Brit. J. Phil. Sci. 40 (1989). 541–555 *Printed in Great Britain*

Connectionism, modularity, and tacit knowledge

MARTIN DAVIES

ABSTRACT

In this paper, I define tacit knowledge as a kind of causal-explanatory structure, mirroring the derivational structure in the theory that is tacitly known. On this definition, tacit knowledge does not have to be explicitly represented. I then take the notion of a modular theory, and project the idea of modularity to several different levels of description; in particular, to the processing level and the neurophysiological level. The fundamental description of a connectionist network lies at a level between the processing level and the physiological level. At this level, connectionism involves a characteristic departure from modularity, and a correlative absence of syntactic structure. This is linked to the fact that tacit knowledge descriptions of networks are only approximately true. A consequence is that strict causal systematicity in cognitive processes poses a problem for the connectionist programme.

1 *Tacit knowledge*
2 *Modularity*
3 *Connectionism*
4 *Syntax*
5 *Tacit knowledge again*
6 *Conclusion*

1 TACIT KNOWLEDGE

It is natural to introduce the notion of tacit knowledge through Chomsky's work. In *Aspects of the Theory of Syntax*, he wrote ([1965], p. 8):

> Obviously, every speaker of a language has mastered and internalised a generative grammar [i.e. a system of rules] that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he could become aware of them . . .

This notion of tacit knowledge of the rules, principles, or generalizations of language recurs throughout his work; and several different pieces of terminology are used to express the same fundamental point.

Thus ([1965], p. 8), 'what the speaker actually knows' is equated with the speaker's *competence*. Then ([1976], pp. 164–5), in order to sidestep what are argued to be irrelevant objections based on intuitive connections—for example, between knowledge and justified belief, and between competence and ability—the technical term *cognize* is introduced, and is explicitly linked with tacit knowledge ([1980], pp. 69–70):

> The particular things we know, we also cognize. . . . Furthermore, we cognize the system of mentally-represented rules from which the facts follow. . . . And finally we cognize the innate schematism, along with its rules, principles, and conditions. . . . Thus 'cognizing' is tacit or implicit knowledge . . . [C]ognizing has the structure and character of knowledge, but may be and in the interesting cases is inaccessible to consciousness.

(See also [1988], pp. 9–12.)

Ordinary speakers know—in the familiar everyday sense—and also cognize facts about, for example, what various complete sentences mean. In addition, they cognize—even though they do not know in the ordinary sense—the facts from which those first facts follow.

We might think of the first facts as stated by the theorems of a systematic theory. If we continue to focus on facts about what complete sentences mean, the systematic theory will be a semantic theory. Then, the basic idea would be this. Ordinary speakers cognize and know the facts stated by these theorems. They also cognize—even if they do not know in the ordinary sense—the facts stated by the axioms from which the theorems are derived in the theory.

If we think of the issue in these terms, then it is easy to raise a major question which confronts any friend of the notion of tacit knowledge.

There will always be extensionally equivalent theories: distinct sets of axioms from which we can derive the same theorems about, say, the meanings of whole sentences. Given that fact, does it make any empirical sense to suppose that an ordinary speaker tacitly knows, or cognizes, or has internalized, one set of axioms, rather than an alternative set from which just the same theorems of the relevant kind can be derived? Does it make any sense to suppose that one theory is psychologically real, rather than another extensionally equivalent theory? This is essentially Quine's challenge [1972] to the empirical credentials of the notion of tacit knowledge.

Following a suggestion of Evans [1981], I would aim to respond to this challenge by construing tacit knowledge as a certain kind of causal-explanatory structure which underlies, or is antecedent to, the pieces of knowledge that the speaker has concerning complete sentences.

We can make the main idea clear enough if we follow Evans in considering two semantic theories for a very simple little language *L*. This language has just one hundred sentences, constructed out of ten names and ten predicates. The names are '*a*', '*b*', . . ., '*j*', and the predicates are '*F*', '*G*', . . ., '*O*'.

