CHAPTER

5

Computational Models of Higher Cognition

Leonidas A. A. Doumas *and* John E. Hummel

Abstract

Process models of higher cognition come in three basic varieties: traditional symbolic models, traditional connectionist models, and symbolic-connectionist models. This chapter reviews the basic representational and processing assumptions embodied in each of these approaches and considers the strengths and limitations of each.

Key Words: computational models, process models, symbolic models, connectionist models

Models in cognitive science span all three of Marr's (1982) levels (see Holyoak & Morrison, Chapter 1). Normative systems (see Chater & Oaksford, Chapter 2), including the Bayesian framework (see Griffiths et al., Chapter 3), address the level of computational theory; neural models (see Morrison & Knowlton, Chapter 6) are specified at the implementation level; and information-processing models, or process models (perhaps the most common type of model in cognitive science), are specified at the level of representation and algorithm.

Process models of higher cognition come in three basic varieties: traditional symbolic models, traditional connectionist models, and symbolic-connectionist models. This chapter reviews the basic representational and processing assumptions embodied in each of these approaches and considers the strengths and limitations of each.

Symbolic Models

The earliest process models of human cognition were *symbolic* in the traditional sense of the term (e.g., Anderson, 1982; Newell & Simon, 1972), and traditional symbolic modeling continues to be important today (see Gentner & Forbus, 2011; Taatgen & Anderson, 2008, for reviews). Although the details of specific symbolic models differ, at their core they share the underlying assumption that the mind is a symbol system that is best modeled using symbolic operations on symbolic data structures (see also Fodor & Pylyshyn, 1988).

Symbolic Representations

Any representational system consists minimally of a vocabulary of representational elements (e.g., symbols in a symbolic model or nodes in a neural or Bayesian network) and a set of rules for inferring new statements from existing statements (see Markman, Chapter 4). In order for a representational system to count as symbolic, it must also make it possible to combine its basic representational elements into complex structures capable of expressing an open-ended set of relations (Pierce, 1879, 1903; see also Deacon, 1997).

Traditional symbolic models use various representational formalisms, the most common being propositional notation (or labeled graphs). Propositional notation takes the general form *predicate* (argument$_1$, argument$_2$,... argument$_n$), where *predicate* specifies some property or relation, arguments $1 \dots n$ are the arguments of that predicate,

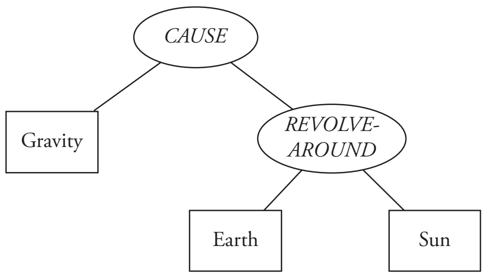

Fig. 5.1 A labeled-graph representation of the higher order relation, *cause*(gravity, *revolve-around*(earth, sun)). Ovals represent relations; rectangles, objects; lines, arcs.

and *n* is typically three or less. For example, *gave* (John, Mary, book) species that John gave Mary a book, and *heavy* (book) specifies that the book is heavy. Labeled graphs specify the same information as propositional notation (i.e., the two systems are isomorphic). In graphical form, nodes represent predicates and their arguments and arcs represent the bindings of arguments to roles of the predicate (see Fig. 5.1). Both formalisms are symbolic in the sense described earlier because they make it possible to form an open-ended (indeed, infinite, as both formalisms permit recursion) set of relational statements with a finite vocabulary of predicates and objects.

Processes

Symbolic representations such as propositions and labeled graphs provide a powerful representational platform that makes many kinds of processes convenient to perform. For example, Forbus, Gentner, and their colleagues (Falkenhainer, Forbus, & Gentner, 1989; Forbus, Gentner, & Law, 1995) have demonstrated that *graph matching*—a process of finding isomorphic substructures in pairs of (potentially very large) systems of labeled graphs—provides an excellent basis for simulating analogical reasoning (i.e., the process of reasoning about a novel *target* domain based on a more familiar *source* domain; see Holyoak, Chapter 13).

Similarly, John Anderson and his colleagues have used propositional notation in their various ACT models (e.g., Anderson, 2007; Anderson & Lebiere, 1998) to very successfully simulate aspects of memory, learning, and inference. The principles embodied in ACT have even been used to develop intelligent tutoring systems that model the learner during the learning process itself, in order to optimize instruction and learning (Anderson, Betts, Ferris, & Fincham, 2010).

In contrast to the graph matching algorithms that Gentner, Forbus, and colleagues have used to model analogical reasoning, the ACT models are based on *production systems*: systems of symbolic rules that operate on propositional knowledge representations to guide action and generate inferences (see Fig. 5.2). Like the graph-matching algorithms of Forbus, Gentner, and colleagues, production systems derive much of their computational power from the fact that they operate on symbolic knowledge structures. For example (see Fig. 5.2), one rule that might be part of a production system is *if* (*larger* (*x*, *y*) *and larger* (*y*, *z*)) *then larger* (*x*, *z*). Because this rule is symbolic, it can automatically be applied to any *x*, *y*, or *z*, regardless of their semantic content (e.g., it could infer that a battleship is larger than a submarine from the fact that a battleship is larger than a cruiser and a cruiser is larger than a submarine, and with equal facility make the same inference in the context of a Rottweiler, a housecat, and a mouse). As trivial as this ability might appear, it is a powerful one that cannot not be taken for granted (as we shall see in the context of connectionist models in the next section).

In our previous example, the production rule was completely abstract, defined over empty variables, *x*, *y*, and *z*. But production rules can also be defined over specific arguments (as in, *if* (*see* (me, neighbor's dog), *then run*(me)). Moreover, the rules need not be accurate. For example, a production system used to model the behavior of a young child might include a rule like *if* (*moves*(*x*), *then* (*has-legs*(*x*)), reflecting

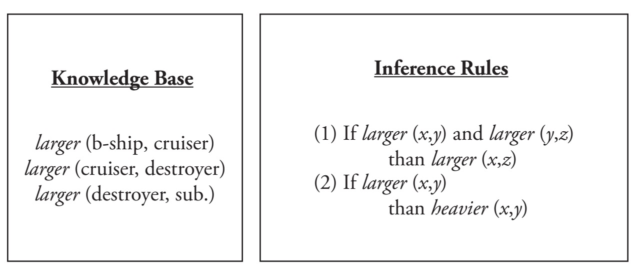

Fig. 5.2 An example of a simple production system.

the child's inaccurate belief that all moving things have legs (e.g., Sheya & Smith, 2006).

