Lewis, Loar and the logical form of attitude ascriptions

Simon Beck
Philosophy Department, Rhodes University, P.O. Box 94, Grahamstown 6140, Republic of South Africa

Received July 1987

In this article, the attempts by David Lewis and Brian Loar to make perspicuous the logical form of sentences ascribing propositional attitudes to individuals are set out and criticized. Both work within the assumption of the truth of 'type' physicalism, and require that logically perspicuous attitude ascriptions be compatible with the demands of such a doctrine. It is argued that neither carry out this task successfully — Lewis's perspicuous ascriptions have counter-intuitive implications, while Loar's avoidance of these undermines type physicalism itself.

Die pogings van David Lewis en Brian Loar om die logiese vorm van sinne wat proposisionele houdings aan individue toeskryf oorsigtelik te maak, word in hierdie artikel uiteengesit en beoordeel. Altwee lê hulself die beperkings op wat voortvloei uit die aanname dat 'soort'-fisikalisme waar is, en vereis dat logies oorsigtelike houdingtoeskrywings met die eise van 'n derglike teorie versoen kan word. Daar word aangevoer dat geen van beide daarin slaag nie — Lewis se oorsigtelike toeskrywings hou implikasies in wat bots met ons intuïsies, en Loar vermy hierdie implikasies slegs ten koste van die soort-fisikalisme self.

David Lewis and Brian Loar put forward similar versions of a way of making perspicuous the logical form of sentences ascribing propositional attitudes (beliefs, desires, etc.) to individuals. Both Lewis and Loar hold versions of 'type' physicalism and this places certain constraints on what will count as adequate renderings of attitude ascriptions. If an account is to be acceptable, given the truth of some form of physicalism, the logically perspicuous sentences it yields must be *extensional*, and they must contain quantification over only *physical*, and not irreducibly mental, states.

Both philosophers attempt to meet these constraints by espousing a functionalist view of the mind, and interpreting propositional attitude ascriptions functionally. Thus perspicuous attitude ascriptions will attribute states with certain causal (functional) roles to the relevant individuals; and the states which play these roles will in fact be physical states. But to avoid perspicuous attitude ascriptions having implausible entailments, both suggest that clauses narrowing the scope of the causal roles involved need to be introduced into those sentences.

At this point, it will be argued, problems arise. For Lewis, once he starts introducing extra clauses, considerations similar to those prompting the initial move dictate that the scope of the causal roles involved be limited further: he just doesn't go far enough. Loar takes this process to its logical conclusion, but in doing so undermines the very strategy of introducing extra clauses in the first place. On top of this, it will be argued that he undermines a central tenet of his (and Lewis's) philosophy of mind — type physicalism. Thus in the attempt to provide the correct logical form of attitude ascriptions, the strategy of Lewis and Loar ends up in conflict with their expressed views on the mind.

Section I

Lewis and Loar are both type-physicalists. That is, they accept not only the doctrine that all mental states happen to be physical states (token physicalism), but also some form of the doctrine that *types* of mental state can be reduced to types of physical state, usually expressed in the claim that mental state-types *are* physical state-types. This view of mental states has important consequences with regard to the sentences we use to ascribe mental states to people; for the suggestion of physicalism is that if we are to make ourselves clear, then we should use only the language of physical science. This would have two immediate effects: firstly, if sentences ascribing mental states were to be expressed using only the language of physical science, then mental terms would not appear in them. And secondly, the language of physical science is generally regarded as taking the form of the first-order predicate calculus which is exclusively extensional, whereas the sentences we use to ascribe certain central mental states are notoriously intensional.

Both Lewis and Loar believe that by interpreting mental ascriptions functionally the requirements of exclusively physical terminology and extensionality can be met. The functionalist view sees mental states as being essentially functional states — states whose definitive characteristics are their causal roles. These definitive roles are made up of the ways in which the states in question interact causally in a system of environmental inputs, other mental states and behavioural outputs. The upshot is that to ascribe a mental state to someone is to ascribe a state with a certain typical causal role.

Functionalists generally, and Lewis and Loar in particular, believe that the states which in fact fill the causal roles so outlined are physical states: but functionalism is in itself 'topic neutral'. It merely refers to states by means of their second-order properties — i.e. the roles they play — without committing itself on their first-order properties. Sentences ascribing mental states will then, when rendered perspicuous, ascribe functional states which need by no means be irreducibly mental.