Consequently, the sentences are '*Fa*', '*Fb*', . . ., '*Fj*', '*Ga*', '*Gb*', . . ., '*Oj*'. These sentences have meanings which—as we theorists can see from the outside—depend in a systematic way upon their construction. Thus, all sentences containing '*a*' mean something about John; all sentences containing '*b*' mean something about Harry; all sentences containing '*F*' mean something about being bald; all sentences containing '*G*' mean something about being happy; and so on.

The two semantic theories that we are to consider are both theories of truth conditions for *L*. They assign just the same truth conditions to the sentences of *L*; but they differ in their internal or derivational structure. (We could just as well consider theories of meaning strictly so called; but theories of truth conditions have the advantage of familiarity.)

The first theory, T_1, is the *listiform* theory. It simply has one hundred axioms, one specifying the truth condition of each sentence of the language. The axioms of T_1 thus include:

'*Fa*' is true if and only if John is bald
'*Ga*' is true if and only if John is happy

and so on.

The second theory, T_2 is a structured or *articulated* theory. It has an axiom assigning a semantic value to each name of the language; and likewise, an axiom for each predicate. For the name '*a*', for example, we have

'*a*' denotes John

and for the predicate '*F*', for example, we have

a sentence coupling a name with the predicate '*F*' is true if and only if the object denoted by the name is bald.

From the twenty axioms of T_2, we can derive just the same truth condition specifications as those that can be derived trivially from the axioms of T_1. The two theories are extensionally equivalent; though they are not, of course, logically equivalent.

Suppose that there is a speaker who uses the sentences of *L*, with the truth conditions which both theories agree in assigning. What evidence can we imagine having, which would incline us to attribute to that speaker tacit knowledge of the articulated theory T_2, rather than merely of the listiform theory T_1? This is the question with which Quine's challenge confronts us. But more important than this evidential question is a constitutive one. What would it be for a speaker to have tacit knowledge of T_2, rather than merely of T_1?

Evans himself gave a constitutive account of tacit knowledge in terms of dispositions ([1981], p. 328):

> I suggest that we construe the claim that someone tacitly knows a theory of meaning as ascribing to that person a set of dispositions—one corresponding to each of the expressions for which the theory provides a distinct axiom.

He added that, for the account to work as intended, the notion of disposition must be understood 'in a full-blooded sense'. Given such an understanding (p. 330):

> the ascription of tacit knowledge of T_2 . . . involves the claim that there is a single state of the subject which figures in a causal explanation of why he reacts in this regular way to all the sentences containing [a given expression].

Thus, according to Evans's account, ascription of tacit knowledge of T_2 involves the attribution to the subject of twenty distinct dispositions, and twenty distinct causal explanatory states—one for each name and one for each predicate of the language.

It is helpful to think of Evans's basic idea in the following way. In theory T_2, but not in theory T_1, the derivations of truth condition specifying theorems for the sentences '*Fa*' and '*Ga*' involve a common factor; namely, the axiom for the name '*a*'. Likewise, the derivations of theorems for the sentences '*Fa*' and '*Fb*' involve a common factor. For tacit knowledge of T_2, and not merely of T_1, we require that where there is, in the theory, a derivational common factor there should be, in the speaker, a causal common factor. Roughly, for a speaker to have tacit knowledge of a particular articulated theory, there must be a causal-explanatory structure in the speaker which mirrors the derivational structure in the theory.

This rough idea requires a number of refinements (Davies [1987]). But for present purposes, it is sufficient to observe two attractive features of any refinement of the basic idea. The first attractive feature is that there can certainly be empirical evidence for or against a particular kind of causal structure in a subject. If attributions of tacit knowledge are basically attributions of structures of causal-explanatory states, then such attributions make perfectly good empirical sense; and they can, in principle, be grounded in empirical evidence. Thus, we meet Quine's challenge.

