Hummel, J. E., & Holyoak, K. J. (2003). 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.

Relational Reasoning in a Neurally-plausible Cognitive Architecture: An Overview of the LISA Project

John E. Hummel Keith J. Holyoak

University of California, Los Angeles

John E. Hummel
Dept. of Psychology
University of California
405 Hilgard Ave.
Los Angeles, CA 90095-1563
jhummel@psych.ucla.edu
(310) 206-9262 (office)
(310) 206-5895 (fax)

Abstract

This paper reviews our work simulating human thinking with the LISA model. Human mental representations are both flexible and structure-sensitive—properties that jointly present challenging design requirements for a model of the cognitive architecture. LISA satisfies these requirements by representing relational roles and their fillers as patterns of activation distributed over a collection of semantic units (achieving flexibility) and binding these representations dynamically into propositional structures using synchrony of firing (achieving structure-sensitivity). The resulting representations serve as a natural basis for memory retrieval, analogical mapping, analogical inference and schema induction. In addition, the LISA architecture provides an integrated account of effortless "reflexive" forms of inference and more effortful "reflective" inference, serves as a natural basis for integrating generalized procedures for relational reasoning with modules for more specialized forms of reasoning (e.g., reasoning about objects in spatial arrays), provides an a priori account of the limitations of human working memory, and provides a natural platform for simulating the effects of various kinds of brain damage.

Introduction

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; Holyoak & Thagard, 1995), 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., Palmer, 1978; Goldstone, Medin & Gentner, 1991; Hummel, 2000; Hummel & Stankiewicz, 1996).

Relational thinking is so commonplace that it is easy to assume that the psychological mechanisms underlying it are relatively simple. But this would be a mistake. The capacity to form and manipulate relational representations appears to be a late evolutionary development (Robin & Holyoak, 1995), closely tied to the substantial increase in the size and complexity of the frontal cortex in the brains of higher primates, and most dramatically in humans (Stuss & Benson, 1986). 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). Relational thinking also develops relatively late in childhood (see Smith, 1989).

This paper presents an overview of our attempts to understand and simulate the underpinnings of human relational (i.e., symbolic) thought. Our goal in this work is to understand how relational perception and thinking can be accomplished in a cognitive architecture that is both psychologically and neurally plausible. Ultimately, our aim is to understand the neurocomputational basis of symbolic thought. Two general properties of human relational thinking jointly present extremely challenging design requirements. First, relational thinking is, by definition, *structure-sensitive*: It depends on the capacity to code and manipulate relational knowledge, with complex structures emerging from the systematic recombination of more primitive elements (Fodor & Pylyshyn, 1988). Second, thinking is *flexible* in the way in which knowledge is accessed and used. People apply old knowledge to new situations that are similar but by no means identical, somehow recognizing and exploiting useful partial matches. A primary challenge for a theory of the human cognitive architecture is to satisfy both these design requirements; doing so permits a surprising number of phenomena characterizing human relational thinking to emerge as a natural consequence.

This paper is organized as follows. First, we elaborate briefly on the joint requirements of structure-sensitivity and flexibility, and their implications for models of the human cognitive architecture. Next, we present LISA (*Learning and Inference with Schemas and Analogies*)—a model explicitly designed to meet these requirements—in broad conceptual terms (for details of the LISA algorithm, including equations and parameters, see Hummel & Holyoak, 1997, 2003). We describe LISA's operation and review its account of several core cognitive phenomena, including memory retrieval, analogical mapping, analogical inference, schema induction, transitive inference, relations between relatively effortless ("reflexive") and more effortful ("reflective") forms of inference, and the effects of brain damage on relational reasoning.

Design Requirements for a Model of the Human Cognitive Architecture

The power of relational thinking resides in its capacity to generate inferences and generalizations that are constrained by the *roles* elements play, rather than simply by the properties of the elements themselves. In the limit, relational reasoning yields universal inductive generalization from a finite and often very small set of observed cases. An important example of relational reasoning is analogical inference. Suppose someone knows that John loves Mary, Mary loves Sam, and John is jealous of Sam (a *source* analog). The reasoner then observes that Sally loves Tom, and Tom loves Cathy (a *target* analog). A plausible analogical inference is that Sally will be jealous of Cathy. This inference is based on the relational correspondences (i.e., participation in corresponding relational roles) of Mary to Tom, John to Sally, and Sam to Cathy. Making a plausible inference thus requires the reasoner to discover those correspondences, even in the face of potentially competing correspondences based on shared semantic features (e.g., Mary corresponds to Tom based on their shared roles in the *loves* (*x*, *y*) relation, even though she is more similar to Sally by virtue of being female).

This same analysis can be extended to understand reasoning based on the application of schemas and rules. Consider a person who has learned a schema or rule of the general form, "If person1 loves person2, and person2 loves person3, then person1 will be jealous of person3," and now encounters Sally, Tom and Cathy. By mapping Sally to person1, Tom to person2 and Cathy to person3, the reasoner can infer that Sally will be jealous of Cathy by extending the mapping to place Sally in the (unrequited) lover role (person1), Tom in the both-lover-and-beloved roles (person2), and Cathy in the beloved role (person3). When the schema has an explicit "if-then" form, as in this example, the inference can readily be interpreted as the equivalent of matching the left-hand side ("if" clause) of a production rule and then "firing" the rule to create an instantiated form of the right-hand side ("then" clause), maintaining consistent local variable bindings across the "if" and "then" clauses in the process (Anderson, 1993; Newell, 1973).

There is also a close connection between the capacity to make relational inferences and the capacity to learn variablized schemas or rules. A plausible bottom-up basis for acquiring the "jealousy" schema would be induction from multiple specific examples. For example, if the reasoner observed John being jealous of Sam, and Sally being jealous of Cathy, and was able to establish the relational mappings and observe the commonalities among the mapped elements (e.g., John and Sally are both people in love, and Sam and Cathy are both loved by the object of someone else's affection), then inductive generalization could give rise to the jealousy schema.

The computational requirements for discovering these correspondences—and for making the appropriate inferences—include the following (Hummel & Holyoak, 2003):

1. The underlying representation must bind fillers (e.g., John) to roles (e.g., lover). Knowing only that John, Mary and Sam are in various love relations will not suffice to infer that John will be jealous of Sam (e.g., as opposed to Sam being jealous of John).
2. Each filler must maintain its identity across multiple roles (e.g., the same Mary is both a lover and a beloved), and each role must maintain its identity across multiple fillers (e.g., *loves* (*x*, *y*) is the same basic relation regardless of whether the lovers are John and Mary or Sally and Tom). That is, the representation of a role must be independent of its fillers, and vice versa (Holyoak & Hummel, 2000; Hummel & Biederman, 1992; Hummel & Holyoak, 2003). If the representation of *loves* (*x*, *y*) depended on who loved whom (e.g., with one set of units for *Sally-as-lover* and a completely separate set for *John-as-lover*), then there would be no basis for appreciating that Sally loving Tom has anything in

common with John loving Mary. And if the representation of Mary depended on whether she is a lover or is beloved, then there would be no basis for inferring that John will be jealous of Sam: it would be as though John loved one person, and some other person loved Sam.

