

The Presence of a Symbol

ANDY CLARK

The image of the presence of symbols in an inner code pervades recent debates in cognitive science. Classicists worship in the presence. Connectionists revel in the absence. However, the very ideas of code and symbol are ill understood. A major distorting factor in the debates concerns the role of processing in determining the presence or absence of a structured inner code. Drawing on work by David Kirsh and David Chalmers, the present paper attempts to re-define such notions to begin to reflect the inextricability of code and process.

KEYWORDS: Connectionism, code, explicit representation, symbol.

1. A Slippery LOT

The received philosophical understanding of a language of thought (Fodor, 1985, 1987) and of a symbol system (Fodor & Pylyshyn, 1988) embodies a confused reliance on the idea of a symbol's *physical presence*. Arguments in favour of classicism (Fodor & Pylyshyn, 1988) and arguments from connectionism to eliminativism (Ramsey *et al.*, 1991) trade on this confusion. Their proponents are victims of a pathology which we may label 'code-fixation', i.e. they believe they have a clear conception of the conditions under which a semantic item is physically present as a symbol in an inner code. It is a false clarity, encouraged by an uncritical reliance on our intuitions about the information carrying properties of written sentences. Such intuitions are tacitly driven by characteristics of the processor (the human being!) of the sentences. The presence of a symbol is always processor-relative. Once this is understood, the basis for some common intuitions concerning connectionism, classicism and the language of thought (LOT) is removed,

The strategy of the paper is as follows. I begin (Section 2) by reviewing some features of the language of thought hypothesis, in particular, its commitment to *explicit representations* in an inner code. I then discuss (Section 3) a trenchant critique of the code-oriented explicit-implicit distinction, developed by David Kirsh. Once the code/process gestalt switch has been achieved, we try (Section 4) to shed new light on the old chestnut "Do connectionist systems use explicit representations?" and (Section 5) to address some pressing arguments concerning compositionality and systematicity. Section 6 plots an interesting consequence of the process-oriented

view, viz that the question whether a system explicitly represents a given content can only be answered relative to an environment in which it is situated.

2. The Pocket Fodor

There are three ingredients to the proprietary Fodorian language of thought theory-mix. First, propositional attitudes are computational relations to *mental representations*. Second, the mental representations form a *symbol system*. Third, mental processes are *causal* processes involving the explicit tokening of symbols from the symbol system. Expansion is almost certainly in order.

The idea that propositional attitudes are computational relations to mental representations goes back a long way (Fodor, 1975). A currently fashionable way of expressing the claim is to introduce the idea of a *belief box*, *hope box*, *desire box*, etc. The box talk just indicates a kind of role in a complex functional economy. To be in a particular propositional attitude state is then to have a representation of the content of the proposition tokened in a functional role appropriate to that attitude. Thus:

to believe that such and such is to have a mental symbol that means that such and such tokened in your head in a certain way; it's to have such a token 'in your belief box' as I'll sometimes say. (Fodor, 1987, p. 17.)

To hope that P is thus to token, in a suitable functional role, a mental symbol that means that P. The same symbol, tokened in a different functional role, might cause effects appropriate to the fear that P or the longing that P and so on.

So far, then, we have a requirement that there should *be* mental symbols (i.e. items which can be non-semantically individuated but which are consistently the vehicles of a certain kind of content) and that the recurrence of such symbols in different functional roles should explain the content commonalities between various attitudes to a single proposition. As it stands, however, these mental symbols could be unique and unstructured. That is, there might be one symbol for each and every proposition. This has seemed empirically unattractive since we seem capable of an infinite or, at least, very large number of distinct thoughts. Hence the second feature, that such representations form a *symbol system*.

A symbol system is a collection of symbols (non-semantically individuable items which are consistently the vehicles of a particular content) which is provided with a syntax which allows for *semantic compositionality*. In such a system we will find atomic symbols and *molecular* representations. A molecular representation is just a string of symbols such that the content of the string is a direct function of the meanings of its atomic parts and the (syntactic) rules of combination. Thus a very simple symbol system might consist of the atomic symbols 'A' 'B' and 'C', and a rule of concatenation such that the content 'A and B' is tokened as 'AB', the content 'A and B and C' as 'ABC', the content 'C and B' as 'CB', and so on. Such symbol structures are supposed to "correspond to real physical structures in the brain" and their syntactic (combinatorial) properties to correspond to real 'structural relations'. For example, just as the *symbol* 'A' is literally part of the complex molecule 'AB', so the brain state which means that A could be literally part of the brain state which means that AB (see Fodor & Pylyshyn, 1988, p. 13). The advantages of deploying a symbol system include the ease with which we can specify that certain operations can be applied to *any* string of a given syntactic form, e.g. for any string, you may derive any member from the string. Thus AB implies A, ABC implies A, CAB

