Compositionality and the Explanation of Cognitive Processes

Tim van Gelder
Department of Philosophy
Indiana University

Abstract: Connectionist approaches to the modeling of cognitive processes have often been attacked on the grounds that they do not employ compositionally structured representations (e.g., Fodor & Pylyshyn 1988). But what exactly is compositional structure, and how does such structure contribute to cognitive processing? This paper clarifies these questions by distinguishing two different styles of compositionality, one characteristic of connectionist modeling and the other essential to mainstream symbolic or "Classical" approaches. Given this distinction, it is clear that connectionist models can employ compositionally structured representations while remaining, both conceptually and in practice, quite distinct from the Classical approach; moreover, it can be shown that certain central aspects of cognition, such as its *systematicity*, are at least in principle amenable to Connectionist explanation.

1. STYLES OF COMPOSITIONALITY

One point of general agreement among cognitive scientists of diverse theoretical persuasions is that sophisticated cognitive processing requires the internal representing of complex structured items or situations. To give an obvious example, engaging in conversation requires, at some level, the ability to represent the sentences used. It is generally agreed, moreover, that it is not sufficient merely to represent such items as a whole; it is essential that the internal *structure* of the items be represented and hence accessible to the system. Thus there is generally little point in representing a sentence with a single letter, for this in itself conveys no information about the syntactic structure of the sentence, and so is of almost no help in determining how the sentence should be processed.

One approach, perhaps the most obvious, to the problem of representing structured items is to use representations that themselves exhibit a compositional structure. In the most general sense, any representation is appropriately said to have a compositional structure when it is built up, in a systematic way, out of regular parts drawn from a certain determinate set; those parts are then the primitive *constituents* of the representation. Constituents are set in systematic correspondence with parts or features of the item to be represented, and various relationships among those parts or features are indicated by the structural relationships among the representation's constituents.

Though this description may have seemed transparent enough, there are in fact a number of fundamentally different ways in which a representation can be "built up" out of parts, and a number of corresponding notions of "part" or "constituent". In particular, we can distinguish the compositional formal structure of the representation itself, which is a fact about its concrete physical design, from a wider sense of compositionality that we get if we consider only its constituency relations (i.e., the constituents the representation happened to be constructed out of and into which it could in turn be broken down), disregarding the particular internal formal configuration of the representation token itself.

To make this more precise: a compositional representation is one that belongs to a compositional scheme, where a compositional scheme is one that satisfies the following conditions:

(a) There is a (typically finite) set of primitive types $\{P_1, ..., P_n\}$; for each type P_i, there is an unbounded number of tokens of each type, p_i.

(b) There is a (typically unbounded) set of expression types R_i; for each type R_i, there is an unbounded number of tokens of those types, r_i.

(c) There is a set of abstract transitive and non-reflexive *constituency* relations over these primitive and expression types. $C(R_i, R_j)$ means, for example, that expressions of type R_j have as constituents expressions of type R_i.

Since in most interesting cases of compositional schemes there is an unbounded number of expression types, *specifying* such a scheme requires recursive rules determining the allowable expression types in terms of their constituency relations; a set of rules of this kind is a *grammar* for the scheme.

Note that these conditions are framed primarily in terms of primitive and expression *types*, and do not yet say anything at all about how *tokens* are actually to be instantiated. In other words, they place no constraints on the formal *sign design* of representations in the scheme. This is a matter needing further specification. Given that there is an unbounded number of expression types to deal with, how is it possible to specify what tokens of each type should look like? Clearly this task must also be carried out recursively. The way in which it is standardly done is by providing (1) actual *samples* of each primitive class, which implicitly (or "ostensively") provide criteria for any physical item's counting as an instance of that primitive class; and (2) a concrete *mode of combination*, which, operating in conformity with the abstract grammatical rules, is used to construct expression tokens out of sets of primitive tokens. Consequently, knowing what kinds of things count as tokens of the primitives, and knowing the systematic effects of the mode of combination, we can determine the characteristic physical makeup of any arbitrary expression.

