Getting Symbols out of a Neural Architecture

John E. Hummel (jehummel@illinois.edu)
Department of Psychology, 603 E. Daniel St.
Champaign, IL 61820 USA

Abstract

Traditional connectionist networks are sharply limited as general accounts of human perception and cognition because they are unable to represent relational ideas such as *loves* (John, Mary) or *bigger-than* (Volkswagen, breadbox) in a way that allows them to be manipulated as explicitly relational structures. This paper reviews and critiques the four major responses to this problem in the modeling community: (1) reject connectionism (in any form) in favor of traditional symbolic approaches to modeling the mind; (2) reject the idea that mental representations are symbolic (i.e., reject the idea that we can represent relations); and (3) attempt to represent symbolic structures in a connectionist/neural architecture by finding a way to represent role-filler bindings. Approach (3) is further subdivided into (3a) approaches based on varieties of conjunctive coding and (3b) approaches based on dynamic role-filler binding. I will argue that (3b) is necessary to get symbolic processing out of a neural computing architecture. Specifically, I will argue that vector addition is both the best way to accomplish dynamic binding and an essential part of the proper treatment of symbols in a neural architecture.

Keywords: connectionism; neural computation; symbolic representation; relational processing

Introduction

Along with our ability to use language, the major factor distinguishing human thinking from the cognitive abilities of all other primates is our ability to reason about relations (Oden, Thompson & Premack, 2001; Penn, Holyoak & Povinelli, 2008). Relational thinking is thinking that depends on the relational roles in which objects (and other relations) are engaged, rather than just the literal features of the objects themselves. The ability to think explicitly about relations is central to mathematics, science, art, engineering and everything else uniquely human (Holland et al., 1986; Holyoak & Thagard, 1995), including activities as mundane as planning a meal, making an analogy and judging basic perceptual similarities (Gentner, 1983; Goldstone, Medin & Gentner, 1991; Hummel, 2000; Palmer, 1978; Taylor & Hummel, 2009).

Relational thinking is so commonplace that it is easy to assume that the computational mechanisms underlying it must be relatively simple. They are not. Relational thinking is a late evolutionary development tied closely to the size and complexity of the human frontal cortex (Robin & Holyoak, 1995; Stuss & Benson, 1986). It is also late to develop in childhood (see Smith, 1989) and highly susceptible both to the capacity limits of working memory (WM) (see Hummel & Holyoak, 1997, 2003) and to brain insult, especially to the frontal cortex (Morrison et al., 2004; Viskontas et al., 2004; Waltz et al., 1999).

Requirements for Relational Thinking

Why should relational thinking pose such a challenge to evolution (note that it appears to have taken roughly four billion years to evolve) and to the cognitive architecture? The answer lies in part with the fact that learning in neural networks is inherently *spatial*, in the sense that a change in the strength of a synapse alters the relation between the neurons sharing that synapse but has no direct effects on the rest of the brain: Any learning that takes place between a pair of neurons can only affect the mappings in which that pair of neurons takes part. As a result, if a neural network learns a response (the post-synaptic neuron) to a stimulus (the pre-synaptic neuron), then it can generalize that response to new stimuli that activate the same pre-synaptic neuron, but it cannot, without additional learning, generalize to new stimuli that fail to activate the pre-synaptic neuron. When the pre-synaptic neuron represents, say, an object feature, this kind of learning provides for a kind of automatic generalization to objects similar to those used in training.

Note, however, that this kind of generalization depends critically on what, exactly, our pre-synaptic neuron represents. If our pre-synaptic neuron represents a basic feature such as *red*, independent of the context in which *red* appears, then it provides the potential for generalization to any new red object. But if the neuron turns out to represent something more like *red square*, then the learned response will only generalize to red squares.[1] (If it truly represents "red square" then it could be expected to generalize to red squares at any size, location, etc., but only to red squares.) In other words, we can only expect the learning to generalize to all red things to the extent that our pre-synaptic neuron represents red *independently* of the object's other properties. Or in still other words, a neuron's *range of generalization* is identical to its *range of response invariance*.

