Relational Reasoning in a Neurally Plausible Cognitive Architecture

An Overview of the LISA Project

John E. Hummel and Keith J. Holyoak

University of California, Los Angeles

ABSTRACT—*Human mental representations are both flexible and structured—properties that, together, present challenging design requirements for a model of human thinking. The Learning and Inference with Schemas and Analogies (LISA) model of analogical reasoning aims to achieve these properties within a neural network. The model represents both relations and objects as patterns of activation distributed over semantic units, integrating these representations into propositional structures using synchrony of firing. The resulting propositional structures serve as a natural basis for memory retrieval, analogical mapping, analogical inference, and schema induction. The model also provides an a priori account of the limitations of human working memory and can simulate the effects of various kinds of brain damage on thinking.*

KEYWORDS—*cognitive architectures; neural networks; reasoning; symbolic thought*

A fundamental aspect of human intelligence is the ability to acquire and manipulate concepts defined by systematic relationships among multiple objects. Many people view the Iraq situation in 2005 as "the same kind of thing" as Vietnam in 1968 because they perceive a variety of relational parallels: the American military bogged down in low-level conflict in a distant theater, an unfriendly local populace, difficulty distinguishing friend from foe, and lack of an exit strategy. Relational concepts abound in social understanding (e.g., a love triangle is defined by relations of affection among three people), law (e.g., breach of contract, based on a relations among two or more parties to an agreement), religion (e.g., atonement for sins, which relates a person's actions and their belief in a deity), science (e.g., force is a relation between mass and acceleration), and indeed even basic perception (e.g., recognition of arrangements of objects in scenes).

Relational thinking involves the ability to see analogies between superficially disparate situations and to form more general schemas, or relationally defined concepts (Gick & Holyoak, 1983). For example, force is an abstract relation between mass and acceleration whose properties remain constant across the object's other properties (e.g., it takes the same force to accelerate a 10-pound bowling ball from 0 to 10 mph as to accelerate a 10-pound rock or 10 pounds of water). Similarly, if person A loves person B but person B loves person C, it is reasonable to conjecture that A will be jealous of C, regardless of who A, B, and C are: The jealousy relation is suggested by the unrequited-love relation, not by the features of the people involved.

Relational thinking is so commonplace that it is easy to assume that the psychological mechanisms underlying it are simple. But the capacity to form and manipulate high-level relational representations appears to be a uniquely human ability, a late evolutionary development that develops relatively late in childhood. The power of relational thinking resides in its *structure sensitivity*—the ability to generate inferences and generalizations that are determined by the roles elements play in relational concepts rather than simply by the features of the elements themselves (e.g., it is the fact that person A is a lover who is not also a beloved, rather than the particular features of A, that suggests that he or she will be jealous).

In this article, we review our attempts to model the underpinnings of human relational thinking using a computational system (i.e., a computer program) called LISA (Learning and Inference with Schemas and Analogies). Many computational models of analogy have been developed (for reviews, see Doumas

Address correspondence to John E. Hummel, Department of Psychology, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095; e-mail: jhummel@psych.ucla.edu.