implies A and so on (see Fodor & Pylyshyn, 1988, p. 13). Another advantage is the ease with which such systems yield a *systematic mental life*. A being, deploying the simple symbol system described above, who can think (i.e. token) AB can *ipso facto* think (i.e. token) BA. This systematicity is echoed, so Fodor & Pylyshyn claim, in a distinctive feature of human mental life, i.e. that, for example, humans who can think that Mary loves John can *also* think that John loves Mary. *This a posteriori* argument for a language of thought is the mainstay of Fodor & Pylyshyn (1988). It is worth noticing that, for the argument to have any force, the symbols which feature in the public language ascriptions of thoughts must have recombinable correlates in the internal code. They need not constitute *atomic* items in such a code, but the code must support recombinable content-bearing structures whose syntactic combinatorics match the semantic combinatorics highlighted by Fodor & Pylyshyn.

Finally, and before the wood vanishes beneath the foliage, we should touch base with the mental causation issue itself. The content-faithful causal powers of our mental states, according to Fodor, are nothing but the causal powers of the physical tokens in the inner symbol system. Thus consider the two characteristic kinds of effect (according to folk psychology) of a mental state of believing that P. One kind of effect consists in the belief's bringing about an action. The other, in its bringing about some further mental state. In both cases, Fodor's motto is "No Intentional Causation without Explicit Representation" (Fodor, 1987, p. 25). The idea is that a particular propositional attitude that P can act as a cause only when there occurs a token of the syntactic kind that means that P and when that token causes either an appropriate action or a further thought content Q (or both). By understanding the way a symbol's syntactic properties (in the context of a particular functional economy and symbol system) determine its causal powers, we can see one way in which content and the physical world can march in step. The Fodorian vision is thus sold as:

a vindication of intuitive belief/desire psychology [insofar as it] shows how intentional states could have causal powers; precisely the aspect of common-sense intentional realism that seemed most perplexing from a metaphysical point of view. (Fodor, 1987, p. 26.)

Fodor rounds off the account with a few subtleties meant to take the sting out of familiar objections. Thus consider the case of emergent rule following. A classic case of emergent rule following is Dennett's example of a chess-playing program which is described as "wanting to get its queen out early" even though:

for all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with 'I should get my queen out early' explicitly tokened...I see no reason to believe that the relation between belief-talk and psychological-process talk will be any more direct. (Dennett, 1985, p. 107.)

Fodor's response to this worry is to introduce an idea of *core cases*. The vindication of common sense psychology by cognitive science requires, he suggests, only that:

tokenings of attitudes must correspond to tokenings of mental representations when they—the attitude tokenings—are episodes in mental processes. (Fodor, 1987, p. 25.)

The core cases, then, are cases in which a given content (e.g. the belief *that it is raining*) is supposed to figure in a mental process or to constitute an 'episode in a mental life'. In such cases (and only in such cases) there must, according to Fodor, be an explicit representation of the content which is at once a syntactic item (hence a bearer of causal powers) and an item which gets a recognizably folk-psychological interpretation (e.g. as the belief about rain). The chess program 'counter-example'

is thus said to be defused since (Fodor insists) “entertaining the thought ‘Better get the queen out early’ never constitutes an episode in the mental life of the machine” (Fodor, 1987, p. 25). By contrast:

The representations of the board—of actual or possible states of play—over which the machine’s computation are defined *must* be explicit, *precisely* because the machine’s computations *are* defined over them. These computations constitute the machine’s ‘mental processes’, so either they are causal sequences of explicit representations or the representational theory of chess-playing is simply false of the machine. (Fodor, 1987, p. 25.)

The claim then is that the *contents* of our thoughts must be tokened in an explicit inner code. However, the ‘laws of thought’—the rules which determine how one content yields another or yields an action—need not be explicitly represented. In the familiar form of words, “programs...may be explicitly represented but ‘data structures’...*have to be*”. (Fodor, 1987, p. 25, original emphasis.)

This may seem a little confusing since in his 1975 work, Fodor wrote that:

What distinguishes what organisms do...is that a representation of the rules they follow constitutes one of the causal determinants of their behaviour. (Fodor, 1975, p. 74, n. 15.)

