Compositionality

in Neural Systems

Elie Bienenstock

Division of Applied Mathematics

Brown University, Box F

Providence RI 02912

and CNRS, Paris, France

Stuart Geman

Division of Applied Mathematics

Brown University, Box F

Providence RI 02912

June 1994

In

The Handbook of Brain Theory and Neural Networks,

M. Arbib, ed.

MIT/Bradford Books, Elsevier, 1995

pp. 223–226

[1]Supported by Army Research Office contract DAAL03-92-G-0115 to the Center for Intelligent Control Systems, National Science Foundation grant DMS-8813699, and Office of Naval Research contract N00014-91-J-1021.

1 Compositionality in cognition

Compositionality refers to our ability to construct mental representations, hierarchically, in terms of parts and their relations. The "rules" of composition are such that (*i*) we have at our disposal an infinite repertoire of hierarchically constructed entities built from relatively small numbers of lower-level constituents—a phenomenon sometimes referred to as (infinite) *productivity*—and (*ii*) allowable constructions nevertheless respect specific constraints, whereby overwhelmingly most combinations are made meaningless.

Example 1. In *language*, no more than twenty-six characters and a half-dozen additional symbols—alternatively no more than a few dozen phonemes— are required to compose a story on any subject one can possibly imagine. Moreover, the collection of all texts that have ever been spoken or written in the English language is an infinitesimal fraction of what can *possibly* be composed. At the same time, most arrangements of symbols that are possible *a priori* from a mere combinatorial point of view are illegitimate as linguistic constructions. The number of character strings of length 1,000 that make up a proper English text is vanishingly small when compared to the number of all possible strings of such length. Thus, while infinitely productive, language is at the same time severely constrained.

When observed from the "surface," the composition mechanism in language appears simple. Individual characters are assembled into syllables, which are themselves assembled into words, further composed into phrases, sentences, etc. One text differs from another text in the same language only by the *relative positioning* (relations) among the constituents (symbols), and not for instance by the frequencies of occurrence of each symbol; these frequencies are about the same for any sufficiently long text. Yet, encoded within this apparently simple surface structure is a considerable amount of highly complex organization: syntax, semantics, pragmatics, etc. Linguistic constraints are partly arbitrary and conventional (a fact sometimes viewed as a violation to the principle of compositionality), and partly obey some language-specific regularities. Chomsky (e.g. 1986) further makes the controversial statement that some "universal," *language-independent*, regularities, may result from specific properties of our brains that are largely genetically determined.

Example 2. In the perception and production of *visual imagery*, primitive elements, that serve analogously to letters or phonemes, are combined in a highly specific relational manner to form composite entities suitable themselves for relational bindings into even more specific "high-level" structures. Edge elements combine, with "rules" about gradient magnitudes and directions, to form curve elements, which in turn can combine, end-to-end, to build the cartoon-like boundary description of a scene. Surface elements piece smoothly together, in a manner consistent with boundary-determined discontinuities, to form three-dimensional shapes, which are themselves combined into the objects of everyday imagery. There is, furthermore, an infinite, but nevertheless topologically and logically restricted, repertoire of object placements to produce a meaningful scene. Grenander (1993) has built a mathematical theory of the composition of patterns, characterized by these basic principles of productivity and restrictedness; Biederman (1987) has exploited a compositional description of objects using a small repertoire of volumetric shapes called "geons," in his psychological theory of object recognition.

Example 3. *Procedures* can be systematically decomposed from broadly-defined goals, to be achieved over minutes or hours, recursively into simple motor actions with durations of fractions of a second. These basic units, analogous to phonemes or simple shapes, can be effectively combined to generate an infinite variety of goal-achieving activities.

Organization by composition is in fact so ubiquitous as to suggest that it is fundamental to cognition. This being the case, there are several implications worth highlighting:

Disambiguation. The problems of interpreting an image or understanding spoken language are sometimes said to be ill-posed. Yet they really are very well-posed, as attested to by the spectacular recognition performances achieved by humans under ordinary circumstances. It is true however that auditory and visual data are often *ambiguous* at all but the most global levels of interpretation. Isolated phonemes or even words and phrases spliced from a continuous speech signal can be impossible to interpret. In fact, the mere segmentation of a speech signal into phonemes or words is difficult, if not impossible, in the absence of a simultaneous global interpretation. This apparent ambiguity persists at any given level, or at several levels at once (acoustic, phonetic, lexical, syntactic). Analogous considerations apply to scene analysis.