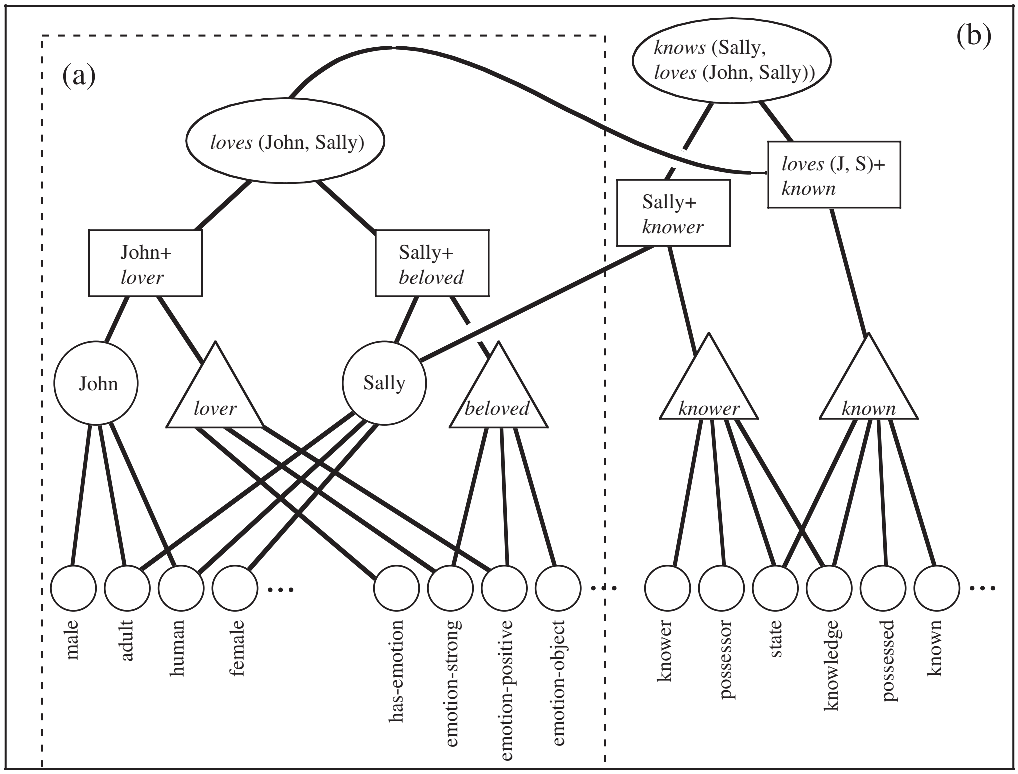

Fig. 1. Learning and Inference with Schemas and Analogies (LISA) representation of the propositions (a) ''John loves Sally,'' and (b) ''Sally knows that John loves Sally.'' Objects in these propositions (e.g., John, Sally; large circles) and relational roles or predicates (triangles) are represented as patterns of activation distributed over semantic units (small circles at the bottom of the hierarchy). For example, John might be represented by features such as *human*, *adult*, and *male*, and Sally might be represented as *human*, *adult*, and *female*; roles these objects may play (e.g., *lover*, *beloved*, *knower*, and *known*) are also represented by units capturing their semantic content. Subproposition units (SPs; rectangles) represent bindings of objects to roles. Separate SPs are bound into complete propositions via proposition (P) units (ovals). Simple propositions such as ''John loves Sally'' may serve as role fillers in other propositions, forming hierarchical propositions such as ''Sally knows that John loves Sally'' (as in b).

& Hummel, 2005; Holyoak, 2005), but few have been based on what is known about the brain's neural structure, and (perhaps for that reason) few have incorporated a general mechanism for learning new schemas. Our goal is to understand how relational thinking can be accomplished in a cognitive architecture that is both psychologically and neurally plausible. Ultimately, we aim to understand the neurocomputational basis of symbolic thought.

THE LISA MODEL

Knowledge Representation: LISAese

Our proposal is a form of *symbolic connectionism*: a computational system that codes relational structures within a neural network. LISA represents both objects and relational roles in a distributed fashion—that is, as patterns of activation over units (roughly analogous to neurons) representing the objects' or roles' semantic features. For example, the object John might be represented by features such as *human*, *adult*, *male*, etc.; Sally by *human*, *adult*, *female*, etc.; the *lover* role by *emotion*, *positive*, *strong*, etc.; and *beloved* by *emotion-object*, *positive*, etc. Objects are dynamically bound to roles by synchronizing the firing of their distributed representations (e.g., John is bound to the *lover* role by synchronizing the firing of units representing John and *lover*); separate role–filler bindings fire out of synchrony with one another (e.g., the John+*lover* set fires out of synchrony with the Sally+*beloved* set). The resulting propositional representations (e.g., "John loves Sally") are stored in long-term memory (LTM) using a hierarchy of units that collectively represent objects, relational roles, and their arrangement into complete propositions (see Fig. 1). This architecture (LISAese) achieves both the representational power and the structure-sensitivity of traditional symbolic systems and the flexibility of feature-based (e.g., connectionist) systems.