3. Elements of the source and target can be mapped on the basis of their relational (i.e., role-based) correspondences. For example, we know that Sally maps to John because she also is a lover, despite the fact that she has a more direct similarity to Mary by virtue of being female.
4. The inference depends on an extension of the initial mapping. Because Sally maps to John on the basis of her role as an unrequited lover, it is inferred that she, like John, will be jealous.
5. The symbols that compose these representations must have semantic content. In order to induce a general *jealousy* schema from the given examples, it is necessary to know what the people and relations involved have in common and how they differ.

Together, requirements (1) - (4) imply that the human cognitive architecture is a symbol system, in which relations and their arguments are represented independently but can be bound together to compose propositions. (5) implies that the symbols in this system are distributed representations that explicitly specify the semantic content of their referents. The capacity to satisfy these requirements is a fundamental property of the human cognitive architecture, and places strong constraints on models of that architecture (Hummel & Holyoak, 1997).

The LISA Model

LISA's knowledge representations, and the operations that act on them, are explicitly designed to satisfy the five requirements outlined above. Our proposal is a form of *symbolic connectionism*: a cognitive architecture that codes relational structures by dynamically binding distributed representations of roles to distributed representations of their fillers (Holyoak & Hummel, 2001; Hummel & Holyoak, 1997, 2003). The resulting architecture far exceeds the capacity of non-symbolic architectures (such as traditional connectionist architectures; see Marcus, 1998, 2001) to capture the power of human relational reasoning. It also exceeds the capacity of traditional symbolic architectures (such as production systems; e.g., Anderson, 1993) to represent the semantic content of roles and their fillers, and to exploit those representations in the service of flexible automatic generalization (e.g., as exhibited by distributed connectionist representations).

Knowledge Representation

LISA's knowledge representations are based on a hierarchy of distributed and localist units that collectively represent the semantic features of objects and relational roles, and their arrangement into complete propositions (see Figure 1). At the bottom of the hierarchy, *semantic units* (small circles in Figure 1) represent objects and relational roles in a distributed fashion. For example, Tom might be represented by features such as *human*, *adult*, and *male* (along with units representing his profession, personality traits, etc.), and Sally 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 1) represent objects and relational roles in a localist fashion, and share bidirectional excitatory connections with the corresponding semantic units. *Sub-proposition* units (*SPs*; rectangles in Figure 1) represent bindings of relational roles to their arguments (which can either be objects, as in Figure 1a, or complete propositions, as in Figure 1b). 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 1). 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, 2003).

Please Insert Figure 1 about here

A complete analog (i.e., story, situation or event) is represented by the collection of P, SP, predicate, object and semantic units that code its propositional content (see Figure 2). 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., Tom is represented by the same unit in both *loves* (Sally, Tom) and *loves* (Tom, Cathy)), but a given element is represented by separate units in separate analogs. The localist units thus represent *tokens* of individual objects, relations or propositions in particular situations (i.e., analogs). By contrast, the same semantic units represent an object or relational role in all the analogs in which that element plays a part: Although the localist units represent tokens, the semantic units represent *types* (Hummel

& Holyoak, 2003).

Please Insert Figure 2 about here

The hierarchy of units depicted in Figures 1 and 2 represent propositions both in LISA's long-term memory (LTM) and in its working memory (WM). In this representation, the binding of roles to fillers is captured by the localist SP units. When a proposition enters WM (i.e., when it becomes active), its role-filler bindings are also represented *dynamically*, by synchrony of firing (Hummel & Holyoak, 1992, 1997). When a P unit becomes active it excites the SPs to which it is connected. Separate SPs under the same proposition 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* (Sally, Tom) would be represented by two such patterns, one binding the semantic features of Sally to the semantic features of *lover*, and the other binding Tom to *beloved*. The proposition *loves* (Tom, Sally) 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 (with *lover* firing in synchrony with Tom, and *beloved* with Sally) (see Figure 3).

Please Insert Figure 3 about here

The resulting representations satisfy the five requirements on relational representations outlined above. The explicit coding of semantic content (requirement 5) is captured by the distributed representation of objects and relational roles on the semantic units. Roles are bound to their fillers (requirement 1) both in terms of the localist SP units and by synchrony of firing. And although the conjunctive role+filler SP units violate role-filler independence (requirement 2; note that the same role, bound to different fillers, requires different SPs), this requirement is satisfied both at the level of the semantic units and at the level of the object and predicate units. As detailed in the next section, this representation permits LISA to map objects based on their relational roles (requirement 3), to make inferences by extending these mappings (requirement 4), and to use the resulting mappings to induce generalized schemas from examples.

Augmented with a few simple operations, LISA's symbolic-connectionist knowledge representations provide a natural basis for memory retrieval, analogical mapping, analogical inference and schema induction. They also permit LISA to communicate with specialized modules for perceptual processing and other specialized functions, such as comparison and causal induction.

Memory Retrieval

LISA performs analog retrieval—retrieving a familiar source analog or schema from LTM given a novel target as a cue—as a form of guided pattern recognition (Hummel & Holyoak, 1997). Analogs are divided into two mutually exclusive sets: A *driver*, which is the current focus of attention, and one or more *recipients*. During analog retrieval, the driver is assumed to reside in *active memory*—a subset of LTM that is less active than WM, but primed for retrieval into WM (Cowan, 1995)—and the recipients are assumed to reside in LTM. One at a time, propositions in the driver become active, generating mutually de-synchronized patterns of activation on the semantic units (one pattern for each role filler binding). These patterns excite units in LTM, bootstrapping memory retrieval. For example, the semantic patterns generated by *loves* (Sally, Tom) will tend to strongly excite other propositions about females loving males, and more weakly excite propositions about males loving females (as well as propositions about other relations between women and men). Propositions in the recipient analog(s) compete via lateral inhibition to become active in response to the patterns generated by the driver, and analogs compete with one another to be retrieved into active memory for the purposes of mapping.

Analogical Mapping

Once a recipient analog has been retrieved into active memory, it becomes a candidate for analogical mapping—discovering the relational correspondences between the elements of the driver and those of the recipient. For the purposes of mapping (as well as analogical inference and schema induction), LISA updates the weights on *mapping connections* between units of the same type in different analogs. Every P unit in the driver has the potential to learn a mapping connection to every P unit in the recipient; likewise, SPs, predicate and object units can learn mapping connections across analogs. The weights on these mapping connections grow more positive when the units they connect are active at the same time, and serve both to keep track of which units in the driver correspond to which in the recipient and to allow correspondences discovered early in mapping to constrain mappings discovered later. For example, consider the following simple analogy. Analog 1 states that Tom loves Sally, and Tom bought a Honda. Analog 2 states that Mike loves Emily, Mike bought a Mazda, and Bill bought a Toyota. Considering only the propositions *buy* (Tom, Honda), *buy* (Mike, Mazda) and *buy* (Bill Toyota), the mappings of Tom to Mike vs. Bill, and Honda to Mazda vs. Toyota are ambiguous (assuming that Tom is roughly as similar to Mike as to Bill, and that a Honda is roughly as similar to a Mazda as to a Toyota). But when *loves* (Tom, Sally) and *loves* (Mike, Emily) are taken into consideration, it is clear that Tom corresponds to Mike and therefore the Honda corresponds to the Mazda.