Strengths

Symbol systems have been used to successfully model a wide range of cognitive phenomena. The successes of the symbolic approach derive from the flexibility of symbolic representations, in particular the fact that predicates are free to vary in their arguments. As a result, anything that is learned about a predicate (e.g., in the form of a production rule) can, in the limit, be automatically applied to any new argument(s) taken by that predicate.[1] This kind of open-ended generalization—which permits extrapolation beyond the examples from which the rule is learned—is a very powerful inductive mechanism (Holland, Holyoak, Nisbett, & Thagard, 1986; Hummel & Holyoak, 2003). Without it, we would be limited to only those inferences and generalizations that can be found by interpolating over the training examples (Marcus, 2001). That is, without the ability to learn variablized rules (i.e., rules defined over predicates that are free to vary in their arguments), having learned that relative size is a transitive relation (i.e., the "*if larger*..." rule) in the context of the naval vessels from the example above, we would be at a complete loss to infer *larger* (Rottweiler, mouse) given *larger* (Rottweiler, housecat) and *larger* (housecat, mouse) (see, e.g., Doumas, Hummel, & Sandhofer, 2008; Holyoak & Hummel, 2000; Hummel & Holyoak, 1997, 2003). Penn, Holyoak, and Povinelli (2008; see Penn & Povinelli, Chapter 27) have argued convincingly that this kind of relational generalization is the most important ability distinguishing human cognition from the cognitive abilities of our closest primate cousins.

It is this same capacity for variablized, relational thinking that allows us to make analogies and to learn and reason from abstract schemas (Holland et al., 1986; Holyoak & Hummel, 2001; Hummel & Holyoak, 2003). Gick and Holyoak (1983) demonstrated that when people draw analogies between similar stories (i.e., specifying which elements of one correspond to which of the other), they may induce generalized schemas describing the shared aspects of those stories. For example, given one story in which Mary is the enemy of Bill and Bill is the enemy of Ted, so Mary regards Ted as her friend, and another story in which Dirk is the enemy of Roger and Roger is the enemy of Jake, so Dirk regards Jake as his friend, one might infer a schema of the general form, if *enemy-of* (person-1, person-2) and *enemy-of* (person-2, person-3), then *friend-of* (person-3, person-1). If one were then to come across a situation in which Joe was the enemy of Tim, and Tim the enemy of Bill, one could map Joe onto person-1, Tim onto person-2, and Bill onto person-3, and subsequently infer that Bill is a friend of Joe. Notice that this sort of schema-driven inference is formally equivalent to inference based on a production rule in which the antecedent condition (the *if* portion of the rule) is fulfilled and the consequent (the *then* portion of the rule) fires (Anderson & Lebierre, 1998; Newell, 1990).

In the limit, the capacity to represent relational rules (or schemas) makes it possible to represent and reason about universally quantified functions (see Marcus, 2001). For example, consider the rule "$\forall x,y,z$ *if* (*pass* (x, y), *collect* (x,z))" (or, for all x, if x passes y, then x collects z). The rule can apply to any set of x, y, and z, such as a player (x) passing the "GO" square (y) in Monopoly and collecting $200 ($z$), or a person ($x$) passing the border of a country (y) and collecting a passport stamp (z). Without the power of symbolic representations—specifically, without the ability to represent predicates that are free to vary in their arguments—this kind of flexibility would be impossible.

Weaknesses

Despite their successes, a number of criticisms have been leveled against traditional symbolic models of cognition. A very basic one concerns the question of how (and even whether) it is possible to learn symbolic representations. How, for instance, would we learn a relational predicate like *larger*(x, y)? This question is made difficult, in part, by the very property that makes these representations powerful, namely, the fact that they are free to vary in their arguments: Although *larger* (battleship, cruiser) and *larger* (Rottweiler, housecat) express different ideas, they nonetheless express, if not identical, then at least very similar relations. It is our appreciation of this similarity that allows us to grasp that the battleship corresponds to the Rottweiler rather than the housecat.[2] In turn, the predicate's ability to remain (at least largely) invariant over changes in its arguments renders it challenging to learn: We are never exposed to disembodied "*larger*-ness"; rather, all those cases where we have had the opportunity to observe an instance of the *larger* relation have presented it in the context of one specific thing being larger than some other specific thing. Given this type of input, how do we ever learn to represent *larger* in a way that is (even partially) independent of its arguments (Doumas et al., 2008; Kellman, Burke, & Hummel, 1999)? The difficulty of this question—along with the fact that rather than answering it, traditional symbolic modelers have often simply hand-coded their symbolic representations—has led some to wonder whether it is even possible to learn symbolic representations, such as predicates, from nonsymbolic inputs (e.g., early perceptual representations), a concern that has been cited as one of the most significant shortcomings of the symbolic approach (see, e.g., Leech, Mareschal, & Cooper, 2008; Munakata & O'Reilly, 2003; O'Reilly & Busby, 2002; O'Reilly, Busby, & Soto, 2003). Certainly, with tools of traditional symbolic models it is unclear how these representations can be learned in the first place. However, as elaborated later this shortcoming does not mean that symbolic representations cannot be learned at all (e.g., Doumas et al., 2008).

A second limitation of the traditional symbolic approach, as an account of the human cognitive architecture, is that human mental representations have semantic content: They are *about* things, and they somehow naturally capture how those things are similar to and different from one another. By contrast, traditional symbolic approaches to cognition do not (and, worse, *cannot*; Doumas & Hummel, 2004) capture similarity relations among the entities to which symbols refer. For example, the symbols *Rottweiler* and *battleship* fail to specify what these concepts have in common (precious little besides being objects found on Earth) and how they differ (e.g., in animacy, size, and function). Based on the symbols alone, one's best guess about what battleships and Rottweilers have in common is that both require 10 letters to spell (11 in the plural). Moreover, the meanings of various relations seem to apply specifically to individual relational *roles*, rather than to the relation as a whole. As a result, it is easy to appreciate that the agent (i.e., killer) role of *murder* (*x*, *y*) is similar to the agent role of *attempted-murder* (*x*, *y*), even though the patient roles differ (i.e., the patient is dead in the former case but not the latter); and the patient role of *murder* (*x*, *y*) is similar to the patient role of *manslaughter* (*x*, *y*), even though the agent roles differ (i.e., the act is intentional in the former case but not the latter).

The semantic properties of human mental representations manifest themselves in countless ways in human cognition, influencing memory retrieval (e.g., Gentner, Ratterman, & Forbus, 1993; Ross, 1987; Wharton, Holyoak, & Lange, 1996), categorization, and reasoning (Bassok, Wu, & Olseth, 1995 Krawczyk, Holyoak, & Hummel, 2005; Kubose, Holyoak, & Hummel, 2002; Ross, 1987). The meanings of relations and their arguments also influence which inferences seem plausible from a given collection of stated facts. For instance, upon learning about a culture in which nephews traditionally give their aunts a gift on a particular day of the year, it is a reasonable conjecture that there may also be a day on which nieces in this culture give their uncles gifts. This inference is based on the *semantic* similarity of aunts to uncles and nieces to nephews, and on the semantics of gift giving, not the syntactic properties of the *give-gift* relation. Given the important role of semantics in the mental representation of relational roles and the objects that fill those roles, an important criterion for a general account for the human cognitive architecture is that the representations on which it is based be able to capture (or at least approximate) that semantic content.

On this point, traditional symbolic models based on varieties of propositional notation and labeled graphs fare poorly. It has been known for a long time that such representational schemes have difficulty capturing shades of meaning and other subtleties associated with semantic content. This limitation was a central focus of the influential critiques of symbolic modeling presented by the connectionists in the mid-1980s (e.g., Rumelhart, McClelland, & The PDP Research Group, 1986). A review of how traditional symbolic models have handled this problem (typically with look-up tables of one sort or another) also reveals that the question of semantics is, in the very least, a thorny inconvenience (see Doumas & Hummel, 2004, for an argument that the problem is more than simply an inconvenience).

