XIII*—MORE TROUBLE FOR FUNCTIONALISM

by Alan Weir

ABSTRACT In this paper I highlight certain logical and metaphysical issues which arise in the characterisation of functionalism—in particular its ready coherence with a physicalist ontology, its structuralism and the impredicativity of functionalist specifications. I then utilise these points in an attempt to demonstrate fatal flaws in the functionalist programme. I argue that the brand of functionalism inspired by David Lewis fails to accommodate multiple realisability though such accommodation was vaunted as a key improvement over the identity theory. More standard accounts of functionalism allow, by contrast, for far too much multiple realisability. Specifically, functionalist structures will be massively reduplicated in the human brain; so functionalism yields the absurd consequence that each human harbours large numbers of minds and exemplifies virtually all mental states.

Modern physicalism ... is behaviouristic ... In the case of cognitive concepts like 'knowing', 'believing', 'understanding', 'remembering' ... there can be little doubt ... that an analysis in terms of dispositions to behave is fundamentally sound. On the other hand, there would seem to be an intractable residue of concepts clustering round the notions of consciousness, experience, sensation, and mental imagery, where some sort of inner process story is unavoidable ... I shall assume that ... statements about pains and twinges, about how things look, sound, and feel, about things dreamed of or pictured in the mind's eye, are statements referring to events and processes which are in some sense private or internal to the individual of whom they are predicated.1

And of course Place went on to claim that these inner processes are constituted by neural processes in the brain and nervous system. What is striking here is how sympathetic this pioneer of the identity theory (more explicitly formulated in


*Meeting of the Aristotelian Society, held in Senate House, University of London, on Monday, 21st May, 2001 at 4.15 p.m.
terms of identity by J. J. C. Smart) was to behaviourism. Behaviourism, it seems, was right for most cognitive states and attitudes, it just needed supplementing with the identity theory to cover occurrent mental processes such as perceptual experience, imagination and sensation. And yet, not long after Place wrote this, behaviourism was as dead as a dodo and the identity theory was on the critical list, the sophisticated new doctrine of functionalism replacing both of these primitive theories as the theory of choice for every up-and-coming philosopher of mind.

What did for the behaviourist analysis of mental states as input-output dispositions was the objection, already clearly put by Chisholm and Geach a year or two after Place,2 that no mental state can be (or can be nomologically correlated with) a mechanism pairing sensory inputs and motor outputs independently of the state’s links with other mental states. Belief that the drink is poisoned may cause one to knock it out of the hands of one’s drinking partner, but on the other hand may not, if, for example, one has a strong desire that one’s partner die. What did for the identity theory was Putnam’s multiple realisability argument: if being in pain is the very same thing as having one or other pattern of neural firings, each constituted by specific chemical interactions involving neurons of particular types, then nothing which does not have a brain of a kind pretty similar to the human brain can have pains of that type. But this seems implausibly species-chauvinistic. Even if, e.g. octopuses, do not actually experience pains, even if there are no extra-terrestrial sufferers, there seems no reason to suppose that there could not have been conscious life with a very different physical basis.3

---


3. See e.g. Hilary Putnam, ‘The Mental Life of Some Machines’, in his Philosophical Papers: vol. II (Cambridge: Cambridge University Press, 1975), 408–28. To be sure in the balmy pre-Kripkean days of the nineteen fifties, ‘central state materialists’ such as J. J. C. Smart held that identity statements involving mental language on the one side and physical language on the other expressed contingent, not necessary, identities. Even if that view is tenable, the consequence is that in any possible situation in which intelligent, physically heterogeneous creatures inhabit disconnected sectors of the universe, any such creature which adopted Smart’s actual viewpoint—thereby
In this paper, I will argue that functionalism is as fatally flawed as the Place-style combination of behaviourism and the identity theory was taken to be.\(^4\) In \(\S I\) I recap the main features of functionalism as I understand it, looking in a little detail at logical and metaphysical issues important for the arguments which follow; I then press my objections in \(\S I I - IV\).

I

Functionalism. Functionalists typically assume there is a true body of sentences containing terms \(M, M_2, \ldots, M_n\), which stand for mental states, properties or events and also terms for sensory input \(I\) and motor output \(O\).\(^5\) We might schematically represent this body of truths—'Functionalist Theory'—by

\[
\text{FT: } \forall x \forall T(x, M, M_2, \ldots, M_n, I, O),
\]

where the individual quantifier ranges over the set of mind-endowed individuals (or perhaps the set of all those with minds whose powers are comparable to those of a typical human mind). The theory \(FT\), in other words, sets out the interrelations among mental states and properties with each other and with sensory input and motor output which must hold true of everyone who can be in any of those mental states or possess any of those properties.\(^6\)

According to one version of functionalism, dubbed psycho-functionalism by Ned Block,\(^7\) \(FT\) is a standard empirical theory parochially restricting mental states to beings with its own physical make-up—would fall into error, and this might seem to show that the identity theory cannot get the nature of psychological states right.

4. Indeed, though there is no space to argue this here, I believe it is definitely inferior to the Place-style theory which can, in contrast to functionalism, be fine-tuned so as to surmount the objections.

5. More fully, the relations will be with a number of different types of input \(I_1, \ldots, I_i\), and a number of different types of output \(O_1, \ldots, O_j\), perhaps further indexed by temporal indices indicating, e.g., how many time-periods apart various inputs and outputs occur.

6. Or at any rate can be in a sufficient number to count as possessing a mind of the right type. A weaker and perhaps more plausible version holds that there is a theory \(FT\) which holds of most individuals with minds, or holds of most individuals when they are functioning in a 'typical' fashion. These qualifications will not bear on the argument of this paper.

so that some of the psychological terms $M_k$ may be scientific terms, not part of ordinary parlance. For instance, many forms of the computational theory of mind are psychofunctionalist theories in which the $M_k$ include terms from computer science or from areas of cognitive psychology heavily influenced by computer science. Note that, in this case, FT cannot be all that complex a theory in that it will have to be comprehensible to individual scientists. The conceptual functionalist, by contrast, limits the psychological concepts expressed in FT to those available to the lay person, the ‘folk psychologist’. Nonetheless some varieties of conceptual functionalism do not require that FT be comprehensible to individual thinkers. Rather, the theory may be a system of meaning-constitutive or analytic truths, perhaps a disjunction of huge conjunctions of folk-psychological platitudes, or a revision of a comprehensive such conjunction which maximises, in a reflective equilibrium, coverage of the original set of platitudes, on the one hand, with coherence on the other. On this view, FT may be too unwieldy for us to know as a whole. Nonetheless we (or most of us) might know each conjunct even though we cannot bring them all into view. Or, more complexly, FT can be broken down into fragments such that competent users of psychological concepts could be brought readily to see the truth of each fragment—rather as the slave in the ‘Meno’ is led by Socrates to see geometric truths—and to see how ‘adjacent’ fragments are related. But note that even in the case of varieties of conceptual functionalism in which FT is too complex for any individual to take in at a glance, there are upper bounds on its complexity, bounds which will be important later on in the argument. For if FT is a system of platitudes in common currency among speakers of typical linguistic communities then, given the limits on vocabulary, modes of sentence composition and length of sentences we can grasp, FT could not comprise, for example, $10^{11}$ comprehensible natural language sentences, indeed it could be nowhere near that long.