LISA discovers these mappings as follows. Let Analog 1 serve as the driver. When *loves* (Tom, Sally) fires, the semantic features activated by *Tom* will tend to excite *Mike* and *Bill* equally. However, the features of *lover* and *beloved* in Analog 1 will excite *lover* and *beloved* in Analog 2 much more strongly than they excite the roles of the *buy* relation (likewise, the features of Sally will excite Emily much more than they excite either Mazda or Toyota). As a result, the units representing *loves* (Mike, Emily) will become active, inhibiting *buy* (Mike, Mazda) and *buy* (Bill, Toyota) to inactivity. At the level of specific role bindings, the semantic patterns generated by *lover+Tom* in the driver will activate *lover+Mike* in the recipient, and *beloved+Sall*y will activate *beloved+Emily*.. As a result, LISA will learn mapping connections between *Tom* and *Mike*, between *Sally* and *Emily*, and between corresponding predicate, SP and P units. When *buy* (Tom, Honda) fires, the semantic pattern activated by *Tom* will again equally excite both *Mike* and *Bill*, and the pattern generated by *Honda* will equally excite *Mazda* and *Toyota*. However, the mapping connection between *Tom* and *Mike* will cause *Tom* to excite *Mike* directly (i.e., in addition to *Mike*'s "bottom-up" semantic excitation). *Mike* will therefore become active, drive *Bill* to inactivity, and allow *buy* (Mike, Mazda) to drive *buy* (Bill, Toyota) to inactivity: LISA will learn that, in this simple example, the Honda corresponds to the Mazda rather than the

Toyota. Similar principles govern LISA's discovery of analogical mappings in much more complex domains. And although the Honda-Mazda example is extremely simple, it also illustrates LISA's ability to map objects based on shared relational roles (requirement 3 above): The semantics of Toyotas vs. Mazdas are not sufficient (in this example) to disambiguate their mapping to the Honda, but LISA can use the Mazda's role in the *buy* (Bill, Mazda) relation, along with Bill's previously established mapping to Tom, to discover that it maps to Honda.

As this example illustrates, the same process of semantically-based guided pattern recognition that drives analog retrieval also plays a central role in LISA's algorithm for analogical mapping. Indeed the primary difference between retrieval and mapping is that LISA is allowed to learn mapping connections during mapping, but not during retrieval. Hummel and Holyoak (1997) showed that this single difference allows LISA to account for a complex pattern of similarities and differences between retrieval and mapping.

Analogical Inference

Augmented with a simple algorithm for *self-supervised* learning—learning that is neither externally supervised like back propagation (Rumelhart, Hinton & Williams, 1986) nor purely statistical like traditional unsupervised learning (e.g., Marshall, 1995)—LISA's algorithm for analogical mapping also serves as a basis for analogical inference (Hummel & Holyoak, 2003). When a unit, i, in the driver learns an excitatory mapping connection to a given unit, j, in the recipient, it also learns a global inhibitory mapping connection to and from all units other, $k \neq j$, in the recipient. Similarly, j learns a global inhibitory connection with all units, $l \neq i$, in the driver. These global inhibitory connections play an essential role in self-supervised learning. Recall the "jealousy" analogy discussed previously, in which John loves Mary, Mary loves Sam, and John is jealous of Sam (Analog 1), and Sally loves Tom, but Tom loves Cathy (Analog 2). Let Analog 1 be the driver, and imagine that LISA has mapped *loves* in Analog 1 to *loves* in Analog 2, John to Sally, Mary to Tom, and Sam to Cathy (see Hummel & Holyoak, 2003, for details on how LISA discovers these mappings).

Once these mappings have been discovered, every unit in Analog 2 will have a positive mapping connection to *some* unit in Analog 1; as a result, every unit in Analog 2 will have a global *inhibitory* mapping connection to all *other* units in Analog 1. For example, the *lover* role unit in Analog 2 will have an excitatory mapping connection with *lover* in Analog 1, but it will have inhibitory mapping connections with *beloved* in Analog 1, and with the roles of the *jealousy* relation. Importantly, nothing in Analog 2 maps to *jealous-of* (John, Sam) in Analog 1. Therefore, when *jealous-of* (John, Sam) fires, everything in Analog 2 will be globally inhibited (except for Sally, which is excited by John, and Cathy, which is excited by Sam): The roles of the *jealous-of* relation will inhibit all the predicate units in Analog 2; the SPs for *jealous*+John and *jealousy-target*+Sam will inhibit all the SPs in Analog 2; and the P unit for the proposition as a whole will inhibit all the P units in Analog 2.

This kind of generalized inhibition occurs only when (a) all the units in the recipient already map to some unit in the driver, and (b) none of the units in the recipient map to the currently active unit(s) in the driver. That is, generalized mapping inhibition indicates that nothing in the recipient maps to whatever is currently active in the driver. LISA therefore interprets it as a signal to initiate self-supervised learning: Whenever all units of a given class (object, predicate, SP or P) are inhibited in the recipient, LISA recruits a new unit of that class to correspond to whatever (unmapped) unit in the driver is causing the inhibition. In the case of the current example, LISA will recruit a new P unit, two new SPs, and two new predicate units.

Newly recruited units are assigned positive mapping connections with the driver units that initiated their recruitment, and they are connected to one another, to existing units in the recipient, and (in the case of predicate and object units) to semantic units, by simple Hebbian learning. That is, these units learn connections to whatever other units they happen to be co-active with. The two new predicate units in the current example are initiated by *jealous* and *jealousy-target* in the driver. As a result, each new predicate unit will fire in synchrony with one of those roles, and thus with its semantic features: One new predicate unit will learn connections to the semantic features of *jealous*, and the other will learn connections to the semantic features of *jealousy-target*. In this way. the newly-recruited predicate units come to represent the *jealousy* relation. Directly analogous operations cause one of the newly recruited SPs to learn excitatory connections to *jealous* and Sally, the other SP to learn excitatory connections to *jealousy-target* and Cathy, and both SPs to learn excitatory connections to the newly recruited P unit. Together, these units encode the proposition *jealous-of* (Sally, Cathy): Based on the analogy to John, Mary and Sam, LISA has inferred that Sally will be jealous of Cathy (see Hummel & Holyoak, 2003, for details).

