We are impressed by the attempt by van der Velde & de Kamps (vdV&dK) to take seriously the challenge of capturing the complexity of human language in a neurally plausible model. Their model makes it possible to ask questions about the encoding of the details of sentence structure that it was difficult even to ask previously. This is no mean achievement. Nevertheless, we are concerned that the authors’ model avoids one of the most fundamental properties of sentence structure and that this could seriously restrict the scope of the model. Although many of the figures in the target article bear a superficial resemblance to the phrase structure trees of linguistics, the sentence structure representations in the neural model lack the hierarchical constituent structure encoded in phrase structure trees. Phrase structure trees encode bindings between primitive elements (words) that create constituents and also bindings between constituents that form larger constituents. In vdV&dK’s model, in contrast, only bindings between the basic word-level structural assemblies are encoded. A verb’s theme subassembly may be temporally bound to a noun’s theme subassembly to form the equivalent of a simple verb phrase, but the verb phrase does not itself combine with other subassemblies to form larger constituents. The S and C structural assemblies that are employed in the encoding of main clauses and embedded clauses, respectively, do not delimit clause-sized constituents. Rather, they are word-level structural assemblies whose subassemblies bind with the subassemblies of other word-level units.

The binding of words and phrases to form hierarchically organized constituent structures is a property shared by a wide variety of linguistic models that differ in many other respects (e.g., Berwick 2001; Chomsky 1985; Frank 2002; Goldberg 1985; Poliard & Sag 1994; Steedman 2000), and it plays a crucial role in explanations of many linguistic phenomena. These include the following:

i. Coordination rules. In most cases, like categories can be combined with the conjunction and to form a larger instance of the same category: nouns coordinate with nouns, verbs with verbs, verb phrases with verb phrases, and so on. In the absence of a mechanism for encoding recursive constituent structures in vdV&dK’s model, it is difficult to capture the fact that John and Mary is a noun phrase that governs plural verb agreement, or the fact that The managers and the pilots who supported the strike is a noun phrase in which the relative clause may modify only pilots or both managers and pilots.

ii. Anaphoric relations. Languages make extensive use of anaphoric expressions that are interpreted as taking the meaning of another constituent in the sentence or discourse. Pronouns such as he or them may cohere with another noun phrase constituent, and forms like it or so may be anaphoric to a clause-sized constituent, as in The sun was shining, but Sue couldn’t believe it. The expression so do so is anaphoric to a verb phrase, which may be a larger constituent, as in Bill finished his homework on Tuesday and Sally did so too, or a smaller constituent, as in Bill finished his homework on Tuesday and Sally did so on Thursday. It is difficult to capture such dependencies in a model that lacks hierarchically organized constituents.

iii. Long-distance dependencies. The long-distance dependencies found in wh-questions, topicalization, relativization, passivization, raising, and scrambling structures consistently involve the appearance of a constituent in a noncanonical position. It is difficult to capture such rules without constituents.

iv. Scope relations. Recursive formation of constituents makes it straightforward to capture the fact that the expression second-longest American river refers not to the Amazon—the second-longest river and also an American river—but rather to the Mississippi-Missouri, which is the second longest among American rivers.

v. Command relations. Many syntactic relations are restricted to constituents that stand in a command relation, the relation that holds between a constituent and its sister and all subparts of its sister. For example, negative polarity items such as ever and any must be co-commanded by a negative expression. This constraint captures the acceptability of Nobody thinks that Bill ever sleeps and the unacceptability of "A man that nobody likes thinks that Bill ever sleeps."

The absence of constituents in vdV&dK’s model makes it more difficult to capture structural generalizations of this kind.

vi. Recursive modification. Modifier expressions such as adjectives and relative clauses may be freely combined with the categories that they modify, in any quantity, as in six big red India rubber balls. In grammars with hierarchically organized constituents, this can easily be captured using a recursive rule such as N’ → Adj N’. In vdV&dK’s model, however, modifier expressions are bound to the categories that they modify by means of dedicated subassemblies, and multiple modifiers require multiple dedicated subassemblies. It strikes us as inefficient to require all noun structural assemblies to include a special adjective subassembly that is exploited only in noun phrases with six or more adjectives.

vdV&dK correctly note that combinatorial productivity and recursive productivity are separable issues. Combinatorial productivity can obtain in the absence of recursive productivity so long as there is arbitrary binding between fillers and roles. Recursive productivity, they note, "deals with the issue of processing most complex syntactic structures, such as (deeper) center-embeddings" (sect. 4.2). The above discussion illustrates, we hope, that, at least for natural language, recursive productivity—that is, constituent depth—is at issue even for simple syntactic structures.

