# Beyond Symbolic: Prolegomena to a *Kama-Sutra* of Compositionality[1]

**Timothy van Gelder[2] and Robert Port[3]**
**Cognitive Science Program**
**Indiana University**
**Bloomington, Indiana, 47405**
*tgelder@indiana.edu, port@indiana.edu*

Consider some of the obvious differences between a printed and a spoken version of the very same English sentence. Both contain the same words, but in the printed sentence they are static ink configurations, while in the utterance they are temporally extended sound patterns. In the printed case, words are combined by juxtaposition in space; in the utterance, by juxtaposition in time. The printed words are combined in a very discrete way - there is, quite literally, space between each - and all occurrences of a given word are effectively identical, whatever words happen to surround them. In the utterance, by contrast, juxtaposition is not discrete and context free; words flow into each other, and their physical shape is typically affected by their neighbors.

Printed and uttered tokens of the very same sentence, then, are physically constructed or built up in quite different ways. These days, the most common technical term for building up a complex representation out of parts is *composition*, and so we can say that the two sentences exhibit very different kinds of *compositionality*, despite being syntactically and semantically identical[4]. This point has two very important consequences. First, compositionality itself comes in at least two, and possibly many, different kinds. Second, the very notion of compositionality can be a subject of study, quite independent of syntax and semantics. Whereas syntax and semantics focus on a language in the abstract, the study of compositionality focuses on the concrete implementations. How many kinds of compositionality are there? What are the fundamental issues? And, if our ultimate interests happen to be in cognitive science: what kinds of compositionality are typically found in mainstream artificial intelligence, in connectionism, and in natural cognitive systems such as

[1]To appear in *Symbol Processing and Connectionist Models in Artificial Intelligence and Cognition: Steps Toward Integration.* edited by Vasant Honavar and Leonard Uhr (Academic Press, 1993).

[2]Also Department of Philosophy, Indiana University

[3]Also Departments of Linguistics and Computer Science. The authors are grateful especially to Sven Anderson, Devin McAuley, Fred Cummins and David Chalmers for helpful discussion of this topic. This work was supported in part by the Office of Naval Research, Grant N00014-91-J-1261.

[4]Note that this use of *compositionality* is not to be confused with another common use: where basic constituents make approximately the same semantic contribution in every context in which they appear.

ourselves?

This paper is an initial foray into the surprising range of exotic forms that composition may take. We list six key issues in thinking about kinds of compositionality. We do not pretend that our list is complete, and each of the issues raised could be explored in much greater detail. Nevertheless, even as it stands the list provides a healthy new perspective on compositional representation in cognitive science. It turns out that there is, in fact, a wide range of kinds of compositional representation. We prefer to speak, in a metaphorical fashion, of a large *space* of forms of compositional representation, a space whose dimensions are the fundamental properties that give rise to different forms of compositionality. For the most part, cognitive scientists building models of cognitive functioning have utilized representations drawn only from one narrow region of this space, a region centered on static, concatenative kinds of representation such as printed sentences and LISP data structures. Recently, various connectionists have been exploring some other regions, building models with representations exhibiting very different kinds of compositionality. But there are vast tracts of the space of possible compositional representations which remain almost totally unexplored in cognitive modeling.

In the first section, we briefly discuss the *concept* of compositionality, and then point out some of the key dimensions along which various actual schemes differ. In the second section, we focus on one general kind of compositionality that is particularly salient from the perspective afforded by the list, and very different from conventional symbolic compositionality. We regard this kind, possessed by what we call *dynamic* representations, as both neglected and, potentially, of great significance for modeling human cognitive systems. Finally, in section three, we summarize our observations with a table of examples and propose some 'natural kinds' of compositionality.

## 1. Compositionality

Intuitively speaking, a representation is compositional if it is systematically constructed out of basic compoundable units. In what follows, we will refer to the basic compoundable units as *symboloids*, to emphasize that they include, but are much more general than, standard symbols such as LISP atoms or printed words.

For a more precise definition of compositionality, we begin by adapting the minimal set of sufficient conditions for compositionality proposed by van Gelder (1990). These conditions capture the idea that the essence of compositionality is hierarchical relations of constituency among representations and their components:

(a) There is a set of *primitive* types $P_i$; for each type, there is available a potentially unbounded number of actual physical instances or tokens (*symboloids*);

(b) There is a (possibly unbounded) set of *compound* types[5], $R_i$ ; likewise, for each type, there is available a potentially unbounded number of actual physical instances or tokens; and

(c) There is a set of transitive and non-reflexive *constituency* relations over these

[5]Expression types here are not to be confused with their grammatical category. In the current sense, there can be many different expression types which are all, for example, noun phrases.

primitive and compound types[6].

Any scheme of representation is compositional (in our sense) if it satisfies these very abstract conditions, and a representation is compositional if it belongs to such a scheme. For example, the sentence token

*Jill believes that the cat is white.*

(i.e., the actual ink configuration which appears on this very page) is a compositional representation because it is one of the unbounded number of possible instances of the compound type $R_{\textit{Jill believes that the cat is white}}$, which stands in constituency relations to the compound type $R_{\textit{The cat is white}}$ and to the primitive type $P_{\textit{cat}}$ , all of which belong to standard written English.