Despite appearances, this is consistent with the current claim. The idea must be that in any case where the *consideration of a rule* is meant causally to explain a judgement or action, *then* the rule must be explicitly tokened. Otherwise not.¹ Thus a novice chess player whose action involved a train of thought in which the rule figured would have had (according to Fodor) to token the rule in an inner code (distinct from her public language, see Fodor, 1975).

The trouble with all this, I shall now suggest, is that it depends on a fundamentally unclear notion of *explicit representation*. Once we see this the whole project of vindicating common-sense psychology via a syntax/semantics parallel in some simple internal code is called into question.

3. On Being More Explicit

The idea of an explicit representation has been seen to bear considerable weight. The language of thought is, precisely, the syntactic vehicle of explicit representation and it is explicitness which is (for Fodor) essential to the vindication of the folk’s use of propositional attitude talk. Propositional attitudes pick out causally efficacious items in mental processes just in case their contents are explicitly tokened. However, David Kirsh (1991) offers some persuasive reasons for caution. He argues that a good account of explicit representation should *not* focus on the form of an inner code but on the combination of information bearing states *and processors*. The quality of explicitness, on that model, is always relative to the *usability* of information rather than (directly) to its form. What misleads us, according to Kirsh, is the “bewitching image of a word printed on a page” (Kirsh, 1991, p. 350).

Kirsh argues that our intuitions about explicitness are inconsistent, insofar as we are inclined to postulate two sets of criteria which may come into conflict.

The first set of criteria is *structural*. In this sense, a representation is seen as explicit if it is ‘on the surface’ of a data structure. Thus the word *cat* is explicit in the list {*cat*, *dog*, *fly*}. However, what exactly *is* the intuition here? Is ‘*cat*’ equally explicit if the word is “hidden in a tangle of other words”. Furthermore, we think that if a word is encrypted in a hard to decipher code, it is not explicitly represented. However, what is the difference between the encryption case and the tangle-of-

words case? It very quickly begins to look as if the structural notion of explicitness is trading on a *processing* notion according to the ease with which it is recovered and put to use. Being the kind of processor we are (as human readers of text) we find it easier to extract the cat information from a typed list than from a jumbled up tangle of words. However, if we were a different kind of processing 'tool', we might have no difficulty with the 'tangle'—hence the cat information ought (relative to such a tool) to count as explicit. We thus move towards a second set of criteria according to which information is explicit if it is ready for immediate use by an embedding system.

If we now consider some standard ideas about explicit representation, we shall see that they rely too heavily on the *structural* notion of explicitness—a notion which is illegitimately building in the idea of visibility to inspection by a human agent as the mark of immediate usability. Thus we find ourselves reflecting on the idea of a word on a page as a paradigm of explicitness. But that case, if we are not aware of the processing dimension, will mislead. Thus consider the following properties, drawn from Kirsh (1991, pp. 350–358).

(i) Words are *localized* in space. This is a great aid to a human visual system. However, all that is required if meaning is to be easily extracted is that the overall system have the power to spot the relevant information without extensive computational effort. Thus spatially superposed information may be visible, with minimal computational effort, given the right 'reading tool'. Examples of such tools include parallel distributed processing (PDP) systems, telephone message de-coders (where several conversations run on a single line), and colour filters (which separate out spatially superposed wavelengths of light—see Kirsh, 1991, p. 350).

(ii) Words are *movable*. 'Cat' means the same in the sentence "The cat sat on the mat" and in the sentence "The cat ate the budgerigar". Why should this matter? It matters because the extent to which meaning depends on context determines (in part) the amount of effort involved in recovery of meaning. Thus the numeral 5 carries meaning in a context-dependent way—the 5 in 501 means 500 whereas the 5 in 51 means 50. To extract the significance of the 5 we need to survey the context; that takes effort (and hence is a move away from total explicitness, i.e. easy visibility). However, once we foreground the *processing* measure, we can see that the *extent* to which context-dependence defeats explicitness is relative to the ease with which context is taken into account by the processor. And some kinds of processor (e.g. connectionist ones) are very well adapted to the fluent processing of contextual data. Thus the requirement of total movability as a constraint on explicitness is revealed as an artifact of lack of attention to the processing roots of the requirement and the richness of the space of possible processors.

Kirsh goes on to develop a precise computational account of ease of processing which allows us to say that information is explicit if it can be retrieved and made available for use in constant time.² The moral is that it is really the processing requirements which drive our intuitions and that the structural vision is a vision of these requirements which is distorted by the image of words on a page. On the Kirsh model:

Explicitness really concerns how quickly information can be accessed, retrieved, or in some other manner put to use. (Kirsh, 1991, p. 361.)