Schema Induction

LISA's self-supervised learning algorithm also plays a central role in its ability to induce generalized schemas or rules from examples. The key difference between schema induction and analogical inference is that schema induction selectively encodes what the examples have in common, placing less emphasis on details on which they differ. In the case of our simplified jealousy example, the schema to be induced consists of the propositions *loves* (person1, person2), *loves* (person2, person3) and *jealous-of* (person1, person3), where person1, person2 and person3 are generic people rather than specific individuals.

LISA induces this schema as follows. The schema-to-be starts out as a third "analog" that is initially devoid of units. Self-supervised learning in the schema works exactly like self-supervised learning in any other analog (e.g., for analogical inference): New units are recruited in response to evidence that none of the existing units map to the currently-active units in the driver, and they learn connections to one another and to the semantic units based on their co-activity. Object and predicate units learn connections to semantic units in such a way that the connection weight tends to mimic the activation of the semantic unit: The more active the semantic unit, the more strongly the predicate or object unit will connect to it (the learning algorithm is quite simple; see Hummel & Holyoak, 2003).

During analogical mapping, excitation in the recipient is allowed to flow both "bottom-up", from objects and predicates to SPs and from SPs to P units, and "top-down", from P units to SPs, from SPs to predicates and objects, and from predicates and objects to semantic units. Semantic units update their activations by simply normalizing their inputs: The activation of a semantic unit is its input divided by the largest input to any semantic unit. As a result, semantic units that are shared by active objects or predicates in both the driver and the recipient tend to become about twice as active as semantic units that are unique to one analog or the other. For example, consider what happens when LISA maps *loves* (John, Mary) onto *loves* (Sally, Tom). Imagine that the object unit for John is connected to semantic units for *human*, *adult*, *male* and *John*, and Sally is connected to *human*, *adult*, *female* and *Sally*. When *lover+John* fires in the driver, *lover+Sally* responds in the recipient. As a result, *human* and *adult* each receive two sources of input (i.e., from both John and Sally), whereas *male*, *John*, *female* and *Sally* each receive one. Due to the normalization, *human* and *adult* will end up with activations close to 1.0, and the other units will have activations close to 0.5. The newly recruited object unit in the schema that corresponds to John/Sally will therefore learn connections of about 1.0 to the semantic units *human* and *adult*, and connection strengths near 0.5 to *male*, *female*, *John* and *Sally*. The resulting unit is one that strongly prefers to be activated by human adults, weakly prefers people named "Sally" and "John", and weakly (and equally) prefers both males and females. In other words, this unit is more a generic "person" unit than a unit for any particular person. Analogous operations will cause the schema to learn generic "person" units for Mary/Tom and Sam/Cathy.

LISA's algorithm for schema induction performs a kind of intersection discovery. At the level of individual objects and predicates, it performs intersection discovery on the semantic representation of those objects and predicates. LISA also performs intersection discovery at the level of entire propositions. If a given proposition in the driver never fires (e.g., because it is pragmatically unimportant; Spellman & Holyoak, 1996; see also below), then it will never map to the recipient, and no corresponding proposition will be learned in the emerging schema. In this way, propositions describing unimportant details do not find their way into the emerging schema. Running LISA's schema-induction algorithm iteratively—by mapping schemas onto other schemas—results in the induction progressively more abstract schemas, and in the limit can yield completely abstract, universally quantified rules (Hummel & Holyoak, 2003).

Working Memory and the Flow of Control

Dynamic binding of roles to their fillers in WM is inherently capacity-limited: It is only possible for LISA to keep a finite number of role-filler bindings simultaneously active and mutually *out* of synchrony with one another. This capacity limit is a fundamental mathematical consequence of the algorithm LISA uses to keep separate bindings out of synchrony (see Hummel & Holyoak, 2003, Appendix A). Human WM is similarly limited, with a capacity of about four to five role bindings (see Cowan, 2000, for a review). Because of this capacity limit, processing in LISA is necessarily sequential: LISA can hold at most two or three propositions in WM simultaneously, so it must map large analogies in small pieces. As illustrated with the Honda/Mazda example, the mappings LISA discovers early can have a large effect, for better or worse, on the mappings it discovers later. In order to map multiple propositions in parallel, LISA must place them into WM together. Accordingly, LISA's mapping performance varies greatly as a function (a) the order in which it maps propositions, and (b) which propositions, if any, it places into WM together. This is especially true for relationally complex analogies, in which the correct mappings cannot be discovered solely on the basis of the semantic features of the stated relations.

The order in which propositions are chosen to fire in LISA is stochastic, but there are three constraints on the probability, p(i), with which proposition i will be chosen to fire at any given time. The first constraint is based on a proposition's *pragmatic centrality*, c(i), (Holyoak & Thagard, 1989; Spellman & Holyoak, 1996)—the degree to which the fact stated by the proposition is central to the system's processing goals: p(i) increases with c(i), so that more important propositions tend to fire earlier and more often than less important ones. The second constraint is based on *support* relations between propositions. Propositions tend to support one another if they are on the same causal chain (e.g., j will support i if j describes a cause and i describes the effect of that cause), if they are both arguments of a single higher-order relation, or if they share arguments. To the degree that j supports i, j's firing at time t increases the likelihood that i will fire at time t+1. Support relations cause propositions to fire in an order that honors the principles of text coherence. That is, they tend to fire in an order in which they would appear in a well-written narrative. Finally, in order to keep LISA from repeatedly firing the same propositions, p(i) is inversely proportional to the recency with which i has fired in the past.

LISA's mapping algorithm is relatively powerful, although it is much less powerful than explicit graph mapping, as performed by models such as SME (Falkenhainer, Forbus & Gentner, 1989) or parallel constraint satisfaction, as performed by the ACME model (Holyoak & Thagard, 1989). LISA can solve many analogies with ease, provided it is allowed to fire propositions in a suitable order. As a result, much (but by no means all) of the variability in LISA's performance as an analogical mapping engine is due to the constraints governing the order in which it fires propositions (see Hummel & Holyoak, 1997, 2003, for details). A general prediction of the LISA model is that similar principles should apply in human cognition: The order and manner in which one thinks about a problem should have profound effects on one's ability to solve the problem. Kubose, Holyoak and Hummel (2003) provide preliminary support for this prediction.

An important open problem in the development of LISA is to specify constraints on whether two or more propositions should be placed into WM together. As a general default, we have assumed that LISA places one proposition into WM at a time whenever possible, based on the assumption that cognitive effort increases with the number of role bindings that must be represented simultaneously, and that people tend to minimize cognitive effort. When it is necessary to place multiple propositions into WM simultaneously in order to solve a problem, we must directly tell the model to do so. We have yet to develop routines that will allow LISA to decide for itself when it is necessary to consider multiple propositions in parallel.