Yet a third limitation of traditional symbolic approaches, also cited by the connectionists in the mid-1980s, is that they make no obvious contact with Marr's (1982) third level of analysis, *physical implementation*: It is not at all clear how something like a graph-matcher or a production system would or could be implemented in the brain (but see Anderson, Qin, Jung, & Carter, 2007, for some progress in this direction).

Connectionist Models

Connectionist neural-network models (also referred to as *parallel distributed processing*, or PDP) were motivated in large part by the perceived limitations of the traditional symbolic approach. Neurally inspired

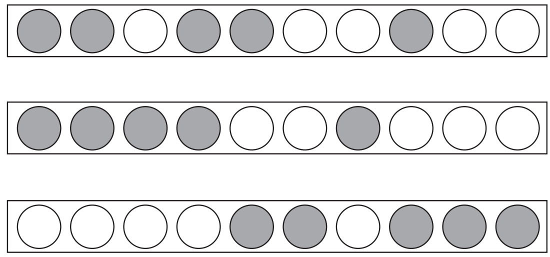

Fig. 5.3 Examples of distributed representations of the concepts, (*a*) dog, (*b*) cat, and (*c*) shark. Note that the same set of units is depicted in each row.

models date back to at least the 1940s (McCulloch & Pitts, 1943 Rosenblatt, 1958). However, their more recent appeal as a general account of the human cognitive architecture—and as a serious alternative to traditional symbolic models—was launched by the work of Rumelhart and colleagues in the mid-1980s (see McClelland et al., 1986; Rumelhart et al., 1986).

Like symbolic models, a variety of connectionist models have been proposed to simulate a wide range of cognitive phenomena; also like symbolic models, the diverse models in the traditional connectionist approach share some basic assumptions about how information is represented and processed.

Representations

Connectionist models consist of collections of simple processors, represented by nodes or *units*, that are connected to form large networks. Units in connectionist networks take activations, typically in the range 0–1. Representations are patterns of activation across *units* in the system. These patterns might correspond to a perception, a thought, a memory, a concept, or any other cognitive state. For example, in the very simple network depicted in Figure 5.3, the pattern of activation on the units depicted in Figure 5.3a might correspond to the concept "dog," the pattern of activation depicted in Figure 5.3b might correspond to "cat," and the pattern of activation depicted in Figure 5.3c to "shark." Other concepts like "animal," "large," or "food" would be represented as other patterns.

An important distinction in connectionist models is the distinction between *localist* and *distributed* representations. A localist representation is one in which individual units have meaning (e.g., a unit for "dog"); a distributed representation is one in which the meaning of a concept is carried by a pattern of activation across many separate units (e.g., the concept dog being represented by units for "animal," "mammal," "canine," "domesticated," etc.). As illustrated by this example, whether a representation is localist or distributed is most often a function of the relationship between the concept and the representation: With respect to the concept "dog," units representing "animal," "mammal," and so on constitute a distributed representation; but with respect to each of the more general concepts "animal," "mammal," and so on, those same units constitute a localist representation. Thus, although the terms "localist" and "distributed" are most commonly used to describe representations (without regard for the entities they represent), they are better thought of as two-place predicates of the form: representation *R* is localist (or distributed) with respect to concept *C* (i.e., *distributed* (*R*, *C*)). The one exception to this generalization of which we are aware is some vector-symbolic architectures (such as those based on holographic reduced representations; e.g., Plate, 1991), in which the meaning of a unit in any pattern of activation depends entirely on the activity of the other units in that pattern. Such representational schemes are entirely distributed in the sense that no unit in the pattern can be interpreted without reference to the others.

One very appealing aspect of distributed representations (although not those that are entirely distributed) is that they very naturally capture the similarities of different concepts. In our simple network, the concepts of "dog" and "cat" are similar to the extent that their representations overlap. Thus, the network naturally captures the fact that dogs are more similar to cats than to sharks as a natural consequence of its representational scheme. Although there is debate about the utility of distributed and localist codes in connectionist systems (see, e.g., Page, 2000), most connectionist models use some combination of the two (e.g., O'Reilly & Busby, 2002; Rogers & McClelland, 2004).

Processes

Units in a connectionist system are densely interconnected via weighted *connections*. Positive weights act as excitatory connections, so that activity on one unit tends to activate the other, and negative weights act as inhibitory connections, so that activity in one unit tends to reduce activity in the other.

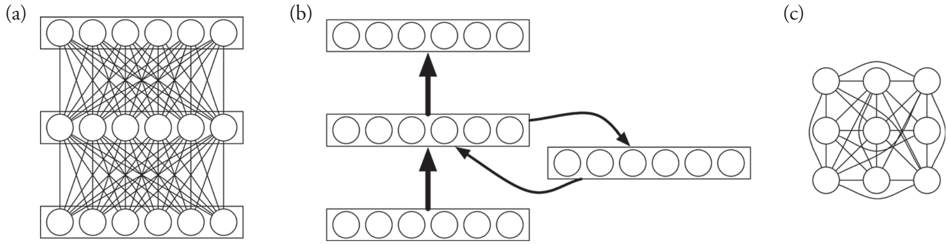

Fig. 5.4 Examples of (*a*) feed-forward, (*b*) recurrent, and (*c*) auto-associative connectionist systems. See text for details.

The architecture of a connectionist network is defined by the manner in which units are connected to one another. The most common architectures are feed-forward (e.g., McClelland & Rummelhart, 1985), recurrent (Elman, 1990), auto-associative (Hopfield, 1982), and hybrids thereof (e.g., O'Reilly, 2006). In a feed-forward architecture, units are arranged in layers such that the units in one layer pass activation (both excitation and inhibition) to the units in the next layer, but units do not typically pass activation to other units in the same layer or in earlier layers (see Fig. 5.4a). Recurrent networks (Fig. 5.4b) are much like feed-forward networks, except that the activation pattern on a "hidden" layer at time t serves as part of the input at time $t + 1$, thereby providing a kind of memory for past states of the network (see, e.g., Elman, 1990). In an auto-associative architecture (Fig. 5.4c) every unit typically has connections to every other unit.

Despite their differences, connectionist networks (at least traditional, nonsymbolic ones as opposed to symbolic ones, as described later) all share the property that the entire currency of computation is activation: Units excite or inhibit one another and over time the state of the network settles into some stable pattern of activation, either on the output units (in the case of feed-forward and recurrent networks) or over the network as a whole (in the case of auto-associative networks). The final or output pattern of activation is interpreted as the network's response to its input.

Strengths

The strengths of the traditional connectionist approach are numerous. Perhaps most obviously, connectionist networks offer a link between cognitive phenomena and their potential neural underpinnings: It is easy to see how the computational principles embodied in a connectionist network could be realized in neurophysiology. In addition, connectionist networks, with their representations based on distributed patterns of activation, provide for a natural kind of automatic generalization. If the representation of, say, *dog*, consists of one pattern of activation (e.g., on the input units of some network) and the representation of *cat* consists of a similar pattern of activation, then many things learned about dogs will generalize automatically to cats (see McClelland et al., 1986). This kind of automatic generalization accounts for the semantic richness of human mental representations in a way that traditional symbolic representations simply cannot (Doumas & Hummel, 2005).

