of happiness or of anger or of nostalgia.[8]

The case of pains is a little more complicated. However, I think that we can analyse pains in the way we do colours. In that case, the sentence, 'I have a sensation of pain' becomes logically akin to 'I have a sensation of colour'. The term 'burning', then, in the report, 'I have a sensation of a burning pain', modifies the term 'pain' and not the term 'sensation', just as the word 'red' in the report, 'I have a sensation of a red colour', modifies the term 'colour' and not the term 'sensation'. So, just as 'red' is not a property of sensations, neither is 'burning'. (The colloquial locutions, 'I have a burning sensation', and 'I have a painful sensation' can easily be treated in the same way as I have treated 'red sensation' above.) I think that we can now treat 'red after-images' in the same way as I have treated 'burning pains'. So, 'I am having a red after-image', can be more accurately rendered as, 'I am having a sensation of a red after-image'.

The notion of 'having' that occurs in reports such as, 'I have a sensation of redness', is no more problematical than the same notion that occurs in reports of physical states of our bodies, such as, 'I have a head of hair'. If the mind is the brain, as the identity theorist supposes, then what is reported when we report states of our mind will in fact be numerically identical with states of our brain.

I think this relatively simple semantic analysis of sensation reports is true independently of the truth or conceptual coherence of the identity theory. If this is so, then it cannot be a valid criticism of a theory that identifies sensations with brain states that on opening up people's brains for inspection we do not find red or burning or angry brain states.[9]

[8] Here, I think J. J. C. Smart [1971: 87], in defending the identity theory, had made a tactical error in replying to Rollins' 'phenomenal properties' objection (see footnote 7 above) by admitting that 'some brain processes are nostalgic'. I have never met an experience, and hence a brain process, that was homesick. Experiences do not even have homes to be homesick about.

[9] The 'phenomenal properties' objection can receive replies within J. J. C. Smart's topic-neutral account of sensation reports and D. M. Armstrong's Causal Theory of Mind. See Smart [1977: 59–61] and Armstrong [1968: 116f]. To the extent that these replies differ from the one that I have given above, I cannot concur, for I consider Smart's and Armstrong's translations of sensation reports to be seriously mistaken (see below). J. Cornman has offered the interesting suggestion that the law of identity of indiscernibles does not apply to 'cross category' identities. Whether Cornman's thesis is true or not, I do not believe that it is relevant in the case of sensations, for I have argued here that sensations do not have phenomenal properties. See Cornman [1977: 123–9].

3. The Problem of Phenomenal Qualities

I have so far argued that sensations do not have phenomenal properties. Nonetheless, there is something about sensations that provides a major stumbling block to the acceptance of a materialist identity theory of mind. This stumbling block is illustrated in the now well-known 'inverted spectrum' argument. Let us briefly consider J. J. C. Smart's and D. M. Armstrong's versions of the materialist identity theory. Smart's account of colour perception goes something like this. He first defines a 'normal percipient' as, roughly, a member of the class of people that can make the most colour discriminations. Smart [1977: 59f] then claims:

> that 'This is red' means something roughly like 'A normal percipient would not easily pick this out of a clump of geranium petals though he would pick it out of a clump of lettuce leaves'. Of course, it does not exactly mean this: a person might know the meaning of 'red' without knowing anything about geraniums, or even about normal percipients. But the point is that a person can be trained to say 'This is red' of objects which would not easily be picked out of geranium petals by a normal percipient.

Let us ignore the obvious point that it is difficult to see how an ordinary person can mean by 'This is red' something about the discriminatory powers of normal percipients when it is not necessary for the person to know anything about normal percipients. This is so irrespective of what Smart says about what the person is 'trained' to say in particular stimulus situations.

Here is a version of the 'inverted spectrum' argument in which a scenario is drawn such that the definiendum, in this case, 'This is red', is rendered false while the purported definiens is rendered simultaneously true. The purpose of this example is to demonstrate that the two sentences cannot be synonymous. The scenario is this. Everybody wakes up tomorrow morning to find that their colour perception has undergone a reversal. The objects that were previously yellow are now seen as blue, and *vice versa;* those previously red are now seen as green, and *vice versa,* and so on. Furthermore, upon investigation, no change is found in our neurophysiology or in the optical properties of objects. Nor has our ability to discriminate between colours diminished.

In this new situation, what we had previously called a red rose petal, we would now say of it that it is green. It would remain true, though, that we would not be able to pick it easily out of a clump of geranium petals (because they are now green also). So, in this situation, the sentence, 'A normal percipient would not easily pick this [rose petal] out of a clump of geranium petals though he would pick it out of a clump of lettuce leaves', would be true, but we would deny emphatically that 'This [rose petal] is red'. (A person who held Smart's semantic theory would be in the absurd position of denying that the rose had changed colour.) The two sentences, therefore, could not even roughly be equivalent in meaning.