The requirement of extensionality appears problematic especially with regard to ascriptions of propositional attitudes. These ascriptions typically take the form

Z F-s that *p*

where '*F-s*' is replaced by 'believes'/'intends' etc, and '*p*' by a proposition.[1] Sentences of this form are intensional; that is, in them the normal rules for substitutivity break down — substitution of terms by identical terms and of predicates by co-extensive predicates is not guaranteed to preserve truth. But once sentences of propositional attitude are rendered functionally, that is, as ascribing states with certain causal roles rather than particular propositional content, the ascriptions will no longer have their characteristic intensional structure.

Precisely what form they will take must now be made clear, but it at least appears that the functionalist programme fits neatly with the tenets of strong physicalist theory.

Section II

Lewis proposes a method for defining the theoretical terms of new (i.e. unfamiliar) theories using only the old, already understood terms occurring in the theory.[2] Theoretical terms are those introduced by a theory, and Lewis points out that our only clues to their meaning are the sentences of the theory in which they occur. It is from the relation of a theoretical term to other theoretical terms, and ultimately to the terms we already understand that we gain understanding of the term itself: it derives its meaning from the way in which it is embedded in the theory. The theory itself thus provides an implicit definition of its theoretical terms. Lewis exploits this belief in providing explicit definitions of theoretical terms, and ultimately in making perspicuous the logical form of attitude ascriptions, as will be spelt out in this section.

The mental terms which occur in our everyday speech, holds Lewis, are to be treated as the theoretical terms of scientific theories. The theory from which they derive their meaning is commonsense or 'folk' psychology — that is, the theory made up of all 'the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses' as well as 'all the platitudes to the effect that one mental state falls under another — "toothache is a kind of pain", and the like' (Lewis 1972: 212). These are generalizations we all know and know that others know, etc.: the principles we use in explaining and predicting behaviour. Since the theory of commonsense psychology envisaged by Lewis is a causal theory, this method of definition fits in with Lewis's functionalism: mental terms are defined via their roles in a causal system.

Formally, the treatment of folk-psychology to arrive at explicit functional definitions of mental terms runs along the following lines. Firstly, all the platitudes of folk-psychology are conjoined into one long sentence, which is called the 'postulate' of the theory and symbolized as

$$T(t_1 \ldots t_n)$$

where t_1 to t_n represent the theoretical terms of theory T, as they occur in that theory. What happens next is that each theoretical term is replaced by a variable — a distinct variable for each term — wherever it occurs in the postulate of T. In order to facilitate the replacement of terms by variables, Lewis advocates the introduction of uniformity to the theoretical terms by nominalizing all of them. That is, no matter whether they appear in the theory as predicates, functors, or whatever, they are converted into names. This causes no important alterations to the theory, since, as he points out, 'names can purport to name entities of any kind' (Lewis 1970: 80). Such nominalization allows the use of only one type of variable for all the theoretical terms of the theory.

Once the nominalized terms are all replaced by variables, we will have

$$T(x_1 \ldots x_n)$$

which is the same formula as the previous one, only devoid of all theoretical terms, containing only terms already understood and variables. This modified postulate then has existential quantifiers prefixed to it, enough to bind each of the variables just introduced. The result

$$(\exists x_1 \ldots x_n)T(x_1 \ldots x_n)$$

is a sentence containing no theoretical terms. Assuming that there is a standard interpretation of the 'old' terms, this sentence can be read as stating that there are states which play such-and-such causal roles.

Lewis's account continues to become more complex, but from what we have so far we have enough to glean a method for explicating separate ascriptions of mental states. Before this can be done, however, what is needed are explicit definitions of mental terms. Here Lewis brings an important assumption into play: the assumption that scientific theories (and folk-psychology is to be counted amongst these) are *uniquely realized*. The realization of a theory will be the n-tuple of entities which are the values of its bound variables (assuming that it is uniquely realized). Lewis (1970: 84) suggests that there is no good reason why we should not expect theories to have unique realizations. And one advantage of unique realization is the facility it brings to the explicit definition of terms. For we can define a term — t_1 for instance — as the first entity in the series which uniquely realises T (and, likewise, t_i as the *i-th* member of that series, and so on).