A third strength of the traditional connectionist approach is that, in contrast to the traditional symbolic approach, it provides a way to seamlessly integrate questions of representation with questions of learning: Connectionist networks are capable of learning their own representations for things, both in the "hidden" layers of feed-forward and recurrent networks, and in classes of unsupervised learning models (which represent a hybrid of feed-forward and auto-associative architectures; see, e.g., Marshall, 1995).

Still a fourth strength of the connectionist approach is that, like biological neural networks, connectionist networks degrade gracefully with damage. As such, they provide a natural platform for simulating the effects of brain damage, normal aging, and even cognitive development (e.g., Colunga & Smith, 2005; Joanisse & Seidenberg, 1999; Li, Lindenberger, & Sikstrom, 2001).

A fifth, less often cited, strength of connectionist models is their mathematical simplicity. In contrast to symbolic models, which are typically complex enough to defy proofs of their computing abilities, connectionist networks are comparatively simple, typically being constrained to using a fixed activation function in a fixed architecture. This makes it possible to prove, for example, that a three-layer (that is, three layers of connections, or a layer of input units, two hidden layers of units, and a layer of output units) feed-forward nonlinear network (where "nonlinear" refers to the activation function of the units in the network) is capable of computing any computable mapping.

Weaknesses

Despite their strengths, traditional connectionist models have some important limitations as models of higher cognition in humans. Most notably, they lack symbolic competence. Many authors have written extensively on this limitation (e.g., Doumas et al., 2008; Hummel, 2000; Hummel & Biederman, 1992; Hummel & Holyoak, 1997, 2003; Marcus, 1998, 2001; von der Malsburg, 1981, 1999), but its importance remains underappreciated.

Although a connectionist network can compute any computable mapping, it effectively must be trained on each individually. Although a properly trained NN can interpolate between learned input-output mapping, it cannot extrapolate to mappings that lie outside of its training set (e.g., Marcus, 1998; St. John, 1992; St. John & McClelland, 1990). Symbolic systems by contrast including humans extrapolate easily (e.g., as in the case of human relational reasoning; Hummel & Holyoak, 2003).

The fundamental reason why connectionist models fail to achieve symbolic competence is that traditional connectionist representations simply do not have enough degrees of freedom—they is, they are formally too weak (Hummel et al., 2004). A symbolic representation, such as propositional notation, has two degrees of freedom with which to expresses information. The first is the choice of which symbols to use: If a modeler wishes to represent that John loves Mary, she would use the symbols *loves*, John and Mary; to represent that John hates Mary, she would use *hates*, John and Mary. The corresponding degree of freedom in a connectionist representation is activation: To represent that John loves Mary, a connectionist network might activate units (or, in the distributed case, collections of units) representing *loves*, John and Mary.

The second degree of freedom in a symbolic representation specifies the bindings of arguments to the roles of the relation: To represent that John loves Mary, it is traditional to place John in the first slot inside the parentheses following the relation and Mary in the second slot, forming *loves* (John, Mary); to specify that Mary loves John, the modeler would place the *very same symbols* in the opposite slots to form *loves* (Mary, John). (An explanation for the italics appears two paragraphs later.) There is no analogous second degree of freedom in a traditional connectionist representation. Activating units or patterns for *loves*, John and Mary, could equally represent "John loves Mary" or "Mary loves John" (or, in the case of a distributed representation, a statement about a narcissistic hermaphrodite; Hummel & Holyoak, 1997; von der Malsburg, 1981), so it is impossible to tell which is the intended meaning of the representation. The reason is that this representation fails to specify the bindings of the arguments to the roles of the relation. Unlike the symbolic representation, the traditional connectionist representation lacks a degree of freedom with which to specify this information.

The most common response to this problem is to use varieties of *conjunctive coding* to carry binding information. Under conjunctive coding, units represent, not objects or relational roles, but *conjunctions* of objects in relational roles. For example, rather than representing *loves*, John and Mary, the units in a conjunctive code might represent conjunctions such as John+*lover*, Mary+*beloved*, Mary+*lover*, and John+*beloved*. Now, to represent that John loves Mary, the network activates John+*lover* and Mary+*beloved*; to represent that Mary loves John, it activates Mary+*beloved* and John+*lover*. Specific varieties of conjunctive coding include tensor products (Halford, Wilson, & Phillips, 1998; Smolensky, 1990), holographic reduced representations (Plate, 1991), and spatter codes (Kanerva, 1998). These approaches vary in their particulars, but they all share the property that units represent specific conjunctions of roles and arguments rather than representing the roles or arguments individually.

Models based on conjunctive coding have been applied with varying degrees of success in various domains. And indeed, some kind of conjunctive coding is certainly necessary for encoding bindings in long-term memory (Doumas et al., 2008; Hummel & Biederman, 1992; Hummel & Holyoak, 1997, 2003; Shastri, 2002). But as a general, or the only, solution to the binding problem in connectionist or neural networks, conjunctive coding is sharply limited. Recall the italicized *very same symbols* wording from two paragraphs earlier. It is the fact that the symbols in *loves* (John, Mary) are the same as those in *loves* (Mary, John) that allows you to know what these statements have in common: Both are about loving, John and Mary. The same is true for the statements *jonks* (grummond, steplock) and *jonks* (steplock, grummond): Although these statements are largely meaningless, because they use and reuse the same symbols, in the very least we know that, whatever "jonking" is, the grummond is doing it (or else stands in that relation) to the steplock in one case and the steplock is doing it (or stands in that relation) to the grummond in the other.

This ability to preserve the identity of symbols across different relational structures is a characteristic of symbolic representations, made possible by the fact that they have two degrees of freedom, that is absolutely essential for symbolic thought—such as reasoning based on analogies, schemas, or rules (Holyoak & Hummel, 2002; Hummel & Holyoak, 1997, 2003; Penn et al., 2008). It is for this reason that there are no successful traditional connectionist models of analogy (see Leech, Mareschal, & Cooper, 2008, for an attempt, the shortcomings of which illustrate the need for that second degree of freedom).

A number of connectionist models have been developed that appear to solve the problem of relational processing. We argue that these models in fact fail to simulate human relational reasoning and instead simulate a proxy to relational reasoning.

A connectionist model developed by O'Reilly and Busby (2002) illustrates what it is possible (and what it is not) without the second degree of representational freedom enjoyed by symbolic systems. O'Reilly and Busby's model is designed to answer questions about the spatial relations among objects. The model consists of "input/output" units representing *(a)* object features at each of 4x4 locations in the visual field, *(b)* the locations of those objects, *(c)* the objects' identities (independent of location), and *(d)* relations among the objects (e.g., above/below). There are also "query units" associated with the input/output units (used for querying the model after training). These units are connected to (and communicate via) a set of "hidden" units, forming a large auto-associative network. The model is trained by pairing patterns of activation across the input/output units and learning connections between those units and the hidden units using the Leabra learning algorithm (O'Reilly, 1996). After training, the model can be presented with objects in various locations and be asked questions about them. For example, in order to ask, "What is at location 3,2?" the location query unit would be activated, along with the unit for location 3,2. No units in the "object" or "relation" arrays would be activated. The model's task would be to activate the object units corresponding to the distributed representation of whatever object resides at location 3,2. Or if asked, "What is above?" the "above" query unit would be activated, and the model's task would be to activate the representation of the object that is above the other, i.e., that in location 2,2 (see Fig. 5.5).