This independence requirement is straightforward in the context of basic object properties such color and shape. (At this point, you may be thinking "Of course a neural network

[1] For clarity, I am describing the neuron's response as though it were all-or-none. Exactly the same logic applies if we replace "if the neuron represents" with "to the extent that the neuron represents", but the wording becomes more cumbersome.

can only respond to something to the extent that it actually represents that thing!") But the independence requirement introduces complexity in the case of relational generalization—i.e., generalizations and inferences that depend on the relational roles in which objects are engaged rather than just the features of the objects themselves—i.e., precisely those kinds of generalizations and inferences that are the hallmark and power of relational thought—because it implies that, in order to make inferences based on relational roles, a neural network must be able to represent those roles independently of their arguments (Hummel & Holyoak, 1997).

That people can represent relational roles independently of their arguments is evidenced by our ability to understand how "John loves Mary" is similar to "Susan loves Bill" or to understand how the lower left pair of shapes is similar to the upper pair of shapes in Figure 1 (Hummel, 2000). In turn, representing relational roles independently of their arguments makes it necessary to specify the binding of roles to their arguments *dynamically*, i.e., as needed and in a way that does not affect the representation of either the roles or their fillers (Hummel & Biederman, 1992; Hummel & Holyoak, 1997).

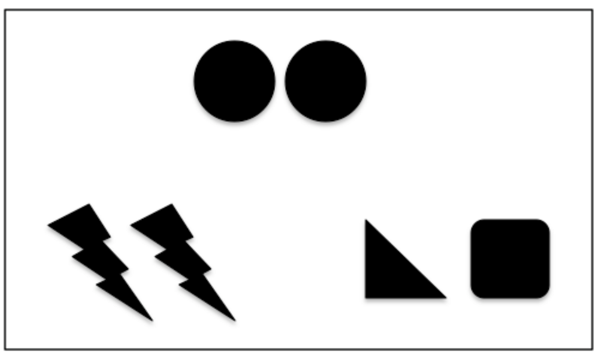

Figure 1. Illustration of our ability to appreciate relational similarities. The left pair of lower objects is similar to the top pair of objects, even though they are different shapes, because both pairs satisfy the *same-shape* (x, y) relation.

There is substantial evidence that the need to dynamically bind relational roles to their arguments is at least one of the reasons why relational thinking is slow to develop in childhood, dependent on attention and WM and susceptible to brain damage (in brief, because it depends on attention, which depends on inhibition; for reviews, see Cowan, 2001; Hummel & Holyoak, 2003). It is tempting to speculate that the need for dynamic role-filler binding is also one of the reasons why relational thought was so late to evolve: It is simply not obvious (certainly to many neural network modelers, and so perhaps not to evolution, either) how to *do* dynamic role-filler binding in a neural architecture. Additional reasons it might have been late to evolve are that (a) it is possible, in many finite domains, to asymptotically approximate the power of relational thinking in a non-relational architecture, so perhaps there has historically been insufficient selection pressure to make the great leap from non-relational to relational thought; and (b) learning abstract relational concepts (such as *same-as*, or *insufficient-to*, or even *larger-than*) from concrete examples itself poses a complex computational problem that may rely on the power of analogical mapping for its solution (Doumas, Hummel & Sandhofer, 2008).

It is interesting to note that the requirements for relational representation (and thought)—the ability to represent relational roles independently of their arguments and the ability to bind representations of roles dynamically to representations of their arguments—are precisely the same as the requirements for symbolic representation and thought (Hummel, 2010). As noted by Peirce (1879, 1903; see also Chomsky, 1959; Deacon, 1997), a symbolic representation is one that makes it possible to combine and recombine a finite number of elements in order to express an open-ended number of relations. In order to do so, the elements (symbols) must be independent of one another and the symbol system of which they are a part must be capable of specifying how they are bound together in any given expression. For this reason, I shall use the term "symbolic" and "explicitly relational" interchangeably.