Symbolically, this can be put as follows. Firstly unique realization is expressed so:

$$(\exists y_1 \ldots y_n)(x_1 \ldots x_n)\ [T(x_1 \ldots x_n) \text{ iff } (y_1 = x_1 \,\&\, \ldots \,\&\, y_n = x_n)]$$

To define t_1, we need only prefix this formula with (ιy_1) and alter the existential quantifiers to $(\exists y_2 \ldots y_n)$, so:

$$(\iota y_1)(\exists y_2 \ldots y_n)(x_1 \ldots x_n)\ [T(x_1 \ldots x_n) \text{ iff } (y_1 = x_1 \,\&\, \ldots \,\&\, y_n = x_n)]$$

This final formula is to be read as saying that there is precisely one state which stands in the causal relation to other states as outlined (and that state is the referent of t_1). Each theoretical term, then, is co-extensive with a definite description picking out a state with a particular causal role. Applied to folk-psychology, the account entails that every mental ascription assigns, to the individual concerned, a state defined by its causal role. For any such ascription, made logically perspicuous, would simply involve the last sentence above with a clause added to the effect that the individual concerned was in that state, thus:

$$(\iota y_1)(\exists y_2 \ldots y_n)(x_1 \ldots x_n)\ \{[T(x_1 \ldots x_n) \text{ iff } (y_1 = x_1 \,\&\, \ldots \,\&\, y_n = x_n)] \,\&\, Z \text{ is in } y_1\}$$

— 'there is precisely one state which plays the role outlined, and Z is in that state'. A sentence of this form would replace any sentence of the form 'Z believes that p', etc. Its expression in the first-order predicate calculus shows it to be, as required, extensional; and Lewis assures us that the values of its variables will turn out to be physical states.[3]

Section III

Unfortunately, sentences of propositional attitude rendered in this way have implausible implications. For, in ascribing to Z the belief that p one ascribes to Z the (physical) state which plays such-and-such a causal role; and this implies that there is a particular type of physical state which plays that causal role in all individuals which have beliefs, regardless of their physical make-up. The logically perspicuous attitude ascription

$(\imath y_1)(\exists y_2 \ldots y_n)(x_1 \ldots x_n)$ {[$T(x_1 \ldots x_n)$ iff $(y_1 = x_1$ &. . .& $y_n = x_n)$] & Z is in y_1}

implies

$(\imath y_1)(\exists y_2 \ldots y_n)(x_1 \ldots x_n)$ [$T(x_1 \ldots x_n)$ iff $(y_1 = x_1$ &. . .& $y_n = x_n)$]

In other words, the implication of Lewis's perspicuous attitude ascription (given the assumption that the values of the variables are in fact physical states) is that the belief that p (that mental state-type) is identical to a specific type of physical state.

That this follows from Lewis's account should not be surprising, since it is one way of expressing the type-physicalist thesis, but it is extremely implausible. There is just no reason *why* creatures of all species in this world[4] should have to share our neuro-physical make-up if they are to have the mental states we have.[5] Lewis (1969) acknowledges this in his later writings, and suggests that functional definitions of mental states should be limited by relativizing them to species.[6] Thus to say that Z believes that p would be to say that Z is in the state which, *for species* S (of which Z is a member), plays causal role R (the role outlined by folk-psychology). This would allow members of other species to share our propositional attitudes without necessarily sharing our physical states, and yet still allow type-type psychophysical identities (within species): the unwelcome implications of the Lewis method appear to have been avoided.

But once the problem has been raised with regard to species, we are on a slippery slope: Lewis's problems don't stop here. Certainly we can't hope for more than species-relative type physicalism, but can we even hope for that much? Intra-species type identities are not much more plausible than their inter-species counterparts. The problem is that it is quite possible, indeed probable, that different individuals within a species have different physical states which play the required functional roles. Take for example the mental state of believing that p. Different individuals reach their beliefs in quite different ways — one may learn the hard way and infer that p from a number of experiences, one may perceive p to be the case in one incident, or one may be told it by some authority (or one may reach it in any number of other ways). There is no reason why a belief which is reached in such diverse ways, and against a background of different beliefs in the case of each individual, should be realized by precisely the same physical state in each individual. Individuals' physical states could be, and most probably are, quite different despite the fact that each plays the causal role used in the definition of the belief concerned.