Transitive Inference: Integrating Mapping with Other Perceptual/Cognitive Modules

A relation, *r*, is transitive if $r\ (x, y)$ and $r\ (y, z)$ jointly imply $r\ (x, z)$. For example, the relation *larger-than* is transitive because *larger-than* (x, y) and *larger-than* (y, z) jointly imply *larger-than* (x, z). Relations over metric dimensions are transitive. Because transitive relations are logically well-behaved (in the sense that $r\ (x, y)$ and $r\ (y, z)$ always imply $r\ (x, z)$), it is reasonable to assume that people might reason about them by applying rules of logical inference. However, numerous findings suggest that people do not reason about transitive relations in this way. In particular, some transitive inferences are easier to make than others (e.g., Clark, 1969; DeSoto, London & Handel, 1965; Holyoak & Patterson, 1981; Huttenlocher, 1968; Sternberg, 1980): The order in which the premises are stated matters (e.g., reasoning about premises stated in the premises in the order *x*, *y*, y *z* is easier than reasoning about them stated in the order *y, z* x, y), and reasoning from premises stated in terms of "unmarked" relations (e.g., "larger than" rather than "smaller than") is easier than reasoning about premises stated in either marked terms or a combination of marked and unmarked terms (see, e.g., Sternberg, 1980). These and related findings suggest that people do not reason about transitive relations simply by applying rules of deductive logic. Instead, human performance in reasoning about transitive relations has led numerous researchers to hypothesize that we reason about transitive relations by mapping objects onto positions in a metric array based on their stated relations (e.g., given *larger-than* (x, y) and *larger-than* (y, z), mapping *x* to the top, *y* to the middle and *z* to the bottom of the array), and then "reading off" additional relations (e.g., *larger-than* (x, z), *largest* (x), *smallest* (z)) based on their locations in the array (e.g., Holyoak & Patterson, 1981; Huttenlocher, 1968).

This idea is intuitive and has received some general empirical support (e.g., Holyoak & Patterson, 1981; Huttenlocher, 1968; Sternberg, 1980; Woocher, Glass, & Holyoak, 1978), but it was not fleshed-out with a process model until Hummel and Holyoak (2001) augmented LISA with a "Metric Array Module" (MAM) for reasoning about transitive relations. The MAM is a specialized module that uses a set of visual routines borrowed from Hummel and Biederman's (1992) model of object recognition, along with a collection of heuristics, to map objects to locations in a metric array based on stated categorical relations, and to "read off" additional relations based on those locations. The resulting model provides a better fit to the extensive data of Sternberg (1980) than his own mathematical model does (and Sternberg's model does not provide an algorithmic account of his data). More importantly for our current purposes, LISA+MAM provides a concrete illustration of how it is possible to integrate the LISA architecture with specialized modules for computing particular functions.

Reflexive and Reflective Reasoning

Analogical reasoning of the kind LISA performs in *reflective*, in the sense that it requires effort in the form of attention and WM resources. Although reflective reasoning clearly underlies many important types of inference, people also make myriad inferences that appear much more automatic and effortless. For example, if told that Tom sold his car to Mary, one will with little apparent effort infer that Mary now owns the car. Shastri and Ajjanagadde (1993) refer to such inferences as *reflexive*. Even more reflexive is the inference that Tom is an adult human male and Mary an adult human female (Hummel & Choplin, 2000). Reflexive reasoning of this sort figures centrally in event and story comprehension (see Shastri & Ajjanagadde, 1993), and also manifests itself as context effects on the interpretation of objects and predicates. For example, told that Tom loves Mary, one is likely to interpret the meaning of “loves” differently than one would told that Tom loves chocolate: The indirect object (Mary vs. chocolate) acts as a context that influences the interpretation of the verb “loves”.

Hummel and Choplin (2000) describe an extension of LISA that accounts for such context effects, as well as other kinds of reflexive inferences. The model does so using the same operations that allow intersection discovery for schema induction—namely, feedback from recipient analogs to semantic units. The basic idea is that instances (i.e., analogs) in LTM store the semantic content of the objects and relations instantiated in them (i.e., the object and predicate units in LTM are connected to specific semantic units), and the population of such instances serves as an implicit record of the covariation statistics of the semantic properties of roles and their typical fillers. In the episodes in which a person loves a food, the "loving" involved is of a culinary variety rather then a romantic variety, so the *loves* predicate units are attached to semantic features specifying "culinary love"; in the episodes in which a person loves another person, the love is of a romantic or familial variety so the predicates are attached to semantic features specifying romantic or familial love. Told that Tom loves Mary, the reasoner is reminded of more situations in which people are bound to both roles of the loves relation than situations in which a person is the lover and a food the beloved, so instances of romantic love in LTM become more active than instances of culinary love. Told that Tom loves chocolate, the reasoner is reminded of more situations in which a food is bound to the beloved role, so instances of culinary love become more active than instances of romantic love.

The Hummel and Choplin (2000) extension of LISA instantiates these ideas using the same recipient-to-semantic feedback that drives schema induction. The only difference is that, in the case of reflexive inferences, the recipient analogs are "dormant" in LTM (i.e., not yet in active memory and therefore not candidates for learning mapping connections), whereas in the case of schema induction, the analog(s) in question are in active memory. Feedback from dormant analogs causes semantic units to become active to the extent that they correspond to instances in LTM. When "Tom loves chocolate" is the driver, many instances of "culinary love" are activated in LTM (because a food is the object of the relation), so the semantic features of culinary love become more active than the features of romantic love; when "Tom loves Mary" is the driver, more instances of romantic love are activated (because a person is the object), so the features of romantic love become active (see Hummel & Choplin, 2000, for details). For the purposes of reflexive inference, predicate and object units in the driver are permitted to learn connections to active semantic units, thereby "filling in" their semantic interpretation. When the driver states that Tom loves Mary, the active semantic features specify romantic love, so loves is connected to—i.e., interpreted as—an instance of romantic love. When the driver states that Tom loves chocolate, the active semantic features specify culinary love, so loves is interpreted as culinary love.

Note that the difference between reflexive and reflective reasoning in LISA is directly analogous to the difference between analog retrieval and analogical mapping: Retrieval is performed in the same manner as mapping, except that mapping connections are learned during mapping but not during retrieval; and reflexive inference is like reflective inference (both rely on feedback from recipient analogs to semantic units), except that reflective inference depends on mapping connections, whereas reflexive inference does not. The principles that allow LISA to integrate analog retrieval with analogical mapping are thus largely the same as those that allow it to integrate reflexive with reflective inference.

Neural Underpinnings of Symbolic Thought

One of the central theoretical tenets of the LISA model is that analogical mapping and all the other processes that depend on it (including analogical inference, schema induction, specialized modules such as MAM.) depend on WM. This claim is supported by extensive empirical evidence. For example, experiments utilizing dual-task methodology have shown that the processes of binding and mapping used in analogical reasoning require WM (Waltz et al., 2000). Other studies have shown that WM is at least in part realized in subareas of prefrontal cortex, which have been implicated both in tasks that involve simple short-term maintenance and in more complex manipulation tasks characteristic of the central executive (see Miller & Cohen, 2001, for a review). For example, Kroger et al. (2002) used neuroimaging techniques to link complexity of relational processing in a reasoning task to activation of areas in dorsolateral prefrontal cortex.