Achieving Independence and Role-Filler Binding in a Neural Architecture

In traditional symbolic systems, such as propositional notation and labeled graphs, role-filler independence and binding are provided by the syntax of the system. For example, in propositional notation, the binding of arguments to roles is specified by the order of the arguments in the parentheses, so that *loves* (John, Mary) means "John loves Mary" whereas *loves* (Mary, John) means "Mary loves John". These systems thus satisfy the requirements for symbolic representation *by assumption*. However, as pointed out forcefully by the connectionists in the mid-1980s (Rumelhart et al., 1986), these traditional symbolic approaches to knowledge representation suffer sharp limitations as accounts of knowledge representation in human cognition. Among other problems, they fail to capture the semantic content of the things they represent (e.g., the symbols "John" and "Mary" do not specify what John and Mary have in common or how they differ; Doumas & Hummel, 2005, showed that this limitation is fundamental to the approach and cannot be remedied using the tools of propositional notation or its isomorphs) and it is not at all clear how they could be implemented neurally.

Instead, these and other researchers proposed that human cognition is *sub-symbolic*—that the seemingly symbolic character of human thought is really just an epiphenomenon of the less-than-symbolic passing around of activation in large networks of neurons. To a first approximation, this claim must be correct: The mind is a consequence of the activity of the brain. But to a second approximation, it is the much stronger claim that *the mind is not symbolic*. And indeed, the vast majority of connectionist models in the literature are consistent with this stronger claim, in the sense that they lack the capacity to represent relational roles independently of their arguments and to bind roles to their arguments dynamically (for reviews, see Bowers, 2009; Bowers, Damian & Davis, 2009; Hummel & Holyoak, 2003; Marcus, 1998).

Responses to the non-symbolic nature of traditional connectionist networks have come in three broad flavors. The first, exemplified by Fodor and Pylyshyn (1988), is to acknowledge that connectionist networks are non-symbolic, to argue that human mental representations are symbolic and to conclude that connectionism is therefore fundamentally ill-equipped to account for human cognition, preferring instead traditional symbolic approaches. The problem with this approach is that it is left with all the limitations of traditional symbolic approaches that motivated the connectionists in the first place (Doumas & Hummel, 2005; Hummel & Holyoak, 1997, 2003).

The second broad response to the non-symbolic nature of traditional connectionist networks is to argue that human mental representations are also non-symbolic and conclude that connectionism is therefore fundamentally well-equipped to account for human cognition (e.g., Elman, 1990; McClelland, McNaughton & O'Reilly, 1995; Reynolds & O'Reilly, 2009; O'Reilly, Busby & Soto, 2003; Rogers & McClelland, 2004, 2008; St. John & McClelland, 1990). The limitation of this response is that it fails to account for all those aspects of human cognition that depend on relational/symbolic representations—that is, everything uniquely human (and, arguably, most interesting). This shortcoming is exemplified by the limitations of a recent connectionist model (Leech, Mareschal & Cooper, 2008) of a quintessentially relational process, the making of analogies. In brief, this model "simulates the development of analogical mapping" by failing to accomplish analogical mapping—a failure that is unsurprising given that the model lacks relational representations (see, e.g., the commentaries of Doumas & Richland, 2008; French, 2008; Holyoak & Hummel, 2008; Morrison & Cho, 2008; Petrov, 2008; and Sloutsky, 2008, among others).

The third broad response to the non-symbolic nature of traditional connectionist networks is to attempt to bridge the gap between connectionist and symbolic models, specifying what is needed to render a neural network symbolic. Central to this enterprise is developing a neurally-plausible solution to the role-filler binding problem. Recall that the goal is to bind relational roles to their fillers in a way that allows the representation of a role to remain independent of its filler and vice-versa.

Within this broad approach to representing symbols in a neural architecture are two more specific approaches, one based on varieties of conjunctive coding (including tensor products and variants of them, such as holographic reduced representations, or HRRs) and the other based on varieties of dynamic binding (including both synchrony and asynchrony of neural firing).