The solution, of course, is not to choose other paradigm cases of red and green objects, nor to include in one's definition a lengthy list of such objects, for the problem lies in Smart's tying of the meaning of colour words to an operationally defined 'normal percipient'. However, as the above counterexample demonstrates, an operational definition

of colour terms misses the qualitative character of colours. This same 'inverted spectrum' objection applies equally to Armstrong's [1968: 248] account of colour perception as 'the coming to be of a state of the person apt for the bringing about of certain sorts of discriminative behaviour'.[10]

We can now see why Smart's topic-neutral account of avowals is so unsatisfactory. His view was that:

> when a person says, 'I see a yellowish-orange after-image', he is saying something like this: '*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.'
>
> [Smart 1977: 60]

The problem with this rendering is that it is logically possible that oranges will change their colour overnight, with no corresponding structural change to anything else in the universe. And the same could possibly occur, logically, to any or all other yellowish-orange coloured things. This argument shows that oranges are only *contingently* paradigmatic instances of yellowish-orange. There is no logical connection here, for if our perceptual experience of oranges were to change radically tomorrow, the last thing that we would do is to continue to call them 'yellowish-orange'.[11] Armstrong's [1968: 254] account of 'sensation' as 'the stimulation of the mind by the world in complete abstraction from the further effects of this stimulation on behaviour or impulses towards behaviour' is an extrapolation from his defective theory of perception, so I shall not consider it further.[12]

Smart and Armstrong, in their versions of the materialist identity theory, sought to define sensations in terms of (but not logically reducible to) physical stimuli and bodily responses. With these definitions, they had the specific intention of leaving a logical gap concerning the nature of the effects of stimuli and causes of behaviour that could be filled definitively at a later date by statements about brain states and processes. I have argued that the 'inverted spectrum' case shows that any such attempt is doomed to failure. This failure arises not because mental states are not identical with brain states, but because any such account ignores the phenomenal qualities of many mental states. The meaning of many 'mental' terms, especially 'sensations', such as 'sensation of redness' and 'sensation of

[10]Even though Smart held a Lockean 'powers' view of colours, while Armstrong adhered to a direct realist view, we cannot help but note the essential similarity of their views on colour perception in that both define colours in terms of discriminative behaviour. Armstrong tried to meet the colour reversal objection, albeit in a different form, in his [1968: 256–60]. I think that his reply is unsatisfactory, for he seems to have misconstrued the nature of the objection. For that part of the objection that Armstrong [1968: 256] correctly stated as, 'it seems logically possible that there might be no evidence to show that my perceptions were reversed in this way', he appears to have misinterpreted it as meaning 'no *behavioural* evidence'. Armstrong then, expectedly, argued that the differences in colour perception could be revealed by the discovery of a systematic difference in the nature of the causes that enable the same behaviour in the two groups. But this is beside the point, for the objection is that it seems logically possible that there might be no evidence *simpliciter* for a reversal of colour perceptions.

[11]There are other, what I consider to be, decisive objections to Smart's topic-neutral analysis that are not directly relevant to the point that I am making here. See, for example, Cornman [1977: 126] and Shaffer [1977: 135f].

[12]For an effective criticism of Armstrong's theory of perception and sensation, see Ellis [1967: 148–57].

burning pain', can only be indicated ostensively. This traditional view is only all too familiar.[13] It is this brushing aside of the phenomenal character of mental states in Smart's and Armstrong's formulations of their 'logical gap' semantic theories that make me think that these materialistic philosophers have not taken the mind-body problem seriously.[14]

An incisive objection to the 'ostensive definition' view of mental terms is Wittgenstein's criticism of the idea of a private mental language. How then do I handle these two specific Wittgensteinian [1953: part 1] objections:

a) if some mental terms refer to something essentially private, then it is not possible to learn their meaning from public demonstrations, for all that we can observe from such demonstrations is outward behaviour, and

b) if the referents of some mental terms are not open to public examination, then we cannot know whether or not other people are ever in these mental states.

Here, I will not attempt a comprehensive response. I will content myself with the less ambitious task of indicating how such a reply would look. The most promising place within which to bed a convincing reply is within an expanded Lakatosian evolutionary epistemology. I think his framework throws light on the nature of language acquisition. From this perspective, we see that the child learns 'mental' terms through a process of unconsciously generating hypotheses about the meaning of a word and testing such hypotheses in *novel* circumstances. However, as with our scientific theories, the child's word-meaning hypotheses do not meet the test of experience in isolation. They are always tested in conjunction with certain background assumptions (the auxiliary hypotheses).

In the case of learning 'mental' terms, a critical background assumption for the child is that his private experiences form certain causal connections with publicly observable events. For example, that the child's private experience of a sensation of redness is causally correlated with the publicly observable rose petal. It is this background assumption, when conjoined with a particular hypothesis about the meaning of a new word (such as 'red'), that proves the most fruitful for the child in predicting the outcomes of new situations. One such successful anticipation of new events could be the approval that the child receives the first time he describes a postbox as 'red'.

An inextricable part of the background assumptions also being tested by the child as they acquire their 'mental' language is the theory that other people have similar private experiences to their own. He posits that these experiences also bear similar causal connections with publicly observable events as his own. The child's learning that he has a private mental life happens concurrently with his learning that others similarly have private mental lives. In this way, his learning of a 'mental' language is inextricably linked with learning about the world and other minds—a world of interactions between private experiences and public events.