The second attractive feature is that the basic idea, and refinements of it, do not require that in order to have tacit knowledge of an articulated theory a speaker must conceptualize the axioms or rules of the theory. The basic idea leaves room for a distinction between tacit knowledge and propositional attitudes like belief (see Davies [1989]).

In fact, the account does not require that there be any explicit representations—doxastic or subdoxastic, personal or subpersonal—of the axioms or rules that are tacitly known. Tacit knowledge can be realized by the presence of a processor rather than the presence of a collection of representational states, provided that the processing exhibits the requisite causal structure. The fact

that the account does not require explicit representation in no way trivializes it; for there is all the difference in the world between processing with a structure that mirrors the derivational structure in T_2 and processing whose structure merely mirrors the derivational structure in T_1. For example, a processor with an autonomous component for each sentence of L would meet the latter condition, but not the former.

2 MODULARITY

The notion of modularity can also be introduced via Chomsky's work. In *Knowledge of Language* [1986], he recommends distinguishing between internalized language—that is, I-language—and grammar. I-language is 'some element of the mind of the person who knows the language' (p. 22); a grammar, in contrast, is a theory of I-language. A grammar is not a cognitive structure; it is a linguist's theory. Now, Chomsky describes grammars as modular (p. 71); and his exposition of the current state of linguistic theory is under the heading 'Modules of grammar' (p. 160). In this use of the term, a *module* is a subtheory of a linguist's theory of I-language.

If a particular grammar is a correct theory of the I-language of a speaker, then the language faculty of that speaker can be characterized—at one level of description—by that grammar. If the grammar is modular, then the language faculty that it characterizes can itself be said to be modular; it is an articulated information processing system. Thus Chomsky says ([1986], p. 204):

> The general idea that the language faculty involves a precisely articulated computational system—fairly simple in its basic principles when modules are properly distinguished, but quite intricate in the consequences that are produced—seems reasonably well established.

A module within the language faculty will be a subsystem that is characterized by a module of the grammar.

For a grammar to be a correct theory of I-language is for it to be psychologically real, or tacitly known. And what this requires—according to Section 1—is that there should be causal common factors lying behind pieces of linguistic knowledge in the speaker, exhibiting a pattern that mirrors the way in which there are derivational common factors lying behind the pronouncements of the grammar.

Thus far, we have two different notions of a module. We could label these notions. On the one hand, there are modules in the *analytical* sense: constituent subtheories of a theoretical characterization of a cognitive task. Chomsky's modules of grammar are modules in this sense. On the other hand, there are modules in the *processing* sense: constituent subsystems of a cognitive system. Fodor's modules (Fodor [1983]) are modules in this sense; although Fodorian

modularity involves characteristics not required simply by modularity in the processing sense.

But there are more notions of modularity than just these two. For just as the processing sense of modularity is the result of projecting the analytical notion to the tacit knowledge level of description, so we can introduce notions of modularity at other levels of description too.

Suppose, for example, that we have identified—at the level of theoretical characterization—two components C and D of some cognitive task. Suppose further that, as a matter of empirical fact, the cognitive system under study does perform the task in question by having *inter alia* component subsystems that carry out the subtasks C and D. This would be a highly non-trivial empirical fact. But it would leave open the further empirical question whether the parts of the brain that subserve the performance of task C are distinct from the parts of the brain that subserve the performance of task D.

In fact, there are a number of more precise questions that we can ask when we move to the *neurophysiological* level of description. One which is of some importance is the question whether the geographical region of the brain implicated in task C overlaps, or is disjoint from, the region implicated in task D. The importance of this question is that, the more the respective regions overlap, the less likely it is that the brain in question could, in practice, be damaged in such a way as to disturb the performance of one task while leaving the performance of the other task intact.

We thus have three different notions of modularity, belonging at three different levels of description and explanation. (These correspond very roughly with Marr's three levels: [1982], pp. 24–5.) There is a clear enough distinction between the analytical notion and the processing notion, even though they are closely related: one kind of modularity is a feature of theories, the other is a feature of systems. Both the processing notion and the neurophysiological notion specify an empirical feature of systems, but the distinction between these two notions is crucial nevertheless.