Smolensky (1990) was the first to propose tensor products as a solution to the binding problem for neural networks. The basic idea is to represent relations (or relational roles) and their arguments as vectors (call them **r** and **a**, respectively) and to represent bindings of these roles and arguments as tensors, **ra**, where **ra** is simply the outer product of **r** with **a**. That is,

$$\mathbf{ra}_{ij} = \mathbf{r}_i \mathbf{a}_j \tag{1}$$

Eq. (1) represents a tensor product binding a relational role (more accurately, a *predicate*) to a single argument. Bindings of relations to multiple arguments can be represented by higher-rank tensors. For example, the tensor binding relation **r** to arguments **a** and **b**, can be formed as a straightforward generalization of (1):

$$\mathbf{rab}_{ijk} = \mathbf{r}_i \mathbf{a}_j \mathbf{b}_k \tag{2}$$

Halford (1992) argued that the fact that the dimensionality of a tensor (i.e., the number of vector elements in the tensor) scales exponentially with the rank of the tensor (i.e., the number of vectors that get multiplied together to form the tensor) provides a natural account of the capacity limit of WM. The basic idea is that the greater the number of role bindings, the greater the number of neurons required to represent the resulting tensor, so the number of role bindings we can represent (i.e., the rank of the tensor) is necessarily limited.

Holographic reduced representations (Plate, 1991; see also Gayler, 2003; Gayler & Levy, 2009) are a variant of the tensor product coding approach to binding that reduces the dimensionality explosion problem (which Halford viewed as a strength) by summing the resulting tensors over reverse diagonals to reduce them to vectors whose dimensionality is equal to the dimensionality of the original vectors from which the tensors were formed. For example, whereas as an N X N X N tensor would be N^3 dimensional, an HRR computed over this tensor would only be N dimensional.

A limitation of this approach is that information is lost in the dimensionality reduction: Just as information is lost in the mapping from a 3-dimensional world to a 2-dimensional image on the retina (resulting in the unsolvable "inverse optics problem"), so is information lost in the mapping from an N^D-dimensional tensor product to an N-dimensional HRR. (This problem is a special case of the fact that a sum [in this case, an HRR] underdetermines its addends [in this case, the tensor product].) It is for this reason that models based on HRRs typically use sparse vectors in a very high-dimensional vector space: It helps them to avoid mistaking one role-binding for another just by chance.

But by far the most important limitation of this approach is that it is ultimately based on tensor products for role-filler binding, and tensor products violate role-filler independence. As summarized previously, the fact that learning in neural networks is spatial implies that generalization in a neural network depends on the *similarity*—that is, the *vector* similarity—of trained patterns to new ones. And the similarity, i.e., the inner product, of two tensors scales as the product of the inner products of the basis vectors from which the tensors were formed. For example, if **ra** is the tensor product of **r** with **a** and **r'a'** is the tensor product of **r'** with **a'**, then:

$$\mathbf{ra} \bullet \mathbf{r'a'} = (\mathbf{r} \bullet \mathbf{r'})(\mathbf{a} \bullet \mathbf{a'}), \qquad (3)$$

where the • symbol denotes the "dot" or inner product. Equation 3 implies that even if **r** and **r'** are identical, if **a** and **a'** are very different (e.g., with an inner product of zero), then the tensor **ra** will be very different (e.g., with an inner product of zero) from **r'a'**: Bound to sufficiently different arguments, the same role has nothing in common with itself (and conversely, bound to sufficiently different roles, the same argument has nothing in common with itself).

This behavior represents an extreme violation of independence: Returning to Figure 1, it implies that if the vector representing the shape of a lightning bolt has an inner product of zero (i.e., shares no shape features with) the vector representing the shape of a circle (which is not entirely implausible), then the tensor representing *same-shape* (bolt1, bolt2) would have an inner product of zero—i.e., would have no units in common with—the tensor representing *same-shape* (circle1, circle2). The resulting representations would make it impossible to appreciate that the pair of lightning bolts is *in any way* more similar to the pair of circles than is the pair of mismatched shapes. And even though an inner product of zero represents an extreme case, the point does not lose its force if we relax the assumption and allow the inner product to take a small positive value (see Figure 2).