We have recently begun to apply LISA as a model of the neural basis of analogical reasoning. Morrison et al. (submitted) found that elderly patients with a degenerative disease that had impacted either the frontal or temporal cortex exhibited deficits in analogical reasoning relative to age-matched controls, and that the underlying basis for their deficits appears to differ. In one experiment, participants were asked to solve 2-alternative verbal 4-term analogy problems such as

BLACK:WHITE::NOISY:__?__ (1) QUIET (2) NOISIER

in which the association strength of the analogical response was greater than, equal to, or less than that of the foil. These problems can be solved on the basis of a single relation, and hence even frontal-variant patients are capable of performing the necessary mapping and making the analogical response when its association strength was at least equal to that of the foil. However, frontal patients were selectively impaired at solving problems in which the association strength of the foil exceeded that of the analogical response. According to LISA, such problems require rapid learning of the analogical response so that its activation can be maintained, coupled with inhibition of the more highly-associated foil. Temporal patients, in contrast, were more uniformly impaired across all problem types, as would be expected if the source of their reasoning deficit involved neither analogical mapping nor inhibition, but rather loss of the underlying conceptual knowledge in LTM required to encode the relation between the A and B terms of the analogy.

We were able to simulate the observed pattern of frontal-lobe deficits by impairing the rate of rapid learning of analogical connections, coupled with reduction of inhibitory control. Both rapid learning (Assad, Rainer & Miller, 1998) and inhibition (Miller & Cohen, 2001) appear to be key functions of frontal cortex. LISA simulations demonstrated that “lesions” that compromise rapid learning mechanisms in WM and also inhibitory control yield deficits in analogical reasoning qualitatively similar to those observed with frontal-lobe patients, whereas random destruction of connections within semantic memory yield deficits qualitatively similar to those observed with temporal-lobe patients. We have also used LISA to simulate changes in performance on analogical reasoning tasks due to normal aging (Viskontas et al., submitted). These recent studies constitute an initial step toward the goal of understanding the mechanisms of high-level human reasoning at a level that makes contact with what is known about cortical functioning.

Summary

At the heart of LISA as a theory of the human cognitive architecture are a few core principles concerning the nature of knowledge representation, and the nature of the processes that operate on those representations. In terms of representation, we assume that the mind represents propositions as hierarchical structures that code the semantic content of objects and relational roles in a distributed fashion. The mind's symbolic code also includes tokens for relational roles, their fillers, role-filler bindings, and even complete propositions in a localist fashion. We assume that two levels of this hierarchy—semantic features and tokens for roles and fillers—represent relational roles independently of their fillers in order to permit relational generalization from familiar cases (source analogs and schemas) to novel ones (target analogs). We assume that in working memory, independent representations of roles and fillers are bound together dynamically by synchrony of firing.

In terms of the processes that act on these representations, LISA postulates that memory retrieval and reflexive inference are performed as a form of guided pattern recognition, in which semantic patterns generated by the driver—i.e., the current focus of attention, which at least initially, will be the target analog—activate instances (potential source analogs) in LTM. This activation of instances in LTM serves both to retrieve potential source analogs for mapping (analog retrieval) and to augment the semantic representation of the target itself (reflexive inference). We assume that analogical mapping and reflective inference (including analogical inference and schema induction) operate on driver and recipient analogs that reside in active memory. These two processes exploit the same process of guided pattern recognition that drives retrieval and reflexive inference, and they both depend on the learning of new mapping connections between the elements of the driver (which at the point of mapping may be either the source or the target) and the recipient. We assume that analogical inference is simply a process of analogical mapping augmented with an algorithm for self-supervised learning, and we assume that schema induction is simply self-supervised learning augmented with intersection discovery.

These relatively simple assumptions have allowed us to simulate large bodies of data in the literatures on analogical retrieval and mapping (Hummel & Holyoak, 1997) and analogical inference and schema induction (Hummel & Holyoak, 2003). With a few additional assumptions, which are entirely consistent with the core principles of LISA, we have been able to extend the model to account for related reasoning tasks, such as transitive inference, and we have begun to connect the key processes involved in human reasoning with their underlying neural substrate.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

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

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, *76*, 387-404.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2000). The magical number 4 in STM. *Behavioral and Brain Sciences*. ####

DeSoto, C., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, *2*, 513-521.

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. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

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

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 36-355.

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

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., & Patterson, K. K. (1981). A positional discriminability model of linear order judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *7,* 1283-1302.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295-355.

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

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

Hummel, J. E., & 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 L. R. Gleitman & A. K. Joshi (Eds.) *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 232-237). Mahwah, NJ: Erlbaum.

Hummel, J. E., & Holyoak, K. J. (1992). Indirect analogical mapping. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 516-521). Hillsdale, NJ: Erlbaum.

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. (2001). A process model of human transitive inference. In M. L. Gattis (Ed.), *Spatial schemas in abstract thought* (pp. 279-305). Cambridge, MA : MIT Press.

Hummel, J. E., & Holyoak, K. J. (in press). A Symbolic-connectionist theory of relational inference and generalization. *Psychological Review*.

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.

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, *75*, 550-560.

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.

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of dorsolateral and anterior prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 12,* 477-485.

Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (in press). The role of textual coherence in incremental analogical mapping. *Memory and Language*.

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

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

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

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167-202.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T., Miller, B. N., & Knowlton, B. J. (submitted). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal dementia.

Newell, A. (1973). Production systems: Models of control structures. In W. C. Chase (Ed.), *Visual information processing* (pp. 463-526). New York: Academic Press.

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

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

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.

Rumelhart, D. E., & Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1 (pp. 318-362). Cambridge, MA: MIT Press.

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

Smith, L. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadov and A. Ortney (Eds.) Similarity and analogical reasoning, ###.

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

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology*, *31*, 307-346.

Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, *109*, 119-159.

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

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (submitted). Relational integration and reasoning across the lifespan.

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.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28,* 1205-1212.

Woocher, F. D., Glass, A. L., & Holyoak, K. J. (1978). Positional discriminability in linear orderings. *Memory & Cognition*, *6*, 165-173.

Figure Captions

Figure 1. (a) Representation of the proposition *loves* (Sally, Tom) in LISA. Objects and relational roles are represented as patterns of activation distributed over *semantic* units (small circles) and are coded locally as *object* units (large circles) and *predicate* units (triangles) that share bi-directional excitatory connections with the corresponding semantic units. *Sub-proposition* units (*SPs*; rectangles) bind relational roles to their fillers in a localist fashion, and *proposition* (*P*) units (ovals) bind separate role-filler bindings into complete propositions. (b) LISA's representation of the hierarchical proposition *knows* (Tom, *loves* (Sally, Tom). To represent a hierarchical proposition, a P unit representing the lower-level proposition (here, *loves* (Sally, Tom)) take the place of an object unit under the appropriate SP of the higher-level proposition (here, the SP for *what is known*).