For example, cognitive neuropsychology is the branch of cognitive psychology in which models of normal cognitive processes are evaluated in the light of data provided by observations of people with acquired disorders of cognition. The classical form of argument in cognitive neuropsychology is from an observed double dissociation of deficits to a claim about modularity. The systems X and Y that are responsible for the performance of the tasks A and B are argued to be independent systems or separate modules, on the grounds that performance of each of the tasks can be impaired while performance of the other remains intact.

The cognitive neuropsychologist infers modularity from findings of dissociations. But he does not generally infer absence of modularity from the failure to observe dissociations. Rather, if dissociation between the performance of two

tasks is not found, then the cognitive neuropsychologist considers two possible explanations. One possible explanation is that the cognitive model is incorrect; the two tasks A and B are really performed by a single integrated system. The other possible explanation is that, although there are indeed two independent information processing systems present, psychologically unimportant features of neurophysiology prevent the systems from being damaged separately. (On these issues, see Coltheart [1985].)

The cognitive neuropsychologist is theorizing about modularity at the processing level; but his arguments are complicated by the fact that modularity at that level might not be matched by modularity at the neurophysiological level.

3 CONNECTIONISM

The processing level—as we have so far characterized it—is a level at which the description of a system is an *interpreted* (semantic, cognitive, or content using) description. The interpreted description is cast in the same terms as the theoretical characterization of the task at the analytical level; this is particularly clear if we think of the interpreted description as a tacit knowledge description.

In fact, the simple equation of the processing level with the level of tacit knowledge description is potentially misleading. A description at the tacit knowledge level specifies the information that the system draws upon. But a full description at the processing level should surely do more than that; it should specify, in addition, how the information is drawn upon. To the extent that the processing level is to be identified with Marr's level two—the level of the algorithm—the tacit knowledge level should be distinguished as a slightly higher level of description. (Peacocke [1986] labels it level 1·5.) What we really have is a hierarchy of levels of coarser and more detailed interpreted descriptions of the way in which the task is carried out.

But as well as all these levels of interpreted description, there are also *uninterpreted* descriptions which are still different from descriptions at the physiological level.

For example, those classical computational theorists who favour the symbol manipulation paradigm recognize a level of uninterpreted *syntactic* description. This is not to say that every state which has an interpreted, or semantic, description also has a description as a representational state with a syntax. For a piece of tacit knowledge can be realized by the presence of a computational processor. But, what is insisted upon is that the representational states which constitute the domain of the processor should be syntactically structured states. Thus, Fodor says ([1987], p. 25):

> [The representational theory of mind] says that the contents of a sequence of attitudes that constitutes a mental process must be expressed by explicit

> tokenings of mental representations. But the rules that determine the course of the transformation of these representations . . . need not themselves ever be explicit.

(*Cf.* Fodor [1985], p. 95.)

The friends of parallel distributed processing (PDP) also recognize a level of uninterpreted description which is quite distinct from the physiological level. At this level, the descriptions are in terms of activation at nodes or units, mediated by weights or strengths attached to connections between the units. Let us label this level of formal description of a connectionist system the *network* level.

On the face of it, the availability of this level of uninterpreted description does not count against the validity of interpreted, or semantic, descriptions of a connectionist network.

Indeed, just as the classical theorist recognizes representational states and computational processes as vehicles of semantic content, so too, the connectionist assigns semantic content to two kinds of patterns within networks.

Some of the information in a system is realized by particular patterns of activation of the units (Smolensky [1988], p. 6):

> The entities in the [network] with the semantics of conscious concepts of the task domain are complex patterns of activity over many units. Each unit participates in many such patterns.

And some of the information is realized by patterns of weights attached to connections (p. 13):

> Patterns of activity representing inputs are directly transformed (possibly through multiple layers of units) to patterns of activity representing outputs. The connections that mediate this transformation represent a form of task knowledge . . .

In a connectionist network, then, the bearers of semantic content are complex, structured items.