For the same reason, the list of mental terms $M$ to $M_n$ will not cover all possible mental predicates: it will not contain, for example, “$x$ believes-true $s_1$”, “$x$ believes-true $s_n$”, ... (or “$x$ believes that $p_1$”, etc.), for all sentences $s_i$ of the language. Rather the functionalist will have to distinguish a basic tool-kit of mental concepts out of which she believes any psychological term can
be constructed, and of course different theories will be distinguished, in part, by the choices they make for the basic mental terms.

The next chapter in the functionalist story involves utilising a technique related to one associated with F. P. Ramsey and ‘dep-spsychologising’ FT by replacing each psychological term $M_i$ with a free variable of the appropriate type (different variables for different terms); the result, which I will call the ‘Ramsey Matrix’, will then be schematically represented by

$$RM : T(x, X, Y_2, ..., Y_n, I_{i}, O_{i}).$$

What terms will be left in $RM$? Clearly the terms $I_1, ..., I_i$ and $O_1, ..., O_j$, specifying the different types of sensory input and motor output relevant to the psychology of individuals remain untouched by the variable replacement, else we fall into ‘liberalism’ or ‘anarchism’ and have to allow that any sufficiently complex system of entities counts as a mind. There will also occur, as constants in $RM$, logical and mathematical expressions and terms standing for the sort of interconnection which is supposed to hold between the elements in the functionalist network: causal linkage, perhaps— say the notion: “changes in $x$ cause changes in $y$” — or some idea of transfer of energy or physical material. But though the specification of interconnections, inputs and outputs must not be evacuated of all content in order to avoid the deadly dangers of liberalism, they must nonetheless be characterised in a fairly general fashion in order to avoid the opposing pitfall of species-chauvinism. A mind, for the functionalist, is any system of items (of whatever inner nature) inter-related in ways which a wide variety of properties can be with one another (and with input and output) so as to instantiate structural patterns of the right sort. This compatibility with multiple ‘realisation’ of mental states by organisms physically quite different from humans is, after all, supposed to be one of the main advantages of functionalism over the identity theory. Perhaps there will be

8. From now on I will refer to the terms $M, M_2, ..., M_n$, which are generalised on to form $RM$ as the structural terms of $RT$, the remaining non-variable terms as the constant terms.

9. Or perhaps some notion of a transition between one computational state or another. However this supposition is highly contentious and problematic since many will hold that the notions of computational state and transition thereof are themselves ‘mentalistic’ and so in need of ‘Ramsifying’ away.
further constants but I assume that to add too much more will take us away from the purer structuralist form of functionalism. If we generalise on RM to get:

$$\forall x \exists X, \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I_{ij}, O_{ij})$$

we get the (closed) Ramsey sentence of FT (with the individual variables ranging over the same set of mind-endowed individuals as before). I will extend this terminology by calling any open sentence of the form

$$\forall x \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I_{ij}, O_{ij})$$

the Ramsey sentence for mental state M (replaced by variable X). Similarly

$$\forall x \exists X, \exists Y_2, ..., \exists Y_k, \exists Y_{k+2}, ..., \exists Y_n T(x, X, Y_2, ..., Y_{k+1}, ..., Y_n, I_{ij}, O_{ij})$$

is the Ramsey sentence for mental state M_{k+1}. If assigning the individual t to x and the property P to Y_{k+1} renders

$$\exists X, \exists Y_2, ..., \exists Y_k, \exists Y_{k+2}, ..., \exists Y_n T(x, X, Y_2, ..., Y_{k+1}, ..., Y_n, I_{ij}, O_{ij})$$

true, I will say that P ‘realises’ mental state M_{k+1} in t.

A variant on the closed Ramsey sentence for M which might seem to lessen opportunities for unwanted multiple realisability and which Lewis calls the modified Ramsey sentence for M, is

$$\forall x \exists ! (X, Y_2, ..., Y_n) T(x, X, Y_2, ..., Y_n, I_{ij}, O_{ij}).$$

Here \(\exists ! (Z_1, ..., Z_n) \Phi (Z_1, ..., Z_n)\) is defined by

$$\exists Z_1, ..., \exists Z_n (\forall W_1, ..., \forall W_n \Phi (W_1, ..., W_n)) \leftrightarrow (Z_1 = W_1 \& ... Z_n = W_n).$$

10. The individual universal quantifier must have widest scope elsé we would lose the multiple realisability since we would end up requiring that there are some fixed properties (carbon-based neural properties, say) such that every individual who can instantiate M, must instantiate those properties. For a related point see Sidney Shoemaker, ‘Some varieties of functionalism’, Philosophical Topics (1981), 93–119, see 96.


12. If one is worried about identity relations among properties substitute a weaker equivalence relation, such as coextensionality, for =.
As remarked, then, functionalism is, at root, a structuralist theory of mind. To possess a mind is to instantiate a system of properties whose structure satisfies the matrix RM: the nature of these properties themselves, whether they are properties of carbon-based life, or of some other type, for example, is irrelevant. So it is assumed that RM,

\[ T(x, X_1, Y_2, ..., Y_n, I_0, O_0) \]

is not categorical with respect to properties in the sense that it is not the case that only one system of properties will satisfy it. Different individuals, and indeed the same individual over time, will be in different mental states and so exemplify different mental structures by virtue of having different perceptions, beliefs, desires, cognitive abilities and so forth. The ideal goal of the functionalist will be to approximate closer and closer to a matrix RM which is true of exactly the mind-endowed individuals. The correct matrix RM will be such that all and only mental structures (or rather individuals containing among their internal structures, mental structures) satisfy RM, so that replacing the variables by the corresponding mental predicates yields a true and comprehensive psychological theory.