Evidently, despite the richness of possible scenes or utterances, there are severe constraints restricting plausible interpretations. Recalling that compositional constructs are themselves highly restricted, we may interpret our remarkable auditory and visual perceptual skills as exploiting, in a fundamental way, the restrictedness of compositionality.

Invariance. Relational descriptions are invariant. In *computer vision*, this is the basis for many object recognition algorithms (cf. Dickinson et al., 1992): define the objects of interest relationally, and often hierarchically, in terms of the relative positionings of identifiable subparts. Identification becomes, essentially, a matter of relational graph matching, and is, *a fortiori*, invariant. One may, for example, identify a chair as a planar rectangular surface with four attached more-or-less-identical cylindrical parts, situated near corners and roughly perpendicular to the plane, and a further plane, attached perpendicular to the first plane, on the side opposite the cylinders. The parts, furthermore, may themselves be defined as relational compositions of still more primitive elements.

Analogous, in *language*, is the invariance of meaning over a multitude of almost equivalent expressions: many different word strings can adequately evoke the mental objects and the (partly metaphorical) relations between them that constitute a given intended message. Speech demonstrates other invariances as well: the simple linear relation of constituents confers an invariance with respect to the *rate* of articulation as well as other phonetic parameters, which can be altered in many ways without affecting meaning. By and large, it is the *order* (in English and most other natural languages) and not the duration of constituents that is the primary vehicle of meaning.

In general then, to the extent that cognitive entities are relationally organized, they can be identified and/or described in multiple equivalent ways. Invariance is thereby related to the relational and productive properties of compositionality.

Computation. Artificial interpretation of speech and image data are daunting engineering tasks; difficulties generally manifest themselves as overwhelming computational requirements. A few successes however have been reached—mostly in speech recognition—and these rely on "divide-and-conquer" strategies, exploiting the hierarchical organization of data. Compositional models may be based upon primitive grammars that restrict word sequences, phonetic models for allowable word pronunciations, and acoustic models for the articulation of the phonetic units. This hierarchy is the basis for computationally-feasible algorithms that infer a word sequence from a raw acoustic signal (e.g. Rabiner, 1989).

Such successes, although sparse, strongly suggest that our brains too avoid explosive combinatorial search by exploiting in a recursive manner the compositional organization of mental representations.

2 Compositionality in neural systems

Since compositionality is so central to cognition, it appears reasonable to construe it as an observable manifestation of a property of compositionality in *neural activity*. Drawing from §1, we can identify several features that neural mechanisms for compositionality would likely possess:

Compositional representation through dynamical binding. The neural representation of a composite entity should include the suitably defined composition of those patterns of neural activity that make up the representations of the constituents of this composite entity. A popular simple example is the problem of representing a scene containing a red triangle and a blue square. The mere *coactivation* of four cells (or groups of cells) representing the four elementary features "red," "blue," "triangle," "square" would lead to a "superposition catastrophe" (von der Malsburg, 1987), that is, in this case, the inability to distinguish a scene containing a red triangle and a blue square from a scene containing a red square and a blue triangle. Composition is thus more than coactivation: a *binding* mechanism is required, to attach with each other the neural representations of the entities "red" and " triangle."

Binding needs to be *dynamical*, i.e., reversible, to allow the representation of other constructs at different times.

Relational binding. Binding further needs to be *relational*, that is, qualified in terms of a collection of domain-specific relations among constituents. For example, to account for our *linguistic* ability to assemble six lexical items such as "feed, carve, Elsa, Sophie, pumpkin, cat" into a string such as "Sophie feeds the cat and Elsa carves the pumpkin," a compositional model should use bonds that are qualified in terms of *predicate roles*. Thus, it will be specified that the bond between the neural representation of the item "Sophie" and the neural representation of the item "feed" is of the *subject* type, i.e., that it is Sophie who does the feeding. Only then will the representation of one particular string be distinguishable from the representation of alternative strings, constructed from the same constituents. Note that these alternative constructs can be: (*i*) syntactically and semantically legitimate, such as "Elsa feeds the cat and Sophie carves the pumpkin"; (*ii*) syntactically correct but semantically/contextually unacceptable, such as "Sophie carves the cat and Elsa feeds the pumpkin"; (*iii*) syntactically wrong, such as "Feeds Elsa Sophie carves and the cat pumpkin."

