

Emotions as Weak Biases in Metaphor Formation

William A. Stubblefield

Sandia National Laboratories
P. O. Box 5800
Albuquerque, New Mexico 87185
wastubb@sandia.gov

Abstract

The recurrence of certain basic metaphors across diverse times and cultures is an interesting issue in the debate over the form and extent of innate human knowledge. Are the common themes of metaphor and narrative evidence for strong innate human knowledge, or the result of weaker constraints on learning and development? Machine learning has demonstrated that strong conceptual knowledge may be acquired from experience and weak initial learning biases. Drawing on these results and a computational model of metaphor and analogy formation, this paper examines the ability of emotions to serve as effective biases on the acquisition of useful, domain specific patterns of metaphor. In addition, it considers the mechanisms through which metaphor projects semantic structure onto emotional experience.

Introduction

It may be that the universal history is the history of the different intonations given a handful of metaphors.

Jorge Luis Borges

Why are there so many similarities among the myths, stories and metaphors different cultures use to express their understanding of life, society and nature? Why have literature, painting, religion and even such practical arts as law, built on the same metaphors and narrative motifs throughout history and across diverse cultures? For example, the hero myth not only exists in nearly all cultures, but shows the same basic structure in all its occurrences (Campbell 1949); there is even evidence for its influence on our understanding of science (Landau 1991). Similar universal themes include the youth's coming of age, the proud man's fall, and the friend's betrayal. This recurrence of a small number of themes has long been recognized in the arts, anthropology, and psychology, and is an interesting problem for theories of the form and extent of innate human knowledge.

Lakoff and Johnson (1980) have argued not only for the importance of metaphor in human thought, but also for the idea that metaphors have their foundations in our biological embodiment and experience of the physical world. This paper explores this idea in the context of a formal model of metaphoric and analogical reasoning. SCAVENGER (Stubblefield 1995; Stubblefield and Luger 1996) is a computer program that learns useful patterns of metaphor

and analogy from experience. This program illustrates possible mechanisms through which such general, pre-conceptual aspects of our embodiment as emotions may guide the acquisition of strong, semantically rich patterns of metaphor. SCAVENGER also provides a model of the way in which metaphor projects meaning onto "raw" emotions, enabling the varied and subtle interpretations we give our emotional states. In exploring these questions, this paper considers such issues as:

- The interaction between metaphor and emotion. How does metaphor gain its unique ability to incite emotions? How do emotions influence the development of powerful, culturally universal metaphors?
- The feasibility of emotions and other aspects of our physical embodiment as weak biases on the acquisition of specific patterns of narrative and metaphor.
- The formal properties required of effective biases for metaphor selection and interpretation.

Formal Models of Metaphor

Although artificial intelligence has largely ignored metaphor's relationship to emotion, it has articulated the formal structure of metaphoric inference. Metaphors are inferences of the form, $A \text{ is } B$, where properties of B (the metaphor's *source*) are inferred to apply to A (the metaphor's *target*). For example, in the metaphor, "love is a rose," "love" is the target, "rose" is the source, and properties of roses that may be applied to love include beauty, complexity and fragility. Narratives, particularly those of myth and literature, may be regarded as complex, temporally structured metaphors.

Metaphoric inferences develop across four major stages:

- Inference begins with selection of an appropriate source. The source should enable inferences that are relevant (providing useful information about the target), novel (metaphor's strength is in discovering new knowledge) and valid.
- The selection of a source generally establishes an initial mapping of source properties onto the target. *Metaphoric inference* extends this mapping to new properties and relations. Essentially, a metaphor projects aspects of the source's semantic structure onto the target.
- Metaphoric inferences must be re-interpreted and vali-

dated in the target context. Validation of metaphors may involve techniques ranging from empirical tests (as with scientific metaphors), to subjective, emotional responses to artistic metaphors. Interpretation and validation are closely coupled, with validation failures often leading to re-interpretation (or *repair*) of the metaphor.

- Finally, participants in the metaphor should learn. The metaphor should not only change our understanding of the target, but also should improve our ability to apply the same metaphoric source in the future.