Although the inner product is a particularly relevant measure of vector similarity (since it is the input function for units in the vast majority of connectionist networks), by way of completeness, it is worth noting that the same multiplicative relation holds if we define vector similarity using the cosine rather than the inner product:

$$\cos(\mathbf{ra},\mathbf{r'a'}) = \cos(\mathbf{r},\mathbf{r'})\cos(\mathbf{a},\mathbf{a'}), \qquad (4)$$

where cos(**x**, **y**) is simply the cosine of the angle between vectors **x** and **y**. (Since the cosine of the angle between **x** and **y** is simply **x•y** divided by the product of their lengths, for vectors normalized to have length 1.0, their dot product is equal to the cosine of the angle between them.)

Since HRRs are derived from tensors, a similar, albeit more complex, multiplicative relation also holds between the similarity of two HRRs and the similarities of the basis vectors from which they were formed. These similarity relations are depicted in Figure 2(b). As is clear in the figure, both tensors and HRRs strongly violate role-filler independence, rendering them inadequate as a basis for role-filler binding for the purposes of achieving symbolic neural computation.

Tensor products and HRRs are special cases of *conjunctive coding*, in which each unit codes for a conjunction of a relation (or relational role) and one or more arguments; in the case of tensors and HRRs, the conjunctions are formed by vector multiplication (Eq. 1). Although tensors and HRRs are distributed representations, whereas traditional conjunctive codes tend to be more localist, they are nonetheless conjunctive representations, in the sense that they represent the logical *and* of the entities they bind. Logical *and* is the equivalent of arithmetic multiplication, giving rise to the multiplicative similarity relations expressed in Eq. (3) and (4) and illustrated in Figure 2. That is, although distributed, tensors and HRRs are nonetheless conjunctive.

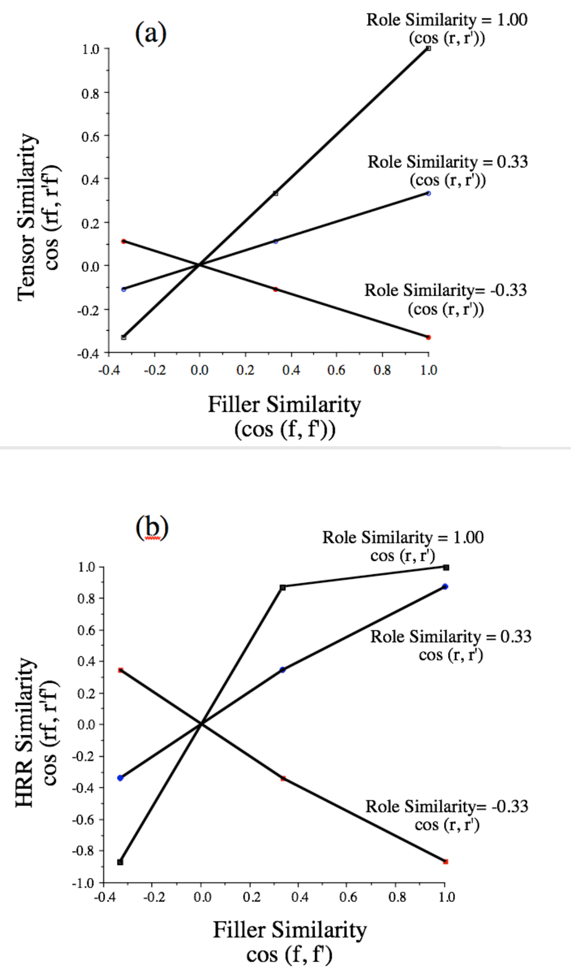

Figure 2. (a) Tensor similarity and (b) HRR similarity as a function of role and filler similarity. In both cases, role and filler similarity interact to determine role+filler binding similarity. That is, under both tensor- and HRR-based binding, roles are not independent of their fillers.