Functionalists have also hoped to provide characterisations not only of what it is to possess a mind but also of the mental states constitutive of minds, characterisations which are not crudely ‘reductive’ yet capture the mental in non-mental terms. Suppose, for example, we are interested in providing a functionalist characterisation of some particular mental state M — believing there is an apple in front of one, feeling mildly anxious, as it may be, depending on the theorist’s choice of basic notion. There are different ways of arriving at a characterisation of M, in purely non-psychological functionalist terms, from the Ramsey sentence for M. The way which is perhaps most natural is to follow David Lewis’ account of the definition of theoretical terms.\(^{13}\) Suppose, to take a fairly simple example, one has a political theory PR of the schematic form

\[ T_{PR}(\text{the president, the prime minister, the treasury minister ...}), \]

setting out the various power relations which hold among the offices of state in some particular political system. Then one gives an explanatory characterisation of the nature of a particular office, that of president say, in terms which are entirely non-political (save for the power relations of subordination of authority and so forth corresponding to the causal or other interlinking relations in functionalism) in the following way. One takes the Ramsey sentence for that concept and forms a definite description (formalisable as usual by means of the iota operator \( \iota x \phi x \)) which is then taken to define the phrase 'The President':

\[ \text{The President} =_{df} \iota x (\exists y_1, \ldots, y_n, T_{PR}(x, y_1, \ldots, y_n)). \]

It is a consequence of the Ramsey account, then, that if no-one satisfies

\[ (\exists y_1, \ldots, y_n, T_{PR}(x, y_1, \ldots, y_n)) \]

or more than one individual does, then 'the president' is an empty term, standing for nothing just like 'phlogiston'. A variant approach applies the iota operator not to the Ramsey sentence but to the modified Ramsey sentence for the appropriate term of the theory:

\[ \text{The President} =_{df} \iota x (\exists ! (y_1, \ldots, y_n) T_{PR}(x, y_1, \ldots, y_n)). \]

In this way one might hope to account for our grasp of theoretical terms such as 'electron' and 'proton' in a non-operationalist fashion: our understanding of the individual terms consists in our grasp of a theory in which they occur, a theory which interrelates them in a holistic fashion.

In the case of our mental property \( M \), which occurs thus in \( FT \),

\[ \forall x T(x, M, M_2, \ldots, M_n, I_\alpha, O_\beta). \]

14. The modified Ramsey sentence approach which Lewis favours requires there to be a unique n-tuple satisfying the matrix RM if the terms are to refer, and does not allow for some structural terms to refer whilst others do not; see 'How to define theoretical terms' 87–8, 'Psychophysical and theoretical identifications', 255. Lewis' account is also slightly more liberal than the above—'How to define Theoretical Terms', 83, 95—in that \( \alpha \) is the referent of the ith theoretical term from theory \( T \) if a sequence \( s \) of the appropriate length, with \( \alpha \) as the ith constituent of \( s \), uniquely satisfies nearby theory \( T^* \) and no other sequence uniquely satisfies a theory as close to \( T \).
the Ramsey approach would go like this:

Mental State $M$

$$=_{df} \mathcal{I} \mathcal{X} \left[ \forall \mathcal{X} \forall Y_2, \ldots, \forall Y_n T(\mathcal{X}, Y_2, \ldots, Y_n, I_{(i)}, O_{(j)}) \right]^{15}$$

Note that both these characterisations, of political positions and mental states, are impredicative, they quantify over all entities in a range from which the term’s referent, if any, is to be drawn. Standard examples of impredicative specifications include terms such as

$$\mathcal{I} \mathcal{X} (\forall y (y \neq x \rightarrow T xy))$$

which might be our formalisation of ‘the tallest person in the room’ and which contains a quantifier—$\forall y$—ranging over a domain from which its referent, if any, is to be drawn. Similarly

$$\lambda x [\forall F ((F 0 \& \forall y (F y \rightarrow F y')) \rightarrow F x)]^{16}$$

is an impredicative specification of the property of being a number ($y'$ standing for the successor of $y$) because it contains a quantifier—$\forall F$—ranging over a domain of properties from among which the referent of the predicate term is drawn.

Though I make no objections to impredicativity itself,\(^{17}\) I will argue in §11 that the role it plays in functionalism critically undermines the latter. For the moment, note that such impredicative specifications are generally described, in the philosophy of mind, as ‘higher-order’ specifications, indeed they are often held to characterise ‘higher-order properties’. This is curious because, although there was an earlier usage, dating from the first discussions of ramified type theory, in which impredicative formulae true of particulars were classed as ‘higher-order’, current logical usage, which I will adhere to, treats the specification above of

---

15. $\mathcal{I} \mathcal{X} (\forall \mathcal{X} \mathcal{Y})$ is a function from sets of first-order properties to first-order properties. Lewis makes do with first-order quantification only, plus operators forming names of properties plus relational predicates such as ‘$x$ has property $y$’. Though the argument below could have been developed in Lewis’ framework, it will be clearer for our purposes if we use second-order quantification in order to distinguish clearly the various categories of item (moreover formulations in a standard second-order logic are generally substantially stronger, in expressive terms, than any first-order ersatz).

16. Read $\lambda x F x$ as ‘$x$ is such that $F x$’ or, nominalising what is really a predicate, ‘being an $x$ such that $F x$’.

being a number as first-order because it is true or false of individuals. Similarly, in current logical terminology the various functionalist formulae which are held to define, e.g., x is in pain, are all nonetheless first-order formulae: it is people who are in pain, not neural properties of people. Formulae satisfied by properties, such as \( \lambda X (\forall y (Xy \leftrightarrow y \text{ is a chemical substance}) \) — being an X such that X applies to chemical substances — are, on the other hand, second-order formulae true or false of (first-order) properties. The terminological confusion here may be no more than a quirk, but perhaps it betokens a degree of blindness among philosophers of mind to the existence of properties instantiated by properties, that is second-order properties, which has led to over-estimation of the strength of the multiple realisability objection to the identity theory and early behaviourism.

For the existence of higher-order properties is perfectly compatible with physicalism: indeed it is arguable that a lot of important physical properties are, or are dependent on, such properties: e.g. a particle has velocity by virtue of second-order properties of its position properties, acceleration by virtue of third-order properties of its second-order velocity properties, and so on. Moreover the functionalist would do well to acknowledge genuinely higher-order properties since appeal to them is the most obvious way to characterise input and output properties in a way general enough to allow for intra-species neural variation and the possible existence of extra-terrestrials, without allowing that every complex system has inputs and outputs, i.e. the most obvious way to solve the anarchism/chaos dilemma for functionalism.