Figure 2. LISA represent of two complete analogies. Analog 1 states that John loves Mary, Mary loves Sam, and John is jealous of Sam. Analog 2 states that Sally loves Tom and Tom loves Cathy. Analogs do not share localist units for objects, relational roles, role-filler bindings (SPs) or propositions. For example, the roles of the loves relation are represented by one pair of units in Analog 1 (L_1 and L_2 under Analog 1) and by a completely separate pair of localist units in Analog 2 (L_1 and L_2 under Analog 2). However, predicate and object units in all analogs are connected to the same set of semantic units. For example, L_1 in Analog 1 would be connected to many of the same (or all of the same) semantic units as L_1 in Analog 2.

Figure 3. Illustration of the use of synchrony of firing for dynamic binding in working memory. Rows represent units of different kinds (e.g., the row labeled "*loves* (Sally, Tom)" represents the

P unit for that proposition). Names of units representing predicates and predicate semantics are written in italics; units for objects and object semantics are not italicized. Each graph depicts the activation of the corresponding unit over time. To represent the proposition *loves* (Sally, Tom), object semantics representing Sally (e.g., human adult and female) and the object unit for Sally fire in synchrony with predicate semantics for *lover* (e.g., *has-emotion* and *emotion-positive*) and with the predicate unit lover, while units for Tom fire in synchrony with units for beloved. The very same object, predicate and semantic units represent the proposition *loves* (Tom, Sally), but the synchrony relations are reversed.

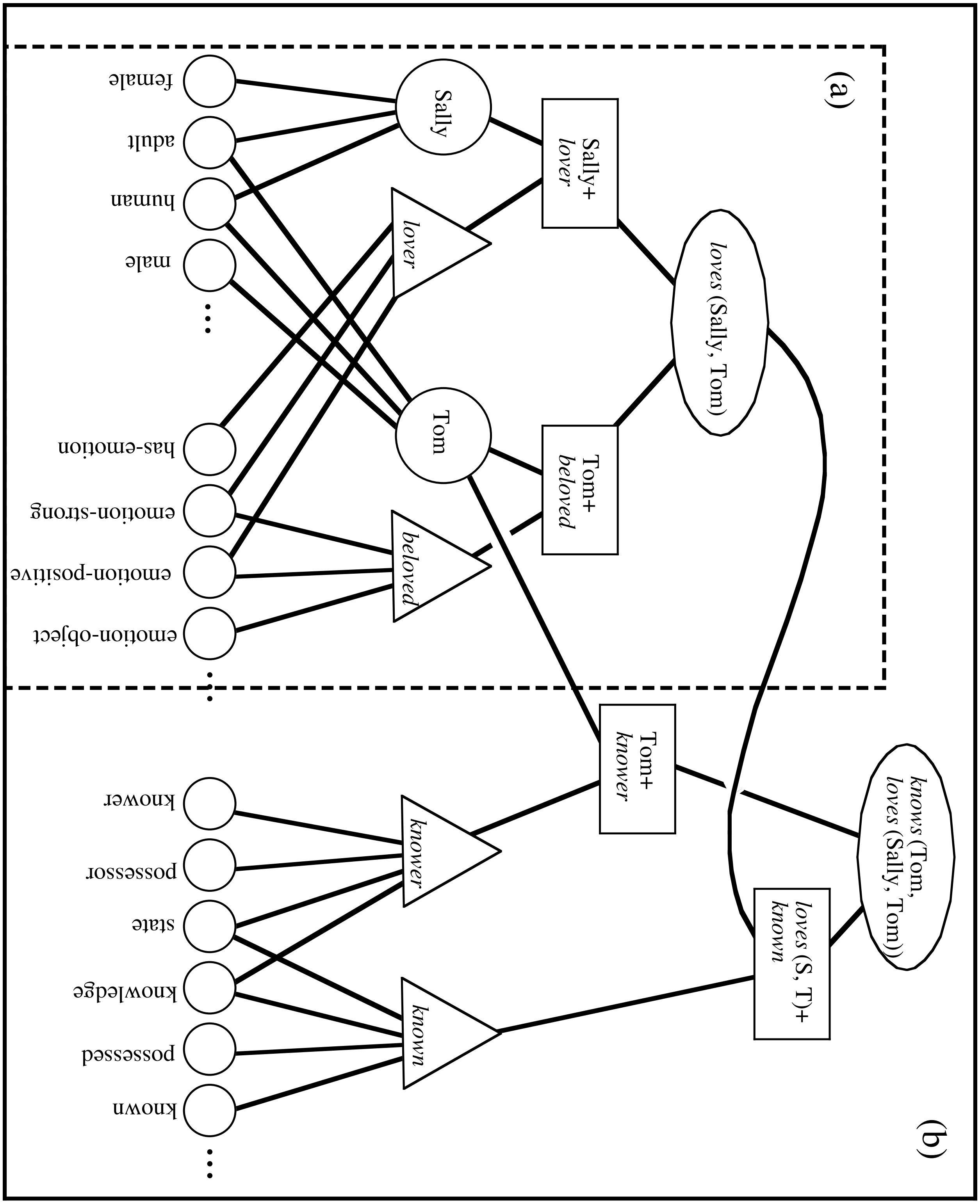
(a)
(b)
female
adult
human
male
has-emotion
emotion-strong
emotion-positive
emotion-object
knower
possessor
state
knowledge
possessed
known
Sally
Tom
lover
beloved
knower
known
Sally+
lover
Tom+
beloved
Tom+
knower
loves (S, T)+
known
loves (Sally, Tom)
knows (Tom,
loves (Sally, Tom))