[13]The case of the ascription of colours to physical objects is a little more complicated, for here, what is being ascribed is a dispositional relational property of the object. So, to say, for example, that 'The post box is red' is to say that a standard observer perceiving the post box under standard viewing conditions will have a sensation of redness (where what constitutes a 'standard observer' and 'standard viewing conditions' is spelled out).

[14]Smart and Armstrong have since gone over to the heir of the identity theory, functionalism. But even here, the 'inverted spectrum' argument tells equally against this new theory.

4. Taming the Phenomenal Qualities Tiger

4.1 Evolution and the Paradox of Non-interactionism

In this final section of this essay, I want to discuss briefly the problem that phenomenal qualities pose for the identity theorists' programme and possible responses. I have argued that we must accept phenomenal qualities at their face value, at least until we are presented with a progressive research programme that allows us to ignore statements about them. An example of the type of successful eliminative programme I am referring to here is our set of current progressive theories on mental illness that now allows us to ignore talk of 'demon possession'. However, if in future we did encounter a research programme that allowed us to deny language about phenomenal qualities, we would be faced with an eliminative materialist view and not an identity theory proper.[15]

The first problem that I want to discuss is a paradox that I, for a while, had considered to be a serious objection to any kind of non-interactionism[16] that, on the one hand, takes seriously the semantic irreducibility of the postulates of theoretical physics to statements about phenomenal qualities and, on the other, takes seriously statements about phenomenal qualities. These include all types of psycho-physical parallelism and epiphenomenalism and those identity theories that are not inclined towards a phenomenalist interpretation of physics,[17] nor to a 'logical gap' theory of 'mental' terms.[18]

It was to be an objection that would make us seriously reconsider interactionism as a viable theory, in spite of the outstanding success of the non-interactionist programme. I include it here not only because I am less than fully confident that the objection can be answered adequately by the non-interactionist, but also because it is a suitable introduction to a consideration of the evidence for the identity theory *vis-à-vis* psycho-physical parallelism and epiphenomenalism.

The paradox is this. The non-interactionist claims that the whole of a person's behaviour, including his speech acts, has a sufficient causal explanation in terms of the entities and forces postulated by modern physics. Phenomenal qualities, he claims, play no part in such a causal account. Now, on this non-interactionist hypothesis, we could imagine a world (depending on the kind of non-interactionist hypothesis we are considering) that was identical to the actual world with the exception that in this possible world:

a) there are no phenomenal events parallel to certain physical events (contra psycho-physical parallelism), or

[15]This is not notwithstanding the fact that some commentators regard the Rorty-Feyerabend version of eliminative materialism to be a type of identity theory.

[16]I mean by 'interactionism' here the view that some human behaviour is the result of the interaction between physical and non-physical forces. Conversely, 'non-interactionism' is the view that there are no such interactions.

[17]An example of an identity theory that appears to me to be so inclined is H. Feigl's early version of his identity theory, or 'pan-quality-ism' as he called it. See his [1961]. Bertrand Russell's neutral monism may be another example. See his [1947: 860f].

[18]Included here in the 'logical gap' versions are Smart and Armstrong's variants of the identity theory.

b) there are no phenomenal events that are epiphenomenal to certain physical events (contra epiphenomenalism), or

c) phenomenal events are not identical with certain physical events (contra mind/brain identity theory).

According to the non-interactionist, in such a world we would do and say exactly the same things that we do and say now, including debating the mind-body problem.

However, this seems paradoxical and counterintuitive, for if there were no phenomenal qualities in this possible world, it is difficult to see how in this possible world we would have been lumbered with and debated about the relationship between mental events and physical events in the first instance. It seems that at least some of the speech acts that we commit while debating the mind-body problem would not have been committed unless there were some conscious mental events, such as conscious sensations of phenomenal qualities.

Statements about phenomenal qualities, then, seem to be a necessary inclusion in any adequate causal account of our speech acts. But if this is true, then the possible world that we have been imagining is not possible at all. That is to say, it is logically impossible, given our talk about phenomenal qualities, that the physical entities and forces currently postulated by physics could fully account for all of our speech acts. Therefore, it is logically impossible that the actual world be the way it is and non-interactionism be true.

How might a non-interactionist reply to such an intuitively convincing argument? The most fruitful reply, I think, is this. The non-interactionist begins by giving a physical account of the existence of a language of phenomenal qualities. (Even this sounds intuitively implausible on an interactionist view.) Such an account would refer to the considerable evolutionary advantage gained by a biological species whose members had the ability to, firstly, monitor their own internal physiological states (this is the basis of all homeostatic mechanisms) and, secondly, convey that information by vocalizing to other members of the same species. Other members of the group can respond appropriately when a member expresses a burning sensation or a feeling of fear, for example. Because the form of such intraspecific co-operation is highly adaptable to changing environmental circumstances, their capacity for co-operation would be greatly enhanced. Reproductive efficiency would be much higher compared with a reliance on the more evolutionary basic instinctual response.[19]

The next step in the argument is to introduce the adaptive advantage of learning. A species whose members can adapt their behaviour to changing environmental conditions has a distinct biological advantage. The relatively simple learning strategies, such as