The view has its price. As Kirsh notes, we shall need to accept that relative to a processor like me, the structure 'add (1.1)' explicitly encodes the information '2'. It

also implies that one and the same static structure may explicitly carry different information relative to different processing environments.

What is perhaps even more problematic is Kirsh's account of *implicit* information. Kirsh is tempted to regard as implicit just that information which "is not explicit in a system but which could be made so" (1991, p. 347). The system, if it is to implicitly encode the information that P, must be able to recover that information and explicitly encode it. However, what exactly can this amount to once we adopt a processing perspective on explicitness? It seems to imply that whatever is truly implicit (i.e. visible only with a high degree of computational effort) must be translatable into something which is explicit (i.e. usable with a low degree of effort). This is either trivial (since to use it *at all*, we must somehow get it into directly usable shape, albeit after some effort) or overly restrictive. Why can we not allow that a system which can only *ever* access certain types of information by a complex deciphering procedure nonetheless *implicitly* encodes that information? It seems more natural (to me at least) to take a processing perspective as arguing for a continuum of cases such that information is implicit relative to the amount of effort needed to bring it to bear.

Finally, I believe Kirsh's account leaves out a fundamental extra dimension: (It is) that truly explicit items of information should be usable in a wide *variety* of ways, i.e. not restricted to use in a single task. The implicit–explicit continuum is, I suggest, better viewed as a two-space whose dimensions are, first, ease of usability of information and, second, variety of *modes of use*. Information which is easily deployed but only in a rigidly circumscribed way is not, *pace* the basic account offered by Kirsh, fully explicit. Our view thus adds to Kirsh's basic account the idea (found in, e.g. Dretske, 1988) that representations which are usable only in a specific context should be regarded (*ceteris paribus*) as more implicit than those whose content is available for many purposes.

To give a simple example which Dretske himself uses (1988, pp. 33–34) consider the case of a rat which has learnt to discriminate safe and poisoned food. Does it rely on an explicit representation of the content "that food is poisoned" to do so? It is certainly relying on an inner state which provides for the easy use of the 'poisoned' information to guide avoidance behaviour. However, the fact that the rat can *only* use the information in that way works *against* the intuition that it is explicitly represented. Likewise, developmental psychologists such as Karmiloff-Smith (1986) depict human cognition as involving a progression from context-bound use of stored information to much more context-flexible uses, and wish to characterize this as a progression from an implicit to an explicit representational form. It does not seem unreasonable, then, to extend Kirsh's basic proposal (i.e. that ease of use in the context of a given processor is the key to explicitness) so that information is increasingly explicitly tokened according to *both* (i) ease of use and (ii) variety of modes of use (relative to an overall system). This maintains the stress on the use of information, but expands our conception of the dimensions of the implicit–explicit continuum.

4. Connectionism and Explicit Representation

Having (hopefully) begun to switch the structure/process gestalt, it is time to return to an old chestnut: do connectionist systems support explicit representations? Now a funny thing happens. Relative to a more *process*-oriented notion of explicitness,

connectionist systems begin to look *very* capable of supporting explicit representations! There is, however, a catch.

The encouraging news is that, within a standard distributed Smolensky-style network, a good deal of stored knowledge is extremely easy to access and use. Thus consider the network which encodes items of information such as “dogs have fur”. It is simplicity itself to access and use this information to answer “yes” to the input question “Do dogs have fur?” More generally, connectionist systems are capable of the swift retrieval of any one of the multiple patterns stored in superpositional distributed style. According to the Kirsh criterion, then, that information must be counted as explicitly represented in the array of weights (in the context of a connectionist input and retrieval system). Indeed, the very distinction between the information store and the retrieval tool is blurred in these cases. The retrieval tool is built into the knowledge representation itself. Thus we read that:

[in connectionist systems] the representation of the knowledge is set up in such a way that the knowledge necessarily influences the course of processing. Using knowledge in processing is no longer a matter of finding the relevant information in memory and bringing it to bear; it is part and parcel of the processing itself. (McClelland *et al.* 1986, p. 32.)

However, at this point our intuitions become tangled once again. For it may seem as if a key feature of (old fashioned, structural) explicitness has now been missed. It is a feature which I shall (somewhat opaquely) label ‘reflectivity’. The idea is simple. It is that one task of the idea of explicit representation was to distinguish reflection and considered action from “mere animal reflexes”. As Fodor once said, what distinguishes me from a paramycium is that when I act there is an intervening stage of representation. In the paramycium, there is (let us assume) no such buffer. It is now obvious how the various intuitions might tangle. For the criterion of process-explicitness is ease of use. The limiting case here is a fast direct input-output link. However, such a link begins to look a lot like a *reflex* response. If you have the intuition that explicit representation should mark the difference between reflex responses and the rest, you have tied yourself into a knot!