Notice that a spoken utterance of that same sentence type would count as a compositional representation for exactly the same reason. Thus, while conditions (a)-(c) might capture the general notion of compositionality as an abstract relationship among primitive and compound types, they say nothing at all about the various implementation issues which distinguish the different *kinds* of compositionality exhibited by various concrete instantiations of a given compositional scheme. In the remainder of this section we lay out a number of these issues. The section can be regarded as a first attempt to specify the ''principal components'' of the space of kinds of compositionality, i.e., the dimensions which do the most to distinguish conceptually the various possible kinds of compositional representation. We have aimed at dimensions that are independent and which clarify the range of possible solutions to the general problem of effective representation in cognitive systems. In illustrating the dimensions, we will make reference both to forms of natural language and to other computational models.

For current purposes, the dimensions belong to two fundamental categories: (1) properties of symboloids themselves; and (2) properties of the manner of combination of symboloids to form compound representations.

### *1.1 Properties of Tokens of Primitives.*

Consider an individual word token from a particular printed sentence of English. It is a classic example of 'symbol'. Some of its obvious properties include the fact it sits still on the page for a long time; that it is straightforward to recognize the symbol-type to which it belongs; and that it has ''meaning'' or semantic significance, yet its physical structure bears no interesting relationship to that meaning. These relatively obvious properties correspond to our three major symboloid dimensions.

Before describing those dimensions in a little more detail, it is worth mentioning that it can often be quite difficult to pin down the list of primitive types of a given compositional scheme of representation. For example, linguists, AI theorists and high-school teachers all assume that individual words are the primitives of English. This is satisfactory as a first approximation. But Bolinger among others has pointed out how

[6]There is a technical reason for stipulating that the constituency relations are defined over types. In some schemes of compositional representation, namely, non-concatenative schemes, compound representation tokens do not contain symboloids corresponding to their constituents. This issue is touched on below in Section 1.2; for further explanation, see (van Gelder 1990) .

difficult it is to nail down the actual list of the particular items that have just the right properties to count as the primitives (Bolinger 1975). Natural language is not just a set of morphemes plus rules of syntax. Speakers use and 'know' linguistic fragments that come in many sizes: from submorphemic ideophones to words, Bolinger's 'collocations', and idioms, as well as 'cliches' and even entire sentences and paragraphs of boilerplate (as in genres like wills and academic recommendation letters). The nature of many of these mysteriously constrained yet flexible 'lexicalized units' lies, in our view, well beyond the grasp of current representational schemes employed in linguistic theory.

Nevertheless, in what follows we will be proceeding on the assumption that the primitive types can be adequately identified for any compositional scheme under consideration.

*(a) Static versus Dynamic*

For our purposes, any representation and any symboloid, is a *concrete physical item* (an ink mark, a sound wave, a collection of electronic bits, a pattern of neural firing, etc.). It counts as a token of a particular type in virtue of some kind of distinctive physical configuration or ''shape''. We can think of this distinctive configuration as a matter of *change* along key physical aspects of the representation. Thus, a printed token of the word *CAT* exhibits change along spatial dimensions in the way ink is placed on the page, change that would be registered by a scanner as it moved along the page. It is precisely the details of this spatial change which makes it different than a token of the word *DOG*. Notice that, by contrast, an instance of *CAT* exhibits no significant change in physical configuration *from one moment of time to the next*. This is what we mean by saying that it ''sits still'' or that it is a static token.

Compare this with an uttered token of the same word. In this case, it is variation in frequency and amplitude *over time* which determines its type-identity. This is an example of a dynamic primitive. The key feature of a *dynamic* symboloid or representation is that *change over time is essential to type-identity*. It follows that, in order to determine to which type the entity belongs, you have to wait long enough for its distinctive changes to unfold. In the case of a static symboloid or representation, by contrast, you have (in principle, at least) enough information at any instant to determine type-identity.

Printed English words, then, are static symboloids, while spoken words are dynamic symboloids. One way to summarize the difference is to say that dynamic symboloids, but not static symboloids, happen in time.

*(b) Digital vs. Analog.*

A key feature of most standard compositional schemes of representation is that they are *digital*. By this we mean that it is possible to produce symboloids in the scheme, and determine the type identity of any given symboloid, *positively* - i.e., with complete success, definitely, perfectly, etc. (Haugeland 1985). Under such conditions we say that the symboloids themselves are digital.

A classic example of a process that can succeed positively is scoring in basketball; the ball either goes through the hoop or it doesn't. Similarly, in standard symbol systems, such as LISP machines, the most basic production, identification and transformation processes can be carried out with complete, unambiguous success. Barring malfunction, the system can always tell whether the symbol in its buffer is *foo* or not; there is simply no question of its being ‘‘somewhat *foo*’’. Obviously, printed English is also a digital scheme, given printing presses or laserwriters to produce the words and normal human readers to type-identify them.

An analog scheme is one that is not digital; analog symboloids are ones that cannot be produced and identified positively. This is not to say that they cannot be *successfully* produced and identified. Consider trying to determine who won a javelin throw by visual inspection. Most of the time, the test will indicate a winner, but sometimes the difference between the throws will be so small that it will be difficult to choose. Unlike basketball scoring, javelin competition is analog. Similarly, spoken English is analog, since, although it is usually possible to produce an identifiable token of any given word, success is always ‘‘more or less’’ rather than perfect. It makes sense, for example, to say that a given uttered sound was rather more like *hardest* than *hottest*.