Some philosophical discussions give the impression that such an interpreted description of a connectionist network is of, at most, heuristic significance; and that the advent of connectionism brings nearer the demise of content using explanations in cognitive psychology. But really, the issue of interpreted or content using description and the issue of connectionism should be regarded as orthogonal. There are four positions that a theorist might occupy.

One quadrant is for those friends of symbol manipulation who insist on the role of content using descriptions (*e.g.* Fodor [1987]). A second quadrant is for friends of symbol manipulation who prescind from content (*e.g.* Stich [1983]). There is a third position that is occupied by enthusiasts for connectionism who would altogether eliminate appeal to semantic content (*e.g.* Churchland [1988]). And there is a fourth box, to be occupied by connectionists who insist that content using descriptions are essential for psychological theory.

The following remark by Smolensky ([1987], p. 101) seems to place him in that fourth box:

> the formal system is at a lower level than the level of semantic interpretation: the level of denotation is higher than the level of manipulation. . . . Both levels are essential: the lower level is essential for defining what the system *is* (in terms of activation passing) and the higher level is essential for understanding what the system *means* (in terms of the problem domain).

But, whether or not any particular theorist clearly occupies the fourth quadrant, if we take Fodor's position as the canonical version of the symbol manipulation paradigm, then the appropriate comparison is with content using connectionism.

So far then, we have no reason to deny that a connectionist system can have a true tacit knowledge description. Nor do we yet have any reason to deny that a connectionist system may exhibit modularity at the processing level, or the tacit knowledge level. For all that this latter requires is that the network should have a true tacit knowledge description cast in the terms of a modular theory (that is, a theory that is modular in the analytical sense). It does not require that the articulation in the formal description of a network should exactly match the articulation of the original modular theory into subtheories. A system that is modular in the processing sense need not also be modular at the network level—the level of description in terms of units and connections—any more than it has to be modular at the physiological level.

Indeed, it is not generally the case that a connectionist network is built from smaller component networks corresponding to the constituent subtheories of a modular theory.

This is not to say that connectionism is committed to the extreme view that there is a single giant network, responsible for all cognitive processes. On the contrary, it is explicit in the work of PDP theorists that specific tasks may be assigned to distinct networks, and that this amounts to an element of modularity (Hinton, McClelland and Rumelhart [1986], p. 79):

> A system that uses distributed representations still requires many different modules for representing completely different kinds of thing at the same time. The distributed representations occur *within* these localized modules. For example, different modules would be devoted to things as different as mental images and sentence structures . . .

There is a potentially misleading mention of '*localized* modules' in this passage: we should not confuse modularity at the network level with the physiological notion of modularity. Rather, the point is that localization or physiological modularity, requires network modularity (though the converse does not hold). To the extent that there is evidence of neural localization of cognitive functions, this is still consistent with the connectionist programme; for that

programme already includes an element of modularity at the network level.

Here it is useful to employ the idea of nested modules, and of coarser and finer grains of modularity. At the analytical level, a theory may be composed of subtheories which themselves have a modular structure. Indeed, we could think in terms of a massive psychological theory, whose subject matter is the whole of cognition, and which is composed of relatively independent subtheories concerning particular cognitive functions. The idea of the language faculty 'with its specific properties, structure, and organization, one "module" of the mind' (Chomsky [1986], pp. 12–13) is a reflection at the processing level of this coarse grained analytical modularity. A component theory, concerning a particular aspect of cognition—such as language, or vision—may itself be modular; and we can pursue this finer grained modularity through the various levels of description of a cognitive system.

The connectionist programme is committed to some coarse grained modularity at the network level; but it is not committed to modularity at the network level matching any finer grained modularity at the analytical level. (*Cf.* the discussion of propositional modularity in Ramsey, Stich and Garon [to appear].)

According to the story that we have told so far, connectionism's characteristic departure from modularity at the network level is compatible with descriptions of PDP systems as embodying tacit knowledge of modular theories. The remaining two sections of the paper will call that compatibility into question.

