CHAPTER 4

Approaches to Modeling Human Mental Representations: What Works, What Doesn't, and Why

Leonidas A. A. Doumas
John E. Hummel

Relational Thinking

A fundamental aspect of human intelligence is the ability to acquire and manipulate relational concepts. Examples of relational thinking include our ability to appreciate analogies between seemingly different objects or events (e.g., Gentner, 1983; Gick & Holyoak, 1980, 1983; Holyoak & Thagard, 1995; see Holyoak, Chap. 6), our ability to apply abstract rules in novel situations (e.g., Smith, Langston, & Nisbett, 1992), our ability to understand and learn language (e.g., Kim, Pinker, Prince, & Prasada, 1991), and even our ability to appreciate perceptual similarities (e.g., Goldstone, Medin, & Gentner, 1991; Hummel, 2000; Hummel & Stankiewicz, 1996; Palmer, 1978; see Goldstone & Son, Chap. 2). Relational thinking is ubiquitous in human cognition, underlying everything from the mundane (e.g., the thought "the mug is on the desk") to the sublime (e.g., Cantor's use of set theory to prove that the cardinal number of the reals is greater than the cardinal number of the integers).

Relational thinking is so commonplace that it is easy to assume the psychological mechanisms underlying it are relatively simple. They are not. The capacity to form and manipulate relational representations appears to be a late evolutionary development (Robin & Holyoak, 1995), closely tied to the increase in the size and complexity of the frontal cortex in the brains of higher primates, especially humans (Stuss & Benson, 1986). Relational thinking also develops relatively late in childhood (see, e.g., Smith, 1989; Halford, Chap. 22). Along with language, the human capacity for relational thinking is the major factor distinguishing human cognition from the cognitive abilities of other animals (for reviews, see Holyoak & Thagard, 1995; Oden, Thompson, & Premack, 2001; Call & Tomasello, Chap. 25).

Relational Representations

Central to understanding human relational thinking is understanding the nature of the mental representations underlying it: How does the mind represent relational ideas such as "if every element of set A is paired with a distinct element of set B, and there are still elements of B left over, then the cardinal number of B is greater than the cardinal number of A," or even simple relations such as "John loves Mary" or "the magazine is next to the phone"? Two properties of human relational representations jointly make this apparently simple question surprisingly difficult to answer (Hummel & Holyoak, 1997): As elaborated in the next sections, human relational representations are both *symbolic* and *semantically rich*. Although these properties are straightforward to account for in isolation, accounting for both together has proven much more challenging.

RELATIONAL REPRESENTATIONS ARE SYMBOLIC

A symbolic representation is one that represents relations explicitly and specifies the arguments to which they are bound. Representing relations explicitly means having primitives (i.e., symbols, nodes in a network, neurons) that correspond specifically to relations and/or relational roles. This definition of "explicit," which we take to be uncontroversial (see also Halford et al., 1998; Holland et al., 1986; Newell, 1990), implies that relations are represented independently of their arguments (Hummel & Biederman, 1992; Hummel & Holyoak, 1997, 2003a). That is, the representation of a relation cannot vary as a function of the arguments it happens to take at a given time, and the representation of an argument cannot vary across relations or relational roles.[1]

Some well-known formal representational systems that meet this requirement include propositional notation, labeled graphs, mathematical notation, and computer programming languages (among many others). For example, the relation *murders* is represented in the same way (and means the same thing) in the proposition *murders* (Bill, Susan) as it is in the proposition *murders* (Sally, Robert), even though it takes different arguments across the two expressions. Likewise, "2" means the same thing in x^2 as in 2^x, even though its role differs across the two expressions. At the same time, relational representations explicitly specify how arguments are bound to relational roles. The relation "*murders* (Bill, Susan)" differs from "*murders* (Susan, Bill)" only in the binding of arguments to relational roles, yet the two expressions mean very different things (especially to Susan and Bill).

The claim that formal representational systems (e.g., propositional notation, mathematical notation) are symbolic is completely uncontroversial. In contrast, the claim that human mental representations are symbolic is highly controversial (for reviews, see Halford et al., 1998; Hummel & Holyoak, 1997, 2003a; Marcus, 1998, 2001). The best-known argument for the role of symbolic representations in human cognition – the argument from systematicity – was made by Fodor and Pylyshyn (1988). They observed that knowledge is systematic in the sense that the ability to think certain thoughts seems to imply the ability to think related thoughts. For example, a person who understands the concepts "John," "Mary," and "loves," and can understand the statement "John loves Mary," must surely be able to understand "Mary loves John." This property of systematicity, they argued, demonstrates that human mental representations are symbolic. Fodor and Pylyshyn's arguments elicited numerous responses from the connectionist community claiming to achieve or approximate systematicity in nonsymbolic (e.g., traditional connectionist) architectures (for a more recent example, see Edelman & Intrator, 2003). At the same time, however, Fodor and Pylyshyn's definition of "systematicity" is so vague that it is difficult or impossible to evaluate these claims of "systematicity achieved or approximated" (van Gelder & Niklasson, 1994; for an example of the kind of confusion that has resulted from the attempt to approximate systematicity, see Edelman & Intrator, 2003, and the reply by Hummel, 2003). The concept of "systematicity" has arguably done more to cloud the debate over the role of symbolic representations in human cognition than to clarify it.

We propose that a clearer way to define symbolic competence is in terms of the ability to appreciate what different bindings of the same relational roles and fillers have in common and how they differ (see also Garner, 1974; Hummel, 2000; Hummel & Holyoak, 1997, 2003a; Saiki & Hummel, 1998). Under this definition, what matters is the ability to appreciate what "John loves Mary" has in common with "Mary loves John" (i.e., the same relations and arguments are involved) and how they differ (i.e., the role-filler bindings are reversed). It does not strictly matter whether you can "understand" the statements, or even whether they make any sense. What matters is that you can evaluate them in terms of the relations among their components. This same ability allows you to appreciate how "the glimby jolls the ronket" is similar to and different from "the ronket jolls the glimby," even though neither statement inspires much by way of understanding. To gain a better appreciation of the abstractness of this ability, note that the ronket and glimby may not even be organisms (as we suspect most readers initially assume they are), but may instead be machine parts, mathematical functions, plays in a strategy game, or anything else that can be named.

This definition of symbolic competence admits to more objective evaluation than does systematicity: one can empirically evaluate, for any f, x, and y, whether someone knows what $f(x, y)$ has in common with and how it differs from $f(y, x)$. It is also important because it relates directly to what we take to be the defining property of a symbolic (i.e., explicitly relational) representation: namely, as noted previously, the ability to represent relational roles independently of their arguments and to simultaneously specify which roles are bound to which arguments (see also Hummel, 2000, 2003; Hummel & Holyoak, 1997, 2003a). It is the independence of roles and fillers that allows one to appreciate that the glimby in "the glimby jolls the ronket" is the same thing as the glimby in "the ronket jolls the glimby"; and it is the ability to explicitly bind arguments to relational roles that allows one to know how the two statements differ. We take the human ability to appreciate these similarities and differences as strong evidence that the representations underlying human relational thinking are symbolic.