Analog compositional schemes are, increasingly, cropping up in connectionist work. Consider, for example, Pollack’s Recursive Auto-associative Memory (RAAM) architecture (Pollack 1990). In this system, a sequence of items may be encoded, one at a time, onto the hidden layer of a ‘narrow-waisted’ feedforward network. By means of successive steps of autoassociative training (where the net is trained to reproduce its input on the output layer), the fixed-width hidden layer may be trained to store an entire list. If one trains such a network to encode an increasingly large structure, the representations of components of the structure become so closely packed in the activation space of the hidden units that eventually the network can no longer positively distinguish one representation from another. Clearly, the composite representation is analog, even though each component symboloid is initially digital.

Intuitively, whether a set of entities is digital or analog in this sense has much to do with whether or not they are ‘‘discrete’’ rather than ‘‘continuous’’. However, it turns out, in practice, to be rather difficult to systematically apply this familiar mathematical idea to schemes of representation. Roughly speaking, in this context, a set of entities is continuous if, between any two entities in the set, there is another entity which also belongs to the set. A discrete set is one such that there are ‘‘gaps’’ between entities. Put differently, a continuous set is densely packed into the space of possible entities of that kind, while a discrete set is sparsely packed into it. However, it turns out that it is difficult to give any more precise formulation of this intuitive idea that really does distinguish those schemes of compositional representation which intuition counts as discrete and those which it regards as more continuous. For this reason, together with the fact that digital vs. analog is clearly more significant from the point of view of the *functional* properties of a scheme of representation, we choose not to emphasize discreteness vs. continuity as a dimension of symboloids.

*(c) Arbitrariness.*

We have noted already that individual symboloids belong to the type they do in virtue of their distinctive physical configuration or makeup. The next dimension is a matter of how the physical makeup of symboloids relate to each other. The next dimension is a matter of how symboloids relate to each other in their physical makeup. A useful way to ask this question is: does fixing the physical configuration of one kind of symboloid in the scheme in any way constrain the configuration of any others?

Most words of English, written or spoken, are basically arbitrary in this sense. Thus, the printed token *CAT* bears no interesting relationship, in its physical configuration, to the printed token *DOG* over and above the fact that they are both ink marks constituted by smaller marks corresponding to letters. In particular, fixing the physical configuration of one word leaves the physical configuration of the other almost completely unconstrained (which is just to say that we could have used a quite different pronunciation and/or spelling in its place; the choice is completely arbitrary). Likewise, the sound [$k^h$@t] bears no interesting physical relationship to the sound [dOg].

The issue of the arbitrariness or non-arbitrariness of symboloids gets particularly interesting when a further complicating factor is introduced: the *meaning* of the individual symboloids. Symboloids always have at least two fundamentally different kinds of properties. On one hand, there are their physical properties, which determine their type identity as tokens. On the other, there is the meaning or semantic significance which each symboloid has and which it contributes to the compound representation that is formed out of it. Consequently, we can ask whether there is any interesting relationship between these two kinds of properties. Does the physical makeup of a given symboloid in any way reflect its semantic properties? If it does, that symboloid is not arbitrary with respect to its meaning; and since meaning is constraining physical configuration, different symboloids with related meanings must be correspondingly related in their physical configuration, and hence the symboloids are not arbitrary with respect to each other.

Clearly, virtually all written or spoken words of English are arbitrary; the printed token *CAT* bears no interesting relationship to feline mammals over and above the semantic one. The same is true of symbols in any standard computational system. This is one deep source of the anxiety that many people feel about ‘‘grounding’’ in computational systems. If the basic symbols are essentially semantically arbitrary, how can they have any ‘‘real’’ or ‘‘intrinsic’’ meaning? Surely, it is often thought, symboloids in *real* cognitive systems must somehow reflect their meanings in their formal configuration more directly.

Some connectionist compositional schemes do utilize semantically non-arbitrary symboloids. For example, in the McClelland and Kawamoto model of case role assignment (McClelland and Kawamoto 1986), input sentences are built out of vectors corresponding to distinct words, and those vectors themselves are essentially just lists indicating the presence or absence of microfeatures in the designated objects. Therefore, since the actual vector that is used to represent a given object (e.g., hammer) is constrained by the features of a hammer -- i.e., by what that vector *means* -- these vector symboloids are semantically non-arbitrary[7].

[7]Of course, the meaning of the *component features* of that vector remains quite arbitrary.

### *1.2. Properties of the Manner of Combination*

Compositional representations are obtained by combining symboloids to obtain compound wholes in accordance with the abstract constituency relations. It is natural and useful (though sometimes a little misleading) to think of this in terms of a mechanical process such that actual individual symboloids are put in one end and out the other comes a compound representation. We can then ask: how does this process work? What process actually combines symboloids to obtain compound representations?

*(d) Mode of Combination.*

An obvious feature of printed sentences of English is that they are made up out of individual words, and that each word appears in the sentence exactly as it would if it were alone on the page. Words are taken ''off the shelf'', as it were, and appear totally unaltered in the resulting sentence. A less obvious feature of *spoken* sentences of English is that this is not generally true: when a word appears in a sentence, it often ends up with a somewhat different shape, depending on the words it is surrounded by, than it has when spoken alone. For example, compare the pronunciation of the word *cut* -- paying special attention to the /t/ -- in *cut Paul, cut some* and *cut out*.[8] Further, somewhat surprisingly, it is possible to systematically generate compound representations out of symboloids in a way that not only changes, but seemingly *destroys* those symboloids completely.