4 SYNTAX

The formal articulation of a connectionist network need not—and typically does not—reflect the articulation in the interpreted description of the system at the tacit knowledge level or processing level. This is why connectionism departs from fine grained modularity at the network level of description.

If we now focus on patterns of activation as vehicles of semantic content, then we can see another consequence of the mismatch between the tacit knowledge description and the network description. The articulation in the network description of a connectionist system is in general not the syntactic articulation that is characteristic of symbol manipulation.

As we have already noted, the symbol manipulation paradigm does not require explicit representation of computational procedures. So the relevant issue is not whether patterns of weights amount to syntactic encodings of tacitly known rules. But symbol manipulation does require syntactic structure in the representational states that lie in the domain of those procedures. The mismatch between the interpreted description and the uninterpreted description of a connectionist network promises a sharp characterization of the difference between the two programmes. For, in general, the connectionist

analogues of syntactic representations—namely, patterns of activation—are structured, but not syntactically structured.

Someone might object to this characterization of the difference. It might be said that the symbol manipulation paradigm can itself recognize levels of description lying between the level of uninterpreted, syntactic, description and the physiological level. Nothing so far shows that the level of formal description of a connectionist network is anything other than one such intermediate level.

There are a number of correct points here. A pattern of activation in units is a structured item; and there is nothing in the idea of such a pattern, as such, which prevents it from having a syntactic description. What is more, the way in which one pattern of activation leads to another pattern at a later time can be specified without reference to what the patterns of activation mean; it is precisely so specified in the formal description of the network. So, transitions between patterns of activation meet a familiar formality condition upon symbol manipulation. Furthermore, it is possible that a system for symbol manipulation should have a description—at a level lower than the syntactic level—as a network of connected units.

But none of this adds up to an argument for regarding patterns of activation *as such* as syntactically structured.

According to Fodor, syntax must meet three conditions. First, 'The syntax of a symbol is one of its higher-order physical properties' (Fodor [1987], p. 18). Second, syntax is systematically related to semantics. Third, syntax is a determinant of causal role (*ibid.* pp. 16–21).

We can agree that the structure in a complex pattern of activation meets the first and third of these conditions. But, the constituent features of a pattern of activation—namely, specific levels of activation at individual units—need not, and typically do not, make a systematic contribution to the semantic content of the overall pattern; they are not like words in a natural language sentence.

The articulation within a pattern of activation does not constitute a syntactic structure, so long as the interpreted description is afforded by the processing or tacit knowledge level.

It is possible to introduce a different level of interpreted description, lining up more neatly with the uninterpreted, formal, network description. This level is sometimes called the *subconceptual* level; the processing level is then called the *conceptual* level. (See Smolensky [1988], p. 3. The terminology is not ideal since it may suggest that tacit knowledge involves conceptualization; see again Davies [1989].) The main difference between these two levels of interpreted description is this. The concepts used in the conceptual level description are the primitive concepts deployed in the theoretical characterization of the task at the analytical level. In contrast, the description at the subconceptual level is in terms of microfeatures.

A consequence of this semantic *dimension shift* (Smolensky's phrase: [1988],

p. 11) between the conceptual and the subconceptual description is that, while the subconceptual interpreted description is a genuinely accurate semantic description of the operation of the network, the conceptual description is an approximation. This consequence calls for a modification to the idea that a pattern of excitation is straightforwardly a vehicle of semantic content.

Suppose that we consider a family of states whose (conceptual level) interpreted descriptions have something in common. Perhaps, they are all states whose contents concern coffee. Or (recalling the language L in Section 1), they might all be states whose contents concern sentences containing the predicate 'F'.

The original idea about bearers of semantic content would suggest that the states in such a family involve a common subpattern of activation which has an interpreted description as being about coffee, or being about a sentence containing the predicate 'F'. But really this is not so, as Smolensky ([1988], p. 17) makes explicit:

> These constituent subpatterns representing *coffee* in varying contexts are activity vectors that are not identical, but possess a rich structure of commonalities and differences (a family resemblance, one might say).