[19]This account has the advantage over interactionist theories in that it derives as a novel fact the relative simplicity of the terms in the language of phenomenal qualities. (For the importance of the successful derivation of novel facts for the confirmation of an explanatory theory, see my Allan [2016b: §4.3]). For effective intraspecific co-operation, a highly complex phenomenal language reflecting the intricate complexity of the monitored brain states is not required. All that is evolutionarily necessary is that the language reflects the gross characteristics of brain states. So, a simple phenomenal language is what is to be expected on this theory. This simplicity is thus explained. At the same time, we would expect an evolutionary development towards greater complexity in the phenomenal language. That this has been the case, with the development of finer discriminations between mental states, is also a plus for this theory.

associative conditioning and operant conditioning, were, in the evolutionary development of the higher mammals, supplemented by the much more powerful strategy of postulating explanatory models of the environment and testing such models in novel situations. (This is akin to what psychologists have called 'insight learning'.)

Given sufficient biological sophistication and enough spare time, such organisms will not only develop models of their external environment, but also their internal environment. More importantly, however, models of the monitored states referred to in the phenomenal language will also arise. It is only a small step from here to the postulation of models showing the relationship between external and internal environments, and the monitored states referred to in the phenomenal language. Hence, we have the rough schemata of a physical explanation of the speech acts that we commit in debating the mind-body problem. Although, I must admit, I am not sure how far we can accept this explanation.

In spite of this intuitive qualm, there seems to me no doubt that what I will call the non-interactionist research programme has made outstanding empirical progress in explaining such things as human origins, mental illness, the effects of drugs on human personality, learning, and so on. I venture to suggest that this programme began with the inception of the mechanist school of biology in the eighteenth century and is now constituted by a number of specialist research programmes, including neurophysiology, neuropsychology, psychopharmacology, evolutionary biology and microbiology.

Research in these disciplines is governed by a methodological principle that states that whenever there is a theoretical problem, always avoid postulating the existence of non-physical entities or forces as a tentative solution. In Lakatosian terms, this principle constitutes the 'negative heuristic' of the programme. As a corollary, the 'hard core' of the programme comprises the postulate that the development and behaviour of all animals, including humans, is the result of the action of physical forces. The programme is non-interactionist in the sense that it bars the idea that such development and behaviour is ever the result of a causal interaction between physical and non-physical events.

The competing interactionist programme, however, which seems to me to be an extension of the vitalist school of biology, appears to have degenerated to the extent that it no longer constitutes a scientific research programme. Proponents of the interactionist programme have failed to develop detailed theories explaining how certain aspects of human development and behaviour are the result of the interaction between physical and non-physical entities or forces. Furthermore, they have failed to deduce and test predictions based on such models. All of my comments here are tentative because, as far as I am aware, no researcher has yet undertaken the task of writing a sorely needed Lakatosian historical account of the interactionist and non-interactionist research programmes.[20]

Where does this leave the identity theory? Even though some researchers working on the non-interactionist programme had pushed for an identity theory,[21] it is my impression that it did not constitute its hard core. If this is true, then the greatest problem

[20]The memorable volume edited by Howson [1976] is the only collection of Lakatosian historical case studies of which I am aware.

[21]For example, the naturalist, George John Romanes, (1885) *Mind and Motion*, in (1964) *Body and Mind*, ed. G. Vesey, London: 183 and quoted in Gregory [1984: 476]; and the psychologist, Place [1977].

for the identity theory is the supply of evidence that will sway us in its favour against its non-interactionist rivals; epiphenomenalism and psycho-physical parallelism.[22]

One way *not* to support the identity theory is through appealing to some principle of simplicity.[23] Although such a principle is necessitated by inductivist and conventionalist epistemologies, within a Lakatosian epistemology it is regarded as being a subjective criterion, and hence inconsistent with an objectivist methodology.[24] With this defence barred to him, the identity theorist can make one of two moves. He can either wait for predictive success from his own research program or claim that the identity in dispute is a brute fact about nature while at the same time pointing out the empirical bankruptcy of the interactionist's programme. I shall consider each of these approaches in turn.

[22]For the purposes of this essay, I have not considered the more recent rival to the identity theory, functionalism.

[23]J. J. C. Smart has pushed this defence for all it's worth. See his [1977: 84–7].

[24]See my Allan [2016b: §4].

4.2 Mind/Brain Identity as a Research Programme

I draw to your attention some notable examples of empirical identities established through scientific research and discovery:

- water with H_2O
- the gene with DNA molecules
- lightning with electrical discharges
- combustion with oxidation

These identities were discovered as the result of progressive research programmes that led to the successful confirmation of novel predictions derived from the respective theories of identification.[25] The predictions were derived from these theories through first postulating the underlying microstructure[26] of the entity or event to be reduced, with the successful confirmations of these predictions constituting the independent evidence for such theories.