RELATIONAL REPRESENTATIONS ARE SEMANTICALLY RICH

The second fundamental property of human relational representations, and human mental representations more broadly, is that they are semantically rich. It means something to be a lover or a murderer, and the human mental representation of these relations makes this meaning explicit. As a result, there is an intuitive sense in which *loves* (John, Mary) is more like *likes* (John, Mary) than *murders* (John, Mary). Moreover, the meanings of various relations seem to apply specifically to individual relational *roles*, rather than to relations as indivisible wholes. For example, it is easy to appreciate that the agent (i.e., killer) role of *murders* (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 like 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 richness of human relational representations is also evidenced by their flexibility (Hummel & Holyoak, 1997). Given statements such as *taller-than* (Abe, Bill), *tall* (Charles), and *short* (Dave), it is easy to map Abe onto Charles and Bill onto Dave, even though doing so requires the reasoner to violate the "*n*-ary restriction" (i.e., mapping the argument(s) and role(s) of an *n*-place predicate onto those of an *m*-place predicate, where $m \neq n$). Given *shorter-than* (Eric, Fred), it is also easy to map Eric onto Bill (and Dave) and Fred onto Abe (and Charles). These mappings are based on the semantics of individual roles, rather than, for instance, the fact that *taller-than* and *shorter-than* are logical opposites: The relation *loves* (x, y) is in some sense the opposite of *hates* (x, y) [or if you prefer, *not-loves* (x, y)], but in contrast to *taller-than* and *shorter-than*, in which the first role of one relation maps to the second role of the other, the first role of *loves* (x, y) maps to the first role of *hates* (x, y) [or *not-loves* (x, y)]. The point is that the similarity and/or mappings of various relational roles are idiosyncratic, based not on the formal syntax of propositional notation, but on the semantic content of the individual roles in question. The semantics of relational roles matter and are an explicit part of the mental representation of relations.

The semantic properties of relational roles manifest themselves in numerous other ways in human cognition. For example, they influence both memory retrieval (e.g., Gentner, Ratterman, & Forbus, 1993; Ross, 1987; Wharton, Holyoak, & Lange, 1996) and our ability to discover structurally appropriate analogical mappings (Bassok, Wu, & Olseth, 1995; Krawczyk, Holyoak, & Hummel, in press; Kubose, Holyoak, & Hummel, 2002; Ross, 1987). They 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.

In summary, human mental representations are both symbolic (i.e., they explicitly represent relations and the bindings of relational roles to their fillers) and semantically rich (in the sense that they make they semantic content of individual relational roles and their fillers explicit). A complete account of human thinking must elucidate how each of these properties can be achieved and how they work together. An account that achieves one property at the expense of the other is at best only a partial account of human thinking. The next section reviews the dominant approaches to modeling human mental representations, with an emphasis on how each approach succeeds or fails to capture these two properties of human mental representations. We review traditional symbolic approaches to mental representation, traditional distributed connectionist approaches, conjunctive distributed connectionist approaches (based on tensor products and their relatives), and an approach based on dynamic binding of distributed and localist connectionist representations into symbolic structures.

Approaches to Modeling Human Mental Representation

Symbol-Argument-Argument Notation

The dominant approach to modeling relational representations in the computational literature is based on propositional notation and formally equivalent systems (including varieties of labeled graphs and high-rank tensor representations). These representational systems – which we refer to collectively as symbol-argument-argument notation, or "SAA" – borrow conventions directly from propositional calculus, and are commonly used in symbolic models based on production systems (see Lovett & Anderson, Chap. 17, for a review), many forms of graph matching (e.g., Falkenhainer et al., 1989; Keane et al., 1994) and related algorithms.

SAA represents relations and their arguments as explicit symbols and represents the bindings of arguments to relational roles in terms of the locations of the arguments in the relational expression. For example, in the proposition *loves* (John, Mary), John is bound to the *lover* role by virtue of appearing in the first slot after the open parenthesis, and Mary to the *beloved* by virtue of appearing in the second slot. Similarly, in a labeled graph the top node (of the local subgraph coding "John loves Mary") represents the *loves* relation, and the nodes directly below it represent its arguments, with the bindings of arguments to roles captured, for example, by the order (left to right) in which those arguments are listed. These schemes, which may look different at first pass, are in fact isomorphic. In both cases, the relation is represented by a single symbol, and the bindings of arguments to relational roles are captured by the syntax of the notation (as list position within parentheses, as the locations of nodes in a directed graph, etc.).

Models based on SAA are meaningfully symbolic in the sense described previously: They represent relations explicitly (i.e., independently of their arguments), and they explicitly specify the bindings of relational roles to their arguments. This fact is no surprise, given that SAA is based on representational conventions that were explicitly designed to meet these criteria. However, the symbolic nature of SAA is nontrivial because it endows models based on SAA with all the advantages of symbolic representations. Most important, symbolic representations enable *relational generalization* – generalizations that are constrained by the relational roles that objects play, rather than simply the features of the objects themselves (see Holland et al., 1986; Holyoak & Thagard, 1995; Hummel & Holyoak, 1997, 2003a; Thompson & Oden, 2000). Relational generalization is important because, among other things, it makes it possible to define, match, and apply variablized rules. (It also makes it possible to make and use analogies, to learn and use schemas, and ultimately to learn variablized rules from examples; see Hummel & Holyoak, 2003a.) For example, with a symbolic representational system, it is possible to define the rule "if *loves* (*x*, *y*) and *loves* (*y*, *z*) and *not* [*loves* (*y*, *x*)], then *jealous* (*x*, *z*)" and apply that rule to any *x*, *y*, and *z* that match its left-hand ("if") side. As elaborated shortly, this important capacity, which plays an essential role in human relational thinking, lies fundamentally beyond the reach of models based on nonsymbolic representations (Holyoak & Hummel, 2000; Hummel & Holyoak, 2003a; Marcus, 1998).

Given the symbolic nature of SAA, it is no surprise that it has figured so prominently in models of relational thinking and symbolic cognition more generally (see Lovett & Anderson, Chap. 17). Less salient are the limitations of SAA. It has been known for a long time that SAA and related 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 et al., 1986). A review of how traditional symbolic models have handled this problem (typically with external representational systems such as lookup tables or matrices of hand-coded "similarity" values between symbols; see Lovett & Anderson, Chap. 17) also reveals that the question of semantics in SAA is, in the very least, a thorny inconvenience (Hummel & Holyoak, 1997). However, at the same time, it is tempting to assume it is merely an inconvenience – that surely there exists a relatively straightforward way to add semantic coding to propositional notation and other forms of SAA, and that a solution will be found once it becomes important enough for someone to turn their attention to it. In the mean time, it is surely no reason to abandon SAA as a basis for modeling human cognition.