This framework is, of course, an abstraction from much more complex and still poorly understood processes. Although alternative models (most notably Black's (1962) interaction theory) exist, this simple framework characterizes the major decisions that must be made in developing any metaphor or narrative. For example, in writing a story, an author must:

- Determine its basic form: is the story a comedy? a tragedy? a tale of a hero's struggles? This is the choice of a metaphoric source. For example, an author may decide that her intention can best be communicated using a "moral agent as embattled hero" metaphor.
- Compose the story through a process of interpreting the basic metaphor and projecting this interpretation onto the story's target theme: Is the hero's struggle a war? a criminal investigation? a journey? How does the author flesh out the basic narrative framework? What are the major traits of the heroes and villains? What are the consequences of their moral choices?
- Interpret and validate the metaphor across different revisions and through the editing process. When the story is published, processes of criticism and discussion contribute to the metaphor's ongoing re-interpretation.
- Learn from the metaphor, from its development across different articulations, interpretations and contexts. This involves both the author and her audience.

Metaphoric inference is complex and could not be performed without some sort of heuristic bias. Following standard artificial intelligence terminology, we may characterize heuristics as either "strong" or "weak":

Strong heuristics involve knowledge of a particular problem domain. An example of a strong heuristic for metaphor formation might be the knowledge that "waves are often a useful metaphor for explaining cyclical natural phenomena." Similarly, the general form of the hero story is a strong heuristic for constructing narratives: cultures often express their key values through stories of a hero's personal struggles to preserve "good" and overcome "evil."

The cognitive science literature supports the role of strong heuristics in metaphor and analogy. For example, Holyoak (1985) has demonstrated that human subjects are better able to interpret a suggested analogy once they have been shown a structurally similar analogy. In this case, the example provides strong clues to the structure of the suggested analogy.

In contrast, *weak heuristics* exploit only general, formal properties of the target or source. *Systematicity*, the ability of a single metaphor to project an entire system of properties and relations onto the target (Black 1962; Gentner 1983; Lakoff and Johnson 1980), is the basis of many such heuristics. For example, the *structure-mapping* theory of analogy favors source-target mappings that transfer more complex systems of relationships to the target (Gentner 1983). Along with such heuristics as parsimony (a preference for simple target-source mappings) and a preference for metaphors that have proven effective in the past, structure-mapping is an often studied example of a weak bias for metaphor formation.

Systematicity also accounts for much of metaphor's importance, its unique ability to define a focused yet flexible context for acquiring and applying knowledge. Once a source metaphor has been selected, learning and problem solving can focus on the source's known properties and relations, greatly reducing the spaces that must be searched in learning and problem solving. Narrative also exhibits systematicity in combining separate events into a coherent, temporally ordered system of facts and relationships.

Although metaphoric reasoning in humans exploits both weak and strong heuristics, it is important to ask whether strong heuristics may be learned from experience starting only with general, weak biases. The relationship of this question to the innateness debate is clear: if it is not possible to learn strong patterns of metaphor from weak initial biases, then the universal themes of human narrative must depend directly upon strong innate knowledge. I.e., the power and universality of the hero myth and other central metaphors would result directly from epigenetic constraints on brain architecture. On the other hand, if weak biases are a sufficient starting point, then we have support for this paper's central conjecture: that basic stories and metaphors develop through the interaction of human experience and weak learning biases, such as may be provided by basic emotional reactions and other aspects of our physical embodiment.

Metaphor, Emotion and Innateness

Again and again I encounter the mistaken notion that an archetype is determined in regard to its content, in other words that it is a kind of unconscious idea (if such an expression be permissible). It is necessary to point out once more that archetypes are not determined as regards their content, but only as regards their form, and then only to a very limited degree. A primordial image is determined as to its content only when it has become conscious and is therefore filled out with the material of conscious experience.

Carl G. Jung (Jung 1963)

One of the justifications for the importance of metaphor is the *inexpressibility hypothesis* (Ortony 1975), which holds that metaphors can express ideas that would be difficult or impossible to communicate using more literal language. This can be given a stronger statement: a primary function

of metaphor is to *impose* a conceptual structure on vague, undifferentiated emotions and experiences.

Cognitive neuroscience supports the idea that our minds impose a structure on emotional experience. Based on experiments with individuals who have had their corpus callosum (the connection between the two sides of the brain) surgically severed, Gazzaniga (1998) argues that the construction of explanations about our experience and behavior is distinct from our ability to recognize patterns of experience, respond to requests, and even carry out complex sequences of actions. Bechera et al. (1997) have shown that in games, people's emotional responses to situations can indicate their decision strategies before they are able to articulate those strategies. LeDoux (1996) distinguishes between *emotional memory*, a conditioned emotional response to a situation, and *memory of an emotion*, a structured, conceptual account of the emotional experience's meaning and circumstance.