Lewis (1969: 233) seems to realize this point in one of his articles, writing that 'pain . . . might even be one brain state in the case of Putnam, another in the case of Lewis', but he does not alter his account to make it consistent with this. The obvious move for him to make would be to relativise definitions of mental states, not only to species, but to *individuals*. 'Z believes that p' would now be expressed as 'Z is in the state which, *for* Z, plays role R'. This sentence implies only that mental types are identical to physical types within individuals.

This is the final amendation that Lewis (in effect) acknowledges. But it is not sufficient to avoid his perspicuous attitude ascriptions having implausible implications; he just doesn't go far enough to avoid his account being undermined. For by reasoning parallel to that which has necessitated the changes so far, the definitions of propositional attitude-terms will have to undergo a further restriction. The physical state which plays the relevant causal role may be replaced by another state of that individual at a different time: the definitions must be relativized to an individual *at a time*. The ascription form then becomes 'Z is in the state which, for Z at time t, plays role R'.

We have still not, however, reached the end of the slippery slope. Attention needs to be focused on Lewis's demand for *unique realization* at this point. It was as a result of taking folk psychology to be uniquely realized that Lewis was able to provide explicit definitions of propositional attitude terms without using mental terminology. With the amendments we have envisaged, the most he could claim is that folk psychology is uniquely realized for an individual at a time; this is still sufficient to allow propositional attitude ascriptions to be expressed as 'Z is in *the state* which . . .'; that is, this is consistent with restricted type-identities (within an individual at a time).

Section IV

The question to be faced now is whether Lewis's hope that folk psychology will be uniquely realized (even for an individual at a time) has any justification. The trouble is that there may be more than one state of Z which would play the relevant causal role for some mental state in Z at t. All a physical state has to do is to be (counterfactually) causally related to Z's other physical states and the external world in the general way which typefies the mental state. There is nothing in this which would prevent a number of Z's possible physical states at t fulfilling the conditions for realizing that mental state. In other words, a given mental state can be multiply realized by physical states, even when we relativize to individuals and times. Lewis's hope for unique realization of folk psychology may well be a vain one; we certainly have no reason to believe otherwise.

The possibility of the multiple realization of the belief that p (as well as any other mental state) has the result that, according to Lewis's account, Z will never believe that p — because Z will never be in *the* state which would play the relevant causal role. Z will never be in that state because there is no such single state as the definite description demands.

At this point, Brian Loar's account of propositional attitude ascriptions becomes relevant. Loar accepts that definitions of mental-state terms must be relativized to persons and times, and he also acknowledges the possibility of the multiple realization of folk psychology. Following Loar, the definite description must be dropped from the modified ascription, which would yield a sentence of the form, 'Z is in *a* state which, for Z at t, plays role R', formally,

$(\exists x_1 \ldots x_n)$ [$T^{z \text{ at } t}(x_1 \ldots x_n)$ & Z is in x_i]

Handled in this way, t_i no longer denotes a single type of state, but a set of states — the set of all *i-th* members of realizations of T (for Z at time t).[7]

This alteration heralds what appears to be a major breakaway from Lewis's account, especially with regard to type physicalism. Loar's version no longer implies psychophysical type *identities*: for identity is a 1:1 relation, whereas this version implies no such relation between mental and physical types. The identity claim which Lewis tried so hard to maintain — to the extent of limiting it considerably by relativization — has been swept aside.

Section V

Does this mean that Lewis-Loar perspicuous attitude ascriptions are not compatible with type physicalism, but only with token physicalism? Loar answers 'no' to this question. He (1981: 15 – 17) claims that, on the contrary, token physicalism together with functionalism *implies* type physicalism — although of a weaker kind than that envisaged by Lewis. He points out that the combined doctrines of token physicalism and functionalism imply that any particular mental state-token (say *Z*'s belief at *t* that *p*) will be identical with a physical state-token which plays a particular causal role. Now, it is in virtue of its possessing certain first-order physical properties that that token can play that causal role. And whatever such first-order properties (i.e. types) enable a state-token to play such a role can be correlated with the type of mental state of which the particular mental state was a token. Thus we have a *correlation* of mental types with first-order physical types, relative to an individual at a time. This is the type physicalism which Loar espouses — a doctrine weaker than that which puts forward type identities, yet stronger than 'mere' token physicalism.