Figure 1

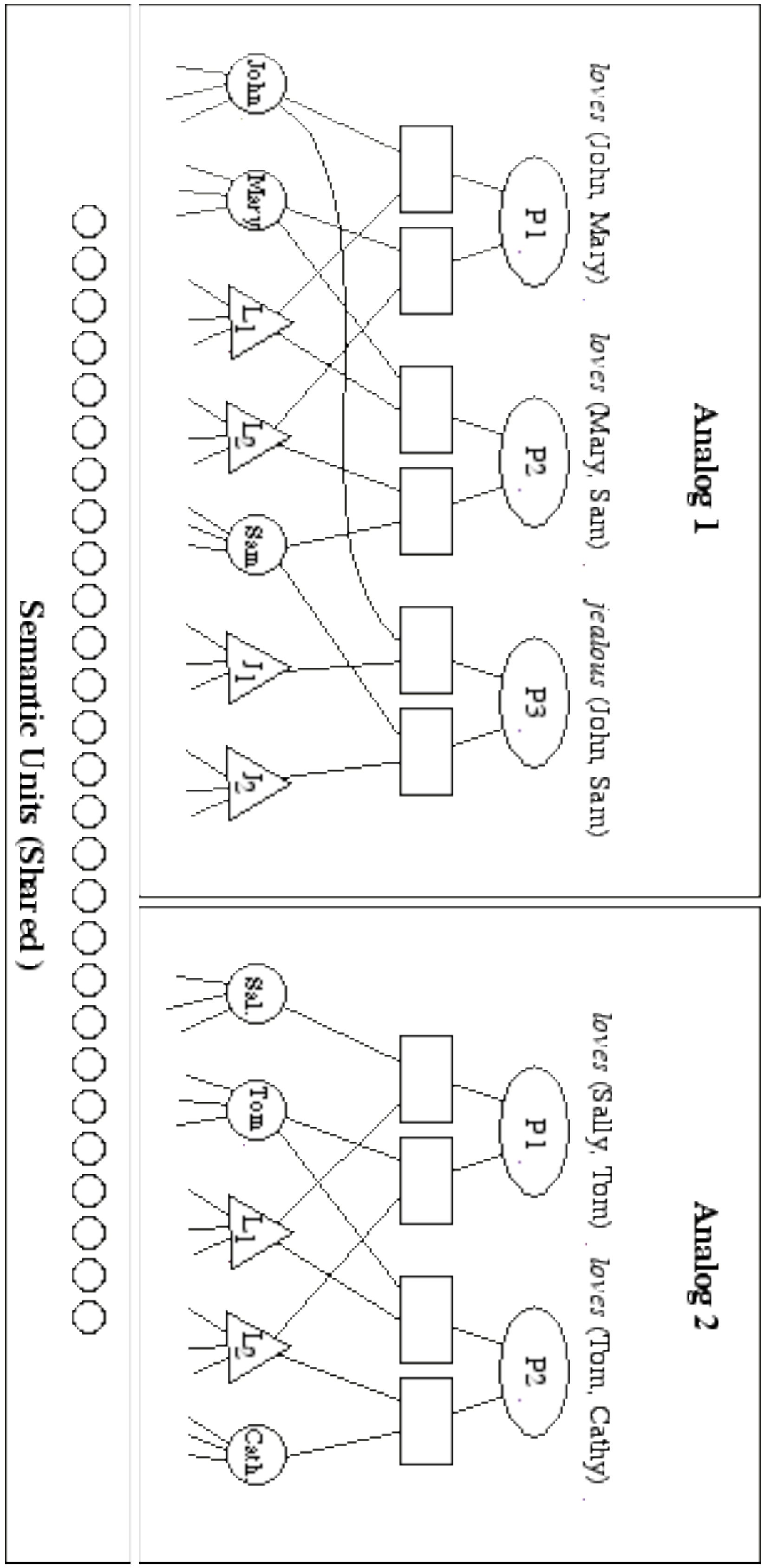
Analog 1
loves (John, Mary)
loves (Mary, Sam)
jealous (John, Sam)
P1
P2
P3
John
Mary
L1
L2
Sam
J1
J2
Analog 2
loves (Sally, Tom)
loves (Tom, Cathy)
P1
P2
Sal
Tom
L1
L2
Cath
Semantic Units (Shared)

Figure 2

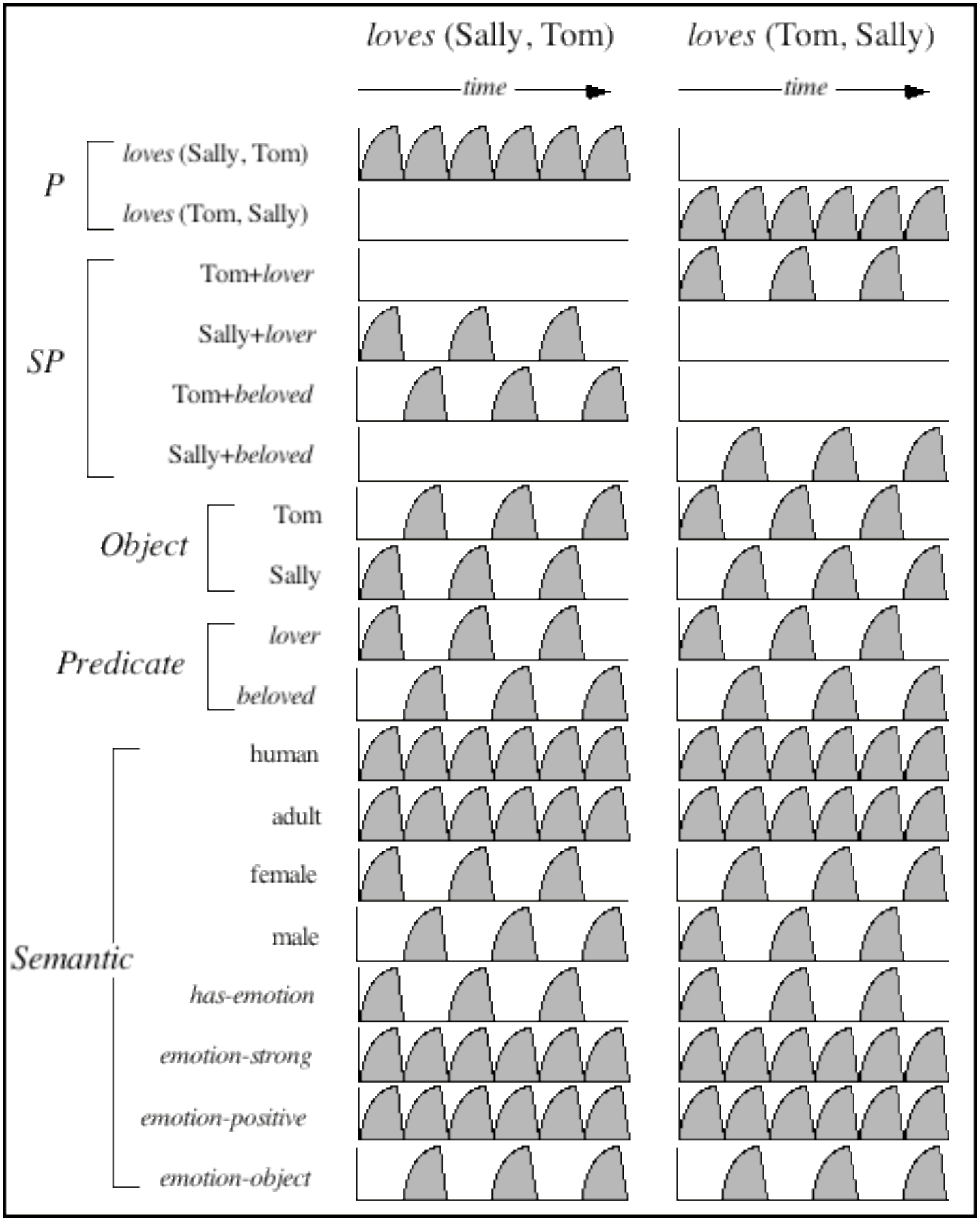
loves (Sally, Tom)
loves (Tom, Sally)
time
time
P
loves (Sally, Tom)
loves (Tom, Sally)
SP
Tom+lover
Sally+lover
Tom+beloved
Sally+beloved
Object
Tom
Sally
Predicate
lover
beloved
Semantic
human
adult
female
male
has-emotion
emotion-strong
emotion-positive
emotion-object

Figure 3