Of course many contemporary philosophers of mind do not acknowledge the existence of any properties, of whatever order, adopting an official policy of Quinean nominalism or anti-realism about universals. But actual practice is hard to square up with the official credo: most contemporary philosophers of mind seem to be up to their necks, if not their brains, in commitment to properties and related categories — states and so forth. At any rate I will attempt to cut a swathe through the metaphysical mire which is contemporary philosophy of mind by assuming dogmatically, for the purposes of this paper, an ontology of mind-independent particular objects, (first-order) properties of those

---

18. I will make no attempt here to tease out an analysis of this notoriously problematic concept.
objects, relations among those objects, second-order properties instantiated by first-order properties of properties and so on. Realism about properties and relations I take to include the idea that just as many singular terms have mind-independent referents which contribute to the truth-aptness of simple predications involving those terms, so, too, many predicates refer to mind-independent properties and this also is necessary if the result of applying the predicate to a referring singular term is to have a truth value.

Physicalism, from within this perspective, can then be defined as the view that all objects are the referents of singular terms from the physical sector of language, or at any rate all objects could be referred to by some suitably extended physical sector of our language, all properties are, or could be, the referents of predicates from the physical sector and so on. Note that the latter condition is stronger than the condition that all properties are co-extensional with predicates from the physical sector, for this could be true if all particulars are physical but the properties divide into two (mysteriously interacting) categories of physical and mental, the latter not designated by any physical predicate; the weaker condition is compatible, in other words, with dualism (of properties) and so is too weak, and I shall take it (though many do not) to count as a physicalist condition.

It might be objected here that a crazily heterogeneous predicate such as \( \lambda x (F x \lor G x \lor \ldots) \), however physically kosher its component expressions, cannot be thought of as standing for a property, or not a causally active property at any rate, since disjunctive properties are not in general causally active per se but only through the good offices of their component disjunct properties (at least given a univocal notion of causation). But this objection is incoherent from the realist perspective which I am presupposing. A mind-independent property can no more be

---

19. The physical sector of language is the sector in which all non-logical constants express purely physical concepts. Some have doubted whether even a rough distinction between physical and non-physical concepts makes sense, e.g. Noam Chomsky, Rules and Representations (Oxford: Blackwell, 1980), 306, Tim Crane and Hugh Mellor, 'There is no question of physicalism', Mind, 1990, XCIX: 185-206. But whether or not we can give an interesting explanation of the difference, it is hard to deny that we can distinguish between hard physical sciences such as mechanical engineering and non-physical domains such as history, a difference which we feel confident we could extrapolate to future cases.
disjunctive (or conjunctive, or negative, or dispositional or categorial) than Benjamin Franklin can be disjunctive, because he is the referent of \( \exists x (x \text{ is the inventor of bifocals} \lor x = \text{x}) \) or conjunctive because he is the referent of \( \exists x (x \text{ is the inventor of bifocals} \land x \text{ is the first US Postmaster general}) \). Logical and grammatical distinctions such as those between conjunctions and disjunctions apply to mind-dependent concepts (in the ordinary, non-Fregean) sense, not mind-independent properties and only on a gross anti-realism about properties can the two be conflated.  

To conclude these remarks on the metaphysical aspects of functionalism, functionalism has often been thought of as at least compatible with a strongly physicalist ontology. The Ramsey approach, for example, seems to show how we can define away mental terms into purely physical ones (if, at any rate, our constant terms for input and output and interconnection fall clearly on the physical side of the conceptual divide) by showing how to turn implicit definitions given in the restricted vocabulary into explicit definitions. It is true that I have developed functionalism within a background framework of an impredicative second-order system, and the Beth Definability Theorem, which tells us that any implicit definition can be turned into an explicit definition, fails for second-order logic (and also for first-order systems in which the notion of admissible model is restricted, e.g. where there are operators such as set-theoretic or property

---

20. Simply to posit a class of basic properties plus complexity-forming operations—corresponding to set-theoretic union, intersection and complementation, perhaps—does nothing to show that this posited structure mirrors something in reality. On the other hand, nothing I have said demonstrates that there is no non-logical or non-grammatical distinction between, say, simple and complex properties. But I doubt whether any very objective distinction along those lines can be made out. Arguably, a physical particular is only simpler than another insofar as it has fewer mereological parts which are salient and of interest to humans. We might class property \( P \) as simpler than \( Q \) if fewer subspecies of \( P \) are of interest to us than those of \( Q \), where \( F \) is a subspecies of \( G \) iff \( F \) refers to a mind-independent property and is such that \( \forall x (F_x \rightarrow G_x) \) is true. But this account also inherits the relativity to human interests of the criterion for physical particulars.

operators taking fixed meanings in all admissible models). But the theorem holds for finite sets of second-order sentences and I am taking it that functionalism is a finite theory, an actual scientific theory or a collection of platitudes current among contemporary speakers of a given natural language. There might be some worries as to whether an impredicative second-order theory (or a first-order theory with a notion of ‘property of’ as primitive) really is physical, even if the basic vocabulary is, but these worries seem to me misplaced if, as I argued, impredicative specifications and the notion of second-order properties are both perfectly legitimate.

II

Ramsification and Multiple Realisability. Thus the Ramseyan approach sketched in the previous section is a very natural one from a structuralist perspective in the philosophy of mind such as functionalism. Unfortunately, it cannot be utilised by functionalists. Multiple realisability, vaunted by functionalists as a great virtue of their theory, stymies a Ramseyan version of functionalism. For, if multiple realisability obtains and mental state \( M \) is realised differently in a number of organisms, then there is no unique property satisfying

\[
[\forall x \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I(i), O(j))]
\]

(nor, a fortiori, the modified Ramsey variant), and so on Lewis’ account there is no mental state \( M \). Ramseyan functionalism plus multiple realisability leads to radical eliminativism!