Damasio (1994) offers additional perspectives on the complementary relationship between emotions and reason. Emotions contribute to reason by supporting a focus of attention, and providing basic selection criteria on alternative choices. Although lacking in propositional content, emotions are essential to our efforts to formulate and act on that content. This ability of emotions to influence our understanding of propositional knowledge supports the view of emotions as a source of weak biases on the learning of symbolic forms of knowledge such as metaphor.

LeDoux (1996) discusses a variety of research findings indicating that emotions are of a few basic types. Drawing on such data as physiological manifestations of emotional states, categorization of human emotional expressions, and analysis of linguistic categories of emotions, researchers have hypothesized a small number of basic emotional states. Although specific taxonomies differ, they all agree that the number of basic emotions is small (< 10) and subject to conceptual specialization. This suggests that the shades of emotion we articulate result from the projection of propositional structure onto these basic emotions.

Metaphor is an important means by which our minds project propositional content onto emotional states. Metaphors have a rich, systematic conceptual structure; emotions are general states lacking specific semantic content. Metaphor and narrative place emotions in a symbolic framework that allows them to be recalled through the experience of art. Emotions, in turn, provide a basic source of values that supports reason in making choices and evaluating the results. This interaction is the basis of emotion's role as weak biases on metaphor formation.

Traditionally, the debate over innateness has been defined by two extreme viewpoints: the view of the human mind as a blank slate, free of any biases, and the idea that we inherit hard-wired instincts for such specific behaviors as greed, aggression and territoriality. The middle view, that our behavior was a product of both nature and nurture, tended to be dismissed as simultaneously obvious and uninteresting. Recently, advances in cognitive science, machine learning and neuroscience have given us a basis

for defining exactly how genetics and experience interact to shape our minds (Elman et al. 1996; Barkow et al. 1992).

In applying developmental psychology and connectionist theories to questions of innateness and learning, Elman et al. (1996) offer such a perspective on the way in which genetics and environment interact to shape the human mind. In their introduction, the authors describe three possible forms of innate mental abilities:

- *Representational constraints* are the strongest possible form of innate knowledge. Representational views hold that specific ideas are genetically determined in the "micro circuitry" of the human brain. Based on neuroscientific evidence about brain development, and the importance of neural plasticity to learning and problem solving, Elman et al. conclude that true representational constraints are rare.
- *Architectural constraints* result from the modular structure of the brain, and are of larger granularity than the micro circuitry required for representational constraints. Architectural constraints include the organization of individual brain modules, and the way those modules are interconnected. Architectural constraints function indirectly, by determining the types of information that can be received in the brain and hence, that can be acquired and stored.
- *Timing constraints* refer to genetically determined differences in the development of various brain structures. Differing rates of maturation among brain structures, changes in the plasticity of brain modules during development and other developmental variations exert biases on the way the brain interacts with the environment and the types of information that shape its development.

Architectural and timing constraints shape our knowledge by functioning as learning biases, heuristics that guide the acquisition of basic knowledge from the similar experiences all humans share.

For example, in proposing an alternative to Chomsky's view that basic grammatical structures are strongly innate, Deacon (1997) presents evidence that weak learning and attention biases are sufficient to support the acquisition of complex grammars. He also argues that these weak innate learning biases interact with the common patterns of human experience to give all human languages their common underlying structure.

Metaphoric inference is one vehicle through which these biases may shape learning. Lakoff and Johnson (1980) argue that the complex patterns of meaning found in human language derive from the basic facts of our physical embodiment through processes of metaphoric extension. Starting with the basic constraints provided by our bodies (standing is better than falling, self is different from other, etc.) as an initial set of metaphoric sources, we derive the idioms of human language by developing successive metaphors to explain experience. Emotions are another source of embodied constraints on metaphor formation.

These findings lead to several hypotheses about the role

of weak biases in metaphor formation:

- Weak biases, such as could be supplied by emotional reactions to experience, can support the acquisition of strong, domain specific patterns of metaphor.
- Metaphor constrains the interpretation of experiential data, imposing semantic content on "raw" sensation and emotion.
- The acquisition of strong patterns of metaphor must occur incrementally. Consequently, the learner be able to represent and exploit partial patterns of metaphor and use them to improve its abilities.