Section VI

Loar has then offered a version of the logical form of attitude ascriptions which appears to be consistent with the demands of a version of type physicalism. What needs to be investigated now is whether the account is indeed adequate.

The first question I wish to ask towards this end is: why does Loar insist on relativizing mental state definitions to persons at times? This question arises because the motivation which led Lewis to relativization was the desire to retain psychophysical type *identities*. Since Loar is no longer concerned with identities but is satisfied with correlations, realizing that more than one state can realize a mental state even with regard to an individual at a time, it is a mystery why he should want to relativise this far. This is the thin end of the wedge; why does he not stop the relativization process with species, since at this level it is also the case that more than one physical state can be correlated with a mental state? Indeed, why even relativize to species? The only reason for Loar's relativization, since he accepts multiple realizability at the lowest level, seems to be that the number of physical states correlated with the mental state will thus be limited. But, leaving aside the weakness of this criterion, it is not clear that it is even relevant. For to be a mental state all that is required is that a state plays a certain very general role, and there is no reason why an indefinitely large number of *Z*'s possible states at *t* could not fulfill this function, at least in some cases. Loar offers no suggestion as to why this should not be so; and in the absence of such a reason, the only apparent motivation for relativizing to persons and times falls away. By giving up the stipulation of uniqueness in perspicuous attitude ascriptions, and opting for type-type correlations rather than identities, Loar undermines the clause he himself insists upon — that relativizing the state concerned to an individual at a time.

It should be noted that Loar's argument outlined above, to the effect that token identities imply relativised type correlations, also offers no motivation for the relativization. For although it is true that any token state we choose is the state of some individual at some time, the function of the assumption of token physicalism in the argument is to ensure that whatever state actually plays the relevant causal role, it will be a *physical* state. That it is a state of *Z* at *t* plays no further part. That the state does belong to a specific individual at a time is quite irrelevant to the general theoretical concept of a belief (for example); the theory of folk psychology applies to *all* individuals with mental states, its conditions and generalizations don't make it apply only to some specific individual.

Thus although Loar is correct in that token identity and functionalism *do* imply relativized type correlations, this argument provides no reason why Loar should insist on relativizing to individuals and times. The argument works equally well for unrelativized type correlations: it implies a correlation between mental state types and the physical types in virtue of which tokens could fulfil the required role in *any* individual at *any* time.

Section VII

What begins to emerge from this discussion is that there appears to be very little difference between Loar's type physicalism and token physicalism. Given Loar's lack of motivation for relativizing mental definitions (and with them, psychophysical correlations) we end up with a correlation between a mental type and a potentially unlimited number of physical types. It would certainly be stretching things to call this a *reduction* of mental types to physical types, which indicates the similarity between this position and the token physicalism from which Loar wishes to distance himself. Token physicalism, as pointed out above, allows that each particular mental state (token) is identical to some or other physical state, but denies that mental states can be reduced to physical states.[8] And even when relativized to persons and times, the potentially large number of physical correlates to a mental type still suggests the similarity between Loar's and token physicalism.

An attempt to alleviate the suspicion that the kind of type physicalism which accompanies Loar's explication of propositional attitude ascriptions is no better than token physicalism might take the following line.[9] What distinguishes Loar's physicalism from token physicalism, runs this response, is that Loar's physicalism associates many physical types with *one* mental type (relative to an individual at a time), whereas token physicalism at most associates many physical types with *many* mental types — the physical properties of a physical token could be correlated with all the mental properties of the identical mental token.