The term we use for the kind of combination that is characteristic of printed sentences is *concatenation*. (For a thorough discussion of concatenation, and its significance in AI as well as the classical conception of cognition and connectionism, see van Gelder, 1990.) For the moment, a relatively informal definition is sufficient: a mode of combination is concatenative just in case it preserves the constituent symboloids in the resulting compound representation; if, in other words, one can literally point to the constituent symboloids in the compound whole. Since this is true for both printed and uttered sentences of English, we need to go on to distinguish three different modes of combination:

(a) *Pure* concatenation, in which the process of combination leaves symboloids utterly unchanged. All standard computer languages and representational schemes in ''classical'' cognitive science utilize pure concatenation.

(b) *Context-sensitive* concatenation, in which each constituent symboloid is recognizably present in the resulting compound, but its shape has been altered by the very process of combination, typically in a way that systematically reflects its context. An interesting connectionist example of context-sensitive concatenation will be described below in Section 2.

---

[8]Thus, both the tongue tip and glottal gestures are distinct in these three phrases for typical American pronunciations. In *cut Paul,* during the tongue tip closure for the [t], the glottis clamps shut (as it would in the isolated phrase *Cut!* ). For *cut some,* the glottis is likely to open up rather than closed during the [t]. In *cut out*, on the other hand, the glottis remains closed and produces voicing while the tongue tip bounces quickly off the roof of the mouth.

(c) A *non-concatenative* compositional scheme is one in which symboloid constituents do *not* literally appear in the compound representations, even though they may have been input into the combination process. Pollack's RAAM architecture is one relatively well-known kind of connectionist model which forms compound representations by non-concatenative combination.

In RAAM, basic symboloids take the form of vectors on the input layer, and they are systematically combined by the network to form compound representations in the hidden layer. The symboloid constituents are not literally present in the compound representation; if you examine the activation vector which is the compound representation, you will find no instance of the vector which was the input symboloid. Nevertheless, the compound representation systematically *reflects*, in its formal structure, its actual constituency relations. This latter point can be shown in two ways. First, further processing in the RAAM network can recreate the symboloid constituents. Second, the compound representations can be used in further processing which is systematically appropriate to its constituent structure (Chalmers, 1990).

*(e) Static vs. Temporal Combination.*

In printed sentences of English, combination is by spatial juxtaposition, and all constituent symboloids are present simultaneously for the entire period of the existence of the whole printed sentence. In *uttered* sentences, by contrast, symboloids are combined temporally; one follows another, and no more than one can be present at a given instant. We call the first kind of combination, in which all constituent symboloids are simultaneously present, *static* combination, while combination by temporal succession is *dynamic*.

Note that dynamic symboloids and dynamic combination are two quite independent properties of compositional representations. The first concerns the nature of the symboloids themselves, the second concerns the way symboloids are combined. Though it is quite natural for dynamic symboloids to be combined dynamically (as in speech) and for static symboloids to be combined statically (as in print), other combinations are possible; thus, dynamic symboloids can be combined statically (like a musical chord), and *vice versa*.

*(f) Syntactic Conformity.*

When basic tokens are combined to form compound wholes, to what extent are the syntactic rules of the scheme observed? A standard assumption is that representations in a compositional scheme are, by definition, grammatically well-formed, i.e., constructed in strict accordance with certain syntactic rules. Any combination of symbols which violates the syntactic rules is junk or ''symbol salad''. We think that it is more useful to regard certain combinations as representations belonging to a given compositional scheme even if inviolate syntactic rules do not exist or if some rules are violated. Many if not most utterances of everyday spoken English are not grammatically well-formed (in the sense that their utterers would wince and edit them if given an opportunity to examine their own speech closely), yet they are still perfectly adequate for representational and communicative functions. Their utility does, however, seem to presuppose at least some loose conformity to the syntax of English. We wish only to claim that there is a rough *spectrum* of cases between those symboloidal schemes in which *strict* conformity is observed (e.g., programming languages, printed English), and schemes in which only *loose* conformity is sufficient (e.g., spoken English). We suspect that, insofar as there are compositional representations underlying general cognitive performance, these representations will, like natural language, exhibit only loose conformity to any relevant syntactic rules.

Having completed our survey of the major dimensions by which compositional schemes may differ, we want to describe one novel kind of representation found in a certain branch of connectionist work in some detail. This kind of representation is different from the kind of representation typically used in cognitive science along *every one* of the six dimensions, and thus illustrates the usefulness of our dimensions and the richness of the space of compositional representations.

### *2.0 Dynamic Memory for Auditory Pattern Recognition*

This representational scheme is found in a model that was developed recently at Indiana University by Port, Sven Anderson and Fred Cummins for an auditory perception task (Anderson et al, 1993). The model was designed to account for how the human (or mammalian) auditory system can recognize familiar sound patterns even when many unrelated patterns may be simultaneously superimposed, and despite variation in their rate of presentation. Humans can track speech or melodies or the sound of a passing car right through klinking dishes, the TV in the next room and laughing children, all at once. Our astonishing skill at segregating a stream of sound into independent components is, in Albert Bregman's terms, largely 'schema-based' (Bregman, 1990), since it depends on familiarity with each sound pattern. This is a generalized version of the so-called ''cocktail party problem'' -- of following a single speaker's voice in a noisy party.