An alternative to binding by vector multiplication is binding by vector addition: Relational roles and objects are represented as activation vectors and a role is bound to an object by adding (rather than multiplying) the corresponding vectors. Binding by synchrony of firing (Hummel & Biederman, 1992; Hummel & Holyoak, 1997, 2003; Shastri & Ajjenagadde, 1993) and by systematic asynchrony of firing (Doumas et al., 2008; Love, 1999) are examples of binding by vector addition.

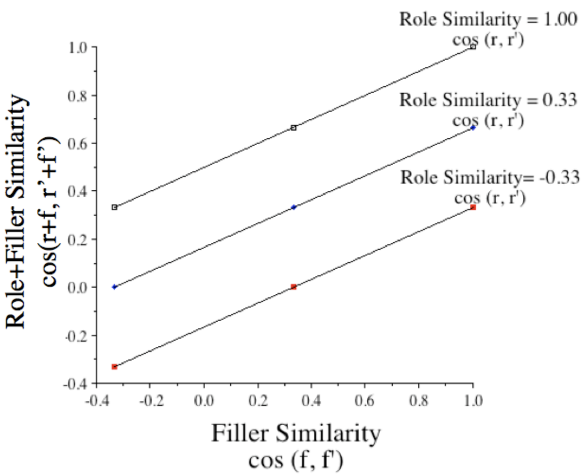

Figure 3. Under additive binding roles and fillers contribute independently (i.e., do not interact) to determine the similarity of bound representations.

In contrast to vector multiplication, which causes role and argument/filler vectors to interact in determining the bound vectors' similarity, vector addition causes roles and fillers to contribute independently to the final similarity of the bound representations (as illustrated in Figure 3). As such, they completely preserve the independence of roles and arguments required for useful relational representations.

One limitation of binding by vector addition is that it is inherently capacity-limited. For example, in the case of binding by synchrony, although there is no necessary limit on the number of neurons that can fire in synchrony with one another, there is a limit on the number of separate groups of neurons that can be active simultaneously while firing mutually *out* of synchrony with one another (Hummel & Biederman, 1992; Hummel & Holyoak, 1997, 2003). This limitation of binding by synchrony (and also systematic asynchrony) of firing provides a very natural account of the limits on human WM (see Hummel & Holyoak, 2003) and visual attention (see Hummel, 2001).

Embedded in the right cognitive architecture, binding by vector addition (i.e., synchrony or asynchrony of neural firing) provides a very natural account of many aspects of human perception and cognition. Keith Holyoak and I showed (Hummel & Holyoak, 1997) that it provides a surprisingly complete account of memory retrieval and analogical mapping (the process of discovering how the elements of one domain correspond to the elements of another for the purposes of reasoning from one to the other).

Later we showed (Hummel & Holyoak, 2003) that the same principles, embodied in the same architecture, account for a great deal of what we know about analogical inference (making inferences from a better-know situation to a less-well-known one) and schema induction (using the two specific situations to learn a more general rule about what situations of that kind have in common). These models account for both the strengths and the limitations of human relational reasoning.

Doumas et al. (2008) later used these same principles to account for a very large number of phenomena in cognitive development (see also Doumas & Hummel, 2010). The same principles have been shown to account for the role of attention and WM in reasoning (Hummel & Holyoak, 2003), the effects of frontotemporal degeneration (Morrison et al., 2004) and normal aging (Viskontas et al., 2004) on relational thinking.

These principles also provide a powerful account of aspects of shape perception and object recognition (Hummel, 2001; Hummel & Biederman, 1992), and have made numerous novel predictions that have been empirically confirmed (Stankiewicz & Hummel, 2002; Stankiewicz, Hummel & Cooper, 1998; Thoma, Davidoff & Hummel, 2007; Thoma, Hummel & Davidoff, 2004; among others; see Thoma & Davidoff, 2007, for a more complete review). In brief, these principles predict that if, and only if, a person attends to visual stimulus will that person generate a relational (i.e., analytic, i.e., symbolic) representation of that stimulus. The findings reviewed in Thoma and Davidoff (2007), which come from numerous laboratories, numerous methodologies and numerous subject populations, all support this prediction.