Similar problems afflict the application to functionalism of what Lewis calls the ‘Carnapian’ approach\(^{22}\) to the characterisation of theoretical terms. On this approach, to take our political example, one defines the President as ‘an arbitrary’ individual who satisfies the Ramsey formula and similarly defines the mental state as ‘an arbitrary property’ which satisfies the functionalist Ramsey formula. Using Hilbert’s epsilon notation we would get this:

\[
M \text{ental State } M \overset{\text{drt}}{=} \exists X \{ \forall x \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I(i), O(j)) \}.
\]

22. Here I develop Lewis’ remarks in ‘How to define Theoretical Terms’, 83, perhaps, in appealing to the epsilon notation, for example, further than he intended.
(Once again there will be a modified Carnapian variant using uniqueness quantifiers rather than ordinary existential quantifiers.) Thus, on the Carnapian approach to definition of physical terms, there will exist muons even if there are many different, disjoint, systems of particles satisfying the interrelations specified by Standard Theory among today’s ‘elementary’ particles.

Lewis wished to avoid the Carnapian approach, if at all possible, and with good reason.23 But even if we can make sense of ‘an arbitrary φ’ referring in a non-magical fashion to one of out many φs, multiple realizability blocks this approach. Since there will be many properties which occupy the X role in the theory T, but only one which is the referent of $\lambda x M x$, most of the individuals who have a property playing the X role in T do not possess the mental state M, and we get almost as radical an eliminativism as before.

The Ramseyan (or Carnapian) functionalist might maintain that her theory could be contingently true because as a matter of contingent fact there is no multiple realizability; she would thereby eschew any ambition to produce an a priori or analytic characterisation of mentality. But even this more modest position is untenable if multiple realizability obtains from individual to individual across our own species.24

Quine’s difficulties with his concept of stimulus meaning point up the very real prospect that this is so. Quine had originally intended his notion of a stimulation, the building block of his stimulus theory of meaning, to be an intersubjective one, and so took stimulations to be physical ‘patterns’ (a Quinean euphemism for ‘property’) occurring just outside appropriately oriented

23. But see his ‘Reduction of Mind’, Papers in Metaphysics and Epistemology, 291-324, where (301) following Field, (‘Theory change and the indeterminacy of reference’, Journal of Philosophy, 1973, 70: 462-81) he is more sympathetic to a variant of the Carnap line in which, whenever there are a plurality of n-tuples satisfying the formula and a plurality of properties occurring in the jth place of these n-tuples, then $M_j$ is ambiguous, denoting all of the properties equally. I will refer to talk of ambiguity in what follows as the ‘Ambigamist’ approach. One problem with this approach is that competent speakers can normally disambiguate ambiguous expressions whereas nothing comparable is in the offering here, and this casts down on the claim of the many-many ‘partial reference’ relation to be a genuinely semantic relation.

24. Similarly, given intra-species multiple realizability, highly plausible in the case of states such as pain given the differences in brain structure between e.g. mammals, reptiles and molluscs such as octopuses, the Ambigamist approach mentioned in the previous footnote has the consequence that our mental language is massively ambiguous, an unpalatable consequence made totally untenable by the further extension of multiple realizability argued for in §IV.
MORE TROUBLE FOR FUNCTIONALISM

sensory organs:

A visual stimulation is perhaps best identified ... with the pattern of chromatic irradiation of the eye. To look deep into the subject's head would be inappropriate even if feasible, for we want to keep clear of his idiosyncratic neural routings ... Ocular irradiation is intersubjectively checked to some degree by society and linguist alike by making allowances for the speaker's orientation and the relative disposition of objects.25

But he later became dissatisfied with this. When a group of people are watching what is intuitively the same scene, which objects are they oriented towards? They will all have somewhat different perspectives, somewhat different anatomies, their sense organs will be in different places and so forth. Moreover, as Quine came to emphasise, the structure of the nerve nets, the detailed architecture of brain and nervous system, differs widely between individuals— as Darwin noted, this is true even of insects.26 If Shug and Senga undergo what is intuitively the same experience at \( t \) and \( t' \) respectively, looking at the same object through a tachistoscope, say, there is unlikely to be any neurally salient similarity relation rendering the neurally specified properties of Shug at \( t \) more like those of Senga at \( t' \) than of Shug at \( t' \) when he is looking at a distinct object.

To be sure different physical mechanisms can share lots of properties: different engines can have the same horsepower and run at the same idling speed.27 But however it is with cars and pocket calculators, the Quine/Darwin observation does strongly suggest that for more complex entities such as humans (perhaps even insects), there will be no intrinsic properties designated by purely neurophysiological predicates which are common among distinct animals of the same species when in the same mental state and only then. Just as distinct species will realise the same functional properties by distinct properties, the distinction being

27. Lewis in 'Reduction of Mind', gives the example (305) of identical calculators calculating the same sums but using different registers to carry over, but here the common property is a functionally specified one whilst the question at issue is whether functional properties will be realised by the same properties, specifiable neurophysiologically, across all or most humans.
expressible in physical terms, so too distinct humans will realise the same properties in different physical ways. Multiple realisability begins at home.

One might try to evade the difficulties caused by inter-human multiple realisability by appealing to different senses in which different physical states may ‘realise’ one and the same mental state. Thus one might class the relation ‘P is a sub-species of property Q’ as a case of the realisation of Q by P. Alternatively one might say that if property P instantiates higher order property Q, Q is thereby realised by P. The first option would require a fairly rich theory of properties since it requires that all the differing neural properties (and non-neural ones, if extra-terrestrial minds exist) of the different minds who are in mental state M be subspecies of a single property. The second option is more plausible since there is no need to claim this single unifying property exists.

But even if these options are viable, can they be used to defend functionalism? Are they not, rather, vindications of the identity theory? In the first case, mental state M is identified with the single property ‘realised’ by different sub-properties in different individuals. In the second case, we argue that being in a state of mild anxiety is identical with the property of possessing a physical property which instantiates a specific higher-order property. Formally the specification of such a property instantiates the grammatical pattern

$$\lambda x (\exists X \Phi X \& X x),$$

a format which impredicative first-order specifications also meet. But in the latter \(\Phi X\) is a formula of second-order logic and thus contains no second-order predicate constants or variables (though it may contain quantifiers ranging over first-order properties and relations as well as individual quantifiers), whereas in the latter there are second-order predicate constants, most simply if it takes the form

$$\lambda x (\exists X F^2 X \& X x).$$

I take the argument of §1 to have shown, then, that, contrary to much recent talk aimed at blurring the distinction between functionalism and the identity theory, functionalism, with its appeal to impredicative first-order properties, is a substantially
different theory from the identity theory, even when the latter appeals to higher-order properties. Multiple realisability, on this view, does not spell the death knell for the identity theory but merely for a crude form which neglects the existence and importance of higher-order physical properties.