The hierarchy depicted in Figure 1 represents the propositions "John loves Sally" and "Sally knows that John loves Sally" in LISA's LTM and active or working memory (WM). In this hierarchy, the binding of roles to objects that fill them is represented by the subproposition (SP) units. When a proposition enters WM—that is, when it becomes active—its roles and fillers are bound by synchrony of firing (Hummel & Holyoak, 1992; Shastri & Ajjanagadde, 1993). When a proposition (P) unit becomes active, it excites the SPs (such as "John+*lover*") to which it is connected. Separate SPs inhibit one another, causing them to fire at different times (e.g., "John+*lover*" does not fire together with "Sally+*beloved*," but rather in sequence). When an SP fires, it activates the object and role units (e.g., John and *lover*) below it, and these in turn activate the semantic units below themselves. At the level of the semantic units, the result is a collection of mutually desynchronized patterns of activation, one for each role–filler binding. For example, the proposition "John loves Sally" would be represented by two such patterns, one binding the semantic features of John to those of *lover*, and the other binding Sally to *beloved*. The proposition "Sally loves John" would be represented by the very same semantic, object, and predicate units, but the synchrony relations would be reversed (with *lover* firing together with Sally and *beloved* with John).

The resulting representations are extremely well suited to support relational thinking: Augmented with a few simple processes, they provide a natural account of memory retrieval, the discovery of analogies, analogy-based inferences, and the generation of schemas from specific examples, as well as the limitations of WM and patterns of deficit following brain damage or degeneration.

Processes: Operations on LISAese

WM Limits

In LISA, the fact that subjects and predicates are bound dynamically by synchrony of firing imposes a hierarchical temporal structure on knowledge representations. Consider how LISA performs analogical mapping—that is, discovers correspondences between elements that play parallel roles in two similar situations (analogs). The two analogs and the emerging correspondences between them are assumed to reside in active memory (i.e., the currently active portion of LTM). Within active memory, a very small number of role–filler bindings can enter the *phase set*—the set of active, mutually desynchronized role–filler bindings representing one or more propositions. Each phase (i.e., temporal period of synchronized firing) in the set corresponds to one role–filler binding; a single phase corresponds to the smallest unit of WM. The phase set corresponds to the current focus of attention and is the most significant bottleneck in the system. The size of the phase set is determined by the number of role–filler bindings it is possible to have simultaneously active but mutually out of synchrony. This number is necessarily limited (see Hummel & Holyoak, 2003, Appendix A). In LISA, the phase set is the WM, so the capacity of the phase set is the capacity of LISA's WM.

There is evidence that, in the brain, binding of roles to fillers is accomplished by synchrony of firing in the 40-Hz (gamma) range, which means that a neuron or population of neurons involved in representing a proposition generates one spike (or burst) approximately every 25 ms. According to Singer and Gray (1995), the temporal precision of spike timing is in the range of 4 to 6 ms. These figures imply that human WM capacity is approximately four to six role bindings. LISA thus yields a principled estimate of the maximum amount of information that a person can process at any given time during analogical mapping: four to six role bindings, or roughly two to three propositions.

Because of the phase set's strongly limited capacity, LISA's processing of complex analogies is necessarily highly sequential: LISA can hold at most three propositions in WM simultaneously, so it must process large analogies in small pieces. The basis of LISA's algorithm for analog retrieval, mapping, inference, and schema formation is a form of guided pattern recognition. At any given moment, one analog (i.e., a set of interrelated propositions) is the focus of attention and serves as the *driver*. One (or at most three) at a time, propositions in the driver become active, generating synchronized patterns of activation of the semantic units (one pattern per SP). In turn, these patterns activate propositions in LTM (during analog retrieval) or in active memory (during mapping, inference, and schema formation).