However, it turns out that it is more than a thorny inconvenience: As demonstrated by Doumas and Hummel (2004), it is logically impossible to specify the semantic content of relational roles within an SAA representation. In brief, SAA representations cannot represent relational roles explicitly and simultaneously specify how they come together to form complete relations. The reason for this limitation is that SAA representations specify role information only implicitly (see Halford et al., 1998). Specifying this information explicitly requires new propositions, which must be related to the original relational representation via a second relation. In SAA, this results in a new relational proposition, which itself implies role representations to which it must be related by a third relational proposition, and so forth, *ad infinitum*. In short, attempting to use SAA to link relational roles to their parent relations necessarily results in an infinite regress of nested "constituent of" relations specifying which roles belong to which relations/roles (see Doumas & Hummel, 2004 for the full argument). As a result, attempting to use SAA to specify how roles form complete relations renders any SAA system *ill-typed* (i.e., inconsistent and/or paradoxical; see, e.g., Manzano, 1996).

The result of this limitation is that SAA systems are forced to use external (i.e., non-SAA) structures to represent the meaning of symbols (or to approximate those meanings, e.g., with matrices of similarity values) and external control systems (which themselves cannot be based on SAA) to read the SAA, access the external structures and relate the two. Thus, it is no surprise that SAA-based models rely on lookup tables, similarity matrices and so forth, to specify how different relations and objects are semantically related to one another: It is not merely a convenience, it is a necessity.

This property of SAA sharply limits its utility as a general approach to modeling human mental representations. In particular, it means that the connectionist critiques of the mid-1980s were right: Not only do traditional symbolic representations fail to represent the semantic content of the ideas they mean to express, the SAA representations on which they are based cannot even be adapted to do so. The result is that SAA is ill equipped, in principle, to address those aspects of human cognition that depend on the semantic content of relational roles and the arguments that fill them (which, as summarized previously, amounts to a substantial proportion of human cognition). This fact does not mean that models based on SAA (i.e., traditional symbolic models) are "wrong," only that they are incomplete. SAA is at best only a shorthand (a *very short* hand) approximation of human mental representations.

Traditional Connectionist Representations

In response to limitations of traditional symbolic models, proponents of connectionist models of cognition (see, e.g., Elman et al., 1996; Rumelhart et al., 1986; St. John & McClelland, 1990; among many others) have proposed that knowledge is represented, not as discrete symbols that enter into symbolic expressions, but as patterns of activation distributed over many processing elements. These representations are distributed in the sense that (1) any single concept is represented as a pattern (i.e., vector) of activation over many elements ("nodes" or "units" that are typically assumed to correspond roughly to neurons or small collections of neurons), and (2) any single element will participate in the representation of many different concepts.[2] As a result, two patterns of activation will tend to be similar to the extent that they represent similar concepts: In contrast to SAA, distributed connectionist representations provide a natural basis for representing the semantic content of concepts. Similar ideas have been proposed in the context of latent semantic analysis (Landauer & Dumais, 1997) and related mathematical techniques for deriving similarity metrics from the co-occurrence statistics of words in passages of text (e.g., Lund & Burgess, 1996). In all these cases, concepts are represented as vectors, and vector similarity is taken as an index of the similarity of the corresponding concepts.

Because distributed activation vectors provide a natural basis for capturing the similarity structure of a collection of concepts (see Goldstone & Son, Chap. 2), connectionist models have enjoyed substantial success simulating various kinds of learning and generalization (see Munakata & O'Reilly, 2003): Having been trained to give a particular output (e.g., generate a specific activation vector on a collection of output units) in response to a given input (i.e., vector of activations on a collection of input units), connectionist networks tend to generalize automatically (i.e., activate an appropriate output vector, or a close approximation of it) in response to new inputs that are similar to trained inputs. In a sense, connectionist representations are much more flexible than symbolic representations based on varieties of SAA. Whereas models based on SAA require predicates to match exactly in order to treat them identically,[3] connectionist models generalize more gracefully, based on the degree of overlap between trained patterns and new ones.

In another sense, however, connectionist models are substantially less flexible than symbolic models. The reason is that the distributed representations used by traditional connectionist models are not symbolic in the sense defined previously. That is, they cannot represent relational roles independently of their fillers and simultaneously specify which roles are bound to which fillers (Hummel & Holyoak, 1997, 2003a). Instead, a network's knowledge is represented as simple vectors of activation. Under this approach, relational roles (to the extent that they are represented at all) are either represented on separate units from their potential fillers (e.g., with one set of units for the *lover* role of the *loves* relation, another set for the *beloved* role, a third set for John, a fourth set for Mary, etc.), in which case the bindings of roles to their fillers is left unspecified (i.e., simply activating all four sets of units cannot distinguish "John loves Mary" from "Mary loves John" or even from a statement about a narcissistic hermaphrodite); or else units are dedicated to specific role-filler conjunctions (e.g., with one set of units for "John as lover" another for "John as beloved", etc.; e.g., Hinton, 1990), in which case the bindings are specified, but only at the expense of role-filler independence (e.g., nothing represents the *lover* or *beloved* roles, independently of the argument to which they happen to be bound). In neither case are the resulting representations truly symbolic.

Indeed, some proponents of traditional connectionist models (e.g., Elman et al., 1996) – dubbed "eliminative connectionists" by Pinker and Prince (1988; see also Marcus, 1998) for their explicit desire to eliminate the need for symbolic representations from models of cognition – are quite explicit in their rejection of symbolic representations as a component of human cognition. Instead of representing and matching symbolic "rules," eliminative (i.e., traditional) connectionist models operate by learning to associate vectors of features (where the features correspond to individual nodes in the network). As a result, they are restricted to generalizing based on the shared features in the training set and the generalization set. Although the generalization capabilities of these networks often appear quite impressive at first blush (especially if the training set is judiciously chosen to span the space of all possible input and output vectors; e.g., O'Reilly, 2001), the resulting models are not capable of relational generalization (see Hummel & Holyoak, 1997, 2003a; Marcus, 1998, 2001, for detailed discussions of this point).

A particularly clear example of the implications of this limitation comes from the story Gestalt model of story comprehension developed by St. John (1992; St. John & McClelland, 1990). In one computational experiment (St. John, 1992, simulation 1), the model was first trained with 1,000,000 short texts consisting of statements based on 136 constituent concepts. Each story instantiated a script such as "<person> decided to go to <destination>; <person> drove <vehicle> to <destination>" (e.g., "George decided to go to a restaurant; George drove a Jeep to the restaurant"; "Harry decided to go to the beach; Harry drove a Mercedes to the beach").

After the model had learned a network of associative connections based on the 1,000,000 examples, St. John tested its ability to generalize by presenting it with a text containing a new statement, such as "John decided to go to the airport." Although the statement as a whole was new, it referred to people, objects and places that had appeared in the examples used for training. St. John reported that when given a new example about deciding to go to the airport, the model would typically activate the restaurant or the beach (i.e., the destinations in prior examples of the same script) as the destination, rather than making the contextually appropriate inference that the person would drive to the airport. This type of error, which would appear quite unnatural in human comprehension, results from the model's inability to generalize relationally (e.g., if a person wants to go location x, then x will be the person's destination – a problem that requires the system to represent the variable x and its value, independently of its binding to the role of *desired location* or *destination*). As St. John noted, "Developing a representation to handle role binding proved to be difficult for the model" (1992, p. 294).