O'Reilly and Busby (2002) trained the model on various subsets (from 1.3% to 25%) of the input/

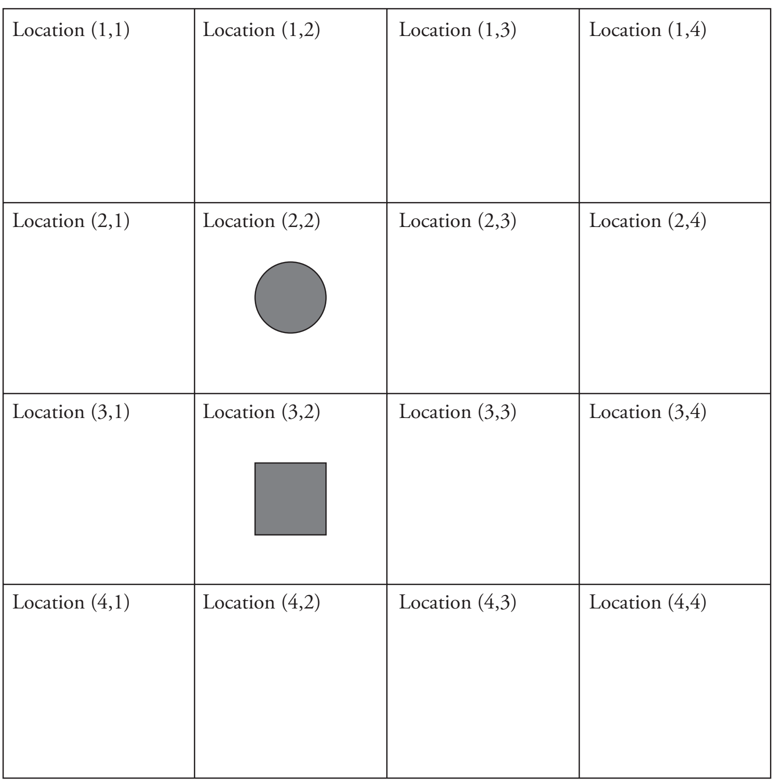

Fig. 5.5 Example of the task simulated by the O'Reilly and Busby (2002) model. A circle is presented in location 2,2, and a square in location 3,2. The model might be asked, "What is above location, 3,2" (which should activate location 2,3).

query conjunctions it is capable of representing, and tested it for its ability to generalize to the untrained input/query conjunctions. After training on 25% of its input space, the model successfully generalized to roughly 95% of the untrained inputs. Based on this performance, O'Reilly and Busby concluded that "... rich distributed representations containing coarse-coded conjunctive encodings can effectively perform binding" (p. 6).

However, this claim, and the simulations on which it is based, deeply underestimate the power of relational perception, thought, and generalization. Being able to answer "What is above?" is not the same as being able to represent "*x* is above *y*," for all *x* and all *y*, and being able to draw inferences based on the latter—for example, if *x* is above (and on) *y*, then *x* could potentially fall off of *y*. Far from this, the O'Reilly and Busby (2002) model can simply say "what is above": it can answer "*x* is above" (provided the features of *x* were part of its training regime), but it cannot say *what x* is above: It lacks even the basic capacity to specify the second argument of the *above* relation.

What makes human relational perception and thought powerful (and difficult to model) is not our ability to answer simple questions of the form "What is above?" (a question the O'Reilly & Busby, 2002, model answers based strictly on associative learning), but our ability to represent relations such as *above* (*x*, *y*), explicitly, to bind arguments to the roles of those relations, and to use that knowledge to make inferences about *x* and *y* for all *x* and *y* (not just those whose features have appeared in our training space). The O'Reilly and Busby model, by contrast, can only answer questions whose answers it has already been taught. Like other traditional connectionist models, the O'Reilly and Busby model can only generalize to new patterns by interpolating among patterns on which the model has been explicitly trained (e.g., patterns that are linear combinations of the examples on which the model has already been trained). And it cannot even represent, much less answer, questions of the form, "If *x* is in top of *y*, can *x* fall off of *y* or can *y* fall off of *x*?," a question even a 3-year-old child would look at you askance for asking her in the first place.

This limitation is not restricted to the O'Reilly and Busby (2002) model, but is true of all models that have tried to tackle symbolic processing using strictly associationist tools (e.g., Rogers & McClelland, 2004, 2008; St. John & McClelland, 1990). As noble as these efforts are, their tools simply are not up to the task.

Symbolic-Connectionist Models

In response to the complementary strengths and weaknesses of the traditional symbolic and connectionist approaches, some researchers have attempted to implement symbolic structures within connectionist architectures with distributed representations. In principle, achieving symbolic competence in a connectionist system should not be difficult: All that is needed is some basis for representing role-filler bindings in a way that allows the representation of the roles and fillers to remain invariant (i.e., to be "reused" as they are in symbolic systems) across different bindings. Proposed solutions to this problem come in two general forms.

Models Based on Vector Multiplication

One approach to implementing symbolic structure in connectionist systems is to use tensor products (Smolensky, 1990; see also Halford et al., 1994; Halford et al., 1998)—or their variants, such as holographic reduced representations (HRRs; Plate, 1991), spatter codes (Kanerva, 1998), or circular convolutions (Metcalfe, 1990)—to represent role-filler bindings. A tensor product is an outer product of two or more vectors (i.e., a matrix) that is treated as an activation vector rather than a matrix (Smolensky, 1990). For example, to bind a one-place predicate, **r** (such as *eats* (*x*) or *runs* (*x*)), to its argument, **f**, the tensor **rf** is formed by multiplying the *i*th element of vector **r**, representing the role, by the *j*th element of **f**, representing the filler (for all combinations of *i* and *j*):

$$\mathbf{rf}_{ij} = \mathbf{r}_i \mathbf{f}_j. \tag{1}$$

There are two ways to bind multiplace relations to their arguments using tensor products. One is to define tensors of progressively higher *rank* (where the rank of a tensor is the number of vectors that come together to define it; see, e.g., Halford et al., 1994). For example, a two-place relation (such as *loves* (*x*, *y*) or *larger-than* (*x*, *y*)) could be represented by the rank three tensor **rfg**:

$$\mathbf{rfg}_{ijk} = \mathbf{r}_i \mathbf{f}_j \mathbf{g}_k, \tag{2}$$

where **r** represents the relation (e.g., *loves* in *loves* (*x*, *y*)), **f** represents the argument bound to the first role of the relation (*x*), and **g** represents the argument bound to the second (*y*). An alternative approach is to designate separate tensors for each role-filler binding and represent the complete proposition as their sum (e.g., Tesar & Smolensky, 1994). For example:

$$\mathbf{rf} = \mathbf{rf}^1 + \mathbf{rf}^2, \qquad (3)$$

where $\mathbf{rf}^1$ is a tensor representing the binding of the first role to its argument and $\mathbf{rf}^2$ represents the binding of the second role to the second argument.