These successful scientific reductions of known phenomena raise the important question: What novel predictions have been derived from the mind/brain identity theory that could not have been similarly derived from either epiphenomenalism or psycho-physical parallelism? There are none of which I know. The type of event that needs to be predicted is either a hitherto unexpected physical event (whether within the brain or without) or a hitherto unexpected mental event. This much is obvious, but what is difficult to conceive is the type of experiment that would yield a result favourable to one theory but not to its rivals.

We may now be in a situation akin to researchers working in the sixteenth century who were aware of the magnetic properties of loadstones, but had no idea how such properties could be reduced to more fundamental physical processes. The successful reduction of emergent magnetic properties had to await a complete revolution in physics. Similarly, the demonstration that mental properties are reducible emergent properties may require another revolution in physics, possibly necessitating the addition of a new fundamental physical force. Once this reduction has been achieved, though, it will have the much coveted advantage of showing that the emergent mental properties of the brain, as with the emergent magnetic properties of macro-objects, are simply a manifestation of the interaction of certain fundamental physical forces in particular, highly organized physical structures. The theoretical advantages to be gained from such a unified ontology would be stupendous. However, the seemingly insurmountable problem for the identity theorist who pins his hopes on this prospect is precisely that this is an extremely tall order for any theory.

[25] For a historical case study on the successful identification of the combustion of substances with the oxidation of these substances, see Musgrave [1976: 181–209].

[26] Einstein's identification of gravitational mass with inertial mass is an exception to this generalization, for this identification resulted from his postulation of the *macro*structure of the universe.

4.3 Mind/Brain Identity as a Brute Fact

I once thought that the only viable option for the identity theorist in demonstrating the veracity of his theory is the approach outlined in the previous section (§4.2). That is, by postulating a unified theoretical system from which novel predictions could be made and later confirmed. I am now more inclined to the view that there is another way open to him. This method consists in simply identifying certain mental states with certain brain states and claiming this identity to be a brute fact about the world requiring no further theoretical explanation. This makes some sense when we consider that the theoretical tools required for completely explaining human evolution and behaviour are already at our disposal. We already think that the theoretical concepts employed in neurophysiology and microbiology are sufficient for this purpose, so why make our job harder than we need to?

This move is not to be confused with the application of Occam's Razor; the principle of simplicity. The principle of simplicity is a *methodological* principle that applies to competing theories entailing the same empirical consequences. A postulation about what are brute facts, however, is a theoretical posit that leaves open the possibility of being disproved through further research. It is, in this sense, an *empirical* claim rather than a methodological manoeuvre. On this approach, then, the rival non-interactionist programmes, epiphenomenalism and psycho-physical parallelism, are to be defeated by historical and critical analysis. That is, by examining how they fared historically as scientific research programs as well as under the microscope of logical scrutiny.

As a lead in, let's start with logical analysis. Consider what I call the 'epistemological paradox' of epiphenomenalism. (This paradox holds for the earlier versions of property and substance phenomenalism as well as for the later interactionist epiphenomenalism. However, I shall state it here in the terms of the earlier property phenomenalism.) The paradox is this. The thesis of epiphenomenalism is that mental phenomena are non-physical properties of brain states that play no causal role in the production of bodily behaviour, including literary and speech acts. This thesis entails that the epiphenomenalist's literary and speech acts would have been and will be exactly the same, irrespective of whether or not there are any such non-physical, epiphenomenal properties; that is, irrespective of whether epiphenomenalism is true or not.

The paradox arises in that the epiphenomenalist is thereby barred from claiming to advance his thesis in literary or vocal form *because* he thinks he is aware of non-physical, epiphenomenal properties. This is so because, on his own thesis, the existence of such properties has not the slightest effect on what he says or writes. The thesis itself is not incoherent; it does not generate a logical paradox. However, it suffers from an epistemological paradox in that the epiphenomenalist is logically precluded, by his own thesis, from advancing reasons for his theory based on his awareness of such non-physical properties.[27]

The identity theorist is now in a position to charge the epiphenomenalist programme with complete heuristic sterility. On an epiphenomenalist account, no information whatsoever can be transmitted about the purported non-physical, epiphenomenal objects

[27] I take it this is what Medlin was trying to say in his [1971: 110f].

or properties; not even that they exist. In comparison, the identity theorist's programme is able to supply a very powerful positive heuristic. For example, the structure and properties of the mind can be tentatively modelled on the known structure and properties of the brain, while clues to the structure and properties of the brain can be gathered from introspectible properties of the mind.

The identity theorist may also claim that epiphenomenalism is an *ad hoc* retreat to a safe domain within the dualist programme (and so, on Lakatosian criteria, a point against epiphenomenalism). To substantiate this claim, though, the identity theorist needs to show that epiphenomenalism can be fruitfully construed as a move within a specifically *dualist* research programme, and so seen as a content reducing move within that programme. I'm unsure how successfully this can be made out. As well as the particular methodological problems involved in applying Lakatos' methodology of scientific research programmes[28] that the identity theorist must contend with, there remains the task of completing a detailed historical analysis of the programmes themselves. Only future research will tell how successful this claim will be.