As developed so far, this model is able to recognize holistic patterns made from a discrete sequence of orthogonal vectors (in acoustic terms, a sequence of pure tones) presented sequentially in time. In contrast to all other recognition schemes, the system does not 'represent' all the elements of the pattern simultaneously in memory -- as though in a 'time window' and then try to recognize the pattern. Instead, it recognizes *only* the whole sequential pattern by means of distinctive properties of the trajectory that the sequence induces in an abstract state space -- a trajectory that leads the state of the system to a privileged region of that space which specifies identification of a particular sequential pattern. Thus only the sequence is recognized, with nothing in the recognition state directly representing the elements that compose the presented pattern.

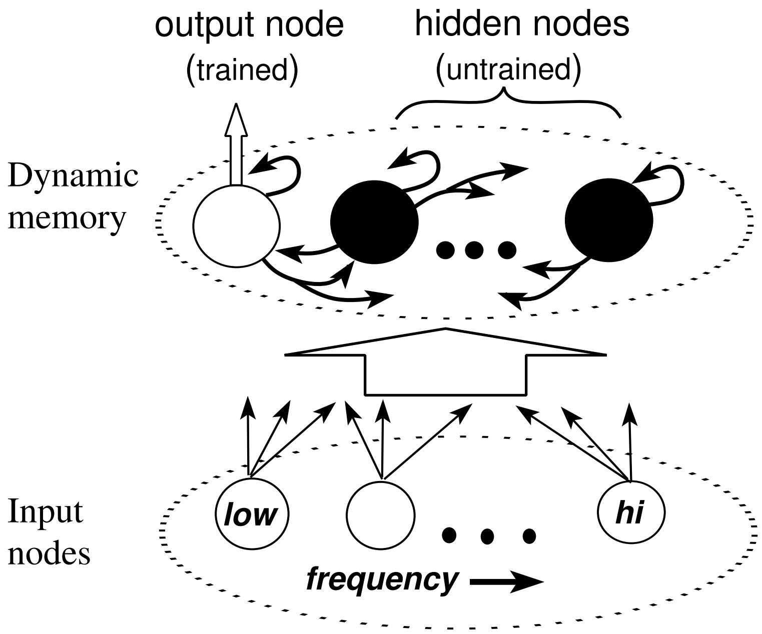
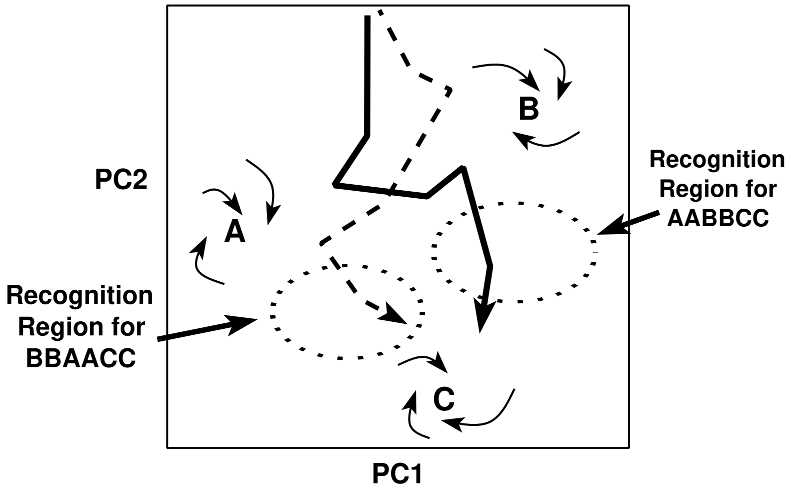



***Figure 1: Network model***. The activation of the Input nodes represents the energy level at each frequency. Most inputs tested so far are pure tones, implying that only one input node has amplitude near 1 while the others have amplitude near 0. All these nodes excite the nodes of dynamic memory on the next time step. These nodes are fully connected with learned weights on each link. One or more of the Dynamic Memory nodes are Output nodes that are trained to output maximum activation (that is, 1) on the last 2 time steps of learned patterns. The other nodes (Hidden nodes) are untrained. (See Anderson, et al, 1993 for further details.)

As shown in Figure 1, the connectionist network that does this job contains a layer of Input Nodes for a limited range of audio frequencies that are fed to all the nodes in Dynamic Memory. The Dynamic Memory layer of nodes, about 8 of them, is fully recurrent, that is, each node feeds output to itself and to all the other nodes in the layer. A gradient descent learning algorithm is used to train one or more nodes in that layer, called Identification Nodes, to move from an activation of 0 to 1 for the last two time steps of a multi-step target pattern. To illustrate, let us suppose that our target patterns consisted of different sequences of sound spectra characterized as A, B and C. Each spectrum type is called an input element. In this discrete-time simulation, differences in the rate of presentation of the sequence will be manifested by the number of time steps for which each element is repeated: thus the input sequence AABBCC could be ‘slowed down’ and presented as AAABBBCCC. But both differ, of course, from BBAACC.