Retrieval and Mapping

Memory retrieval in LISA occurs when a novel problem, the *target*, serves as the driver, cuing the retrieval of a useful source analog (recipient) from LTM: Up to three propositions in the target fire at a time, generating patterns of activation (one pattern for each role–filler binding) of the semantic units; these patterns activate propositions in LISA's LTM, retrieving source analogs into WM for mapping and inference. For example, given the novel target analog "Mary loves Tom and Tom loves Cathy," LISA is likely to be reminded of other analogs in which males and females love one another (as the same semantic units are involved in multiple propositions). LISA's algorithm for analogical retrieval as guided pattern recognition provides an excellent fit to the data on analogical reminding in people (Hummel & Holyoak, 1997).

LISA's algorithm for analogical mapping—that is, for finding the structural correspondences between the elements of two different but similar situations—consists of its algorithm for analog retrieval augmented with a mechanism for learning which elements of one analog tend to activate which elements of the other. Whenever a unit in one analog (e.g., the target) becomes active in response to the semantic patterns generated by a unit in another analog (e.g., the source), LISA updates the weight on a mapping connection (a direct link) between the source unit and the target unit (see Hummel & Holyoak, 2003, for more detail). The resulting mapping connections serve as LISA's hypotheses about which units map to which, and to constrain future mappings based on mappings already discovered (e.g., if LISA makes a mapping connection from A to B on one occasion, then the connection representing that mapping will allow A to excite B directly on future occasions).

This algorithm provides an account of the known strengths and limitations of people's ability to discover analogical mappings. It also correctly predicted previously unknown effects of text coherence (i.e., relational interconnectedness, defined as shared objects linking propositions) on people's ability to find structurally coherent mappings between the elements of two situations (Kubose, Holyoak, & Hummel, 2002). Processing multiple, interrelated propositions together generated more accurate mappings than did processing individual propositions, both for LISA and for human reasoners. LISA also correctly predicted previously unobserved asymmetries in mapping and inference accuracy. When one analog was more internally coherent than the other, inferences generated from the more coherent analog to the less coherent analog were more accurate than those made in the reverse direction.

Relational Inference and Schema Induction

Augmented with a simple algorithm for self-supervised learning, LISA's algorithm for analogical mapping provides an account of relational inference and schema induction (Hummel & Holyoak, 2003). When elements of the driver have no corresponding structures in the recipient, then initially random units in the latter are recruited to represent the "missing" elements; connected together, these units form appropriate inferences. For example, if a driver containing the propositions "John loves Sally," "Sally loves Bill," and "John hates Bill" is mapped onto a recipient stating that "Susan loves Tom" and "Tom loves Cathy," new units in the recipient will be recruited to construct the inferable proposition "Susan hates Cathy." The same operations, augmented with a simple mechanism for discovering what elements of one analog have in common with elements of the other, serves as a basis for relational schema induction. Together, these simple operations form a surprisingly complete account of analogical inference and schema induction.

Scaling Up: LISA Can Map Large Analogs Despite Limited Capacity

One of the surprising successes of LISA is its ability to explain an apparent paradox of human analogical thinking. People have great difficulty mapping some small, spartan analogies, yet easily map other analogies that are much larger and might seem more complex. For example, Hummel & Holyoak (1997) showed that LISA, like college students, is unable to reliably find the correct correspondences between boys and dogs (and their traits) in the following apparently simple analogy:

"Boys" analog	"Dogs" analog
smart (Bill)	*hungry* (Rover)
tall (Bill)	*friendly* (Rover)
smart (Steve)	*hungry* (Fido)
timid (Tom)	*frisky* (Blackie)
tall (Tom)	*friendly* (Blackie)

This tiny analogy problem is hard because the semantics of the two analogs provide no clues to the mapping, and the structural constraints that need to be processed jointly exceed LISA's (and people's) WM capacity. (The only consistent one-to-one correspondences between the two analogs are the following: *Timid* corresponds to *frisky*, Tom to Blackie, *tall* to *friendly*, Bill to Rover, *smart* to *hungry*, and Steve to Fido.)