The deep problem with all these approaches is that, because tensors are a variety of conjunctive coding, they violate role-filler independence. A tensor is a *product* of two or more vectors (see Eqs. 1–3), so the similarity of two tensors—and hence the ability to generalize something learned about one of them to the other—scales with the *product* of the similarities of the simple vectors from which the tensors were created. For example, if vector similarity is defined in terms of the inner ("dot") product, then:

$$\mathbf{rf}_1 \cdot \mathbf{rf}_2 = (\mathbf{r}_1 \cdot \mathbf{r}_2)(\mathbf{f}_1 \cdot \mathbf{f}_2), \qquad (4)$$

where $\mathbf{rf}_1$ and $\mathbf{rf}_2$ are tensors made from $\mathbf{r}_1$ and $\mathbf{f}_1$ and from $\mathbf{r}_2$ and $\mathbf{f}_2$, respectively, and $\cdot$ indicates the dot product. Similarly, if vector similarity is defined in terms of the cosine of the angle between two vectors, then:

$$\cos(\mathbf{rf}_1, \mathbf{rf}_2) = \cos(\mathbf{r}_1, \mathbf{r}_2)\cos(\mathbf{f}_1, \mathbf{f}_2). \qquad (5)$$

Under both these definitions of vector similarity, two tensors will be similar to one another only to the extent that both their roles *and* fillers are similar: Identical roles, bound to completely different fillers, result in completely different tensor products, precluding any generalization from one to the other. This property is true, not only of tensor products, but of any binding scheme based on tensors (such as HRRs; Plate, 1991).

This property of tensors reflects the fact that they are the result of vector multiplication: Since role and filler vectors are multiplied by one another to form role-filler bindings, the similarities of those vectors are multiplied to determine the similarity of the resulting tensors. This observation suggests that one way to avoid role-filler interaction in role-filler binding similarity is to perform role-filler binding by role-filler addition rather than multiplication.

Models Based on Vector Addition

A second approach to implementing symbolic structure in connectionist networks is to bind roles to fillers by vector addition, which can be implemented as synchrony of neural firing (see Hummel & Holyoak, 1997, 2003). The basic idea is that units representing a relational role are added to (i.e., fire in synchrony with) the units representing the argument filling that role; units representing separate role-filler bindings fire out of synchrony with one another. A related proposal is to represent role-filler bindings by systematic *a*synchrony of firing, such that units representing a relational role fire, for example, just before the units representing its filler, and separate bindings fire in more distant temporal relations (see Doumas & Hummel, 2005; Doumas et al., 2008; Love, 1999). In either case, the binding is represented as a kind of vector addition because a given vector always represents a given role or object, regardless of the object or role to which it happens to be bound, and the bindings of roles to their fillers are represented by operations (synchrony or systematic asynchrony of firing) that put roles together with their fillers in an additive rather than multiplicative fashion.

However, synchrony of firing cannot be the whole story concerning the neural basis for binding, because temporal patterns of neural activity are necessarily transient. At a minimum, conjunctive codes are necessary for the purposes of storing bindings in LTM, and for forming localist tokens of roles, objects, role-filler bindings, and complete propositions (Hummel & Holyoak, 1997, 2003). A complete account of the human cognitive architecture must incorporate both dynamic binding (for independent representation of roles bound to fillers in WM) and conjunctive coding (for LTM storage and token formation), and specify how these coding systems are related.

One example of a system that binds via vector addition is a model of analogical reasoning called Learning and Inference by Schemas and Analogies, or LISA (see also the SHRUTI model; Shastri & Ajjanagadde, 1993). In LISA, relational structures are represented by a hierarchy of distributed and localist codes (see Fig. 5.6). At the bottom, "semantic" units represent the features of objects and roles in a distributed fashion. At the next level, these distributed representations are connected to localist predicate-and-object units (POs) representing individual predicates (or relational roles) and objects. Localist role-binding units (RBs) link object and

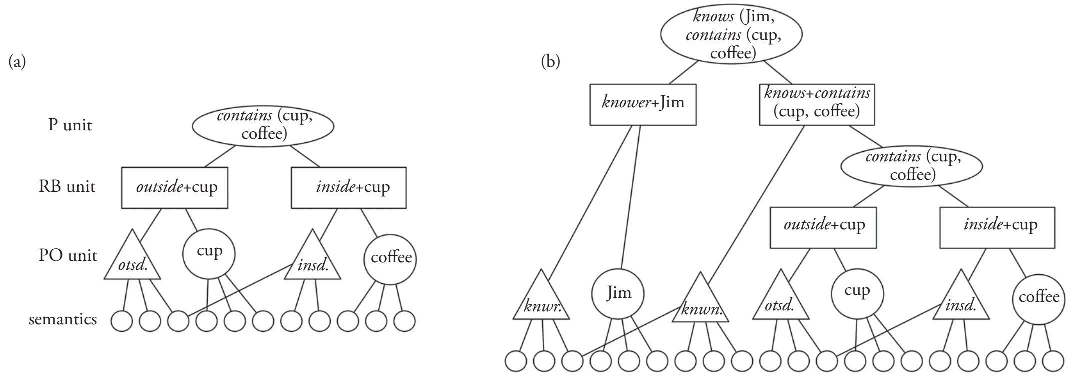

Fig. 5.6 Representation of propositions in LISA (Learning and Inference by Schemas and Analogies). Objects and relational roles are represented both as patterns of activation distributed over units representing semantic features (*semantic units*; small circles) and as localist (PO) units representing tokens of objects (large circles) and relational roles (triangles). Roles are conjunctively bound to fillers by localist role-binding (RB) units (rectangles), and role-filler bindings are conjunctively bound into complete propositions by localist P units (ovals). (*a*) Representation of *contains* (cup, coffee). (*b*) Representation of *knows* (Jim, *contains* (cup, coffee)). When one proposition takes another as an argument, the lower (argument) proposition serves in the place of an object unit under the appropriate RB of the higher level P unit (in this case, connecting contains (cup, coffee) to the RB representing what is known). In the figure we represent these localist units (i.e., POs, RBs, and Ps) with different shapes for the purposes of clarity. Importantly, these units are not different kinds of units (i.e., they do not work differently). Rather, they are simply units in different layers of the network. We use different names for the units in each layer (and different shapes in the figures) only to make them easier to distinguish.

relational role units into specific role-filler bindings. At the top of the hierarchy, localist proposition units (Ps) link RBs into whole relational propositions.

To represent the proposition *contains* (cup, coffee), PO units (triangles and large circles in Fig. 5.5) representing the relational roles *outside* and *inside*, and the fillers cup and coffee, are connected to semantic units coding their semantic features. RB units (rectangles) then conjunctively code the connection between roles and their fillers (one RB connects cup to *outside*, and one connects coffee to *inside*). At the top of the hierarchy, P units (oval) link sets of RBs into whole relational propositions. A P unit conjunctively codes the connection between the RBs representing *outside*+cup and the RB representing *inside*+coffee, thus encoding the relational proposition *contains* (coffee, cup). Each level of the representational hierarchy serves an important purpose. The semantic units capture the semantically rich (i.e., distributed) nature of human mental representations. The three layers of localist units make it possible to treat each level of the hierarchy as an independent entity for the purposes of mapping and inference (Hummel & Holyoak, 1997, 2003).