---

***Figure 2: Trajectories for different sequences.*** The axes are the values of the two most significant principal components in this stylized example. Each point on this plane thus represents a particular pattern of activations on the dynamic memory (DM) nodes. After training, input of one of the pattern elements makes DM into a dynamical system in which the state settles toward a particular fixed point -- indicated by the letter name. Its vector field is suggested by the curved arrows near each fixed point. So when input A is presented, the system state approaches the 'A' fixed point for as many time steps as element A is presented (as in early parts of the solid curve resulting from the sequence AABBCC). But if Bs are input (as in early parts of the dashed curve for BBAACC), the dynamics of the system immediately changes and a new vector field is created with an attractor at 'B'. So, on presentation of AABBCC, the state veers first toward A's fixed point for 2 time steps, then toward B, and finally toward C. The sequence BBAACC yields a very different trajectory -- one that leads through a different Recognition Region. The training regimen has thus resulted in (1) fixed points for pattern elements located appropriately to distinguish target sequences from each other and from nontarget sequences, and (2) Recognition Regions located just where they need to be to distinguish target trajectories from nontarget trajectories.

---

Figure 2 shows stylized trajectories for the two hypothetical target patterns, 'ABC' and 'BAC', displayed on a plane representing the activations of dynamic memory in a network trained to recognize these two patterns. Since there are typically 8 or so nodes in this memory (including the Identification Nodes) -- too many to directly examine all at once -- we can summarize the patterns in the space of activations of dynamic memory by displaying the two most significant 'principal components', PC1 and PC2 of this 8-dimensional space. The principal components are axes through the multidimensional activation space (and thus linear combinations of the activations of the dynamic memory nodes) that are selected to exhibit the greatest amount of variance during presentation of the stimulus sequences.

The two Recognition Regions of this plane turn out to be the regions of this plane in which each trained Identification Node has a value of 1. Thus when examined on this plane, the activation state will pass through one of these regions on the last couple time steps if it correctly identifies the pattern. Otherwise, it will not pass through either recognition region and thus, in effect, classify the stimulus pattern as a 'distractor', that is, as a non-target sequence.

This system is able to recognize patterns pretty well -- not just tell ABC from BAC, but even ABCDEF from BACDEF (where the difference between the patterns lies 5 pattern elements before the end of the pattern). Also, a single small network can learn to simultaneously recognize quite a few different patterns. Furthermore, it does so despite severe superimposed noise (see Anderson *et al* , 1993, for detailed explication). But one might ask why such a strange recognition scheme was developed in the first place. Why not just store the inputs as they come in and then recognize the pattern when the pattern elements are all sitting next to each other in a buffer of some sort?

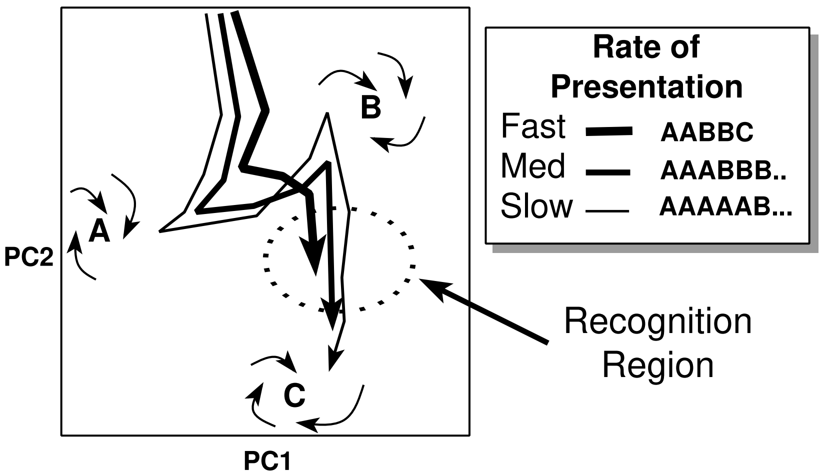



***Figure 3: Rate Variation:*** The fixed point for each pattern element ''attracts'' the system state toward it asymptotically. But the vector field gets weaker as the attractor is approached, so it approaches more and more slowly. Thus modest changes in rate of presentation of a trained pattern (up to a factor of 2, at least) have rather small effects on the shape of the trajectory and have almost no effect on identification performance. Thus, adjustment for rate variation is completely automatic.

The first reason not to use the traditional method is that this scheme recognizes its target patterns presented at a range of rates *without any training* at all on rate variation. The mechanism for this is illustrated in Figure 3. The figure shows that when a pattern, such as ABC, is presented at different rates, the trajectories induced in dynamic memory are very similar -- and they tend to pass through the recognition region at *any* of the rates. And, most surprisingly, it is not necessary to train the network at more that one rate in order to obtain this behavior!

There is a major disadvantage to this kind of system for certain purposes: it is necessary to learn each pattern individually. That is, this system, because of its dependence on the structural characteristics of each specific pattern, has no way to represent the significant properties of an *arbitrary* pattern (unlike standard 'time-window' or 'echoic-memory' models). But the psychological task it was designed to model (a highly relevant one biologically, in our opinion) is not concerned with completely general representation of any sound at all (the way a tape recorder does, for example), but only with familiar sounds -- and in a way that is highly resistant to superimposed noise from competing patterns. (For arguments that human auditory perception does depend, in fact, on such a learned auditory memory rather than a raw 'echoic memory, see Port, 1990.)