Hierarchical computation. A basic tenet of compositionality is that cognitive representations are hierarchically organized. Likewise, it is natural to expect that the computational mechanisms that elicit the sequence of neural events corresponding to a perceptual or motor act—e.g. in visual pattern recognition or in the interpretation or production of spoken language—are hierarchically organized.

It is hardly disputable that, in the sense of the features just outlined, no satisfactory encompassing treatment of neural compositionality is available to date. In particular, models of the cell-assembly type, inasmuch as they address the issue of compositionality, represent each new composite entity by allocating for it *separate* neural machinery rather than by composing the representations of its constituents. This has led some authors (e.g. Fodor and Pylyshyn, 1988) to the strong, and highly controversial, conclusion that modern "connectionism" is wholly inadequate to model cognition at the representational level—the level discussed here.

Nevertheless, models do exist that provide some elements of a theory of neural compositionality (e.g. von der Malsburg, 1987; Shastri and Ajjanagadde, 1993; Smolensky, 1990; Gindi et al., 1991; Hummel and Biederman, 1992; see also the articles of von der Malsburg, Singer, and Shastri, in this volume). An important common feature of most of these models is the use of mechanisms through which a number of neural activity patterns are combined into a composite pattern that preserves, as subpatterns, the original constituent activities. Binding is dynamical: the constituent patterns can either appear by themselves, representing isolated entities, or they can be explicitly bound to represent a composite entity.

Thus, a compositional model will in general *not* allocate a specific cell—or group of cells—for a composite entity such as "red triangle," as a typical feedforward-net model would. Rather, it will employ the already-available machinery, that is, the elementary-feature cells or cell groups, and posit the existence of an *additional degree of freedom* in neural activity. This new degree of freedom will be used to dynamically express the bond between "red" and "triangle," thereby avoiding the superposition catastrophe. Similarly, in the above linguistic example, the primary activity patterns associated with the six lexical items will be preserved, and an additional degree of freedom will be used to express syntactical dynamical bonds. In short, composite patterns will be constructed by suitably arranging constituent activities, thereby providing an explicit representation of parts and their relations. Productivity will then arise, fundamentally, from combinatorics in a space of neural activity patterns; see Damasio (1989) for a similar picture based on neuroanatomical data and lesion studies.

Most compositional models to date follow a suggestion of von der Malsburg (1987), and use *fine temporal structure* of neural activity as the medium for expressing dynamical binding. One currently popular implementation of this idea—not a part of von der Malsburg's original theory—posits that the neurons whose activities are to be bound fire, for some time, periodically or nearly periodically. Each neuron is then viewed as carrying two *independent* variables, a level of firing and a phase. The latter is the additional degree of freedom used to express dynamical binding. At this point, there exists no conclusive neurophysiological evidence for this mechanism. Some support however comes from recent findings in visual cortex (see the article of Singer in this volume).

Shastri and Ajjanagadde (1993) propose a linguistic model along these lines, in which the representation of a predicate such as "carve" would include the oscillatory activity of two distinct neurons (or neuronal populations) for the two roles: subject ("person carving") and object ("thing carved"). When representing "Elsa carves the pumpkin," the person-carving neuron is phase-locked with the oscillating neuron whose activity represents Elsa. The other entity occurring in the representation—namely, the instantiation of "thing carved" as "pumpkin"—uses a different phase. Bindings are propagated between predicates along hard-wired *phase-preserving* lines, which embody long-term *rules*. For instance, the rule "a person (or animal) being fed eats" uses a phase-preserving line from the object neuron of predicate "feed" to the subject neuron of predicate "eat"; it allows the system to *infer*, from the short-term fact that Sophie feeds the cat, another short-term fact, namely that the cat eats. Long-term facts may also be encoded in the system, e.g. "Sophie loves animals," or "children carve pumpkins for Halloween." Shastri and Ajjanagadde show that such a system can perform simple, "reflexive," reasoning. Albeit limited in several ways, this reasoning may access a virtually unlimited store of long-term rules and facts.

In the same spirit of locking the phases of oscillators to express binding, though with the additional assumption that specialized fast links are used for signal synchronization, Hummel and Biederman (1992) propose a model of object recognition from line drawings based upon the compositional approach to object representation of Biederman (1987) mentioned in §1 (see the article of Hummel in this volume for further details).