When a proposition enters working memory, role-filler bindings must be represented dynamically on the units that maintain role-filler independence (i.e., POs and semantic units; see Hummel & Holyoak, 1997). In models using synchrony of firing as a binding tag, roles are dynamically bound to their fillers by synchrony of firing (see earlier). In models using systematic asynchrony of firing as a binding tag, roles and their fillers fire in direct sequence. Binding information is carried in the proximity of firing (e.g., with roles firing directly before their fillers).[3]

An important consequence of the approach is that it allows a solution to the problem of how the structured (i.e., symbolic) representations that underlie symbolic systems may be learned in the first place (as noted earlier, one of the most oft-levied criticisms of the symbolic account of cognition). DORA (*Discovery of Relations by Analogy*; Doumas et al., 2008) is an account of how children and adults learn novel relational concepts from examples and subsequently use those representations in the service of understanding and reasoning about the world. Unlike other models that represent and employ structured representations, DORA *learns* structured relational representations from unstructured (i.e., nonsymbolic and nonrelational) examples. DORA starts with unstructured representations of objects as simple vectors of features. When DORA compares two or more of these objects, it learns explicit representations of any properties they share. Because DORA can use time as a binding tag (see earlier), the resulting representations are effectively single-place predicates (represented in a distributed fashion) that can be bound to novel arguments. DORA then combines sets of these predicates to form representations of complete multiplace relations (where each of the combined predicates serves as a role of the new relation). For example, DORA will learn predicates like *inside*(*x*) and *outside*(*y*) when it compares examples of different objects inside and outside one another (e.g., children outside a house and coffee inside a cup). DORA will then combine the *inside*(*x*) and *outside*(*y*) predicates to form the multiplace relation *contains* (*x*, *y*). Importantly, this is precisely the learning trajectory that children follow when they learn relations (e.g., Smith, 1984). Thus, DORA provides an account of how structured representations can be learned from examples, and so it addresses one of the major criticisms levied at structured models of cognition.[4]

Based on a small set of basic principles (notably comparison-based learning and constructing multiplace relations from sets of single-place predicates), DORA accounts for many phenomena surrounding the development of relational thinking, the development of the shape bias, and the effect of labeling on relational category learning (e.g., Doumas & Hummel, 2010; Doumas et al., 2008; Sandhofer & Doumas, 2008; Son, Doumas, & Goldstone, 2010).

Strengths

Models based on symbolic-connectionist enjoy a range of strengths. Because they are structured (i.e., they have two representational degrees of freedom in that they solve the binding problem), symbolic-connectionist models support the flexible and powerful generalizations of traditional symbol systems. Predicates are free to vary in their arguments, and thus what is learned about a predicate in one context potentially generalizes to other contexts and other arguments (see later discussion). So, like symbolic models, symbolic-connectionist models can extrapolate beyond training examples. The structure sensitivity of symbolic-connectionist models has allowed them to account for a wide range of phenomena in higher order cognition, including analogy making and retrieval from long-term memory (Hummel & Holyoak, 1997), relational generalization and schema induction (Hummel & Holyoak, 2003), learning relational concepts (Doumas et al., 2008), object recognition and the role of attention in shape perception (Hummel, 2001; Hummel & Biederman, 1992), motivation (Sun, 2009), speech production (Chang, Dell, & Bock, 2006), explicit skill learning (Sun, Slusarz, & Terry, 2005), and creative problem solving (Helie & Sun, 2010). There are also models of this sort that are able to learn structured (i.e., symbolic) representations from unstructured examples (Doumas et al., 2008).

Just as their structure sensitivity affords them the advantages of symbolic competence, so the distributed representations in symbolic-connectionist models afford them many of the strengths of traditional connectionist models. Like traditional connectionist models (as discussed earlier), symbolic-connectionist models offer a link between algorithmic models and possible neural underpinnings, allow natural automatic generalization, integrate representation and learning, and degrade gracefully with damage. Symbolic-connectionist models have been used to account for implicit skill learning (Sun et al., 2005), categorization, reflexive inference (Shastri & Ajjanagadde, 1993), and cognitive deficits due to brain damage (Morrison et al., 2004) and normal aging (Viskontas et al., 2004).

Weaknesses

Symbolic-connectionist models also have their share of weaknesses. One of the more serious shortcomings of symbolic-connectionist models is that they are weaker than purely symbolic models. For example, whereas symbolic models can deal with executive function and goal setting, representing and responding to negation and quantification (e.g., universal—for all—and existential—there exists), these phenomena have yet to be successfully modeled within symbolic-connectionist architectures. It remains unclear how (or even whether) symbolic-connectionist models will be able successfully account for them.

A second shortcoming of symbolic-connectionist models is that their representational assumptions (and thus their resulting representations) are more complex than traditional connectionist models. Although symbolic-connectionist models are connectionist in spirit, in that they are built out of collections of simple processing units that are highly interconnected, they include additional processing assumptions (such as sensitivity to temporal patterns of firing) that allow them to solve the binding problem (i.e., that provide them an additional informational degree of freedom comparable to that available to symbolic models). But whereas it might appear that symbolic-connectionist models make more nativist assumptions than traditional connectionist models (in that the former assumes an additional informational degree of freedom by which to carry binding information), we would argue that this appearance is misleading. Symbolic-connectionist models make assumptions about learning rules and activation functions similar to those made by traditional connectionist models. The additional assumption that symbolic-connectionist models (at least those that use time to carry binding information) must make beyond traditional connectionist models is that units are organized in layers and are sensitive to what layers they are in.

Conclusions and Future Directions

An adequate account of human mental representations—and the human cognitive architecture more broadly—must account both for our ability to represent the semantic content of relational roles and their fillers, and for our ability to bind roles to their fillers dynamically without altering the representation of either. Traditional symbolic approaches to cognition fail to specify the semantic content of roles and their fillers—a failing that, as noted by the connectionists in the 1980s, renders them too inflexible to serve as an adequate account of human mental representations. Traditional distributed connectionist approaches have the opposite problem: They succeed in capturing the semantic content of the entities they represent but fail to provide any basis for binding those entities together into symbolic (i.e., relational) structures. This failure renders them incapable of relational generalization, which appears to be required for such human abilities as assessing similarity based on alignment (Goldstone & Son, Chapter 10), analogical reasoning (Holyoak, Chapter 13), and learning problem schemas (Bassok & Novick, Chapter 21).

By contrast, symbolic-connectionist models combine the strengths of both the symbolic and connectionist approaches. These models have the potential to produce representations that are neurally plausible, semantically rich, flexible, and meaningfully symbolic. Armed with these representations, the symbolic-connectionist approach may be able to provide a powerful foundation for understanding human cognition. Nonetheless, models to date have their limitations. Those models based on vector multiplication (using tensor products and other forms of conjunctive coding as the sole basis for role-filler binding) fail to capture the natural pattern of similarities among propositions. There remain many important aspects of human cognition that symbolic-connectionist models have not yet addressed, including planning, quantification, negation, and other aspects of language use. It remains to be seen whether these are simply questions or represent fundamental limitations of the approach.