However, it appears that whether we regard the formulation of epiphenomenalism as a move within a separate dualist programme or as a move within the identity theorists' own non-interactionist programme, the identity theorist possesses a crucial argument against the epiphenomenalist. And this is that regardless of whether, as a matter of historical fact, the construction of the epiphenomenalist theory happened after or before the construction of the identity theory, it is impossible for the epiphenomenalist programme to ever display an empirically progressive problemshift compared with the identity theorists' programme.[29]

Consider the first possibility. If the formulation of epiphenomenalism had happened after the formulation of the identity theory, then this move constituted a degenerating problemshift. It was degenerating because there is no prediction deducible from epiphenomenalism (in conjunction with background assumptions) that is not equally deducible from the identity theory (in conjunction with background assumptions). Epiphenomenalism is *in principle* unable to predict that if we perform such and such experimental procedures, we will discover (the effects of) non-physical properties or substances, for these, on the epiphenomenalists' own hypothesis, can have no discernible effect whatsoever on physical test setups. Of course, the epiphenomenalist can predict that if we perform such and such physical operations, such and such mental events will result, but this is equally deducible from the identity theory.

Alternatively, if the formulation of the identity theory had occurred after the formulation of epiphenomenalism, this would have constituted progress for the identity

[28] An outstanding problem for Lakatos' methodology is the drawing up of detailed criteria for the identification of the boundaries of each research programme. So, in our case, the problem is whether we should identify two programmes; (property and substance) dualism and monism, and treat epiphenomenalism, parallelism, eliminative materialism, and so on, as moves within these two programmes. Alternatively, we could construe the two programmes as (non-physical–physical) interactionism and non-interactionism. Or we could even regard epiphenomenalism, parallelism, substance interactionism, the materialist identity theory, eliminative materialism, and so on, as separate programmes in their own right.

[29] For Lakatos' criteria for evaluating research programmes in terms of problemshifts, see his [1978a: 31ff] and his [1978b: part 2, §8, 170–93].

theorist's programme. This is because adherents to the epiphenomenalist programme could never justifiably assert anything about non-physical, epiphenomenal properties; not even that they exist. The assertion of any such statement, according to their own hypothesis, is equally consistent with there being no such properties. With no programme of research—no positive heuristic—epiphenomenalism is not only robbed of the potential of becoming empirically progressive, it is difficult to see how it could constitute a research programme at all. Compare this situation with the powerful positive heuristic of the identity theorists.

What of psycho-physical parallelism? The only versions that the identity theorist need consider as a serious rival are those that offer an explanation for the correlation between mental events and some neurophysiological events.[30] For explanations that conscript the help of some supernatural being to establish or maintain the synchronicity, the identity theorist can point out the historical fact that this type of explanation is no longer a part of any tenable research programme. This is especially so since the research programme that sought to establish the existence of such a being (that is, theology) has been undergoing a degenerating problemshift since the Enlightenment.[31]

It is possible to maintain a form of psycho-physical parallelism while admitting that although we do not know the explanation for the correlation, nonetheless, there is one to be discovered. I am not aware of any researcher using this thesis as a basis for a research programme. If there is, I do not think that it should be rejected out of hand, for Newton's terrestrial and celestial mechanics programme held a similar status in this respect. The hard core of Newton's programme contained his Universal Law of Gravitation, a law for which Newton admitted he had no explanation. (For Newton, 'action at a distance' was an absurdity.) However, the identity theorist will rightly point out, Newton's programme otherwise enjoyed spectacular empirical success. Even so, following the singular lack of success from the strenuous search for an explanation for the law of gravitation, later Newtonians eventually abandoned the expectation for such an explanation.

The identity theorist would rightly point out that if such a psycho-physical research programme was established prior to the identity theorists' programme, then its lack of success in providing an independent confirmation of an explanation for the correlation signals its degenerating problemshift. If the programme was established after the identity theorists' programme, then it cannot supersede the identity theorists' programme until it furnishes us with an independently confirmed explanation, which it has not done to date.

So, with this second 'brute fact' method of attempting to substantiate his theory, the identity theorist is able to argue that his theory is rationally acceptable because his is a progressive research programme that has in fact not been superseded by any rival programme. As I have said, how far this argument can be pushed depends on the results of a detailed historical analysis of the rival programmes.

[30]The theory that every correlation is a chance correlation is rendered extremely improbable by the application of the probability calculus. The theory that the chain of physical causes and the chain of mental causes run in synchronism by chance alone had undergone a degenerative problemshift and is no longer a viable programme. This occurred primarily because not every mental event was found to have a mental cause.

[31]A Lakatosian historical study of theology is another piece of historiographical research that is sorely needed.

There is one outstanding objection to the identity theorist who takes this second line of argument. And that is that if it is simply a brute fact that certain mental states are strictly identical with certain brain states, as he maintains, then it is no longer obvious that what is being advanced is a materialist version of the theory. (Remember that we have already rejected 'logical gap' versions of the identity theory.) It now seems that there is an irreducible duality of types of physical states; those that are not identical to mental states and those that are identical to mental states.

What seems so implausible about all forms of dualism is that it is very odd to think that at some stage of our evolutionary development from amoeba to *homo sapiens*, and at some stage of our individual biological development from embryo to adulthood, there arises spontaneously, literally *ex nihilo*, irreducible non-physical substances and/or properties. This version of the identity theory seems to replace one form of fundamental dualism with another. It seems the identity theorist has managed to climb out of one deep hole only to have dug himself into another.