In contrast, even large, "messy" analogs can be mapped reliably if rich semantic constraints guide the mapping. For example, college students were able to identify the major plausible correspondences between the Persian Gulf crisis of 1991 (when Iraq, under Saddam Hussein, invaded Kuwait) and World War II (e.g., Hussein and Iraq mapped to Hitler and Germany; President George H.W. Bush and the United States might map either to Franklin Roosevelt and the United States or to Winston Churchill and Great Britain). Although the analogs were large and the mappings were complicated by possible many-to-one correspondences (e.g., Bush might map to either Roosevelt or Churchill), rich semantic connections helped guide the mapping (Spellman & Holyoak, 1992).

To assess LISA's ability to scale up to such large examples, we applied it to several variations of this war analogy, with about 100 propositions describing each analog (Holyoak & Hummel, 2001). LISA's preferred mappings corresponded closely to those most frequently given by people. These results establish that LISA can find sensible mappings for large, ambiguous, and semantically rich analogies of the sort that people are able to map, while operating within psychologically realistic limits on WM.

LISA as a Neuropsychological Model of Relational Reasoning

The processes that control WM resources in LISA provide a natural account of the loss of relational reasoning in people with some forms of brain damage, such as patients with either frontal-lobe or temporal-lobe variants of frontotemporal lobar degeneration (FTLD; Morrison et al., 2004). These two patient groups showed different deficits in picture and verbal analogies: Frontal-lobe FTLD patients tended to make errors due to impairments in WM and inhibitory abilities, whereas temporal-lobe FTLD patients tended to make errors due to semantic memory loss. LISA provides a specific account of how such deficits may arise within neural networks supporting analogical reasoning. We were able to simulate the observed pattern of frontal-lobe deficits by impairing LISA' ability to rapidly learn mapping connections, while also reducing inhibitory control. Both rapid learning and inhibitory control appear to be key functions of the frontal cortex. To model the loss of conceptual knowledge produced by temporal-lobe damage, we destroyed connections between semantic units representing relational roles and the predicate units for those roles.

CONCLUSION

LISA shows promise as a neurocomputational model of symbolic thought. For example, we hypothesize that the mappings in LISA may correspond to neurons in the frontal cortex that have rapidly modifiable synapses. LISA may thus pave the way for integrative theories that better ground high-level human cognition in neurobiology. LISA also models many other aspects of normal cognitive functioning, including relations between effortful, reflective, forms of reasoning and more effortless, reflexive reasoning (Hummel & Choplin, 2000) and the human ability to exploit perceptual representations in the service of more general reasoning (Hummel & Holyoak, 2001).

There are numerous important questions about relational thinking that LISA in its current form does not address. Among the most important of these are issues of relation discovery (how do people learn new relational concepts?) and metacognition (how do people monitor their own progress toward solving a problem?). These issues provide goals for our ongoing research (e.g., Kittur, Hummel, & Holyoak, 2004).

Recommended Reading

Holyoak, K.J. (2005). (See References)

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

Hummel, J.E., & Holyoak, K.J. (1997). (See References)

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

REFERENCES

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

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

Holyoak, K.J. (2005). Analogy. In K.J. Holyoak & R.G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 117–142). Cambridge, England: Cambridge University Press.

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.

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. In *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. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264.

Kittur, A., Hummel, J.E., & Holyoak, K.J. (2004). Feature- vs. relation-defined categories: Probab(alistic)ly not the same. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 696–701). Mahwah, NJ: Erlbaum.

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.

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

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., & Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586.

Spellman, B.A., & Holyoak, K.J. (1992). If Saddam is Hitler then who is George Bush?: Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology*, *62*, 913–933.