It is now possible to make an important distinction among kinds of compositional scheme, according to the manner in which they construct expression tokens. Most compositional schemes we are familiar with also satisfy the following two further conditions:

(d) Primitive tokens are *symbols*, i.e., instantiate a distinct physical pattern, such that primitive token classes are disjoint and digitally separable;

(e) The mode of combination is concatenative.

A concatenative mode of combination is, intuitively speaking, one that preserves primitive symbol tokens in the expression itself. More precisely, suppose the set of constituents of an expression type R_j is the set of primitives or expressions $\{\alpha_i: C(\alpha_i, R_j)\}$. Then a necessary condition for a mode of combination to be concatenative is that any token r_j generated using that mode must literally contain a token of every α_i, in the sense that some part or feature of r_j satisfies the identity criteria for each constituent α_i. Since the set $\{\alpha_i\}$ includes the primitive symbolic constituents, it must be the case that some part or feature of any token r_j must satisfy the criteria for counting as an instance of each of the symbols of which r_i is constructed.

Representations in such a scheme therefore have a characteristic formal structure; they are appropriately described as *symbolic*, since they are built up out of primitive symbols in a very direct sense. An excellent example of a symbolic scheme in this sense is the space of expressions of standard propositional logic. Primitive symbols are the letters "P", "&", "(" etc., and expressions take the form "(P&Q)", "((P&Q)&R)" and so on. Note that expression tokens contain within their boundaries instances of their constituents, including in particular their primitive symbolic constituents, and that this is just a blunt fact about their physical configuration. Just about every compositional scheme we are familiar with is symbolic in this sense; this includes natural languages (by and large), the various formal languages of logic, mathematics and computer science, and knowledge representation formalisms in artificial intelligence.

Symbolic schemes should however be contrasted with schemes that are merely *functionally* compositional. Such schemes relax condition (e); they do not require the use of any concatenative mode of combination, and so are not constrained to preserving tokens of symbolic constituents in expression tokens.[1] For such schemes to count as genuinely compositional however it is crucial that they do at least satisfy the following condition:

(e') there are general, effective and reliable processes for (i) generating expression tokens from their constituents and (ii) decomposing those expressions into their constituents again.

Designing and implementing such processes, without relying on simple concatenation of symbol tokens, is to say the least - a challenging engineering problem; this is one reason why concatenative languages are so pervasive.

A graphic (albeit for a number of reasons quite impractical) example of a formal scheme that is compositional but non-concatenative, and hence non-symbolic in this strong sense, is "Gödelese" the numerals corresponding to the Gödel numbers of the expressions of propositional logic. Imagine that instead of writing down expressions in their normal notation we chose to write down their corresponding Gödel numerals instead. Since under a Gödel numbering scheme every expression is assigned a unique natural number, this new scheme is expressively equivalent to propositional logic; moreover, it is functionally compositional, since there are simple recipes for generating the Gödel numerals of expressions from those of their constituents and vice versa, on the basis of which one can design and implement the relevant composition and decomposition processes. Yet it is a simple fact about the concrete shape of the numerical tokens themselves that, in general, a "complex" Gödelese numeral does not contain within its physical boundaries tokens of its Gödelese constituents. Suppose, for example, that $gn(\sim) = 3$ and $gn(P) = 5$, then $gn(\sim P) = 2^3.3^5 = 1944$. In Gödelese, then, the expression *1944* has, as constituents, *3* and *5*. Yet it is just a blunt fact about the shapes of the ink marks on the page that no part or feature of the token *1944*

[1] There are interesting consequences that follow from relaxing condition (d) as well: see Smolensky (1987b, 1988)

counts, by any reasonable criteria, as a token of either *3* or *5*.[1] Gödelese expressions are built up by multiplication, not concatenation.