So, whatever approach the identity theorist takes to the problem of phenomenal qualities, he is burdened with unpalatable consequences. He may retain an uncompromising materialism, but at the cost of a promissory note regarding future theoretical research—research which may not turn out the way he envisages. Or he may attempt to win the rewards of rational acceptance here and now, but be burdened with the cost of completing a detailed historical analysis of the competing approaches to the mind-body problem. However, in so doing, he may seriously endanger his commitment to a thoroughgoing materialism.

5. Conclusion

In this essay, I have outlined what I consider to be the outstanding difficulties for a materialist version of the identity theory. I began with a consideration of the application of the law of indiscernibility of identicals to phenomenal properties, as this application constitutes a problem for the theory. In showing how an identity theorist may dispose of this objection, I led the discussion into a consideration of the problem of phenomenal qualities. My aim here was to show that in attempting to solve this problem, the identity theorist is faced with two horns of a dilemma. Either he abandons all hope of demonstrating his thesis to be true in the near future, barring a revolution in physics just over the horizon, or he flirts with an irreducible dualism of types. Although the identity theory has some significant advantages over its non-interactionist rivals, epiphenomenalism and psycho-physical parallelism, it is caught somewhat between a rock and a hard place. It is unclear which is the least unpalatable of the alternatives. Perhaps further analysis will reveal one or both horns of the dilemma to be imaginary.

One thing that I have stressed throughout this essay is the necessity of evaluating mind-body theories in their historical context. This is because there is no decisive, once and for all, refutation of any of the current mind-body theories. In this vein, I have considered objections to the identity theory that are not logically decisive, but which should give us reason to pause and consider its problems.

There are other challenges for the theory that I think are important, but less significant than the ones that I have considered here. The doubt concerning the one-to-one identity relationship between mental states and brain states is such a problem. Another problem that I have not considered is the charge of speciesism. I have not discussed these objections here because I think that they can be accommodated on a revised identity theory. On such a revised version of the theory, I think some version of a type-type identity can be salvaged. Also, the charge of speciesism can be deflected by rendering the form of the identity relation open-ended. Such modifications to the theory, of course, leave revisionist identity theorists open to the charge of *ad hoc*ness. But that is another story.

References

Allan, Leslie 2016a. The Existence of Mind-Independent Physical Objects, URL = <http://www.rationalrealm.com/philosophy/metaphysics/mind-independent-physical-objects.html>.

Allan, Leslie 2016b. Towards an Objective Theory of Rationality, URL = <http://www.rationalrealm.com/philosophy/epistemology/objective-theory-rationality.html>.

Armstrong, David M. 1968. *A Materialist Theory of Mind*, London: Routledge and Kegan Paul.

Borst, C. V., ed., 1977. *The Mind-Brain Identity Theory*, London: Macmillan.

Bradley, M. C. 1969. Two Arguments Against the Identity Thesis, in *Contemporary Philosophy in Australia*, eds R. Brown and C. D. Rollins, London: Allen and Unwin: 173–89.

Bunge, Mario 1980. *The Mind-Body Problem–A Psychobiological Approach*, Oxford: Permagon Press.

Bunge, Mario 1981. *Scientific Materialism*, Dordrecht: D. Reidel.

Cade, John 1979. *Mending the Mind*, Melbourne: Sun Books.

Campbell, Keith 1970. *Body and Mind*, New York: Anchor Books.

Campbell, Keith 1983. Abstract Particulars and the Philosophy of Mind, *Australasian Journal of Philosophy* 61/2: 129–41.

Chappell, V. C., ed., 1962. *The Philosophy of Mind*, Englewood Cliffs: Prentice-Hall.

Churchland, Paul M. 1985. *Matter and Consciousness*, Cambridge: MIT.

Cornman, James 1977. The Identity of Mind and Body, in *The Mind-Brain Identity Theory*, ed. C. V. Borst, London: Macmillan: 123–9.

Deshpande, D. Y. 1974. Are There Sensations?, in *Contemporary Indian Philosophy, Series 2*, ed. M. Chatterjee, London: Allen and Unwin: 42–50.

Ellis, Brian 1967. Physical Monism, *Synthese* 17/1: 141–61.

Feigl, Herbert 1953. The Mind-Body Problem in the Development of Logical Empiricism, in *Readings in the Philosophy of Science*, eds H. Feigl and M. Brodbeck, New York: Appleton-Century-Crofts: 612–26.

Feigl, Herbert 1961. Mind-body, Not a Pseudoproblem, in *Dimensions of Mind*, ed. Sidney Hook, New York: Collier: 33–44.

Feyerabend, Paul K. 1981. *Realism, Rationalism and Scientific Method: Philosophical Papers Volume 1*, Cambridge: Cambridge University Press: chs 8–10.

Flew, Antony, 1969. *Body, Mind and Death*, London: Macmillan.

Gregory, Richard L. 1984. *Mind in Science*, New York: Penguin.