In general, although an eliminative connectionist model can make "inferences" on which it has been directly trained (i.e., the model will remember particular associations that have been strengthened by learning), the acquired knowledge may not generalize at all to novel instantiations that lie outside the training set (Marcus, 1998, 2001). For example, having learned that Alice loved Sam, Sam loved Betty, and Alice was jealous of Betty, and told that John loves Mary and Mary loves George, a person is likely to conjecture that John is likely to be jealous of George. An eliminative connectionist system would be a complete loss to make any inferences: John, Mary, and George are different people than Alice, Sam, and Betty (Holyoak & Hummel, 2000; Hummel & Holyoak, 2003a; Phillips & Halford, 1997).

A particularly simple example that reveals such generalization failures is the identity function (Marcus, 1998). Suppose, for example, that a human reasoner was trained to respond with "1" to "1," "2" to "2," and "3" to "3." Even with just these three examples, the human is almost certain to respond with "4" to "4," without any direct feedback that this is the correct output for the new case. In contrast, an eliminative connectionist model will be unable to make this obvious generalization. Such a model can be trained to give specific outputs to specific inputs (e.g., as illustrated in Figure 4.1). But when training is over, it will have learned only the input–output mappings on which it was trained (and perhaps those that can be represented by interpolating between trained examples; see Marcus, 1998): Lacking the capacity to represent variables, extrapolation outside the training set is impossible. In other words, the model will simply have learned to associate "1" with "1," "2" with "2," and "3" with "3." A human, by contrast, will have learned to associate *input (x)* with *output (x)*, for any *x*; and doing so requires the capacity to bind any new number (whether it was in the training space or not) to the variable *x*. Indeed, most people are willing to generalize even beyond the world of numbers. We leave it to the reader to give the appropriate outputs in response to the following inputs: "A"; "B"; "flower."

The deep reason the eliminative connectionist model illustrated in Figure 4.1 fails to learn the identity function is that it violates variable/value (i.e., role/filler) independence. The input and output units in Figure 4.1 are intentionally mislabeled to suggest that they represent the concepts "1," "2," etc. However, in fact, they do not represent these concepts at all. Instead, the unit labeled "1" in the input layer represents, not "1," but "1 *as the input to the identity function*." That is, it represents a conjunctive binding of the value "1" to the variable "input to the function." Likewise, the unit labeled "1" in the output layer represents, not "1," but "1" as output of the identity function. Thus, counter to initial appearances, the concept "1" is not represented anywhere in the network. Neither, for that matter, is the concept "input to the identity function": Every unit in the input layer represents *some specific input* to the function; there are no units to represent *input* as a generic unbound variable.

Because of this representational convention (i.e., representing variable-value conjunctions instead of variables and values), traditional connectionist networks are forced to learn the identity function as a mapping from one set of conjunctive units (the input layer) to another set of conjunctive units (the output layer). This mapping, which to our eye resembles an approximation of the identity function, $f(x) = x$, is, to the network, just an arbitrary mapping. It is arbitrary precisely because the unit representing "1 as output of the function" bears no relation to the unit representing "1 as input to the function." Although any function specifies a mapping [e.g., a mapping from values of x to values of $f(x)$], learning a mapping is not the same thing as learning a function. Among other differences, a function can be universally quantified [e.g., $\forall x$, $f(x) = x$], whereas a finite mapping cannot; universal quantification permits the function to apply to numbers (and even nonnumbers) that lie well outside the "training" set. The point is that the connectionist model's failure to represent variables independently of their values (and vice versa) relegates it to (at best) approximating a subset of the

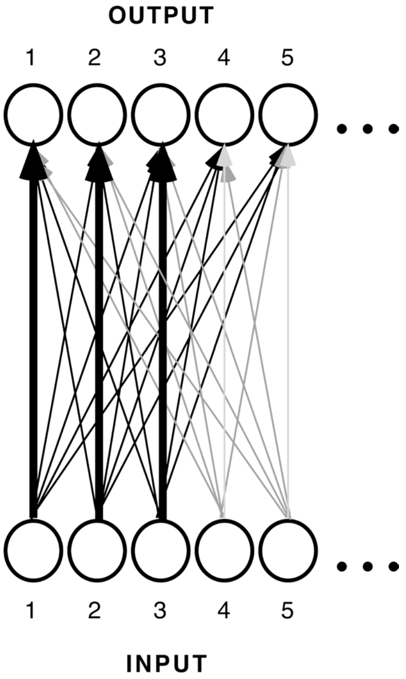

Figure 4.1 . Diagram of a two-layer connectionist network for solving the identity function in which the first three units (those representing the numbers 1, 2, and 3) have been trained and the last two (those representing the numbers 4 and 5) have not. Black lines indicate already trained connections, whereas grey lines untrained connections. Thicker lines indicate highly excitatory connections, whereas thinner lines slightly excitatory or slightly inhibitory connections.

identity function as a simple, and ultimately arbitrary, mapping (see Marcus, 1998). People, by contrast, represent variables independently of their values (and vice versa), and so can recognize and exploit the decidedly nonarbitrary relation between the function's inputs and its outputs: To us, but not to the network, the function is not an arbitrary mapping at all, but rather a trivial game of "say what I say."

As these examples illustrate, the power of human reasoning and learning, most notably our capacity for sophisticated relational generalizations, is dependent on the capacity to represent relational roles (variables) and bind them to fillers (values). This is precisely the same capacity that permits composition of complex symbols from simpler ones. The human mind is the product of a symbol system; hence, any model that succeeds in eliminating symbol systems will *ipso facto* have succeeded in eliminating itself from contention as a model of the human cognitive architecture.

Conjunctive Connectionist Representations

Some modelers, recognizing both the essential role of relational representations in human cognition (e.g., for relational generalization) and the value of distributed representations, have sought to construct symbolic representations in connectionist architectures. The most common approach is based on Smolensky's (1990) tensor products (e.g., Halford et al., 1998) and its relatives, such as spatter codes (Kanerva, 1998), holographic reduced representations (HRRs; Plate, 1994), and circular convolutions (Metcalfe, 1990). We restrict our discussion to tensor products because the properties of tensors we discuss also apply to the other approaches (see Holyoak & Hummel, 2000).

A tensor product is an outer product of two or more vectors that are treated as an activation vector (i.e., rather than a matrix) for the purposes of knowledge representation (see Smolensky, 1990). In the case of a rank 2 tensor, **uv**, formed from two vectors, **u** and **v**, the activation of the *ij*th element of **uv** is simply the product of the activations of the *i*th and *j*th elements of **u** and **v**, respectively: $\mathbf{uv}_{ij} = \mathbf{u}_i \mathbf{v}_j$. Similarly, the *ijk*th value of the rank 3 tensor **uvw** is the product $\mathbf{uvw}_{ijk} = \mathbf{u}_i \mathbf{v}_j \mathbf{w}_k$, and so forth, for any number of vectors (i.e., for any rank).