The SCAVENGER Experiments

SCAVENGER is a computer program that solves problems through metaphoric and analogical reasoning. It looks for similarities between new problems and stored components of previously successful solutions. After selecting a source, it transfers source properties to the target problem and elaborates the result to arrive at a complete solution. Finally, it tests its inferences and uses the results to improve its future performance. This section offers a high level description of SCAVENGER; for details, see (Stubblefield 1995; Stubblefield and Luger 1996).

The SCAVENGER algorithm follows the four-stage structure common to metaphoric and analogical inference:

- Source selection: choosing a set of candidate sources.
- Elaboration: extending each candidate source to fully match the target problem.
- Interpretation and evaluation: testing each source-target mapping for validity.
- Learning to improve its future performance.

Source Selection

SCAVENGER maintains a knowledge base of problem-solving operators. Solutions require the proper sequence of operators. Each operator is described by a set of properties; these are essentially preconditions for its application. Some, but not all of these properties can be observed in target problems.

SCAVENGER organizes its knowledge-base using an hierarchical index as in figure 1. Each node of the index contains properties that are common to a set of problem solving operators. In the figure, operators have labels of the form O_n, and properties of the form p_n.

Beginning with observed properties of a target problem, SCAVENGER walks the tree and composes a set of candidate partial solutions. Each path through the tree represents a set of operator sequences, generated by the cross product of the operators stored at each node. In figure 1, a path through nodes #1 & #2 generates 6 such sequences:

{O1 - O1; O1 - O4; O2 - O1; O2 - O4; O3 - O1; O3 - O4}

Each of these partial candidate solutions combines an operator from node #1 with an operator from node #2. Note

that the index functions as a source of strong constraints on metaphor formation. As will be shown, SCAVENGER constructs its index through inductive learning.

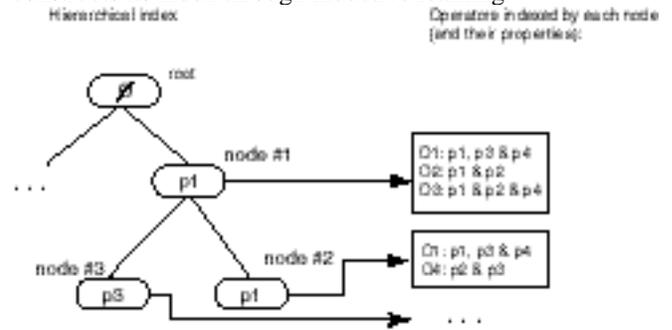


Figure 1

SCAVENGER generates a set of partial solutions by finding all paths through the index that match the target problem. When it encounters a node that does not match the target, it prunes that node and its descendants. It does retain the path leading to the failed node's parent, along with all candidate solutions produced by that path.

Although this follows a common form of search-based problem solving, source selection in SCAVENGER has a number of properties that more closely match the patterns of metaphoric inference. The most important is the algorithm used to determine if an hypothesis matches a given index node. *Assumption-based* matching does not require that all the index node's properties be true for the target problem, only that these properties are not known to be false for the target. In other words SCAVENGER allows the match unless it is known to be false. Consequently, source retrieval in SCAVENGER is not a process of matching known properties, but one of projecting the source operator's properties onto the target problem. We may think of each candidate solution that SCAVENGER forms as a different metaphor, projecting different properties onto an under constrained target problem.

Because SCAVENGER's search of the index tree produces a large number of partial candidate solutions, it relies on heuristics to rank them for further elaboration. SCAVENGER's heuristics include:

- A preference for longer partial solutions (those found deeper in the index tree). This reflects a preference for specificity, since longer operator sequences may be interpreted as making a stronger assertion about the target problem. This also can be viewed as a variation of structure-mapping's systematicity heuristic (Gentner 1983), since longer operator sequences impose more structure on the target problem.
- A preference for simpler solutions. Simplicity measures vary with the problem domain, but a typical metric preferred chains with the fewest different operators. For example, SCAVENGER might prefer the sequence O1-O2-O1 over O1-O2-O3.

After searching the index tree, SCAVENGER produces a

set of partial candidate solutions. Again, because assumption-based matching allowed the target to match index properties that were not known to be either true or false in the target, SCAVENGER's matching algorithm followed metaphor's behavior of projecting information onto poorly structured targets. It is reasonable to think of these as partially interpreted metaphors.