Binding by vector addition is not just a good idea from first theoretical principles. As it happens, it is also a powerful tool for helping us to understand how our minds actually work.

Conclusion

An analysis from first principles, in terms of the goal of representing role-filler bindings without violating role-filler independence, suggests that binding by vector addition (including but not necessarily limited to binding by synchrony and asynchrony of neural firing) is likely to be an important step toward getting relational representations and processes out of a neural computing architecture. The broad successes of various models based on this principle in accounting for aspects of human perception and cognition serve as evidence of its utility as a principle of neural computation. And although controversial, there is also empirical support for the role of synchrony for binding in the neuroscience literature (e.g., Asaad, Rainer & Miller, 1998; Desmedt & Tomberg, 1994). These factors suggest that vector addition is an essential part of the proper treatment of symbols in a neural architecture.

Acknowledgments

Preparation of this paper was supported by AFOSR grant # FA9550-07-1-0147. I am grateful to Ross Gayler and to an anonymous reviewer for very helpful comments on an earlier draft of this manuscript.

References

Asaad, W. F., Rainer, G., & Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron, 21*, 1399-1407.

Bowers, J.S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review, 116*, 220-251.

Bowers, J. S., Damian, M. F., & Davis, C. J. (2009). A fundamental limitation of the conjunctive codes learned in PDP models of cognition: Comment on Botvinick and Plaut (2006). *Psychological Review, 116*, 986-997.

Chomsky, N. (1959). Review of *Verbal behavior*, by B.F. Skinner. *Language* 35: 26-57.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-114.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York, NY: W. Norton and Company.

Desmedt, J., & Tomberg, C. (1994). Transient phase-locking of 40 Hz electrical oscillations in prefrontal and parietal human cortex reflects the process of conscious somatic perception. *Neuroscience Letters*, *168*, 126-129.

Doumas, L. A. A., & Hummel, J. E. (2005). Approaches to modeling human mental representations: What works, what doesn't and why. In K. J. Holyoak and R. Morrison (Eds.). *The Cambridge handbook of thinking and reasoning*, 73-91. Cambridge, England: Cambridge University Press.

Doumas, L. A. A., & Hummel, J. E. (2010). A computational account of the development of the generalization of shape information. *Cognitive Science*, *34*, 698-712

Doumas. L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1 – 43.

Doumas, L. A. A. & Richland, L. E. (2008). Developing structured representations. *Behavioral and Brain Sciences, 31*, 384-385.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

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

French, R. M. (2008). A new manifesto for child development research. *Behavioral and Brain Sciences*, *31*, 339.

Gayler, R. (2003). Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. Paper presented at the Joint International Conference on Cognitive Science, 13-17, University of New South Wales, Sydney, Australia.

Gayler, R.W. and S.D. Levy (2009). A Distributed Basis for Analogical Mapping. *Proceedings of the Second International Analogy* C*onference*. NBU Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155-170.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, *23*, 222-262.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development*, *35*, 193-217.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holyoak, K. J., & Hummel, J. E. (2008). No way to start a space program. *Behavioral and Brain Sciences*, *31*, 388-389.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in human shape perception. In E. Deitrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 157-185). Mahwah, NJ: Erlbaum.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, *8*, 489 – 517.

Hummel, J. E. (2010). Symbolic vs. associative learning. *Cognitive Science*, *34*, 958-965.

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

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427-466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220-264.

Leech, R., Mareschal, D. & Cooper, R. (2008) Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences, 31,* 357-414.

Love, B. C. (1999). Utilizing time: Asynchronous binding. *Advances in Neural Information Processing Systems*, *11*, 38-44.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*, 243-282.