Tensors and their relatives can be used to represent role-filler bindings. For example, if the *loves* relation is represented by the vector **u**, John by the vector **v**, and Mary by the vector **w**, then the proposition *loves* (John, Mary) could be represented by the tensor **uvw**; *loves* (Mary, John) would be represented by the tensor **uwv**. This procedure for representing propositions as tensors – in which the predicate is represented by one vector (here, **u**) and its argument(s) by the others (**v** and **w**) – is isomorphic with SAA (Halford et al., 1998): One entity (here, a vector) represents the relation, other entities represent its arguments, and the bindings of arguments to roles of the relation are represented spatially (note the difference between **uvw** and **uwv**). However, this version of tensor-based coding is SAA-isomorphic; the entire relation is represented by a single vector or symbol, and arguments are bound directly to that symbol. Consequently, it provides no basis for differentiating the semantic features of the various roles of a relation.

Another way to represent relational bindings using tensors is to represent individual relational roles as vectors, role-filler bindings as tensors, and complete propositions as sums of tensors (e.g., Tesar & Smolensky, 1994). For example, if the vector **l** represents the *lover* role of the *loves* relation, **b** the *beloved* role, **j** John and **m** Mary, then *loves* (John, Mary) would be represented by the sum **lj** + **bm**, and *loves* (Mary, John) would be the sum **lm** + **bj**.

Tensors provide a basis for representing the semantic content of relations (in the case of tensors that are isomorphic with SAA) or relational roles (in the case of tensors based on role-filler bindings) and to represent role-filler bindings explicitly. Accordingly, numerous researchers have argued that tensor products and their relatives provide an appropriate model of human symbolic representations. Halford and his colleagues also showed that tensor products based on SAA representations provide a natural account of the capacity limits of human working memory, and applied these ideas to account for numerous phenomena in relational reasoning and cognitive development (see Halford, Chap. 22). Tensors are thus at least a useful approximation of human relational representations.

However, tensor products and their relatives have two properties that limit their adequacy as a general model of human relational representations. First, tensors necessarily violate role-filler independence (Holyoak & Hummel, 2000; Hummel & Holyoak, 2003a). This is true both of SAA-isomorphic tensors (as advocated by Halford and colleagues) and role-filler binding-based tensors (as advocated by Smolensky and colleagues). A tensor product is a product of two or more vectors, so the similarity of two tensors (e.g., their inner product or the cosine of the angle between them) is equal to the *product* of the similarities of the basic vectors from which they are constructed. For example, in the case of tensors **ab** and **cd**, formed from vectors **a, b, c**, and **d**:

$$\mathbf{ab} \cdot \mathbf{cd} = (\mathrm{a} \cdot \mathrm{c})(\mathrm{b} \cdot \mathrm{d}), \qquad (4.1)$$

where the "·" denotes the inner product, and

$$\cos(\mathbf{ab}, \mathbf{cd}) = \cos(\mathbf{a}, \mathbf{c})\cos(\mathbf{b}, \mathbf{d}), \qquad (4.2)$$

where cos(x, y) is the cosine of the angle between x and y.

In other words, two tensor products are similar to one another to the extent that their roles *and* fillers are similar to one another. If vectors **a** and **c** represent relations (or relational roles) and **b** and **d** represent their fillers, then the similarity of the **ab** binding to the **cd** binding is equal to the similarity of roles **a** and **c** times the similarity of fillers **b** and **d**. This fact sounds unremarkable at first blush. However, consider the case in which **a** and **c** are identical (for clarity, let us replace them both with the single vector **r**), but **b** and **d** are completely unrelated (i.e., they are orthogonal, with an inner product of zero). In this case,

$$(\mathbf{rb} \cdot \mathbf{rd}) = (\mathbf{r} \cdot \mathbf{r})(\mathbf{b} \cdot \mathbf{d}) = 0. \qquad (4.3)$$

That is, the similarity of **rb** to **rd** is zero, even though both refer to the same relational role.

This result is problematic for tensor-based representations because a connectionist network (and for that matter, probably a person) will generalize learning from **rb** to **rd** to the extent that the two are similar to one another. Equation (4.3) shows that, if **b** and **d** are orthogonal, then **rb** and **rd** will be orthogonal, even though they both represent bindings of different arguments to exactly the same relational role (**r**). As a result, tensor products cannot support relational generalization. The same limitation applies to all multiplicative binding schemes (i.e., representations in which the vector representing a binding is a function of the product of the vectors representing the bound elements), including HRRs, circular convolutions, and spatter codes (see Hummel & Holyoak, 2003a).

A second problem for tensor-based representations concerns the representation of the semantics of relational roles. Tensors that are SAA-isomorphic (e.g., Halford et al., 1998) fail to distinguish the semantics of different roles of the relation precisely because they are SAA-isomorphic (see Doumas & Hummel, 2004): Rather than using separate vectors to represent a relation's roles, SAA-isomorphic tensors represent the relation, as a whole, using a single vector. Role-filler binding tensors (e.g., as proposed by Smolensky and colleagues) do explicitly represent the semantic content of the individual roles of a relation. However, these representations are limited by the summing operation that is used to conjoin the separate role-filler bindings into complete propositions. The result of the summing operation is a "superposition catastrophe" (von der Malsburg, 1981) in which the original role-filler bindings – and therefore the original roles and fillers – are unrecoverable (a sum underdetermines its addends).

The deleterious effects of this superposition can be minimized by using sparse representations in a very high-dimensional space (Kanerva, 1998; Plate, 1991). This approach works because it minimizes the representational overlap between separate concepts. However, minimizing the representational overlap also minimizes the positive effects of distributed representations (which stem from the overlap between representations of similar concepts). In the limit, sparse coding becomes equivalent to localist conjunctive coding, with completely separate codes for every possible conjunction of roles and fillers. In this case, there is no interference between separate bindings, but neither is there overlap between related concepts. Conversely, as the overlap between related concepts increases, so does the ambiguity of sums of separate role bindings. The ability to keep separate bindings separate thus invariably trades off against the ability to represent similar concepts with similar vectors. This trade-off is a symptom of the fact that tensors are trapped on the *implicit relations continuum* (Hummel & Biederman, 1992) – the continuum from holistic (localist) to feature-based (distributed), vector-based representations of concepts – characterizing representational schemes that fail to code relations independently of their arguments.

Role-Filler Binding by Vector Addition

What is needed is a way to both represent roles and their fillers in a distributed fashion (to capture their semantic content), and simultaneously bind roles to their fillers in a way that does not violate role-filler independence (to achieve meaningfully symbolic representation and thus relational generalization). Tensor products are on the right track, in the sense that they represent relations and fillers in a distributed fashion, and they can represent role-filler bindings – just not in a way that preserves role-filler independence. Accordingly, in the search for a distributed code that preserves role-filler independence, it is instructive to consider why, mathematically, tensors violate it.