III

More Trouble. The failure of the Ramseyan technique to work in the functionalist case indicates that Lewis' refusal to class himself as definitely a functionalist or, alternatively, definitely an identity theorist has deeper grounds than simply the terminological variation among philosophers of mind as to the meanings of ‘functionalism’ and ‘identity theory’. For another response Lewis makes to multiple realisability, appealing to an alleged domain-specificity in scientific ‘reductions’ (heat is the usual reductive target in these examples) is to suggest that different kinds of mental state are experienced by different species if they are realised in different ways: there is pain-for-Martians and pain-for-humans.28 However if, as I have urged, mental states are realised differently in different humans, Lewis will have to go further in this Heideggerian hyphenation and distinguish pain-for-Jack from pain-for-Jill.29 Moreover what are we to say of the concatenation of four letters ‘p’ ‘a’ ‘i’ ‘n’ as it occurs firstly in ‘pain-for-Martians’ and secondly in ‘pain-for-humans’? Is it a mere typographic accident that the same sequence of letters occur in both? Well, in that case multiple realisability is being denied.

Or is there some common content expressed by the word in the two cases? But in that case a property realist will assume there is a common property. Lewis accepts that there is a property common to Martians in pain and humans in pain, but he denies that it is pain! For pain causes behaviour and what does the causing, for Lewis, are the realising states, different in each species, by hypothesis. What is common across the species is the (impredicative, first-order) property of having a property which

29. He appears to be prepared to go down this road in ‘Psychophysical and theoretical identifications’, 257, fn. 12.
'plays the pain role'. But this now takes us well away from Ramseyean or Carnapian functionalism towards what I think is the more standard formulation of the doctrine:

$t$ has mental property $M$ iff

$$\exists X (Xt & \forall x \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I_{(i)}, O_{(j)})).$$

This defines what it is for an arbitrary individual $t$ to have the mental property in question. Using $\lambda$ terms, the standard account then defines the mental state $M$ as

$$\lambda y [\exists X (Xy & \forall x \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I_{(i)}, O_{(j)}))].$$

(There will also be a modified Ramsey variant in which we replace the existential quantifiers with uniqueness quantifiers.) On this version of functionalism, it is no longer the case that mental properties are 'nodes' in functional networks, or places in structures; rather, a mental property is an impredicative first-order property of possessing a node-in-a-network property.

This latest version of functionalism does now permit cross-species and intra-species multiple realisability; doubtless this is partly why it has been so prevalent. Mental state $M$ is now identified with being an $x$ such that $x$ bears a property which plays the $M$-role in theory $T$. There is no reason for a realist to deny that $M$ refers to a definite property just as a realist has no reason to deny that a dispositional specification can pick out a property, one which stands in causal relations to other properties and so on, even when many properties 'realise' the disposition.

Nonetheless there is an asymmetry with the dispositional case, one which does not favour functionalism, at least not conceptual functionalism. For according to that theory, we got $T(x, X, Y_2, ..., Y_n, I_{(i)}, O_{(j)})$ from the FT theory $T(x, M, M_2, ..., M_n, I_{(i)}, O_{(j)})$. But now $\lambda y M y$ is supposed to be

$$\lambda y [\exists X (Xy & \forall x \exists Y_2, ..., \exists Y_n T(x, X, Y_2, ..., Y_n, I_{(i)}, O_{(j)}))].$$

$$\lambda y M_3 y$$ is

$$\lambda y [\exists Y_3 (Yy & \forall x \exists X, \exists Y_2, \exists Y_4, \exists Y_n T(x, X, Y_2, ..., Y_n, I_{(i)}, O_{(j)}))].$$

30. In a slightly extended sense of 'realise': $P$ realises in $t \exists F (F$ typically causes $x$ to $\varphi$ in circumstances $C)$, i.e. (simplifying somewhat) realises the disposition to $\varphi$, just in case assigning $P$ to $F$ and $t$ to $x$ renders $\varphi$ typically causes $x$ to $\varphi$ in circumstances $C$. True.
and so on. There is an evident circularity: the property \( M \) is defined by a formula which includes a variable place—\( X \)—such that assigning \( M \) to \( X \) satisfies the Ramsey formula, relative to other assignments, if our initial pre-Ramsified theory \( FT = \forall x T(x, M, M_2, ..., M_n, I_{(i)}, O_{(j)}) \)—is true.

To be sure, a form of circularity is part and parcel of any impredicative specification. The property of being a number is one of the properties we generalise over in our standard impredicative specification of being a number. If Shug is the only lover in the world and Shug loves two people, Senga and himself, then \( \exists x (\exists y x \text{ loves } y) \) refers to Shug, who is also one of the satisfiers of \( \text{”Shug loves } y \text{”} \). On a realist view of properties, it is not ruled out that the property \( P \) designated by the dispositional schema \( \exists F \text{ } F \text{ typically causes } x \text{ to } \varphi \text{ in circumstances } C \), could itself be the sole realiser, the sole property which causes \( F \)-ing. But suppose many properties realise the disposition to \( \varphi \)? Could one of them also be the disposition itself? Could the property of sounding the alarm whenever an intruder is detected via either ultra-sonic, infra-red or pressure methods itself be a particular one of those detection methods, the infra-red one say? This seems absurd.

In the present case, the property of being an \( x \) such that \( x \) bears a property which plays the \( M \)-role itself plays the \( M \)-role, according to conceptual functionalism. Yet if multiple realisability holds, differing species, indeed differing individuals granted the Quine/Darwin point, will have differing realising properties. Hence there will be an overdetermination of properties playing the \( M \)-role in each individual: in each case the role will be played by both a parochial realiser, and the property of having an \( M \)-role property itself. Each mental state will be doubly realised, at least, in each individual. A Martian and a human who are both in pain both have the property of having a property playing the pain-role, since this is what it is to be in pain. ANd there will be at least two properties playing that role in the human, a neural one and the distinct property of being in pain; likewise for the Martian.

This is surely bizarre.\(^3\) The only way out for the functionalist is to drop conceptual functionalism for a form of psychofunctionalism in which mental states are defined by Ramsifying on a formula \( T(x, S_1, S_2, ..., S_n, I_{(i)}, O_{(j)}) \) in which the \( S \) terms are all terms from

---

3. Lewis, indeed, rules this scenario out tout court on the basis of his assumption that the notion of disjunction makes sense as applied to properties but ‘disjunctive properties’ are causally ineffectual—‘Reduction of Mind’, p. 307.
scientific psychology, not folk psychology. Such a non-conceptual functionalist cannot, of course, explain our grasp of the concepts of folk psychology. The functionalist definition cannot provide an analysis of mental concepts, cannot tease out what we are implicitly aware of in our grasp of mental terms since most of us will not have the concepts $S_1, S_2, \ldots, S_n$. But even if one is untroubled by this limitation, there is a further objection which, I will argue, is fatal for all versions of functionalism considered so far.