But this response doesn't work: token physicalism when viewed in terms of type correlations for a person at a time is not a many:many correlation, but a 1:many correlation, just like Loar's 'type' physicalism. This is because mental tokens are no different from Loar's mental types, when those types are limited to an individual at a time. Loar's are mental types which could only *possibly* have one token, and thus the type/token distinction collapses in this context. A mental type like 'believing that *p*' has different tokens in that it can occur in different individuals or at different times; but once the context is limited to a person at a time, these possibilities are ruled out: '*Z*'s belief at *t* that *p*' is not a term of 'divided reference'. It may be suggested that a belief, for example, that rhubarb is nourishing, may take slightly different forms — with tokens having slightly different contents falling under this one type. But Loar (1981: 58) rules this out by his contention that the contents of the beliefs with which he is concerned are 'fine-grained' propositions: slight differences in content would mean beliefs of different types.

Believing that *p* for *Z* at *t* is thus indistinguishable as to type or token, and so far as token physicalism is concerned, *one type* can be associated the physical properties of corresponding physical tokens — a 1:many correlation. The

suggested defence of Loar's type physicalism fails, leaving the distinction between such type physicalism and token physicalism hazier than ever.

The distinction becomes even more problematic when it is realized that the argument to the conclusion that, in Loar's context, type and token become indistinguishable actually establishes more than just that conclusion. In showing that '*Z*'s belief at *t* that *p*' is not a term of divided reference, what is shown is that it is not a general term — a 'type-expression' — but rather the name of a token. Any attempt to establish a type physicalism with such terms representing the mental correlates must then be in trouble, since the mental correlates would not be types at all, but tokens. Loar's is just such an attempt.

Section VIII

Loar's exposition of propositional attitude ascriptions, though initially an improvement on Lewis's, is nevertheless extremely problematic. The attempt to avoid intuitively incorrect implications by giving up the requirement of unique realization, though unavoidable if he is to retain Lewis's general picture, involves Loar in serious difficulties. As we saw, Lewis was forced to progressively weaken his type physicalism by introducing relativising clauses into perspicuous attitude ascriptions, but Loar's type physicalism turns out to be so weak as to be no better than token physicalism. Thus in attempting to set out the logical form of attitude ascriptions in a way acceptable to type physicalism, Lewis and Loar end up undermining that very doctrine.

Notes

1. Or a sentence, for those averse to propositions.
2. This method is set out in Lewis (1970) and (1972).
3. There is actually a further important amendment which Lewis makes to this definition. He suggests (1972, p. 215 n 13) that '*Z* believes that *p*' should be treated as expressing a relation between *Z* and *p*, and only the relational term 'believes' is to be treated as a theoretical term. I deal with problems emerging from this in my 'Physicalism, functionalism and intentionality' (Unpublished, 1987).
4. Lewis's account is consistent with creatures in other worlds having different states realizing the functional roles — for, in his view, while 'believing' is a rigid designator, picking out a property shared by all creatures with beliefs, 'belief' is a non-rigid designator, picking out different states in different possible worlds.
5. This point is made by Putnam (1967).
6. In (1980) he suggests relativization to a population (p. 126).
7. Loar foresees problems for his account at this point and indulges in some fancy footwork to avoid them — see Loar p. 53.
8. Loar sees the need to defend his account against Davidson's contention that the mental cannot be reduced to the physical because the constitutive force of rationality in the mental realm has no echo in the physical realm. But, tellingly, his argument fails — see J. McDowell, Functionalism and anomalous monism, in E. LePore and B. McLaughlin (Eds.) *Actions and events*, 1985.
9. An argument along these lines was suggested to me by Dr S.R. Miller in correspondence.

Bibliography

Block, N. (Ed.) 1980. *Readings in philosophy of psychology.* Vol. 1. London: Methuen.

Lewis, D. 1966. An argument for the identity theory, in D. Lewis (1983).

Lewis, D. 1969. Review of Putnam, in N. Block (Ed.) (1980).

Lewis, D. 1970. How to define theoretical terms, in Lewis (1983).

Lewis, D. 1972. Psychophysical and theoretical identifications, in N. Block (Ed.) (1980).

Lewis, D. 1980. Mad pain and martian pain, in Lewis (1983).

Lewis, D. 1983. *Philosophical papers.* Vol. 1. New York: Oxford U.P.

Loar, B. 1981. *Mind and meaning.* Cambridge: Cambridge U.P.

Putnam, H. 1967. The nature of mental states, in N. Block (Ed.) (1980).

Tye, M. 1983. Functionalism and type physicalism. *Philos. Stud.*, vol. 44.