Although these attempts exhibit most of the features outlined in the beginning of the section, they do not add up to a fully coherent theory of neural compositionality. Most notably, they use rigid architectures—sometimes hierarchically structured as in Hummel and Biederman (1992). They suggest no convincing hypothesis for the mechanisms underlying the extreme versatility manifested by our brains in linguistic behavior, e.g. in the handling of recursive constructions, or of metaphors, or, more generally, of analogical discourse or reasoning. Furthermore, they fail to address the important issue of how compositional representations are learned and modified, e.g. during language acquisition. This stands in sharp contrast with the wealth of ideas about learning advanced for feedforward connectionist networks.

Some efforts have been made to tackle these issues as well. In particular, von der Malsburg (1987) suggests adopting a developmental/epigenetic approach, stressing the role of processes of self-organization and natural selection in neural compositionality. In this approach, one investigates mechanisms of brain development that could result in the formation of specific spatio-temporal activity patterns providing a suitable medium for highly versatile compositional operations. Bienenstock (1994) has proposed that "synfire chains" (Abeles, 1991 and references therein) may be relevant here, more so than oscillating circuits. Synfire chains are, roughly, large networks that are wired in such a fashion as to support wave-like patterns of activity specified with a millisecond accuracy. Electrophysiological data collected in frontal cortical areas of behaving monkeys are suggestive of the existence of (reverberating) synfire chains (Abeles et al., 1993).

The hypothesis is that these structures could be dynamically bound via weak synaptic couplings; the wave-like activities of two synfire chains could be synchronized in much the same way as coupled oscillators lock their phases. More complex spatio-temporal patterns could arise from reverberation of activity, a form of "folding" of the chains upon themselves. Such complex patterns could exhibit highly specific binding properties (think of the highly specific interactions between folded proteins), providing a suitable medium for both productive and restricted composition. Recursiveness of compositionality could, in principle, arise from the further binding of these composite structures. Here again, however, we are a long way from a completely coherent, much less a comprehensive, theory.

In sum, neural compositionality remains among the most challenging issues in brain theory. Particularly vexing are the computational aspects, related for instance to the problem of graph matching, e.g. for object recognition, in compositional neural models. Although one may expect significant progress in theoretical investigations, such progress is bound to remain largely speculative, until it becomes possible to map cortical activity with high spatial *and* temporal resolution, and to process in a useful way the overwhelming amounts of data that will result.

References

Abeles, M. (1991) *Corticonics: Neuronal circuits of the cerebral cortex*, Cambridge University Press.

Abeles, M., Bergman, H., Margalit, E., and Vaadia, E. (1993) Spatiotemporal Firing Patterns in the Frontal Cortex of Behaving Monkeys. *J. Neurophysiol.*, 70(4):1629–1638.

Biederman, I. (1987) Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147.

Bienenstock, E. (1994) A Model of Neocortex. To appear in *Network: Computation in Neural Systems.*

Chomsky, N. (1986) *Knowledge of Language: Its nature, origin, and use*, New York: Praeger.

Damasio, A. R. (1989) Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition, *Cognition*, 33:25–62.

Dickinson, S. J., Pentland, A. P., and Rosenfeld, A. (1992) From volumes to views: an approach to 3-D object recognition *CVGIP: Image Understanding*, 55:130–154.

Fodor, J. A., and Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: a critical analysis, *Cognition*, 28:3–71.

Gindi, G., Mjolsness, E., and Anandan, P. (1991) Neural networks for model based recognition, in *Neural Networks: Concepts, Applications and Implementations* (P. Antognetto and V. Milutinovic, Eds.), Englewood Cliffs, N.J.: Prentice-Hall, pp. 144–173.

Grenander, U. (1993) *General Pattern Theory: A Study of Regular Structures*, Oxford University Press.

Hummel, J. E., and Biederman, I. (1992) Dynamic binding in a neural network for shape recognition, *Psychological Review*, 99:480–517.

Rabiner, L. R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77:257–286.

Shastri, L., and Ajjanagadde, V. (1993) From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings, *Behavioral and Brain Sciences*, 16:417–494.

Smolensky, P. (1990) Tensor product variable binding and the representation of symbolic structures in connectionist networks, *Artificial Intelligence*, 46:159–216.

* von der Malsburg, C. (1987) Synaptic plasticity as a basis of brain organization, in *The Neural and Molecular Bases of Learning* (J.P. Changeux and M. Konishi, Eds.), John Wiley and Sons, pp. 411–432.