McClelland, J.L., McNaughton, B.L., & O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

Morrison, R.G., & Cho, S. (2008). Neurocognitive process constraints on analogy: What changes to allow children to reason like adults? *Behavioral and Brain Sciences*, *31,* 391-392.

Morrison, R.G., Krawczyk, D.C., Holyoak, K.J., Hummel, J.E., Chow, T.W., Miller, B.L., & Knowlton. (2004). A neurocomputational model of analogical reasoning and its breakdown in Frontotemporal Lobar Degeneration. *Journal of Cognitive Neuroscience*, *16*, 260-271.

Oden, D. L., Thompson, R. K. R., & Premack, D. (2001). Can an ape reason analogically? Comprehension and production of analogical problems by Sarah, a chimpanzee (*Pan troglodytes*). In D. Gentner, K. J. Holyoak & B. N. Kokinov (Eds.), The analogical mind: Perspectives from cognitive science (pp. 471-497). Cambridge, MA: MIT Press.

O'Reilly, R.C., Busby, R. S., & Soto, R. (2003). Three forms of binding and their neural substrates: Alternatives to temporal synchrony. In A. Cleeremans (Ed) *The Unity of Consciousness: Binding, Integration, and Dissociation*, 168-192, Oxford: Oxford University Press.

Palmer, S. E. (1978). Structural aspects of similarity. *Memory and Cognition*, *6*, 91-97.

Peirce, C. S. (1879, 1903). Logic as semiotic: The theory of signs. In J. Buchler (Ed.), The philosophical writings of Peirce (1955) (pp. 98–119). New York: Dover Books.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Brain and Behavioral Sciences*, *31*, 109-178.

Petrov, A. A. 2008. Relational priming plays a supporting but not leading role in adult analogy-making. *Behavioral and Brain Science, 31*, 392-393.

Plate, T. (1991). Holographic reduced representations: Convolution algebra for compositional distributed representations. In J. Mylopoulos and R. Reiter (Eds.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (pp. 30-35). San Mateo: Morgan Kaufmann.

Reynolds, J.R. & O'Reilly, R.C. (2009). Developing PFC representations using reinforcement learning. *Cognition, 113,* 281-292.

Robin, N., & Holyoak, K. J. (1995). Relational complexity and the functions of prefrontal cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 987-997). Cambridge, MA: MIT Press.

Rogers, T. T. & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cam-bridge, MA: MIT Press.

Rogers, T. T. & McClelland, J. L. (2008). Precis of *Semantic Cognition, a Parallel Distributed Processing Approach. Behavioral and Brain Sciences, 31,* 689-749.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1. Cambridge, MA: MIT Press.

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

Sloutsky, V. M. (2008). Analogy is to priming as relations are to transformations. *Behavioral and Brain Sciences, 31,* 396-397.

Smith, L. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadov and A. Ortney (Eds.) Similarity and analogical reasoning (pp. 147-177). Cambridge: Cambridge University Press.

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

Stankiewicz, B.J. & Hummel, J.E. (2002) The role of attention in scale- and translation-invariant object recognition. *Visual Cognition*, *9*, 719-739.

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 732-744.

St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217-257.

Stuss, D., & Benson, D. (1986). *The frontal lobes*. New York: Raven Press.

Taylor, E. G., & Hummel, J. E. (2009). Finding similarity in a model of relational reasoning. *Cognitive Systems Research*, *10*, 229-239.

Thoma, V., & Davidoff, J. (2007). Object recognition: attention and dual routes. In N. Osaka, I. Rentschler & I. Biederman (Eds.), *Object recognition, attention and action*. Tokyo: Springer.

Thoma, V., Davidoff, J., & Hummel, J. E. (2007). Priming of plane-rotated objects depends on attention and view familiarity. *Visual Cognition*, *15*, 179-210.

Thoma, V., Hummel, J. E., & Davidoff, J. (2004). Evidence for holistic representations of ignored images and analytic representations of attended images. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 257-267.

Viskontas, I., Morrison, R., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004) Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging*, *19*, 581 – 591.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*, 119-125.