The reason is that a tensor is a product of two or more vectors, so the value of ij^{th} element of the tensor is a function of the i^{th} value of the role vector *and* the j^{th} element of the filler vector. That is, a tensor is the result of a multiplicative interaction between two or more vectors. Statistically, when two or more variables do not interact – i.e., when their effects are independent, as in the desired relationship between roles and their fillers – their effects are additive (rather than multiplicative). Accordingly, the way to bind a distributed vector, *r*, representing a relational role to a vector, *f*, representing its filler is not to multiply them, but to add them (Holyoak & Hummel, 2000; Hummel & Holyoak, 1997, 2003a):

$$\mathbf{rf} = \mathbf{r} + \mathbf{f}, \tag{4.4}$$

where **rf** is just an ordinary vector (not a tensor).[4]

Binding by vector addition is most commonly implemented in the neural network modeling community as synchrony of neural firing (for reviews, see Hummel & Holyoak, 1997, 2003a), although it can also be realized in other ways (e.g., as systematic asynchrony for firing; Love, 1999). The basic idea is that vectors representing relational roles fire in synchrony with vectors representing their fillers and out of synchrony with other role-filler bindings. That is, at each instant in time, a vector representing a role is "added to" (fires with) the vector representing its filler.

Binding by synchrony of firing is much reviled in some segments of the connectionist modeling community. For example, Edelman and Intrator (2003) dismissed it as an "engineering convenience." Similarly, O'Reilly et al. (2003) dismissed it on the grounds that (1) it is necessarily transient [i.e., it is not suitable as a basis for storing bindings in long-term memory (LTM)], (2) it is capacity limited (i.e., it is only possible to have a finite number of bound groups simultaneously active and mutually out of synchrony; Hummel & Biederman, 1992; Hummel & Holyoak, 2003a; Hummel & Stankiewicz, 1996), and (3) bindings represented by synchrony of firing must ultimately make contact with stored conjunctive codes in LTM. These limitations do indeed apply binding by synchrony of firing; (1) and (2) are also precisely the limitations of human working memory (WM) (see Cowan, 2000). Limitation (3) is meant to imply that synchrony is redundant: If you already have to represent bindings conjunctively in order to store them in LTM, then why bother to use synchrony? The answer is that synchrony, but not conjunctive coding, makes it possible to represent roles independently of their fillers, and thus allows symbolic representations and relational generalization.

Despite the objections of Edelman and Intrator (2003), O'Reilly et al. (2003), and others, there is substantial evidence for binding by synchrony in the primate visual cortex (see Singer, 2000, for a review) and frontal cortex (e.g., Desmedt & Tomberg, 1994; Vaadia et al., 1995). It seems that evolution and the brain may be happy to exploit "engineering conveniences." This would be unsurprising given the computational benefits endowed by dynamic binding (namely, relational generalization based on distributed representations), the ease with which synchrony can be established in neural systems, and the ease with which it can be exploited (it is well known that spikes arriving in close temporal proximity have superadditive effects on the postsynaptic neuron relative to spikes arriving at very different times). The mapping between the limitations of human WM and the limitations of synchrony cited by O'Reilly et al. (2003) also constitutes indirect support for the synchrony hypothesis, as do the successes of models based on synchrony (for reviews, see Hummel, 2000; Hummel & Holyoak, 2003b; Shastri, 2003).

However, synchrony of firing cannot be the whole story. At a minimum, conjunctive coding is necessary for storing bindings in LTM, and forming localist tokens of roles, objects, role-filler bindings, and complete propositions (Hummel & Holyoak, 1997, 2003a). It seems likely, therefore, that an account of the human cognitive architecture that includes both "mundane" acts (such as shape perception, which actually turns out to be relational; Hummel, 2000) and symbolic cognition (such as planning, reasoning, and problem solving) 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 they are related.

The remainder of this chapter reviews one example of this approach to knowledge representation – "LISAese," the representational format used by Hummel and Holyoak's (1992, 1997, 2003a) LISA (Learning and Inference with Schemas and Analogies) model of analogical inference and schema induction – with an emphasis on how LISAese permits symbolic representations to be composed from distributed (i.e., semantically rich) representations of roles and fillers, and how the resulting representations are uniquely suited to simulate aspects