Notes

1. We say "in the limit" here because this statement assumes that the predicate is defined over variables open enough to take any kind of arguments, such as the x, y, and z in the larger . . . production rule. By contrast, a production rule defined over specific objects, such as the "me and my neighbor's dog" rule, would not necessarily be expected to generalize automatically to all new arguments (although used as part of an analogy, it might be expected to generalize very substantially; see, e.g., Holyoak & Thagard, 1995; Hummel & Holyoak, 2003).

2. Note that this correspondence is not based on anything as trivial as the arguments' locations within the parentheses: The same correspondence is also suggested by the propositions larger (battleship, cruiser) and smaller (housecat, Rottweiler)—a correspondence that also suggests that the semantics of the larger and smaller relations reside, not in the relations as holistic entities, but in their individual roles (Doumas & Hummel, 2005; Doumas, Hummel, & Sandhofer, 2008; Hummel & Holyoak, 1997, 2003).

3. Asynchrony-based binding allows role and filler to be coded by the same pool of semantic units, which allows a system to learn representations of relations from representations of objects (Doumas et al., 2008).

4. As noted by Holyoak (Chapter 13), DORA does not provide an account of what the features of relations are. Rather, it provides an account of how those features (assuming they exist) can be extracted from unstructured representations of objects and represented as explicit structures that can take arguments (i.e., as predicates).

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–403.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., Betts, S. A., Ferris, J. L., & Fincham, J. M. (2010). Neural imaging to track mental states while using an intelligent tutoring system. *Proceedings of the National Academy of Sciences USA*, *107*(15), 7018–7023.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., Qin, Y., Jung, K-W., & Carter, C. S. (2007). Information-processing modules and their relative modality specificity. *Cognitive Psychology*, *54*(3), 185–217

Bassok, M., Wu, L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretive effects of content on problem-solving transfer. *Memory and Cognition*, *23*, 354–367.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234–272.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*, 347–382.

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

Doumas, L. A. A., & Hummel, J. E. (2004). A fundamental limitation of symbol-argument- argument notation as a model of human relational representations. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 327–332). Mahwah NJ: Erlbaum.

Doumas, L. A. A., & Hummel, J. E. (2005). A symbolic-connectionist model of relation discovery. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 606–611). Mahwah NJ: Erlbaum.

Doumas, L. A. A. & Hummel, J. E. (2010). A computational account of the development of the representations underlying object recognition. *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.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1–63.

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

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–205.

Gentner, D., & Forbus, K. (2011). Computational models of analogy. *WIREs Cognitive Science*, *2*, 266–276.

Gentner, D., Ratterman, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, *25*, 524–575.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory. Vol. 2: Analogical connections* (pp. 363–415). Norwood, NJ: Ablex.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Brain and Behavioral Sciences*, *21*, 803–864.

Helie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, *117*, 994–1024.

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. (2000). The proper treatment of symbols in a connectionist architecture. In E. Deitrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 229–263). Mahwah, NJ: Erlbaum.

Holyoak, K. J., & Hummel, J. E. (2001). Toward an understanding of analogy within a biological symbol system. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 161–195). Cambridge, MA: MIT Press.

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

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA*, *79*, 2554–2558.

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., & 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*, 220–263.

Hummel, J. E., Holyoak, K. J., Green, C., Doumas, L. A. A., Devnich, D., Kittur, A., & Kalar, D. J. (2004). A solution to the binding problem for compositional connectionism. In S.D. Levy & R. Gayler (Eds.), *Compositional connectionism in cognitive science: Papers from the AAAI fall symposium* (pp. 31–34). Menlo Park, CA: AAAI Press.

Joanisse, M., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences USA*, *96*, 7592–7597.

Kanerva, P. (1998). *Sparse distributed memory*. Cambridge, MA: MIT Press.

Kellman, P. J., Burke, T. & Hummel, J. E. (1999). Modeling perceptual learning of abstract invariants. In *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 264–269). Mahwah, NJ: Erlbaum.

Krawczyk, D. C., Holyoak, K. J., & Hummel. J. E. (2005). The one-to-one constraint in analogical mapping. *Cognitive Science*, *29*, 29–38.

Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (2002). The role of textual coherence in incremental analogical mapping. *Journal of Memory and Language*, *47*, 407–435.

Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences 31*, 357–378.

Li, S. C., Lindenberger, U., & Sikstrom, S. (2001). Aging cognition: From neuromodulation to representation to cognition. *Trends in Cognitive Sciences*, *5*, 479–486.

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*(3), 243–282.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Marshall, J. A. (1995). Adaptive pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity, and scale. *Neural Networks*, *8*, 335–362.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–197.

McClelland, J. L., Rumelhart, D. E., & The PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.

McCulloch, W. A., &Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.
Metcalfe, J. (1990). Composite Holographic Associative Recall Model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General*, *119*, 145–160.
Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, *16*, 260–271.
Munakata, Y., & O'Reilly, R. C. (2003). Developmental and computational neuroscience approaches to cognition: The case of generalization. *Cognitive Studies*, *10*, 76–92.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*, 895–938.
O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science*, *314*, 91–94.
O'Reilly, R. C. & Busby, R. S. (2002). Generalizable relational binding from coarse-coded distributed representations. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (NIPS) 14 (pp. 75–82). 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* (pp. 168–192). Oxford, England: Oxford University Press.
Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*, 443–512.
Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), 109–178.
Pierce, C. S. (1879/1903). Logic as semiotic: The theory of signs. In J. Buchler (Ed.), *The philosophical writings of Pierce (1955)* (pp. 98–119). New York: Dover Books.
Plate, T. (1991). Holographic reduced representations: Convolution algebra for compositional distributed representations. In J. Mylopoulos & R. Reiter (Eds.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (pp. 30–35). San Mateo, CA: Morgan Kaufmann.
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, 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.
Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.
Ross, B. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 629–639.
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.
Sandhofer, C. M. & Doumas, L. A. A. (2008). Order and presentation effects in learning categories. *Journal of Cognition and Development, 9,* 194–221.
Shastri, L. (2002). Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Science*, *6*, 162–168.
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.
Sheya, A. & Smith, L. B. (2006). Perceptual features and the development of conceptual knowledge. *Journal of Cognition and Development*, *7*(4), 455–476.
Smith, L. B. (1984). Young children's understanding of attributes and dimensions. *Child Development*, *55*, 363–380.
Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*, 159–216.
Son, J. Y., Doumas, L. A. A., & Goldstone, R. L. (2010). When do words promote analogical transfer? *Journal of Problem Solving, 3,* 52–92.
St. John, M. F. (1992). The Story Gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science*, *16*, 271–302.
St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217–257.
Sun, R. (2009). Motivational representations within a computational cognitive architecture. *Cognitive Computation*, *1*, 91–103.
Sun R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, *112*, 159–192.
Taatgen, N. A., & Anderson, J. R. (2008). ACT-R. In R. Sun (Ed.), *Constraints in cognitive architectures* (pp. 170–185). Cambridge, UK: Cambridge University Press.
Tesar, B., & Smolensky, P. (1994). Synchronous-firing variable binding is spatio-temporal tensor product representation. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*. Atlanta, GA. August.
Viskontas, I. V., Morrison, R. G., 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.
von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Report No. 81–82). Goettingen, Germany: Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry.
von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, *24*, 95–104.
Wharton, C. M., Holyoak, K. J., & Lange, T. E. (1996). Remote analogical reminding. *Memory and Cognition*, *24*, 629–643.