We can imagine at least two ways in which the neural blackboard architecture could be extended to encode hierarchical constituent structure without sacrificing the main insights of the model. One possibility would be to add new structural assemblies that correspond to nonterminal nodes in a phrase structure tree. For example, assemblies for the categories NP and VP would bind with other categories and not with individual words. All NP assemblies would then need to have a number of subassemblies that would allow them to bind with any potential mother or daughter node of NP. An alternative possibility would be to directly exploit the delay assemblies that are activated in a memory circuit when a pair of subassemblies is bound together. If the delay assembly could double as a structural assembly for a constituent node that could bind with other constituent nodes, then this might allow encoding of hierarchical constituent structure. Indeed, vdV&dK hint that a pair of bound structural subassemblies can themselves be bound to another pair of bound subassemblies when they draw a dotted line between the n and v subassemblies in Figure 10 as a means of capturing subject-verb agreement. Crucially, the model must be able to encode not only the first-order relationships between word-level primitives but the second-order relationships between relationships that characterize constituency in natural language.

Whether these or any other solutions turn out to be most feasible, we suggest that the neural blackboard architecture cannot properly address the challenge of the "massiveness of the binding problem" (Jackendoff 2002) unless it is able to recursively encode constituents and bindings among constituents.

On the unproductiveness of language and linguistics

David M. W. Powers
School of Informatics and Engineering, Indiana University South Australia, Adelaide SA 5001, Australia. David.Powers@indiana.edu
http://www.infoeng.indiana.edu/people/pages/powers_david

Abstract: van der Velde & de Kamps (vdV&dK) present a response to Jackendoff’s four challenges in terms of a computational model. This commentary supports the position that neural assemblies mediated by recurrence and delay indeed have sufficient theoretical power to deal...
Commentary/van der Velde & de Kamps: Neural blackboard architectures of combinatorial structures in cognition with all four challenges. However, we question the specifics of the model proposed, in terms of both neurophysiological plausibility and computational complexity.

It is often assumed that language is vastly different from and largely independent of other cognitive processing, and this modularity fact is apparent even in the target article. Nonetheless, its attempt to explain aspects of language in more general cognitive and perceptual mechanisms exemplified in the examples given. In particular, in introducing the massiveness of the binding problem (sect. 2.1), it is suggested that cat, mouse, and chases activate specialized populations of neurons and distinguish word order in a manner similar to the operation of motion detectors for a vertical bar. But then it is argued that language is fundamentally different because motion detection is able to be specialized for motion of a limited set of possibilities, but language has unlimited combinatorial productivity. In fact, the productivity has nothing to do with language or linguistic theory. Rather the complexity derives from our environment, and even without language, an organism is continually faced with new situations, new entities, and new interactions and interrelationships.

The fact that many things can change a mouse has nothing to do with language. Indeed, in Dumbeldor, we would be introduced with a sentence, whereas in real life he would be introduced verbally and multimodally by a person in a direct personal encounter or might just be seen chasing the mouse without his name or profession being known.

The visual system detects the motion of Dumbeldor, the cat, or anything else irrespective of any linguistic bindings. The fact of motion is detected at a very primitive level, and work such as Hubel's (1988) indicates that there are neural detectors for salient features such as edges in motion. The mechanisms by which an object is recognized, its motion is recognized, and its name is recognized all seem to be similar and are often theorized to be associated by neural synchrony on the basis of empirical evidence of synchronous neural activity (Roesel et al. 1997; Sarnthein et al. 1989; Shastri & Ajjanagadda 1993b; von Steina et al. 1999).

The problem of variables (sect. 2.3) is in essence an artefact of the assumption of rule-based structures, and both are linguistic constructs that probably have no concrete correlate in brain function. Rules and variables, moreover, do not necessarily occur in modern statistical language learning approaches; rules are implicit in supervised approaches involving tree-banks (Marcus 1991), probabilistic grammars (Charniak 1993), and/or data-oriented parsing (Bod 1995) but are supplanted by a more general concept of prosodic, phonological, morphological, syntactic, and semantic patterns in unsupervised approaches (Clancy 2001; Hutchens 1995; Powers 1983). The underlying phenomenon whereby variables get attached to (in a rule-based approach) or abstracted patterns get matched with current sensory-motor or linguistic content is again a matter of binding or association, which is commonly dealt with by theories of synchrony (Weiss et al. 2001).