IV

Intra-Individual Multiple Realisability. The problem lies in the massive difference between the complexity of FT and the complexity of the human brain. Suppose Senga is a fairly typical individual and imagine we have a total description $D$ of all Senga’s mental states at time $t$. Senga will be subject to sensory input $I_t$ and will exhibit output $O_t$. There will also be various dispositional properties she has at $t$, some of which may be expressible in simple form as a conjunction of input:output conditionals $[\text{if } I_{t} \langle i \rangle \text{ then } O_{t} \langle j \rangle ]$. FT plus the definitions of mental states, conjoined with $D$, will entail that Senga exemplifies at $t$ a ‘quasi-abstract’ mental structure $S$ which satisfies the Ramsey matrix $RM$. For each of the mental states $M_1, \ldots, M_o$, which $D$ attributes to Senga there are nodes $N_1, \ldots, N_o$ with each $N_i$ realising the corresponding $M_i$. S is quasi-abstract because it is ‘tied-down’ at the fringes to specified (but not species-specific) input and output and its nodes are linked together by connections of a specified type—the specifications will be physical, for a physical form of functionalism.

To illustrate, if Senga is feeling mildly anxious at $t$, then if FT is correct there will be a property, $N_k$, say, which is the realisation in Senga of mild anxiety. Since Senga is a human, this property is presumably a neural property, as will be the other properties realising her other mental states at $t$. The interconnecting links in $S$ will connect, generally indirectly, at least some of the nodal properties with $I_t$ and $O_t$. FT together with $D$ may well also entail that these nodal properties stand in relations to the input and output properties featuring in the specifications of the dispositional powers Senga has at $t$, such as the properties in the simple input:output conditionals of the form $[\text{if } I_{t} \langle i \rangle \text{ then } O_{t} \langle j \rangle ]$. 
The problem is this: the mental structure $S$ is realised, we have seen, in Senga’s brain and nervous system. But the latter is an incredibly complicated network. Senga will have around $10^{11}$ neurons, (and these in turn we know not to be the only important elements in the brain)\(^{32}\) with an average of 3,000 connections per neuron, thus around $3 \times 10^{14}$ connections overall. Thus even if not all sets of $n$-tuples of brain components determine a really existing relation (that is, even if the second-order comprehension axiom is false) there will be a vast number of real $n$-tuple relations for $n < 3 \times 10^{14}$ which then can be combined in a plurality of ways to form a vast number of brain structures.

Structure $S$, by contrast, since it satisfies the matrix generated from FT by ‘Ramsification’, will be less complex by many orders of magnitude, even if FT is not a comprehensible psychofunctionalist theory but a large collection of platitudes current among the ‘folk’. It is thus well-nigh inevitable that there will be a large number of $S$-structures exemplified in Senga’s brain at time $t$. That is, the brain is so much more complex than FT that there will be many brain structures $S^*$ which are not only isomorphic with $S$ but which also agree with $S$ on all constant, non-structural places in $RM$ (call such a structure a ‘tightly isomorphic’ one). Functionalists identify minds with structures of the right type, structures which satisfy $RM$. Since there will be vast numbers of such structures in Senga’s brain, all linked up in the right way with input and output, it would seem that multiple personality syndrome, about which some have expressed scepticism, is actually exemplified, in buckets, in all of us. I take this to be a reduc-tio ad absurdum of this brand of functionalism.\(^{33}\)

Nor is this all: as well as the multiplicity of neural structures tightly isomorphic with $S$ in Senga’s brain, there will also be huge numbers of other neural structures which satisfy $RM$ exemplified in Senga’s brain. It is plausible, indeed, that just about every mental state will be exemplified in her brain. Suppose that, whilst $M\kappa$ is the state of mild anxiety, $M\iota$ is the state of being stress-free. There is sure to be a structure $L$ in Senga’s brain which

---

\(^{32}\) Perhaps we need to go down to the level of all the potassium, sodium and chlorine ions and so forth, or at least to neuro-transmitters. Perhaps we will eventually have to consider even more complex and holistic quantum properties of the brain.

\(^{33}\) Note that if we utilise instead the modified Ramsey sentence for each state $M$ the result will be radical eliminativism since all the uniqueness clauses will fail.
satisfies RM and is such that the lth property is one which Senga possesses.

Perhaps an analogy will help to make this point more vivid. Consider pinball machines as input:output devices, with the momentum of the ball as the plunger hits it being the input and the momentum as it exits past the bottom flippers the output. Consider, in particular, machines in which more than one ball can be in play at a time so that inputs and outputs are n-tuples of momenta (the analogy of course is with sensory input and motor output, though the number of sensory nerves need not equal the number of motor nerves.) Two machines could be ‘behaviourally’ indistinguishable in that they embody the same input:output function— if balls $b_1, ..., b_n$ in machine one are physically indistinguishable from $c_1, ..., c_n$ in machine two and, \( \forall i, 1 \leq i \leq n, c_i \) is fired with the same momentum as $b_i$, then $\forall i, 1 \leq i \leq n, c_i$ will exit the second machine with the same momentum as $b_i$ exits the first—yet a ‘functionalist’ about pinball machines might distinguish the two machines if the internal structure of bumpers and so forth which the balls collide with differed.

Imagine, next, that we are allowed to describe possible trajectories of balls only by means of flow charts which utilise representations of fewer bumpers than there actually are in the machines. Thus the diagram

\[\text{IN} \rightarrow \text{OUT}\]
represents a possible path of a ball in which it reaches bumper 1 possibly after colliding with several other bumpers, then reaches 2, again perhaps not directly and so on. The above pattern is thus realised twice over in the situation below:

![Diagram of a ball path with multiple bumpers and paths]

Firstly—with 1 = 2 and 2 = 4—by the path marked out by undashed lines into 2 (via 1), across to 6 (via 3), back to 2 (via 5) and finally out (via 6). But it is also exemplified by the dashed line path—now setting 1 = 4 and 2 = 1—the path being first to 4 direct, then directly to 1, then back to 4 (via 5), then out. If there are many more bumpers than we are allowed in our charts, it is highly likely that a given potential path can be exemplified many times over. Similarly it is likely that just about any flowchart will be actually instantiated in a given machine if its complexity is vastly greater than that of the chart.