Now, to return to the issue of composition and representation: How is the input pattern represented in the model? One answer is that the representation is the final *state* of the network, when it is in the recognition region. Another possibility, which we will explore here, is that it is the whole temporally extended trajectory that represents the input sequence[9]. The pattern to be recognized is not represented in the model by a static state, but rather by a movement through its state space -- an entity rather more like a spoken utterance than it is like a printed sentence or a computer data-structure. But if the representation is the whole trajectory, what are its components? What are the basic symboloids in this scheme? One initially tempting answer is that they are the fixed-point attractors corresponding to individual elements of the input patterns. But this cannot be right, since the attracting points are never actually reached; the system is never actually in a state corresponding to the attractors (certainly it need not reach them to recognize the pattern). It does however *move toward* those attractors one after another. For this reason, we take the symboloids in these representations to be *trajectory segments* -- the sequence of states that the system is in as it moves toward an attractor under the influence of a given input. Since the system might approach a given attractor from a wide range of different starting points, these trajectory-symboloids are intrinsically highly variable.

How, then, is this kind of representation classified by our six dimensions? If the above discussion is right, these representations are, like speech, made up of

[9]Of course, the dynamical system that has been proposed (along with the trajectories that occur in it) is only a mathematical model for the changes in the *physical* state of a real nervous system in which there is real semantic content that correlates with those physical states. This metaphysical ambiguity between a mathematical model and the (representational) physical system it simulates is characteristic of all models in cognitive science.

dynamic symboloids (movements in state space) that are dynamically combined. These symboloids are analog and non-arbitrary (since the trajectories are obtained as a result of training which fixes the location of any given attractor in a way that depends crucially on both the location of the other attractors and the input frequency channels). The mode of combination is context-sensitive concatenation, since the actual shape of a given symboloid (the trajectory segment) obviously depends on the symboloid that precedes it. Finally, conformity to any syntactic rules is at best loose, since this model will generate a compound trajectory in response to any sequence of inputs, whether grammatically constrained (that is, in this case, learned) or not.

This model should be viewed as a rather speculative proposal for how familiar but brief sound patterns might be represented. Its appropriateness for this purpose is not relevant here. Our principal interest in this model at the moment is simply as an illustration of a very different, and seemingly exotic, corner of the space of possible compositional schemes that cognitive systems *might* be exploiting, and to demonstrate the possibility that there are many species of compositional systems that could serve practical functions for a cognitive system that have not yet been imagined.

### *3.0 Compositionality: Natural Kinds*

The following table summarizes much of what has been said thus far. It shows how our six dimensions can be used to classify already existing kinds of compositional representations. It reveals some of the diversity in the forms of compositionality that have *already* been devised and put to use. We compare printed and spoken natural language, a paradigm example of compositional representation from classical AI research, and two connectionist schemes. Looking across the rows, it is apparent that each of the dimensions is indeed useful in highlighting differences between various kinds of compositionality.

The six dimensions are, for the most part, independent. Cheap combinatorics, then, would suggest that there are scores of significantly different general kinds of compositional representation based only on these dimensions. It is likely that there are dimensions that we have not noted, and so the space of possible kinds of compositional representation is large indeed. Of course, it is far from clear that all of these different kinds would be practically implementable and usable. It is more likely that there are something like ‘‘natural kinds’’ of compositional representation, in which certain properties cluster together to give rise to particularly useful concrete methods of representation

**Some Types of Compositional Representations**

| | | printed English | spoken English | Hearsay-2 | RAAM | Dynamic Memory |
|---|---|---|---|---|---|---|
| **Tokens** | **Static vs. Dynamic** | *static* | *dynamic* | *static* | *static* | *dynamic* |
| | **Analog vs. Digital** | *digital* | *analog* | *digital* | *analog* | *analog* |
| | **Arbitrary Difference** | *yes* | *yes* | *yes* | *no* | *no* |
| **Combination** | **Concatenation** | *pure* | *context senstv* | *pure* | *non-concatv* | *context senstv* |
| | **Static vs. Temporal** | *static* | *temporal* | *static* | *static* | *temporal* |
| | **Syntactic Conformity** | *strict* | *loose* | *strict* | *strict* | *loose* |

***Table 1.***

Consider, for example, printed English, as shown in the table: it's symboloids are static, digital, and arbitrary, and its manner of combination is concatenative, static and relatively strict (despite massive uncertainty about the syntactic rules themselves). This combination of properties makes it possible for printed text to play its many social communicative roles. Interestingly, our single example of compositional representation in AI, the style of representation found, for example, on the ''blackboard'' of Hearsay-2 (jumping over 2 columns), also has much the same cluster of properties -- which make it possible for those representations to play *their* functional role within the standard computational architecture of an artificial intelligence system (Lesser *et al*, 1975) . It is clear that this cluster of properties constitute one natural kind, probably the most important one for technical purposes. The traditional term used to pick out this kind is *symbolic*. That is, the core notion of symbolic representation is a compositional representation which is static (in both senses), digital, arbitrary, concatenative and syntactically strict (Newell and Simon, 1976; Fodor and Pylyshyn, 1988). Other forms of representation are appropriately said to be symbolic to the extent that they approximate this core notion.