Elaboration and Evaluation

Source selection produced a number of partial candidate solutions, ranked according to heuristic merit. SCAVENGER then completed each candidate solution using standard AI search techniques. Note that a single partial solution could lead to many complete solutions. SCAVENGER evaluates each completed solution by determining whether or not it solved the target problem. Once SCAVENGER found an acceptable solution, it stopped the elaboration and evaluation phase.

Learning

After the elaboration and evaluation of candidate inferences, SCAVENGER was left with one correct solution, a set of incorrect solutions, and a set of untested partial solutions. It used these as training data for improving its ability to select candidate sources. SCAVENGER categorized the correct solution as a positive training example, and the failed solutions as negative examples. It then examined the operators that the elaboration stage added to source selection's partial solution, selecting the properties that best distinguished correct solution from the negative examples. When it selected these properties, it used them to create a child node of the index node that produced the correct solution, storing all matching operators under it.

SCAVENGER chose the properties that best distinguished the correct and incorrect solutions performing an information theoretic analysis to select properties that carry the most information in distinguishing positive and negative training examples (Quinlan 1986).

Note that SCAVENGER uses only weak biases in learning to improve its source selection:

- It relied on its ability to classify solutions as correct or incorrect. This is no more information than could be gained from emotional reactions to a story or metaphor.
- It used an information theoretic evaluation to select those properties of the positive solution that best distinguished it from the set of negative solutions. This sort of distinction is well within the capabilities of even simple connectionist architectures.

Testing SCAVENGER

SCAVENGER was tested on several application domains (Stubblefield 1995), including determining the applicability of LISP functions to new problems from examples of their desired behavior, constructing explanations of observed behaviors in simple physical systems, and inferring the cause of children's mistakes in subtraction, a problem

previously examined by (Brown and VanLehn 1980; VanLehn 1990). I chose these problems because all of them are severely under constrained by the given data. These problems are too difficult to be solved efficiently using only general, weak heuristics; they require strong heuristics for efficient solution. SCAVENGER's task was to learn these strong heuristics.

Figure 2 illustrates SCAVENGER's performance in diagnosing bugs in children's subtraction. The graph shows its improvement across 5 trials of the training data. Due to the many possible combinations of bugs it had to test, the untrained algorithm took approximately 1600 seconds per problem. The trained version took an average of 11 seconds per problem. The untrained version of the algorithm generated an average of 3096 candidate analogies per problem. After training, it generated an average of 78 candidates per problem.

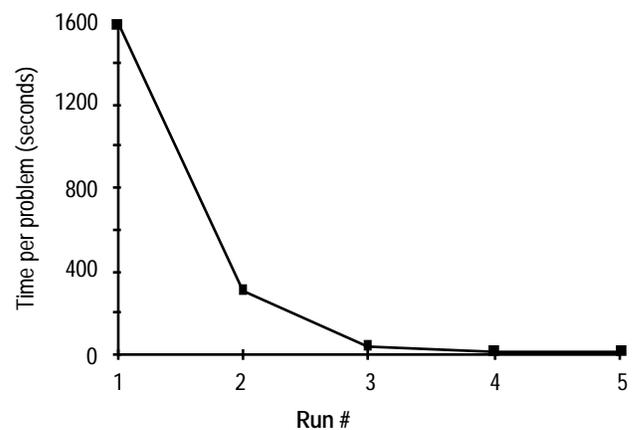


Figure 2

In applying SCAVENGER to the test set (problems that the learning algorithm had not seen), the untrained version took about 1500 seconds per problem. This improved to 76 seconds after training. Although this is a strong result, it is worth noting that it was skewed by the presence of a small number of novel problems that took a very long time to solve. Excluding these, the average time on the test problems drops to under 11 seconds. These results are typical of those achieved in the other problem domains.

Conclusion

Although there are considerable differences between the problems SCAVENGER solved, and the creation of powerful literary metaphors and narrative themes, the SCAVENGER experiments clearly support the ability of weak biases to drive learning of more powerful, domain specific patterns of metaphor. As SCAVENGER extended its index tree, it acquired a number of strong but broadly applicable patterns of metaphor. In doing so, it relied only on the classification of candidate solutions as positive or negative, and the ability to select properties that best distinguished positive and negative solutions. The first of

these could be provided by emotional reactions to experience, and the second is well within the capabilities of connectionist systems.

The role of culture in recording and transmitting metaphors further supports the efficacy of weak biases. Culture greatly augments our own learning abilities by preserving stories and metaphors that have proven effective in the past.