Hook, Sidney, ed., 1961. *Dimensions of Mind*, New York: Collier.

Howson, Colin, ed., 1976. *Method and Appraisal in the Physical Sciences*, Cambridge: Cambridge University Press.

Hurd, D. L. and J. J. Kipling, eds, 1964. *The Origins and Growth of Physical Science, Volume 2*, Harmondsworth: Penguin.

Jackson, Frank 1976. The Existence of Mental Objects, *American Philosophical Quarterly* 13/1: 33–40.

Jackson, Frank 1982. Epiphenomenal Qualia, *Philosophical Quarterly* 32/127: 127–36.

Joad, C. E. M. 1948. *Guide to Modern Thought*, London: Faber and Faber.

Joad, C. E. M. 1963. *Philosophical Aspects of Modern Science*, London: Unwin.

Lakatos, Imre and Alan Musgrave, eds, 1970. *Criticism and the Growth of Knowledge*, London: Cambridge University Press.

Lakatos, Imre 1978a. *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*, eds J. Worrall and G. Currie, Cambridge: Cambridge University Press.

Lakatos, Imre 1978b. *Mathematics, Science and Epistemology: Philosophical Papers Volume 2*, eds J. Worrall and G. Currie, Cambridge: Cambridge University Press.

MacIntosh, J. J. 1983. The Logic of Privileged Access, *Australasian Journal of Philosophy* 61/2: 142–51.

Medlin, Brian 1971. Ryle and the Mechanical Hypothesis, in *The Identity Theory of Mind*, 2nd edn, ed. C. F. Presley, St. Lucia: University of Queensland Press: 94–150.

Musgrave, Alan 1976. Why Did Oxygen Supplant Phlogiston?, Research Programmes in the Chemical Revolution, in *Method and Appraisal in the Physical Sciences*, ed. C. Howson, Cambridge: Cambridge University Press: 181–209.

Nisbett, Richard E. and Timothy DeCamp Wilson 1977. Telling More Than We Can Know: Verbal Reports on Mental Processes, *Psychological Review* 84/3: 231–59.

Place, U. T. 1977. Is Consciousness a Brain Process?, in *The Mind-Brain Identity Theory*, ed. C. V. Borst, London: Macmillan: 42–51.

Presley, C. F., ed., 1971. *The Identity Theory of Mind*, 2nd edn, St. Lucia: University of Queensland Press.

Rollins C. D. 1971. Are Mental Events Actually Physical?, in *The Identity Theory of Mind*, 2nd edn, ed. C. F. Presley, St. Lucia: University of Queensland Press: 21–37.

Rosenthal, David M., ed., 1971. *Materialism and the Mind-Body Problem*, Englewood Cliffs: Prentice-Hall.

Rosenthal, David M. 1980. Keeping Matter in Mind, in *Midwest Studies in Philosophy Volume 5*, eds P. A. French, T. E. Uehling and H. K. Wettstein, Minneapolis: University of Minnesota Press: 295–322.

Russell, Bertrand 1947. *A History of Western Philosophy*, London: Allen and Unwin.

Shaffer, Jerome A. 1968. *Philosophy of Mind*, Englewood Cliffs: Prentice-Hall.

Shaffer, Jerome A. 1977. Mental Events and the Brain, in *The Mind-Brain Identity Theory*, ed. C. V. Borst, London: Macmillan: 134–9.

Smart, J. J. C. 1971. Comments on the Papers, in *The Identity Theory of Mind*, 2nd edn, ed. C. F. Presley, St. Lucia: University of Queensland Press: 84–93.

Smart, J. J. C. 1977. Sensations and Brain Processes, in *The Mind-Brain Identity Theory*, ed. C. V. Borst, London: Macmillan: 52–66.

Smart, J. J. C. 1978. The Content of Physicalism, *Philosophical Quarterly* 28/113: 339–41.

Srzednicki, Jan 1972. Some Objections to Mind-Brain Identity Theories, *Philosophia* 2/3: 205–25.

Stevens, Leonard A. 1973. *Explorers of the Brain*, London: Angus and Robertson.

Taylor, Richard 1963. *Metaphysics*, Englewood Cliffs: Prentice-Hall.

Thompson, Richard F. 1975. *Introduction to Physiological Psychology*, New York: Harper and Row.

Toulmin, Stephen and June Goodfield 1965. *The Architecture of Matter*, Harmondsworth: Penguin: Part III.

Van Gulick, Robert 2014. Consciousness, *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), ed. Edward N. Zalta, URL = <http://plato.stanford.edu/archives/spr2014/entries/consciousness/>.

Weiskrantz, L. *et al* 1974. Visual Capacity in the Hemianopic Field Following a Restricted Occipital Ablation, *Brain* 97: 709–28.

Williams, Moyra 1970. *Brain Damage and the Mind*, Harmondsworth: Penguin.

Wisdom, John 1963. *Problems of Mind and Matter*, Cambridge: Cambridge University Press.

Wittgenstein, Ludwig 1953 (1963). *Philosophical Investigations*, 3rd edn, trans. G. E. M. Anscombe, Oxford: Basil Blackwell.