The key to successful unravelling is, I believe, the notion of a second dimension of assessment of ease of use, which we briefly mentioned in Section 3. For the relative explicitness of information should not be assessed without reference to the *variety* of uses to which the information can be put. The idea of a genuine difference between reflex and considered action is, I suggest, a distorted vision of two genuine considerations. First, some creatures are *consciously aware* of their reasons. Let us bracket this issue. Second, some creatures can use stored information in more flexible ways than others. A neural network which is *only* capable of using the stored information “dogs have fur” to answer “yes” to the input “do dogs have fur?” is not breathtakingly flexible. A human being who knows that dogs have fur can use the information to plan ways of making fur coats, to irritate allergenic neighbours, to predict musty smells in the rain and all the rest. On our account a full-blooded process-oriented account of explicit representation will demand both ease of retrieval *and* flexibility of use. That, of course, is the catch for standard connectionist proposals. For despite the advertising, the information thus stored is often very limited in its range of usability. Particular limitations (see Clark & Karmiloff-Smith, forthcoming) involve the systematic adaptation of the stored knowledge to new tasks, the systematic de-bugging of the stored information and the integration of new information. In short, the reflex/non-reflex distinction may often really be a distinction between flexible and less flexible uses of information. A more process-

oriented account of explicitness, *as long as* it includes the dimension of *variety* of use, will be well able to accommodate it. The fate of connectionism as a locus of such fully explicit representations must therefore remain undecided until the flexibility issue (more on which below) is resolved.

5. Code-fixation: Its Symptoms and Cure

The cash value of a process-oriented conception of representation is its power to treat a common pathology which we may label 'code-fixation'. Sufferers from code-fixation expect explanations of cognitive phenomena to fall naturally out of considerations about the form of an inspectable internal code. A particularly striking example is Fodor & Pylyshyn's (1988) argument against connectionist models of mind.

The argument, briefly rehearsed in Section 2 above, is that humans have a systematic mental life. Those who can think that Margaret and John love Mary can think that Mary loves John, and that Mary and John love Margaret, and that Margaret loves John and so on. This fact, Fodor & Pylyshyn argue, lends strong support to the idea of a language of thought in which internal tokens carry the meanings 'John', 'loves', etc., and are recombinable according to syntactic rules. More generally, Fodor & Pylyshyn insist that

In classical models, the principles by which mental states are transformed or by which an input selects the corresponding output, are defined over structural properties of mental representations. Because classical mental *representations* have combinatorial structure, it is possible for classical mental *operations* to apply to them by reference to their form. (Fodor & Pylyshyn, 1988, pp. 12–13.)

However, what does it really *mean* for a representation to have such combinatorial structure? Fodor & Pylyshyn are certain that distributed connectionist cognitive modes cannot provide it. Yet a variety of recent proposals show that it is possible to *process* connectionist distributed representations in ways which yield the very abilities (to perform structure-sensitive operations) which Fodor & Pylyshyn foreground.

Thus Smolensky (1991), Pollack (1990) and Elman (1992) all offer treatments in which distributed representations sustain various kinds of compositional, iterative and recursive operations. Elman, for example, used a simple recurrent network to learn some grammatical structures, including multiple levels and types of embedding.

To get more of the flavour of such proposals, consider Chalmers (1990) model of active to passive transformations. the model uses a RAAM (Recursive Auto Associative Memory) architecture due to Pollack (1988). This consists of a three-layer feed-forward network with a small number of hidden units, and a larger (and equal) number of input and output units (e.g. 39 input, 13 hidden and 39 output). The net is taught to develop compressed distributed representations of linguistic tree structures. Thus it may be fed inputs coding for the contents of three terminal nodes on a tree by dividing the input units into groups and using one group per terminal node. The network is required to *reproduce* the input tree at the output layer. To do so, it uses the back-propagation learning rule to learn a compressed distributed representation of the tree structure at the hidden unit layer (13 units). These hidden unit patterns are also fed to the network as inputs, thus forcing it to "auto-associate on higher order structures" (Chalmers, 1990, p. 55). The upshot is a network which can *decode* compressed representations of trees of arbitrary depth. To perform the decoding you give the compressed representation direct to the hidden unit layer and

read an expanded version at the output layer. If the expanded version contains only *terminal* tree structures, the decoding is complete. If it does not, any non-terminal structure must again be fed in to the hidden unit layer until they are discharged.