The SCAVENGER experiments also shed light on the question that originally motivated this paper, the reasons behind the universality of a relatively small number of recurring story themes and motifs. In the explanation that accompanied figure 2, I noted that, as it learned, SCAVENGER reduced the number of partial solutions considered by several orders of magnitude. Of course, this reduction depends upon the structure of the problem domain, and the ability of a few source patterns to solve large numbers of problems. Again, it is reasonable to assume that this is true for human stories and metaphors. The small number of basic emotions, the common paths of human development and the universal importance of family, love and self-preservation to our lives, suggest that a small number of basic themes may indeed be adequate to give structure to most human experience.

Another interesting feature of SCAVENGER is its ability to represent partial patterns of metaphor. SCAVENGER's index hierarchy is a very compact representation of a many partial metaphors. The learning algorithm expands these incrementally. Also, like human metaphors, SCAVENGER's partial solutions could be completed in a variety of ways to accommodate different targets.

Perhaps the most important aspect of the SCAVENGER experiments is their support for the role of metaphor in giving propositional content to emotional experience. Assumption-based matching, by allowing source retrieval to assume that source properties were true in the target in the absence of knowledge to the contrary, enabled SCAVENGER to project patterns of meaning that had proven useful in the past onto new situations. Remarkably, in SCAVENGER's test domains, these projections usually proved useful and valid. This idea that meaning can be projected on experience, rather than drawn from it, is a strange notion that not only fits our understanding of art, but also our growing understanding of the human brain (Bechera et al. 1997; Gazzaniga 1998).

This two way interaction between metaphor and emotion also explains the unique power and importance of art. This was beautifully expressed in a scene from Al Pacino's recent film about Shakespeare's *Richard III*, entitled *Looking for Richard*. Rather than presenting the play, the film was a semi-documentary in which scenes from the play were interwoven with discussions about it. In one of the film's most memorable sequences, the film makers asked people in the street about Shakespeare's work. I particularly remember one man who, when asked why Shakespeare was important, replied simply: "because he teaches us to feel."

I hope this paper has helped fill in some of the details behind this man's remarkable insight.

References

- Barkow, J. H., Cosmides, L. and Tooby, J. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press, 1992.
- Bechera, Damasio, A. H., Tranel, D. and Damasio, H. Deciding advantageously before knowing the advantageous strategy. *Science* 275 (1997): 1293-1294.
- Black, M. *Models and Metaphors*. Ithaca, NY: Cornell University Press, 1962.
- Brown, J. S. and VanLehn, K. Repair Theory: a generative theory of bugs in procedural skills. *Cognitive Science* 4 (1980): 379-426.
- Campbell, J. *The Hero With a Thousand Faces*. Princeton, NJ: Princeton University Press, 1949.
- Damasio, A. R. *Descartes Error: Emotion, Reason, and the Human Brain*. New York: Grosset/Putnam, 1994.
- Deacon, T. *The Symbolic Species*. New York: Norton, 1997.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. and Plunkett, K.. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, Mass: MIT Press, 1996.
- Gazzaniga, M. S. *The Mind's Past*. Berkeley: University of California Press, 1998.
- Gentner, D. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7 (1983): 155-170.
- Holyoak, K. J. The pragmatics of analogical transfer. *The Psychology of Learning and Motivation* 19 (1985): 59-87.
- Jung, C. G. *Analytical Psychology: Its Theory and Practice*. London: Routledge and Kegan Paul, 1963.
- Lakoff, G. and Johnson, M. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.
- Landau, M. *Narratives of Human Evolution*. New Haven: Yale University Press, 1991.
- LeDoux, J. *The Emotional Brain*. New York: Simon & Schuster, 1996.
- Ortony, A. Why metaphors are necessary and not just nice. *Educational Theory* 25 (1975): 45-53.
- Quinlan, J. R. Induction of Decision Trees. *Machine Learning* 1 (1986): 81-106.
- Stubblefield, W. A. *Source Selection for Analogical Reasoning: An Interactionist Approach*. Ph.D., University of New Mexico, 1995.
- Stubblefield, W. A. and Luger, G. F. Source Selection for Analogical Reasoning: An Empirical Approach. In *Thirteenth National Conference on Artificial Intelligence in Portland, Ore.*, 1996.
- VanLehn, K. *Mind Bugs: The origins of procedural misconceptions*. Cambridge, Mass: MIT Press, 1990.