Although functionalist networks need not be representable as flow charts (for non-homuncular versions at any rate), the same point holds for the brain. The structure S which the functionalist takes to be Senga’s total mental state at time t will be vastly less complex than her brain. If we can select brain components to play the role of nodes as we wish and select any neural relations, however indirect and convoluted, we like as representing an ‘arrow’ or relation from one node to another, we will be able to discern S, tied down as it is to all the possible inputs Senga at t
is capable of transforming into outputs, many times over in her brain.

Against this, one may wonder whether all or virtually all mental states will be realised in Senga (according to functionalism). Will each constant place in RM be occupied by a property or relation of the right type, in particular by an input or output property or interconnection relation of the right type? Since the interconnection relations have to be specified in a fashion general enough not to exclude e.g. silicon-based life, there is little doubt that a structure can be found in Senga’s brain which fits any given bill. But what about input and output properties? Suppose $M_1$ is not the state of being stress-free but rather the state of seeing a Tyrannosaurus Rex-like creature bearing down on one and Senga in fact is looking intently at a blank piece of paper at time $t$. Even in this case, it is not obvious that there will not be a structure of the right type in Senga’s brain given the non-atomistic nature of mental states, the possibility of hallucination and so forth. Though it is not actually the case that Senga, whilst staring at the blank paper, is hallucinating that a T. Rex is bearing down on her, she might have been. FT will entail that a particular structure $S_1$ would have obtained in this case, a structure which includes various relations to the input and output properties in that counterfactual situation—which happen to be the actual input and output properties—as well as to other possible input and output properties. Given the incredible richness of brain structures, it is likely that in the actual situation there will be a structure $S_2$ which is tightly isomorphic to $S_1$.

Even if there are no tightly isomorphic structures which realise appearing to see a T. Rex etc., focus instead on mental states less directly linked to input and output—Senga’s various beliefs about history, the reliability of her friends, her political attitudes etc. Intra-individual multiple realisability together with functionalism will entail that Senga believes both that Kenneth mac Alpin united the Scots and Picts in 847 AD and that he did not, that Shug is a two-timing cheat and that he is not, that taxation levels should be raised to pay for better welfare services and that they should not. Again I think this reduces functionalism to absurdity. Maybe most individuals are prone to holding implicitly contradictory attitudes concerning taxation and social services, but it is absurd to think that every individual holds contradictory beliefs and conflicting attitudes on virtually every topic.
not closely linked with perception of, and reaction to, the current environment.

Finally consider once again the suggestion of Lewis that in the light of multiple realisability we treat theoretical terms as ambiguously referring to each realiser (see footnote 23), or else treat them as context-dependent terms, like ‘toxic’, picking out different states relative to different contexts, for example relative to different species or individuals. Intra-individual multiple realisability shows that ascriptions of mental states to a given individual will, on the first account, be infected with massive ambiguity, and on the second will have to be relativised to something narrower than even a given temporal stage of an individual—it is very hard to see what. As if that is not bad enough, there will then arise a deep problem for mental causation. I explain Shug’s behaviour as caused, in part, by a strong desire to drink whisky. But ‘the strong desire to drink whisky’ ‘partially refers’ to a large number of brain states, each linked far too indirectly, surely, with behavioural output to count as a cause. If so, the result is epiphenomenalism.

The Ambigamist might appeal to a broadly Millian notion of causation, with its standard examples of lighting matches and so forth.34 When we talk of ‘the cause’ we are usually singling out a salient partial condition, perhaps the one most amenable to human manipulation and control, from within that total state which is a necessary and sufficient condition for the event we are interested in (or for the state containing that event). We pick out the friction on the match as it strikes the matchbox as the cause, not the ambient oxygen. So, the Ambigamist or context-relativist may continue, though there are many structures exemplifying Shug’s psychology at time t and thus many neural states partially referred to by ‘strong desire to drink whisky’ each of these ‘desire’ states must be linked with input and, most relevantly here, output—including the act of drinking—in exactly the same way, relative to the specifications available in the functionalist theory. So each is a part cause of the action and hence each has a perfectly good title to be classed as a cause.35


35. Here I am indebted to Jim Edwards.
Note, though, the strong disanalogy between the matchstick case and the present one, where there is a myriad of ‘partial causes’ each playing exactly the same role in the causal story. This latter case is much more analogous to panic buying in shortages. After news of a danger of fuel shortages, the petrol stations run dry because millions of people fill up their cars when their tanks are quite full rather than letting them run near to empty as they usually do. Does it make sense to say of any one of the (entirely rational!) panic buyers—Hamish H. McSporran, of 42, Glen Road, Auchenshuggle, say—that, just as striking the match caused it to light, he caused the nation-wide petrol shortage (rather than caused a very small part of it)? Once again, a form of functionalism is seen to lead to epiphenomenalism.

To sum up, the Ramseyan version of functionalism is incapable of accounting for multiple realisability. The version I described as standard can accommodate multiple realisability but only at the cost, in the case of conceptual functionalism, of a highly problematic reduplication of realisers of mental states in each individual. Finally all versions of functionalism succumb to intra-individual multiple realisability which has the absurd consequence that human organisms host vast numbers of different minds, and exemplify just about every possible mental state simultaneously. I conclude that functionalism fares, at very best, no better than the behaviourist and central state materialist theories it replaced as the most popular physicalist explanation of mind.37

School of Philosophical Studies
Queen’s University, Belfast
Belfast BT7 1NN
email: a.weir@qub.ac.uk

36. This example also shows why Lewis was wise to shun Carnapian responses to multiple realisation. However useful epsilon terms are in mathematical logic in implementing choice axioms, can we really refer to a particular individual such as Hamish H. McSporran by an English phrase such as ‘an arbitrary panic-buyer’. How do we do it; or does Carnap, up in Hilbertian heaven, do it for us, by making Hamish the output of a magical selection function?

37. Ancestors of this paper appeared as part of talks at Birmingham and Glasgow; my thanks to audiences there, particularly to Jim Edwards and Philip Percival in Glasgow. Thanks also to John Divers, Ernie Lepore and Alex Miller for commenting on earlier drafts of the paper.
MORE TROUBLE FOR FUNCTIONALISM

References