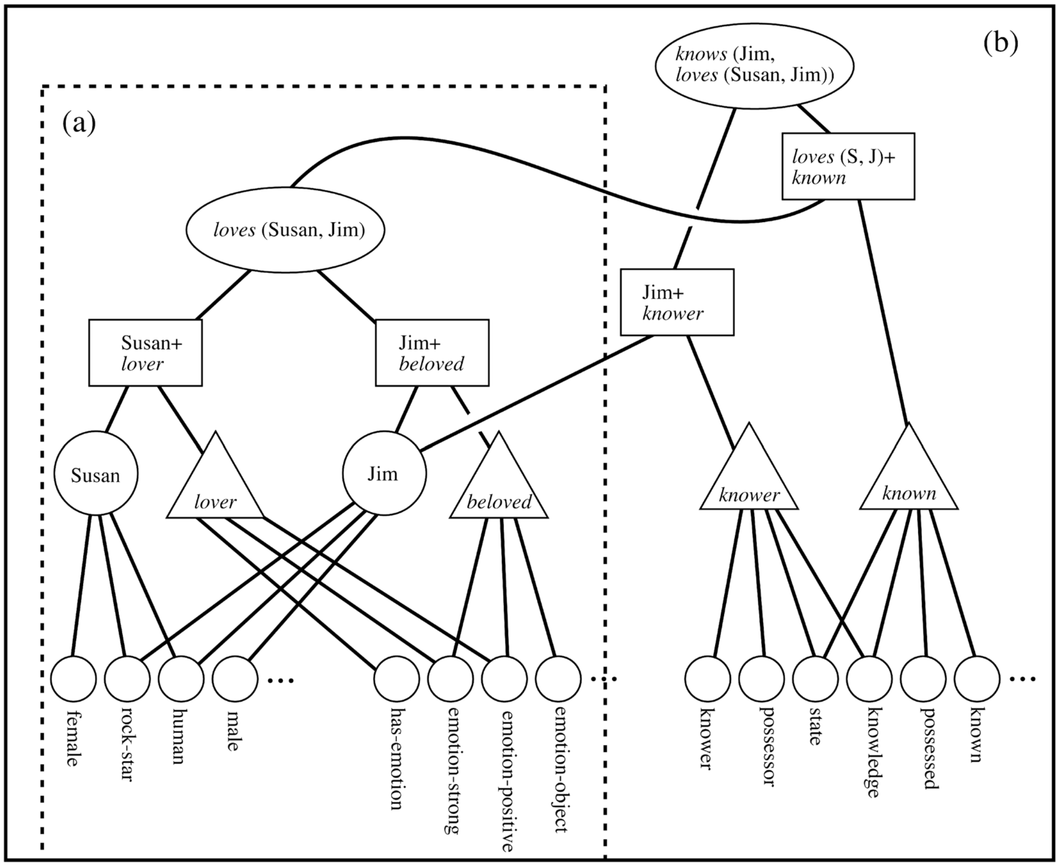

Figure 4.2. Representation of propositions in LISAese. Objects and relational roles are represented both as patterns of activation distributed over units representing semantic features (*semantic units*; small circles) and as localist units representing tokens of objects (large circles) and relational roles (triangles). Roles are bound to fillers by localist subproposition (SP) units (rectangles), and role-filler bindings are bound into complete propositions by localist proposition (P) units (ovals). (a) Representation of *loves* (Susan, Jim). (b) Representation of *knows* [Jim, *loves* (Susan, Jim)]. When one P takes another as an argument, the lower (argument) P serves in the place of an object unit under the appropriate SP of the higher-level P unit [in this case, binding *loves* (Susan, Jim) to the SP representing what is known].

of human perception and cognition (also see Holyoak, Chap. 6).

LISAese is based on a hierarchy of distributed and localist codes that collectively represent the semantic features of objects and relational roles, and their arrangement into complete propositions (Figure 4.2). At the bottom of the hierarchy, semantic units (small circles in Figure 4.2) represent objects and relational roles in a distributed fashion. For example, Jim might be represented by features such as *human*, *adult*, and *male* (along with units representing his personality traits, etc.), and Susan might be represented as *human*, *adult*, and *female* (along with units for her unique attributes). Similarly, the *lover* and *beloved* roles of the *loves* relation would be represented by semantic units capturing their semantic content. At the next level of the hierarchy, object and predicate units (large circles and triangles in Figure 4.2) represent objects and relational roles in a localist fashion, and share bidirectional excitatory connections with the corresponding semantic units. Subproposition units (SPs; rectangles in Figure 4.2) represent bindings of relational roles to their arguments [which can either be objects, as in Figure 4.2(a), or complete propositions, as in Figure 4.2(b)]. At the top of the hierarchy, separate role-filler bindings (i.e., SPs) are bound into a localist representation of the proposition as a whole via excitatory connections to a single proposition (P) unit (ovals in Figure 4.2). Representing propositions in this type of hierarchy reflects our assumption that every level of the hierarchy must be represented explicitly, as an entity in its own right (see Hummel & Holyoak, 2003a). The resulting representational system is commonly referred to as a *role-filler binding* system (see Halford et al., 1998). Both relational roles and their fillers are represented explicitly and relations are represented as linked sets of role-filler bindings. Importantly, in role-filler binding systems, relational roles, their semantics, and their bindings to their fillers are all made explicit in the relational representations themselves. As a result, role-filler binding representations are not subject to the problems inherent in SAA representations, discussed previously, wherein relational roles are left implicit in the larger relational structures.

A complete analog (i.e., story, situation, or event) in LISAese is represented by the collection of P, SP, predicate, object, and semantic units that code its propositional content. Within an analog, a given object, relational role, or proposition is represented by a single localist unit, regardless of how many times it is mentioned in the analog [e.g., Susan is represented by the same unit in both *loves* (Susan, Jim) and *loves* (Charles, Susan)], but a given element is represented by separate localist units in separate analogs. The localist units thus represent tokens of individual objects, relations, or propositions in particular situations (i.e., analogs). A given object or relational role will tend to be connected to many of the same semantic units in all the analogs in which it is mentioned, but there may be small differences in the semantic representation, depending on context (e.g., Susan might be connected to semantics describing her profession in an analog that refers to her work, and to features specifying her height in an analog about her playing basketball; see Hummel & Holyoak, 2003a). Thus, whereas the localist units represent tokens, the semantic units represent types.

The hierarchy of units depicted in Figure 4.2 represents propositions both in LISA's LTM and, when the units become active, in its WM. In this representation, the binding of roles to fillers is captured by the localist (and conjunctive) SP units. When a proposition becomes active, its role-filler bindings are also represented dynamically, by synchrony of firing. When a P unit becomes active, it excites the SPs to which it is connected. Separate SPs inhibit one another, causing them to fire out of synchrony with one another. When an SP fires, it activates the predicate and object units beneath it, and they activate the semantic units beneath themselves. On the semantic units, the result is a collection of mutually desynchronized patterns of activation, one for each role binding. For example, the proposition *loves* (Susan, Jim) would be represented by two such patterns, one binding the semantic features of Susan to the features of *lover*, and the other binding Jim to *beloved*. The proposition *loves* (Jim, Susan) would be represented by the very same semantic units (as well as the same object and predicate units), only the synchrony relations would be reversed.

The resulting representations explicitly bind semantically rich representations of relational roles to representations of their fillers (at the level of semantic features, predicate and object units, and SPs) and represent complete relations as conjunctions of role-filler bindings (at the level of P units). As a result, they do not fall prey to the shortcomings of traditional connectionist representations (which cannot dynamically bind roles to their fillers), those of SAA (which can represent neither relational roles nor their semantic content explicitly), or those of tensors.

Hummel, Holyoak, and their colleagues have shown that LISAese knowledge representations, along with the operations that act on them, account for a very large number of phenomena in human relational reasoning, including phenomena surrounding memory retrieval, analogy making (Hummel & Holyoak, 1997), analogical inference, and schema induction (Hummel & Holyoak, 2003a). They provide a natural account of the limitations of human WM, ontogenetic and phylogenetic differences between individuals and species (Hummel & Holyoak, 1997), the relation between effortless ("reflexive"; Shastri & Ajjanagadde, 1993) and more effortful ("reflective") forms of reasoning (Hummel & Choplin, 2000), and the effects of frontotemporal degeneration (Morrison et al., 2004; Waltz et al., 1999) and natural aging (Viskontas et al., in press) on reasoning and memory. They also provide a basis for understanding the perceptual–cognitive interface (Green & Hummel, 2004), and how specialized cognitive "modules" (e.g., for reasoning about spatial arrays of objects) can work with the broader cognitive architecture in the service of specific reasoning tasks (e.g., transitive inference; Holyoak & Hummel, 2000) (see Hummel & Holyoak, 2003b, for a review).

Summary

An 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 capture the symbolic nature of human relational representations, but they 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, and, as shown by Doumas and Hummel (2004), appears inescapable.

Traditional distributed connectionist approaches have the opposite strengths and weaknesses: 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.

Connectionist models that attempt to achieve symbolic competence by using tensor products and other forms of conjunctive coding as the sole basis for role-filler binding find themselves in a strange world in between the symbolic and connectionist approaches (i.e., on the implicit relations continuum), neither fully able to exploit the strengths of the connectionist approach, nor fully able to exploit the strengths of the symbolic approach.