However, van der Velde & de Kamps (vdVdK) do not see synchrony or recurrence as a panacea for Jackendoff's challenges but rather show how various early models exhibit exactly these problems. They point out that the Shastri et al. solution to the multiple binding problem is duplication and that this then faces problems with nested structures and implies a "one-level restriction." This is technically incorrect, but the argument does indeed imply a "flute levels restriction" which is consistent with Miller's (1985) Magic Number Seven constraints, with the inability of humans to cope with arbitrarily complex embeddings, with phenomena such as subjacency, and with the observation that there is insufficient space in our heads for the infinite stack implied by linguistic theories that see language as strictly more complex than context-free.

Synchrony involves a specific pattern that is present in each neuron triggered as a result of a specific object or event, and this pattern represents a temporal encoding that would seem to encode low (<20 Hz) frequencies as well as information correlated with spatiotemporal location that results in higher-frequency components and evidently has a reverberatory or recurrent origin. A Hebbian neural assembly is intrinsically a spatiotemporal association, and the concept of synchrony adds the potential for overlapping in the sense that the same neurons can synapse into multiple assemblies with different characteristic signatures. The circles or triangles that represent terminal or nonterminal symbols linguistically in vdVdK are in fact intended to represent assemblies neurologically, and these are intrinsically dynamical structures that exhibit synchrony and provide evidence of recurrent processes (Hebb 1949; Pulvermüller 2002), although this is not made explicit in the target article.

There are, moreover, alternatives to the duplication approach as well as refinements such as a homomorph model built upon evidence of a propensity for spatiotemporal receipt fields and projections that reflect information-distorted sensory-motor representations of the opposite half of the body (Powers & Turk 1989). Plausible models should also take account of attention and saccade and the evidence that we maintain a relatively high-level spatiotemporal model of our environment that is informed by attended events and peripheral change (e.g., in the visual domain motion or lightening; in the auditory domain, modulation or softening). The spatiotemporal representations have very much the function of the more abstract phonotactic blackboard metaphor. Powers (1997) envisions the spatiotemporal model as being like the columns of blackboards in a lecture theatre—different columns for different sensory-motor or cognitive modalities, different levels for different times and levels of complexity. In the lecture theatre, a board is pushed up after it has been written on, and a clean board emerges from below. While working on the current layer of boards, one can refer to any board on any higher layer that is still visible.

vdVdK's model is largely consistent with this kind of model but is more reminiscent of sequential digital log circuitry and in fact makes the analogy explicit through the use of gate terminology. Synchrony, reverberation, and recurrence would seem to be important mechanisms in realizing their model, although there is an interacting pair of neural attributes that they neglect: delay and decay. Delay is clearly implicit in reverberatory and recurrent models, but delayed copies of a signal can exist at different parts of the network even without recurrence. These delayed copies create the temporal dimension of a blackboard-like spatiotemporal representation. Hence a neuron is as much a short-term memory element as a processing element and functions something like a dynamic memory cell that maintains its memory while it is being refreshed as relevant. It is a complementary effect that is substantiated by the refractory period and its role in habituation. Powers (1983) used this delay-decay model (as well as a more abstract statistical model) for induction of both grammatical and ontological structure.

Although presented in an unhelpful way that is not standard for both neurons and gates, the vdVdK gating model is similar to sequential digital log circuits, and the resulting model of memory acts like a static memory cell. Whilst the model is sufficient to a priori solutions for Jackendoff's problems, it is considerably more opaque than the simpler model, and there is no direct neurological support for this kind of model except in terms of the ongoing synchronous recurrence between features triggered for the same event that forms the Hebbian assembly.

The elaboration of the vdVdK model is intrinsically syntactic in nature and this extends to their models of short-term (blackboard) and long-term (knowledge base) memory. There is no ontology, no semantics, no meaning captured in distinguishing The cat chases the mouse from The mouse chases the cat. There is no discussion of how cat is recognized or the verb chased is understood, and the representation of words with circles, supposedly representing neural assemblies, fails to capture the
Commentary/van der Velde & de Kamps: Neural blackboard architectures of combinatorial structures in cognition

inherent overlap of the distributed representation of a Hebbian assembly (Palmermiller 2002). A more realistic model involves abstraction of entire scenes and association of words with those scenes in the same way as any other attributes of the objects or events involved, these structures reflecting (Piaget 1954) the multimodal multilevel spatiotemporal short-term delay-decay representation of the current ontological context in the blackboard network. The hearing or reading of a sentence generates a sequence of neural patterns that starts off being perceptual in nature and becomes increasingly abstract and conceptual as well-attested neural processes extract features and relationships. The sentence and its context, both linguistic and ontological, are thus represented in different layers of the network at different complexity levels and time points - the spatiotemporal neural blackboard. The function of recurrence and inhibition is not to implement short-term memory but to allow the representation of complex spatiotemporal relationships.