Similarly, if a connectionist network were to perform the task of assigning a meaning (specified in some format) to each sentence of L, then the constituent subpatterns representing the presence of the predicate 'F' in the varying contexts provided by the sentences 'Fa', 'Fb', and so on, would not be identical.

The argument at the beginning of this section showed that the articulation within a pattern of activation does not constitute syntactic structure, given that the semantic description is cast in the same terms as the theory at the analytical level. Someone might have responded to that argument with the suggestion that we develop a level of syntactic description by taking certain subpatterns of activation to be the primitive syntactic items corresponding to the primitive concepts that are employed in the theory at the analytical level. Because patterns of activation are simply superimposed, this would have been a rather weak suggestion; it would not even preserve the idea of the order of constituents in a syntactically complex expression. But, in any case, we can now see that the suggestion would not work. For there is no single pattern of activation corresponding to each primitive concept; and so there are no candidates for the role of syntactic primitive.

5 TACIT KNOWLEDGE AGAIN

We can now draw an important consequence for the attribution of tacit knowledge to connectionist systems. Recall, once again, the example in Section 1 and the two semantic theories T_1 and T_2.

Suppose that a network that involves a dimension shift between its

conceptual and subconceptual descriptions succeeds in assigning the correct truth conditions to all the sentences of *L*. In particular, it assigns the correct truth conditions to all the sentences containing the predicate '*F*': '*Fa*' is true iff John is bald, '*Fb*' is true iff Harry is bald, and so on.

Suppose too, that this network is not simply made up of a collection of completely autonomous subsystems, one for each of the sentences containing '*F*'. Then we do not have a straightforward instance of tacit knowledge merely of the listiform theory T_1.

Nevertheless, it need not be the case that there is a single pattern of weights on connections which is a causal common factor in all these transitions from representations of sentences to representations of meanings. For a typical connectionist system, we shall be able to say only that the approximate equivalence of the patterns corresponding to the predicate '*F*' in varying contexts results in a considerable overlap in the patterns of connection weights implicated in the several transitions. Consequently, it will not be strictly correct to attribute to the network tacit knowledge of the articulated semantic theory T_2. Such an attribution will be, at best, approximately correct.

This result generalizes. Typically, PDP systems do not strictly embody tacit knowledge of modular theories.

It is not an accident that the absence of accurate tacit knowledge attributions, and the absence of syntactic structure, go in step here. Tacit knowledge does not have to be explicitly represented; it can be realized by the presence of a processor. But tacit knowledge is a matter of strict causal systematicity in the transitions mediated by that processor—causal systematicity mirroring the derivational systematicity in the theory that is tacitly known. And the way to incorporate that causal systematicity is to provide, for the states which are inputs to the processor, a physical articulation or structure which is systematically related both to the interpreted descriptions of those states and to the causal transitions to which the states lead. Given the three conditions upon syntactic descriptions, what this amounts to is providing the input states with a syntax. (For arguments from causal systematicity of process to syntactically structured representations, see Fodor [1987], pp. 135–54, and Fodor and Pylyshyn [1988].)

6 CONCLUSION

What prevents even the most rudimentary syntactic articulation in the states of a connectionist network is the dimension shift between the conceptual and subconceptual level. Given such a shift, the terms deployed in the theory at the analytical level do not figure in any accurate interpreted description of the network.

It is arguable that the apparent empirical inadequacy of some connectionist

models is attributable to this attempt to do without resources which are, in fact, crucial; namely, the categories used in the classical theoretical characterization of the cognitive task in question. (For this issue, see Rumelhart and McClelland [1986] and Pinker and Prince [1988].)

More generally, strict causal systematicity of the kind required for tacit knowledge presents a problem for the connectionist programme. For the absence of a syntactic level of description is characteristic of connectionism. But causal systematicity requires syntactically structured representational states.[1]

Philosophy Department
Birkbeck College
Malet Street
London WC1E 7HX

REFERENCES

CHOMSKY, N. [1965]: *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

CHOMSKY, N. [1976]: *Reflections on Language*. London: Fontana/Collins.