Knowledge representations based on dynamic binding of distributed representations of relational roles and their fillers (of which LISAese is an example) – in combination with a localist representations of roles, fillers, role-filler bindings, and their composition into complete propositions – can simultaneously capture both the symbolic nature and semantic richness of human mental representations. The resulting representations are neurally plausible, semantically rich, flexible, and meaningfully symbolic. They provide the basis for a unified account of human memory storage and retrieval, analogical reasoning, and schema induction, including a natural account of both the strengths, limitations, and frailties of human relational reasoning.

Acknowledgments

This work was supported by a grant from the UCLA Academic Senate. We thank Graeme Halford and Keith Holyoak for very helpful comments on an earlier draft of this chapter.

Notes

[1] Arguments (or roles) may suggest different shades of meaning as a function of the roles (or fillers) to which they are bound. For example, "loves" suggests a different interpretation

in *loves* (John, Mary) than it does in *loves* (John, chocolate). However, such contextual variation does not imply in any general sense that the filler (or role) itself necessarily changes its identity as a function of the binding. For example, our ability to appreciate that the "John" in *loves (John, Mary)* is the same person as the "John" in *bites (Rover, John)* demands explanation in terms of John's invariance across the different bindings. If we assume invariance of identity with binding as the general case, then it is possible to explain contextual shadings in meaning when they occur (Hummel & Holyoak, 1997). However, if we assume lack of invariance of identity as the general case, then it becomes impossible to explain how knowledge acquired about an individual or role in one context can be connected to knowledge about the same individual or role in other contexts.

2. In the most extreme version of this account, the individual processing elements are not assumed to "mean" anything at all in isolation; rather they take their meaning only as part of a whole distributed pattern. Some limitations of this extreme account are discussed by Bowers (2002) and Page (2000).
3. For example, Falkenhainer, Forbus, and Gentner's (1989) structure matching engine (SME), which uses SAA-based representations to perform graph matching, cannot map *loves* (Abe, Betty) onto *likes* (Peter, Bertha) because *loves* and *likes* are nonidentical predicates. To perform this mapping, SME must recast the predicates into a common form, such as *has-affection-for* (Abe, Betty) and *has-affection-for* (Alex, Bertha), and then map these identical predicates.
4. At first blush, it might appear that adding two vectors where one represents a relational role and the other its filler should be susceptible to the very same problem that we faced when adding two tensors where each represented a role-filler binding, namely the superposition catastrophe. It is easy to overcome this problem in the former case, however, by simply using different sets of units to represent roles and fillers so the network can distinguish them when added (see Hummel & Holyoak, 2003a). This solution might also be applied to role-filler binding with tensors, although doing so would require using different sets of units to code different role-filler bindings. This solution would require allocating separate tensors to separate role-filler bindings, thus adding a further layer of conjunctive coding and further violating role-filler independence.

References

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.

Bowers, J. S. (2002). Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. *Cognitive Psychology*, 45, 413–445.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114.

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. (2004). A fundamental limitation of symbol-argument-argument notation as a model of human relational representations. In Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society, 327–332.

Edelman, S., & Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, 27, 73–109.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press/Bradford Books.

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.

Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.

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

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. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306–355.

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

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

Green, C. B., & Hummel, J. E. (2004). Relational perception and cognition: Implications for cognitive architecture and the perceptual-cognitive interface. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 201–223). San Diego: Academic Press.

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.

Hinton, G. E. (Ed.). (1990). *Connectionist symbol processing*. Cambridge, MA: MIT Press.

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. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 229–263). Hillsdale, NJ: Erlbaum.

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 shape perception and object recognition. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 157–185). Hillsdale, NJ: Erlbaum.

Hummel, J. E. (2003). Effective systematicity in, effective systematicity out: A reply to Edelman & Intrator (2003). *Cognitive Science*, 27, 327–329.

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

Hummel, J. E., & Choplin, J. M. (2000). Toward an integrated account of reflexive and reflective reasoning. In *Proceedings of the twenty-second annual conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Hummel, J. E., & Holyoak, K. J. (1992). Indirect analogical mapping. *Proceedings of the 14th annual conference of the Cognitive Science Society*, 516–521.

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. (2003a). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220–263.

Hummel, J. E., & Holyoak, K. J. (2003b). Relational reasoning in a neurally-plausible cognitive architecture: An overview of the LISA project. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 10, 58–75.

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 93–121). Cambridge, MA: MIT Press.

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

Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, 18, 387–438.

Kim, J. J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science*, 15, 173–218.

Krawczyk, D. C., Holyoak, K. J., & Hummel, J. E. (in press). Structural constraints and object similarity in analogical mapping and inference. *Thinking and Reasoning*.

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.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211–240.

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

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, 28, 203–208.

Manzano, M. (1996). *Extensions of first order logic*. Cambridge: Cambridge University Press.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37(3), 243–282.

Marcus, G. F. (2001). *The algebraic mind*. Cambridge, MA: MIT Press.

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., Holyoak, K. J., Hummel, J. E., Chow, T., Miller, B., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, 16, 1–11.

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.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, 13, 1199–1242.

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: Oxford University Press.

Oden, D. L., Thompson, R. K. R., & Premack, D. (2001). Spontaneous transfer of matching by infant chimpanzees. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind* (pp. 471–497). Cambridge, MA: MIT Press.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral & Brain Sciences*, 23, 443–512.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Erlbaum.

Phillips, S., & Halford, G. S. (1997). Systematicity: Psychological evidence with connectionist implications. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth conference of the Cognitive Science Society* (pp. 614–619). Hillsdale, NJ: Erlbaum.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model. *Cognition*, 28, 73–193.

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.

Plate, T. A. (1994). *Distributed representations and nested compositional structure*. Unpublished doctoral dissertation, Department of Computer Science, University of Toronto, Toronto, Canada.

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.

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.

Saiki, J., & Hummel, J. E. (1998). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 227–251.

Shastri, L. (2003). Inference in connectionist networks. *Cognitive Studies*, 10, 45–57.

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.

Singer, W. (2000). Response synchronization, a universal coding strategy for the definition of relations. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 325–338). Cambridge, MA: MIT Press.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science*, 16, 1–40.

Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortoney (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.

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.

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

Tesar, B., & Smolensky, P. (1994, August). Synchronous-firing variable binding is spatio-temporal tensor product representation. *Proceedings of the 16th annual conference of the Cognitive Science Society*, Atlanta, GA.

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, 24, 363–396.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature*, 373, 515–518.

van Gelder, T. J., & Niklasson, L. (1994). On being systematically connectionist. *Mind and Language*, 9, 288–302.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (in press). Relational integration, attention and reasoning in older adults.

von der Malsburg, C. (1981). *The correlation theory of brain function*. Internal Report 81–2. Department of Neurobiology, Max-Plank-Institute for Biophysical Chemistry, Gottingen, Germany.

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.

Wharton, C. M., Holyoak, K. J., & Lange, T. E. (1996). Remote analogical reminding. *Memory & Cognition*, 24, 629–643.