Are there other natural kinds? An obvious candidate is apparently the cluster of properties possessed by *spoken* natural language (moving to the left in the table): dynamic and analog tokens, and context-sensitive, temporal and loose combination. The critical role of spoken language as a communicative medium testifies to the utility of a form of compositional representation with this set of properties -- at least for human beings. We describe this general kind of compositionality as *dynamic*, to emphasize that, unlike the standard symbolic form, representations of this kind are essentially temporal; i.e., they happen over time. One other example of compositionality that seems to have the much same general properties is *music* -- in more-or-less every musical culture and genre. The recursive autoassociative memory

(RAAM) has static tokens and static combination, but the tokens are analog and combined by nonlinear superposition. While it illustrates a novel corner of the space of compositional systems, it is not obvious that such a model will turn out to be 'natural kind'.

We assume that, insofar as there are compositional representations in human cognitive activity at all, it is still an open question exactly how those representations are implemented, and hence what kind of compositionality they exhibit. The standard assumption seems to have always been that they are symbolic (in the current sense), or at least, static. We suspect that this assumption is due in part to a common failure to appreciate, first, the importance of the physical implementation of representations[10], and second, the diversity of possible kinds of compositionality. We want to suggest the possibility that at least some compositional representations in human cognitive systems are dynamic (though of course other possibilities may exist as well).

There are at least two very general considerations which suggest that it might be fruitful to focus research attention in the direction of dynamic representation (see Port and van Gelder, in press, for deeper exploration of this possibility). First, many of the phenomena with which cognitive systems must cope are themselves dynamic. To take just one pertinent example, in perceiving speech, we cope with a temporally extended speech event. It is clearly the way sound changes over time that is the basis of recognition and understanding. Now it may turn out that the first thing our cognitive system does is to generate a static representation for that dynamic entity (certainly that is the bet that standard linguistic theory has made). But another possibility is that the representations remain dynamic -- in temporal lockstep with auditory input. A second very general consideration is just the simple observation that real neurons and neural systems are in constant activity. They never do nothing (as one might expect to find in a system based on static representations), but rather are always firing, changing their firing rate and modulating the firing rate of other neurons. Now, this constant flux may simply reflect the set of processes that manipulate and transform static representations. But another suggestion that seems more plausible to us is that the representations themselves are changing patterns of neural activity and that complex thoughts and cognitive structures are manifested as rapidly modulated dynamic events -- that the mind is, in effect, constantly obeying 'cognitive laws of motion'.

Static, concatenative styles of compositionality are currently the most widely used in modeling in cognitive science. Such forms of representation certainly offer some functional advantages, but we suspect that their ubiquity also owes much to other factors such as the prevalence of the written sentence as a model, the widespread reliance on standard digital computer architectures in cognitive modeling, the great conceptual difficulties involved in thinking about and modeling with entities that happen in time, and, not least, the force of tradition and a lack of imagination. At this

[10]For example, why are the major differences pointed out here between spoken and printed English completely missed in the discipline of linguistics? A major reason is that it is assumed that time is simply another parameter -- that , for the purposes of a theory of language, it can be simply modelled with a sequence of static symbols. But any attempt to physically implement language or speech functions (either for production or perception) forces one to face the logical differences between time as a parameter and time as the historical matrix of physical cognitive processing.

stage we urge only that cognitive scientists push beyond standard positions and explore a greater range of forms of compositionality. In cataloguing at least some of these forms, and the relevant foundational issues, this paper is intended as, if not a completed Kama-Sutra of compositionality, at least a prolegomena to one.

***References***

Anderson, Sven, Robert Port, Fred Cummins and Devin McAuley (1993) Dynamic memory: a model for auditory pattern recognition. (unpublished manuscript).

Bolinger, Dwight (1975). *Aspects of Language*. New York: Harcourt Brace Jovanovich.

Bregman, Albert (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge: Bradford/MIT Press.

Chalmers, David (1990) Syntactic transformations on distributed representations. *Connection Science 2*, 53-62.

Fodor, Jerry and Zenon Pylyshyn (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition: International Journal of Cognitive Science 28*, 3-71.

Lesser, V. R., R. D. Fennel, L. D. Erman and D. R. Reddy (1975) Organization of the Hearsay-II speech understanding system. *IEEE International Conference on Acoustics, Speech and Signal Processing 23* , 11-23.

McClelland, J. L.and A. H. Kawamoto (1986). Mechanisms of sentence processing: Assigning roles to constituents of sentences. In McClelland and Rumelhart (1986), pp. 272-326.

McClelland, J. L.and D. E. Rumelhart (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge MA: Bradford/MIT Press.

Newell, Allen and Herbert Simon (1976) Computer science as empirical inquiry. *Communications of the Association for Computing Machinery 19*, 113-126.

Pollack, Jordan B. (1990). Recursive distributed representations. *Artificial Intelligence 46*, 77-105.

Port, Robert (1990) Representation and recognition of temporal patterns. *Connection Science 2*, 151-176.

Port, Robert and Tim van Gelder (in press) *Mind as Motion: Dynamics, Behavior and Cognition.* Cambridge: Bradford/MIT Press.

van Gelder, Tim (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science 14*, 355-384.