CHOMSKY, N. [1980]: *Rules and Representations*. Oxford: Blackwell.

CHOMSKY N. [1986]: *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

CHOMSKY, N. [1988]: *Language and Problems of Knowledge*. Cambridge, Massachusetts: MIT Press.

CHURCHLAND, P. M. [1988]: 'On the nature of theories: A neurocomputational perspective', in *Minnesota Studies in the Philosophy of Science, Volume 14*. Minneapolis: University of Minnesota Press.

COLTHEART, M. [1985]: 'Cognitive neuropsychology and the study of reading', in M. I. Posner and O. S. M. Marin (*eds.*), *Attention and Performance XI*, pp. 3–27. London: Erlbaum.

DAVIES, M. [1987]: Tacit knowledge and semantic theory: Can a five per cent difference matter?' *Mind*, **96**, pp. 441–62.

DAVIES, M. [1989]: 'Tacit knowledge and subdoxastic states', in A. George (*ed.*), *Reflections on Chomsky*, pp. 131–52. Oxford: Blackwell.

EVANS, G. [1981]: 'Semantic theory and tacit knowledge', in *Collected Papers*, pp. 322–42. Oxford: Oxford University Press (1986).

FODOR, J. [1983]: *The Modularity of Mind*. Cambridge, Massachusetts: MIT Press.

FODOR, J. [1985]: 'Fodor's guide to mental representation: The intelligent Auntie's vademecum', *Mind*, **94**, pp. 77–100.

FODOR, J. [1987]: *Psychosemantics*. Cambridge, Massachusetts: MIT Press.

[1] An earlier version of this paper was written while I was visiting the Research School of Social Sciences, Australian National University in late 1987, and was presented at the conference *Cognition et Connaissance* held in Toulouse, in March 1988. I am grateful to Ned Block for comments on a more recent version.

FODOR, J. AND PYLYSHYN, Z. [1988]: 'Connectionism and cognitive architecture: A critical analysis', *Cognition*, **28**, pp. 3–71.

HINTON, G. E., MCCLELLAND, J. L. AND RUMELHART, D. E. [1986]: 'Distributed representations', in D. E. Rumelhart, J. L. McClelland and the PDP Research Group, *Parallel Distributed Processing, Volume 1*, pp. 77–109. Cambridge, Massachusetts: MIT Press.

MARR, D. [1982]: *Vision*. New York: W. H. Freeman and Co.

PEACOCKE, C. [1986]: 'Explanation in computational psychology: Language, perception and level 1·5', *Mind and Language*, **1**, pp. 101–23.

PINKER, S. AND PRINCE, A. [1988]: 'On language and connectionism: Analysis of a parallel distributed processing model of language acquisition', *Cognition*, **28**, 73–193.

QUINE, W. V. O. [1972]: 'Methodological reflections on current linguistic theory', in D. Davidson and G. Harman (*eds.*), *Semantics of Natural Language*, pp. 442–54. Dordrecht: Reidel.

RAMSEY, W., STICH, S. AND GARON, J. [to appear]: 'Connectionism, eliminativism, and the future of folk psychology'.

RUMELHART, D. E. AND MCCLELLAND, J. L. [1986]: 'On learning the past tenses of English verbs', in J. L. McClelland, D. E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing, Volume 2*, pp. 216–71. Cambridge, Massachusetts: MIT Press.

SMOLENSKY, P. [1987]: 'Connectionist AI, symbolic AI, and the brain', *Artificial Intelligence Review*, **1**, pp. 95–109.

SMOLENSKY, P. [1988]: 'On the proper treatment of connectionism', *Behavioural and Brain Sciences*, **11**, pp. 1–74.

STICH, S. [1983]: *From Folk Psychology to Cognitive Science*. Cambridge, Massachusetts, MIT Press.