Expressions in a merely functionally compositional scheme typically do have formal (i.e., non-semantic) structure of a kind. Indeed, their possessing a certain systematic internal physical configuration is essential to the possibility of real, implementable generation and decomposition processes. Still, the crucial point is that they are without formal *symbolic* structure, since they do not in general contain tokens of their constituents.[2]

2. COMPOSITIONALITY AND COGNITIVE PROCESSES

This distinction between strictly concatenative compositionality on the one hand and merely functional compositionality on the other can be used to clarify the difference between mainstream symbolic or "Classical" approaches to cognitive modeling and the emerging connectionist alternatives. Briefly put, while Classical approaches are committed, in both theory and practice, to concatenative compositionality, connectionists tend to abjure such strict constraints.

For example, according to Newell and Simon in their classic formulation of the symbolic approach, it is both necessary and sufficient for a system to exhibit intelligent behavior that it be a *Physical Symbol System*, where

> A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one being next to another). At any instant of time the system will contain a collection of these symbol structures...[3]

This definition encapsulates conditions (a) through (e). In particular, it is manifestly committed to (e) rather than (e') because symbol structures are composed of *tokens* of symbols that are *related in some physical way*. Since there can be no relations without relata, the symbol tokens must be present themselves, and not merely extractable by further processing (which would take us to some *further* instant of time). This strong view has recently received further authoritative endorsement from Fodor & Pylyshyn, who claim that

> In the Classical machine, the objects to which the content A&B is ascribed (viz., tokens of the expression 'A&B') literally contain, as proper parts, objects to which the content A is ascribed (viz., tokens of the expression 'A')....In short, it is characteristic of Classical systems...to exploit arrays of symbols some of which are atomic (e.g., expressions like

[1] This point is not at all impugned by the fact that it always takes some finite amount of work for us to *determine* whether or not there are any instances of the symbol *5* in tokens of the expression *1944*, and that some *idiot savant* might, with the same amount of effort, be able to extract the prime factors and hence produce (in his mind or elsewhere) some *other* token of *5*. What matters here is simply the physical shape (and hence the causal properties) of the tokens themselves. All *5* inscriptions have a flat top, and nothing in *1944* has a flat top.

[2] We should not be misled by the fact that Gödel numerals *are* constructed concatenatively from the tokens "0", "1", ... "9" The crucial point is that these tokens are not properly described as the constituents in the Gödel numeral scheme. The essentially orthographic rules of numeral construction are entirely different from the grammatical rules governing generation of the space of *Gödel* numeral expressions.

[3] Newell and Simon 1975 p.40.

'A') but indefinitely many of which have other symbols as syntactic and semantic parts (e.g. expressions like 'A&B').[1]

The presence of symbolic structure in the current strong sense is essential to the conception of cognitive processing which underlies the symbolic approach, for it is symbols which mediate between the semantic and the physical constraints on the behavior of the system. On one hand, the law-governed physical behavior of the system is explained by reference to the *causal* role of symbol tokens themselves, while on the other the system is *interpreted* by means of semantic assignments to those symbols. This is just the classical conception of computation, and the symbolic approach to cognitive modeling asserts that cognition *is* computation.[2]

Does this analysis confuse properties of representations at the cognitive level (the level of the *functional* architecture of the system) with details of the actual implementation? Might not a symbolic theorist be satisfied with merely functionally compositional representations at the implementation level? Suggestions like this do a disservice to the symbolic approach by conceding too much. Classical theorists have always insisted on the concreteness of their concatenatively structured symbolic representations. Thus, for Newell and Simon, symbol structures are physical entities within which symbol *tokens* are related "in some physical way;" Fodor & Pylyshyn, likewise, have stressed that the combinatorial structure of Classical representations must be mapped directly onto structures in the brain. It is this fact which makes possible the Classical explanation of cognitive processes by reference to the causal role of the internal syntactic structure of the representations themselves. If you deny that representations are concatenative at the implementation level, you must have up your sleeve an *independent* explanation of how cognitive processes are engineered. But, according to the true symbolic theorist, you will not - as a matter of contingent, empirical fact - be able to provide such an explanation.