Linguists tend to get this backwards. It is not that language determines who is the subject or the object in a sentence, but rather that in the environment there is an actor and an undergrounder (Pike & Pike 1977). The reality has primacy, and the underlying representation is arguably there to support language but to represent past, present, and potential experience. The issues Jackendoff raises are not primarily problems of linguistics but matters of causation. In dealing with problems that Jackendoff poses from a linguistic perspective, vdVcdK tend to force their cognitive modules into a linguistic mold rather than producing a proper ontological model and showing how linguistic relationships, semantic, phonetic, morphological, and syntactic, can be represented - or, indeed, can be emergent.

It is therefore highly appropriate that vdVcdK conclude by looking at how their blackboard architecture maps onto vision, which is arguably representative of the entire array of sensory-motor and cognitive modalities. This would have been a starting point, as understanding this kind of feature-binding model can potentially lead to a better understanding of syntactic and semantic binding.

Comparing the neural blackboard and the temporal synchrony-based SHRUTI architectures

Lokendra Shastri
International Computer Science Institute, Berkeley, CA 94707.
shastri@icsl.berkeley.edu  http://www.icsl.berkeley.edu/~shastri

Abstract: Contrary to the assertions made in the target article, temporal synchrony, coupled with an appropriate choice of representational primitives, leads to a functionally adequate and neurally plausible architecture that addresses the massiveness of the binding problem, the problem of 2, the problem of variables, and the transformation of activity-based transient representations of events and situations into structure-based persistent encodings of the same.

Table 1 compares two sets of solutions to the challenges posed by Jackendoff (2002). One set of solutions is provided by the SHRUTI architecture, which uses temporal synchrony for encoding dynamic bindings (Mani & Shastri 1993; Shastri 1999; Shastri & Ajjanagadde 1993b; Shastri & Wendellen 2000), and the other by the target article. This comparison is clearly at odds with what is stated in the target article. The following discusses the bases of the comparison and point out some of the factual errors and faulty analyses underlying the flawed evaluation of the temporal synchrony approach proposed in the target article.

The SHRUTI architecture represents relations (or predicates), types, entities, and causal rules using focal-clusters. Figure 1 depicts focal-clusters for relations (e.g., give), types (e.g., Person), entities (e.g., John), and the rule give(x, y, z) (own(y, z)). Within a focal-cluster, the activity of the + node represents a degree of belief, the activity of the + node represents querying of information, and the synchronous firing of a role node (e.g., giver) and an entity's (or type's) + node represents the dynamic binding of the role and the entity (or type). Type + nodes are further differentiated to encode quantification (e for existential and e for universal). Thus the sustained activity of + give together with the firing of +: John, +: Mary, and +: e: Book in synchrony with giver, recipient, and give-object, respectively, encodes the active belief: "John gave Mary a book." This activity immediately leads to the inference "Mary owns a book" because of synchronous activity propagating along connected nodes (e.g., owner synchronizes with recipient and hence with +:Mary).

Contrary to what is claimed in the target article (sect. 3, para 2, sect. 3.3, para 1), no preexisting fact nodes or synchrony detectors are required for the activity-based encoding of "John gave Mary a book," and no such nodes/detectors are required for drawing inferences based on this fact. Furthermore, as long as SHRUTI is told that Dumbledore is an entity or an instance of an existing type, it will have no problem encoding the novel event "John gave Dumbledore a book" and productively inferring that Dumbledore owns the book.

This brings out the first major error in the authors' understanding of the temporal synchrony-based SHRUTI architecture. Contrary to their claim, SHRUTI does not require prewired fact nodes for all possible facts. SHRUTI requires fact nodes (actually, fact circuits) only for encoding memorable facts in its long-term memory.

Problem of 2. Although the simple network shown in Figure 1 permits an entity simultaneously to fill multiple roles in different relations, it cannot simultaneously encode multiple instances of the same event-type (e.g., "John gave Mary a book" and "Mary gave John a pen") without binding conflicts. Shastri and Ajjanagadde (1993b) present a solution to this problem within the temporal synchrony framework. The solution requires having a small number of copies of each relational focal-cluster and encoding rules by interconnecting antecedent and consequent focal-clusters via switching circuits (Mani & Shastri 1993; see Wendellen & Shastri 2004 for an alternate solution).

The authors correctly point out that the use of multiple copies of focal-clusters makes it difficult to learn regularities between relations. Because an occurrence of a situation wherein a give event leads to an own event would engage only one focal-cluster each of give and own, respectively, only this pair of focal-clusters will learn the causal link between give and own.