Chalmers trained a RAAM architecture to encode tree structures representing sentences of 'active' form (e.g. John love Michael) and 'passive' form (e.g. Michael is love by John). Forty sentences of each type were used. As expected, the network learned to decode the compressed representations of the sentences which it formed at the hidden unit layer. Chalmers then went on to train a further network to take as input the compressed representation of its passive correlate. The point of this exercise was to show that a standard network could transform the RAAM representations from active to passive form *without* first decomposing the RAAM representation into its constituent structures. The experiment was a success. The new network learned the transformation of the training cases *and* was then able to perform quite well even on *new* sentences. Thus new active sentences, once compressed by the RAAM network, were transformed into appropriate passives with *at least* 65% accuracy (Chalmers, 1990, p. 59).

It is worth quoting Chalmer's own discussion of his results at some length. He claims that the experiment shows that:

Not only is compositional structure *encoded* implicitly in a pattern of activation, but this implicit structure can be utilised by the familiar connectionist devices of feedforward/back-propagation in a meaningful way. Such a conclusion is by no means obvious *a priori*—it might well have turned out that the structure was 'buried too deeply' to be directly used, and that all useful processing would have had to proceed first through the step of extraction. (Chalmers, 1990, p. 60.)

He goes on to suggest that the model constitutes a disproof of the idea (see Fodor & McLaughlin, 1991) that the only way to support genuinely structure sensitive operations is by deploying representations which concatenate explicit tokens of the constituent parts of the structure. To quote once more:

[According to Fodor & McLaughlin] If a representation of 'John loves Michael' is not a concatenation of tokens of 'John' 'loves' and 'Michael'...then later processing cannot be sensitive to the compositional structure that is represented. The results presented here show that this conclusion is false. In the distributed representations formed by RAAM there is no such explicit tokening of the original words. ...Nevertheless the representations still support systematic processing. *Explicit* constituent structure is not needed for systematically; implicit structure is enough. (Chalmers, 1990, p. 61.)

According to our analysis of explicit representation, this characterization of the situation is nonetheless slightly misleading. For Chalmers is still relying on what we (following Kirsh) have called the structural notion of explicitness, in which a constituent part is explicitly represented if it is easily visible, to a human theorist, in an informational structure. However, rather than rest content with this ill-motivated notion, we might do better to expand our idea of explicit representation of structure. If we do this along the lines sketched in Section 3, then the ease of use of the structured information (the constant time active-passive transformations) gives us some cause to regard the RAAM representations (in the context of the transformation net) as *explicit representations* of constituent structure. The fact that such structure is not directly visible to the human theorist is neither here nor there.

It would be tempting to stop here and conclude that the sense in which a suitably advanced connectionist system must fail to account for structure sensitive processing and systematicity is in failing to do so *transparently*, by means of an inner code in which the semantic constituents of complex representations are *easily visible*. The kind of *concatenative* (see van Gelder, 1990) compositionality which Fodor &

Pylyshyn seem to seek is, indeed, one which is often nicely transparent to us. However, it is not the only kind available. The *functional* compositionality exhibited by the new wave of connectionist models is proof that at least some degree of compositional structure can be preserved in distributed representations *and readily exploited*.

Such a conclusion, would, however, be premature. For the second functional dimension of explicitness (i.e. variety of use) has still to be addressed.

Thus suppose we ask what it is about, for example, the representation of information as a string of LISP atoms, which (relative to the processing tool provided by a standard CPU) inclines us to view such a string as an explicit and highly structured representation? According to the Kirsh-style account developed above, the answer lies in the fact that information thus encoded is (relative to the usual embedding processing environment) nicely available for use. We saw, however, that the same could be said of the structural information contained in the RAAM encoded sentence representation. Relative to the embedding environment of the transformation network, the structural information is nicely tee-d up and ready for use. However, now notice a revealing (or so I believe) difference. Call the LISP string environment provided by the CPU a CPU environment and call the RAAM representation environment provided by the transformation network a T environment. The difference is then this; that relative to the T environment the target information is easily usable but *only* in one specific way, i.e. to mediate an active-passive transformation. Whereas relative to the CPU environment, the target information is easily usable for an open-ended variety of purposes. Thus given the LISP representation it is a simple matter to write additional programs which recursively refer to the elements of the target data structure, which treat particular elements as variables, or which re-organize the elements in new ways for other purposes. This is in stark distinction to the one-track usability of the RAAM representation relative to the T environment.