The theoretical framework governing connectionist approaches is by comparison almost completely undeveloped, but we can nevertheless discern an increasing tendency in more recent work to reject precisely this commitment of the mainstream approach. Some relatively well-known work utilizing compositional but non-symbolic methods of representation are Smolensky's tensor product formalism, Hinton's techniques for the representation of hierarchical structures via reduced descriptions, and Pollack's Recursive Auto-Associative Memory (RAAM).[3]

Pollack, for example, devised a way to represent variable-sized data structures such as standard linguistic sequences in the form of stacks, where each stack is a distinct pattern of activity over a bank of hidden units in a three layer (i.e., one hidden layer) network. These stack representations exhibit functional compositionality since elements of a sequence can be stored and recovered, in appropriate order, quite reliably. Yet careful analysis of the stack representations themselves (the patterns over the hidden units) does *not* reveal features that could possibly count as

---

[1] Fodor & Pylyshyn 1988 p.16.

[2] Pylyshyn 1984.

[3] Smolensky 1987a, Hinton 1988, Pollack 1988.

tokens of the various primitive elements of the original sequence (nor those of any other symbolic scheme). Constituents of the represented sequences are effectively *stored* in stacks without being *instantiated* there. Consequently, while it is certainly appropriate to say that stacks are compositional representations *of* symbolic structures, they are not, strictly speaking, symbolic representations themselves. Pollack is therefore being somewhat misleading when he describes his stack representations as "compositional in the strictest sense;"[1] they are compositional, but only functionally so, not in the stricter concatenative sense.

One way to understand the disagreement here is to see that for the Classical theorist, any non-trivial representation of a complex structured item must itself have a parallel structural complexity in its internal compositional configuration. ("Non-trivial" means that the details of the internal *structure* of the item, and not just the item as a whole, are being effectively represented.) Connectionist representations, by contrast, eschew such internal compositionality in favor of a merely functional substitute. Insofar as we are concerned with *compositional* formal internal structure, then, the general point can be put in terms of the following handy slogan: for the connectionist, representations of structure need not be structured representations.

3. IS CLASSICAL COMPOSITIONALITY NECESSARY?

If it is true that connectionism is properly characterized as utilizing compositional but non-symbolic representations, a number of important consequences follow. First, as Smolensky has already pointed out, it is clear that connectionists can employ compositional representations without thereby committing themselves to the strict Classical approach.[2] Compositionality, in other words, is by no means the exclusive prerogative of the Classical paradigm.

Second, we can show that connectionism is well-equipped for the task of explaining certain aspects of cognition which, it has been argued, are beyond the explanatory reach of any model which refuses to employ strictly Classical representations and processes. In their recent influential critique, Fodor & Pylyshyn argued that cognition is *systematic*, and that only by postulating Classical representations and processes is there any real hope of explaining this phenomenon. Systematicity consists in such mundane facts as the following: that the ability to entertain one kind of thought always goes along with the ability to entertain systematically related thoughts (if you can think *John loves the girl* you can think *the girl loves John*); and that the ability to perform one kind of inference always goes along with the ability to perform systematically related inferences (if you can infer P from $P\&Q$ you can also infer P from $P\&Q\&R$).

How does utilizing Classical representations help us in generating an explanation of systematicity? Fodor & Pylyshyn summarize as follows:

---

1 p.37.

2 See Smolensky 1987b. Though we agree on this point, we differ in emphasizing different ways in which connectionist representations are non-Classical. As mentioned above, Smolensky focuses on relaxing condition (d) while I focus on condition (e).