It might reasonably be objected that we are not comparing like with like, i.e. that the proper comparison is, for example, between a LISP representation relative to the environment provided by a *specific program* (e.g. to take active into passive voice) and the RAAM-representation relative to the T environment. However, that is to miss the essential point. We may *agree* that relative to a specific program environment the LISP string is no more explicit than the RAAM representation, but still the usual CPU environment provides a tool capable of cheaply exploiting the LISP string information in a very flexible set of ways. Perhaps it will one day be possible to create an environment in which the structural information encoded in the RAAM representation is just as easily and variously exploitable. As things stand, however, this does not seem to be the case. To exploit the RAAM representation for a different purpose (e.g. to take present into past tense) would at the moment require the extensive training of a wholly separate transformation network. In short, the structural information in the RAAM encoding is indeed easily usable (relative to the T environment), but in a very *domain-specific* manner. Until such domain specificity is overcome, connectionist representations will fail to live up to even our functionally-oriented criterion of explicitness.

In sum, the move towards a functional or process-based vision of explicitness and structure, combined with recent demonstrations like Chalmers (1990), goes some way towards de-fusing the force of Fodor & Pylyshyn's worries. However, until the deeper issue of the multi-track useability of the information encoded by a network is fully resolved, we must bring back an open verdict.

6. All the World's a Processor

Sections 2–5 argued for a processing device relative account of explicitness. A representation is explicit if it is both cheaply and multiply deployable by the system in which it is embedded.

Once we have embraced such a processing-device relative view of explicitness, however, it becomes necessary to ask *what counts* as a processing device. Kirsh raises, but does not pursue, the claim that:

information can be implicit in a system because that system is embedded in a particular environment. (Kirsh, 1991, p. 12.)

The case which he has in mind, it seems, is one in which, in a certain sense:

A system well adapted to its environment contains information about that environment and about its momentary relations to that environment even though the information is built into the design of the system and so is in principle inaccessible. (Kirsh, 1991, p. 12.)

Thus (I suppose) someone might say that in a certain sense a fish's shape embodies information concerning the hydrodynamics of sea water or that the visual system, since its processing uses heuristics which rely on certain properties of the distal environment, implicitly carries information about these properties.

Consider, however, a somewhat different range of cases; cases in which a system can (unlike the fish/visual system cases) in fact *access* certain information (generate an internal representation of it), but *only* in virtue of some *wider* processing environment than that constituted by its on-board processing and storage apparatus. Thus I may be able to exploit the distinct informational elements represented in some inner code *only* if, for example, I am augmented by some external memory (like paper and pencil). I may also be able to *computationally cheaply* retrieve and deploy some specific item of information only in a particular external setting (e.g. one in which it is cued by a written reminder). It seems to me that in those cases we have to allow that relative to the broader processing tool of me plus my environment, information which would *otherwise* count as unstructured and/or inexplicit should count as structured and/or explicit! It is unclear why the skin should constitute the boundary of the processing environment relative to which such questions are to be decided.

To see this, consider the case where my brain is augmented with an add-on mechanical processing device which increases my short-term memory span. There seems little doubt that in such a case the processing tool, relative to which the internal representational states are to be judged as structured, explicit, etc., has been *altered*. However, why is this different to taking the original processor (of brain and body) and setting it in front of an external environmental device (the pad and pencil) which likewise allows the augmentation of my short-term memory? I conclude that to take *seriously* our picture of structure and explicitness as processing environment relative properties of inner states is *necessarily* to allow that both the nature and ultimately the content (a structured content is different from an unstructured one, after all) of our inner states is always a joint function of their intrinsic nature and the broader environment in which they exist. In short, there is just *no answer* to the questions "What is the content of that state?", "Is it explicit?" and so on, independent of considerations involving the processing capacities of the local system as currently embedded in some wider environment.

7. Conclusions: From Code to Process

A familiar image depicts mental processes as the logico-manipulative transformation of static symbols in a concatenative and recombinative inner code. Commensurate with such an image is a model of intentional causation as involving the explicit tokening of content-bearing symbol strings as an intervening layer between input and action. Connectionist approaches do not lend themselves easily to interpretation in these terms. As a result we find ourselves pushed towards a more liberal (process-oriented) understanding of the key notion of explicit representation. Such a revised understanding should, I have argued, build in two dimensions of functional assessment. The first (ably canvassed by Kirsh) highlights ease of useability of information. The second (neglected by Kirsh, but vital to our intuitions concerning the difference between considered action and reflex) highlights the multi-track useability of stored information. The second dimension reveals a problem with current connectionist representations of structured information. Despite some demonstrations of single-track fluency in the exploitation of such information, the issue of its wider deployability remains unresolved. The explicit representation of structure thus remains a key research target, even once the text-based image of the inner code is abandoned.