> all the arguments we've been reviewing...are really much the same: If you hold the kind of theory that acknowledges structured representations, it must perforce acknowledge representations with *similar* or *identical* structure... So, if your theory also acknowledges mental processes that are structure sensitive, then it will predict that similarly structured representations will generally play similar roles in thought.[1]

By this reasoning, if connectionism is to be able to explain systematicity, it must also utilize representations that can have *similar* or *identical* structures, such that mental processes can be sensitive to that structure. The crucial question, then, is whether such structural relations can obtain among non-Classical representations, or whether *only* strictly symbolic representations can exhibit the relevant structural similarities. Fodor & Pylyshyn, of course, clearly prefer the latter view. Any non-Classical representations, they assume, must be completely unstructured; consequently, there can be no structural similarity relations for mental processes to pick up on, and so these processes must be purely "associationist."[2]

As I pointed out above, however, compositional but non-concatenative representations must be internally structured, though of course they are not *symbolically* structured. The distinctive structure of a given RAAM stack representation for example is found in the particular distribution of activity levels over the hidden units. Somewhat surprisingly, Fodor & Pylyshyn have simply left this whole class of representations out of consideration entirely. Since these representations are structured, it follows that they can stand in structural similarity relations. Indeed, it is increasingly common practice in connectionist modeling to analyze (using e.g. cluster analysis) sets of representations in order to uncover the order of similarity relations among the representations themselves. Further, these similarity relations tend to be systematic in that they reflect the constituency relations of the representations. Representations that were constructed in a grammatically similar fashion end up as neighboring points in the relevant high-dimensional vector space.

Consequently, in this respect connectionists have at least the raw resources for generating an explanation of the systematicity of cognition; their representations exhibit what Fodor & Pylyshyn themselves argue is the essential ingredient in such explanations. The task that remains is to devise processes, implemented in connectionist architectures, for manipulating these representations in a way that is systematically "sensitive to" (i.e., causally influenced by) their internal structure, and thus respects the compositionality-based similarity relations among representations. In this way, non-symbolic connectionist representations can be manipulated in a way that is systematically "sensitive to" (i.e., reflects) the complex structure of the items they represent.

It is important to realize that connectionists have scarcely begun this difficult task. Perhaps the best argument for the strictly Classical approach is that such processes will prove to be infeasible in the general case, i.e., with respect to the eventual goal of accounting for the full systematic complexity of human cognitive performance. Nevertheless it is also important to realize that, as far

---

1 1988 p.48.

2 See, e.g., p.32. An associationist mental process is one that is sensitive only to prior correlations in experience, and not "to features of the content or the structure of representations per se."

as can reasonably be predicted at this stage, there is (contra Fodor & Pylyshyn) no *principled* barrier to success in that enterprise. Connectionist approaches which are genuinely non-Classical have at least the basic resources to produce systematic performance, and the hypothesis that such representations in fact underlie our cognitive capacities does *not* render the systematicity of thought a mystery.

REFERENCES

Fodor J.A. & Pylyshyn Z.W. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition*; 28: 3-71.

Hinton G.E. (1988) Representing part-whole hierarchies in connectionist networks. Proceedings of the Tenth Annual Conference of the Cognitive Science Society. Montreal, Quebec, Canada: 48-54.

Newell A. and Simon H. (1975) Computer science as an empirical inquiry. *Communications of the Association for Computing Machinery*; 19: 113-126.

Pollack J. (1988) Recursive auto-associative memory: Devising compositional distributed representations. Proceedings of the Tenth Annual Conference of the Cognitive Science Society. Montreal, Quebec, Canada.

Pylyshyn Z.W. (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge MA: MIT Press.

Smolensky P. (1987a) On variable binding and the representation of symbolic structures in connectionist systems; Technical Report CU-CS-355-87, Department of Computer Science, University of Colorado.

— (1987b) The constituent structure of mental states: A reply to Fodor and Pylyshyn. *Southern Journal of Philosophy*; 26 Supplement: 137-160.

— (1988) Connectionism, Constituency, and the Language of Thought. Technical Report CU-CS-416-88, Department of Computer Science, University of Colorado.