Notes

1. Indeed, as Fodor concedes (1987, p. 23), not *all* the rules *can* be merely explicitly represented or else the machine could not actually *do* anything!
2. Constant time is a complexity measure according to which the number of steps needed to solve a kind of problem is a constant; for example, if all instances of that problem type can be solved in three steps. This is a very strong requirement which he is forced to water down a little. The details of this proposal are, however, not important for our argument.

References

- Chalmers, J. (1990) Syntactic transformations on distributed representations. *Connection Science*, 2, 53–62.
- Clark, A. & Karmiloff-Smith, A. The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind and Language*, in press.
- Dennett, D. (1985) A cure for the common code? In D. Dennett (ed.), *Brainstorms: Philosophical Essays on Mind and Psychology*. Sussex: Harvester Press, pp. 90–108.
- Dretske, F. (1988) *Explaining Behaviour: Reasons in a World of Causes*. Cambridge, MA: MIT Press/Bradford Books.
- Elman, J. (1992) Structured representations and connectionist models. In G. Altmann (ed.), *Computational and Psycholinguistic Approaches to Speech Processing*. New York: Academic Press.
- Fodor, J. (1975) *The Language of Thought*. New York: Crowell.
- Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press/Bradford.
- Fodor, J. & McLaughlin, B. (1991) Connectionism and the problem of systematicity: why Smolensky's solution doesn't work. In T. Horgan and J. Tienson (Eds), *Connectionism and the Philosophy of Mind*. Cambridge, MA: MIT Press/Bradford Books.
- Fodor, J. & Pylyshyn, Z. (1988) Connectionism and cognitive architecture. *Cognition*, 28, 3–71.
- Karmiloff-Smith, A. (1986) From meta-processes to conscious access: evidence from children's meta-linguistic and repair data. *Cognition*, 23, 95–147.
- Kirsh, D. (1991) When is information explicitly represented? In P. Hanson (Ed.), *Information, Language and Cognition*. Vancouver, BC: UBC Press, pp. 340–365.
- McClelland, J., Rumelhart, D. & Hinton, G. (1986) The appeal of parallel distributed processing. In J. McClelland, D. Rumelhart & PDP Research Group (Eds), *Parallel Distributed Processing*, Vol. 1, Cambridge, MA: MIT Press/Bradford Books, pp. 3–44.

- Pollack, J. (1988) Recursive auto-associative memory: devising compositional distributed representations. In *Proceedings of the 10th Annual Conference of the Cognitive Science Society*, Montreal, Canada, pp. 48–54.
- Pollack, J. (1990) Recursive distributed representations. *Artificial Intelligence*, 46, 77–105.
- Ramsey, W., Stich, S. & Garon, J. (1991) Connectionism, eliminativism and the future of folk psychology. In W. Ramsey, S. Stich & D. Rumelhart (Eds), *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence Erlbaum, pp. 199–228.
- Smolensky, P. (1991) Connectionism, constituency and the language of thought. In B. Loewer & G. Rey (Eds), *Meaning in Mind: Fodor and his Critics*. Oxford: Blackwell.

[IU home](#)
[IUCAT/Databases](#)
[IU home](#)
[Ask a librarian](#)
[Site Search](#)
[Indiana University Bloomington Libraries](#)

Media Services

Success!

Your request has been received and will be processed as soon as possible. Scroll down to see the information you submitted.

If you would like to place another request you may either:

- Hit your BACK button and change only the information that needs to be changed, then hit the submit button again (Don't worry this won't effect your original request).
- Select one of the following:
 - Request another photocopy from a journal to be placed on reserve.
 - Request a photocopy from a book to be placed on reserve.
 - Request a book to be placed on reserve

Patron Information

Name: Andy Clark

Dept: COGS

Course #: Q540

Request Information

Journal: Brit J Phil Science

Call #: n/a

Volume: 41

Issue #:

Date: 1990

Article Author: Andy Clark

Article title: Connectionism, Competence, and Explanation

Pages: 28

No. of Copies: 1

Meet copyright? Yes

Need by: Fall '03

Email librsvp@indiana.edu if you have any questions or problems.

Media & Reserve Services
Copyright 1995-2000, The Trustees of Indiana University

Form processed by Transform version 3.0