Chapter 2: Review of the Literature

Introduction and Overview

In the latter half of the 20th century research in two fields, categorization and analogy, has challenged long held classical views of these topics. Up until the nineteenth century, discussions on analogy, and in particular metaphor, were insistent on one theme: analogies are decorations of speech; they do not contribute to the cognitive meaning of the discourse, but instead lend it color, vividness, emotional impact, or accessibility. Thus it was characteristic of the Enlightenment philosophers and their predecessors, such as Hobbes and Locke, to insist that though philosophers may sometimes have good reason to communicate their thoughts with metaphors, they should do their thinking entirely without metaphors. Only by using nothing but unambiguous, literal language could knowledge be gained and communicated properly (Hobbes, 1651):

> To conclude, the light of humane minds is perspicuous words, but by exact definitions first snuffed, and purged from ambiguity; reason is the pace; increase of science, the way; and the benefit of mankind, the end. And, on the contrary, metaphors, and senseless and ambiguous words are like ignes fatui; and reasoning upon them is wandering amongst innumerable absurdities; and their end, contention and sedition, or contempt.

Similarly, John Locke, in the Essay Concerning Human Understanding, criticized imprecise and ambiguous "civil" language and proposed a proper and well-defined philosophical language that "may serve to convey the precise notions of things, and to express in general propositions certain and undoubted truths, which the mind may rest upon and be satisfied with in its search after true knowledge" (Locke, 1686).

What characterizes almost all theories of metaphor from the time of the Romantics up through the twentieth century is the rejection of this theme. Metaphors, it has been argued, are not cognitively dispensable decorations. They contribute to the cognitive meaning of our discourse and they are indispensable, not only to philosophical discourse, but to ordinary, and even scientific discourse. Nietzsche went so far as to argue that all speech is metaphorical and truth is "a mobile army of metaphors" (Nietzsche, 1873). Lakoff and Johnson's seminal work, *Metaphors We Live By* (1980), begins with the statement of traditional interpretations of metaphor and summarily reject this interpretation:

> Metaphor is for most people a device of the poetic imagination and the rhetorical flourish – a matter of extraordinary rather than ordinary language… We have found, on the contrary, that metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical. (p3)

This shift in understanding of the importance of analogies – from ornamental to fundamental – has resulted in the emergence of research in the structure and comprehension of analogy. The results of this research will be addressed later in this chapter.

Categorization, too, has experienced a dramatic reinterpretation in the 20th century. As Lakoff summarizes in *Women, Fire and Dangerous Things: What Categories Reveal about the Mind* (p 6):

> From the time of Aristotle to the later work of Wittgenstein, categories were thought [to] be well understood and unproblematic. They were assumed to be abstract containers, with things either inside or outside the category. Things were assumed to be in the same category if and only if they had certain properties in common. And the properties they had in common were taken as defining the category.
>
> This classical theory was not the result of empirical study. It was not even a subject of major debate. It was a philosophical position arrived at on the basis of a priori speculation. Over the centuries it simply became part of the background assumptions taken for granted in most scholarly disciplines. In fact, until very recently, the classical theory of categories was not even thought of as a *theory*.

Current theories on categorization have shifted from this Aristotelian view towards what has become known as the "prototype" view, the importance of which was first established by Rosch (1973). This shift in our understanding of categorization can be traced to the Whorfian hypothesis, which claims that language is not merely a medium for the expression of our thoughts but that "linguistic patterns themselves determine what the individual perceives in this world and how he thinks about it. Since these patterns vary widely, the modes of thinking and perceiving in groups utilizing different linguistic systems will result in basically different world views" (Fearing, 1954). Rosch's tests of the Whorfian hypothesis led her to study the nature of color categories among speakers of languages without a blue-green color distinction. In these studies, she identified structure within categories, with some members of a category being seen as more or less prototypical than others. This recognition, with its associated research paradigm, launched a field of study into categorization that has come to be characterized as the "prototype" view. More recent developments in this field have extended the types of categories studied from "common, or stable, categories to *ad hoc* categories" (Shen, 1992) – from Rosch's study of color categories to Barsalou's study of "things to take from your house during a fire." The results of this research and its implications on cognitive structure will be discussed later in this chapter.

It has not gone unnoticed that these two research programs, analogy and categorization, are both studies of similarity and each may have insights that might inform the other, but as late as 1992 it was remarked: "Despite the obvious affinity of these two fields of research, the link between them has received little attention in cognitive psychology or in other disciplines" (Shen, 1992). Shen and others (most notably Glucksberg and Keysar), have, during the 1990's, dedicated study to linking

these two research enterprises. However, the emphasis of this research has been on metaphors and their interpretation, developing "a coherent and unified framework by assuming that metaphor interpretation is, in fact, a process of ad hoc category formation" (Shen 1992), and "a basis for a theory of metaphor comprehension, and also clarifies why people use metaphors instead of similes" (Glucksberg and Keysar, 1990). It is the aim of this thesis, in the context of scientific discourse, to extend the link between metaphor and categorization to include analogical statements and to argue that the statement of an analogy – not merely its comprehension – is an assertion of categorization. Below I will outline the existing research on analogies and categorization, along with the research that attempts to link these two fields of study, in more depth.

Analogies

Analogies in philosophy and the philosophy of science

The modern views of analogy can be traced to the influences of philosophers Max Black (1962) and Mary Hesse (1966). Black's work built on that of Ivon Richards' (1936) work in the field of rhetoric, who proposed a set of useful terms for talking about metaphors (the "topic" or "tenor," the "vehicle," and the "ground") and a theory of how metaphors function. This theory, called the tensive view, emphasized the conceptual incompatibility, or "tension," between the terms (the topic and the vehicle) in a metaphor (Ortony, 1993). Black's interpretation of metaphor, referred to as the interaction model, claims that metaphor is a cognitively irreducible phenomenon that works not at the level of word combination, but much deeper, arising out of the interactions between the conceptual structures underlying words. In comparing two concepts, the significance of one concept is not merely projected onto another, but the two interact, altering the perception of both concepts. Explaining this theory in metaphorical terms, Black offers the following analogy (1962):

> Suppose I look at the night sky through a piece of heavily smoked glass on which certain lines have been left clear. Then I shall see only the stars that can be made to lie on the lines previously prepared on the screen, and the stars I do see will be seen as organized by the screen's structure. We can think of the metaphor as such as screen and the system of 'associated commonplaces' of the focal world as the network of lines upon the screen.

In his theory, Black claims that the following are true of all metaphors (synopsis from Ortony, 1993):

- Metaphors consist of primary and secondary components. (In the statement "*X is like Y" X* is referred to as the primary and *Y* the secondary. Other literature refers to these as the tenor, vehicle, target, etc., and there is no standard convention across disciplines. I will use *base* and *target*, recognizing that these make implications about the nature of analogical reasoning.)
- The significance of the base subject is not so much as a "thing" as it is a system.
- The associated implications of the base are projected onto the target.

- This projection selects, organizes, emphasizes and suppresses features of the target component. Through this interaction of the two subjects there is a selection of properties, an implication on the target, and, reciprocally, an implication back on the secondary.

This interpretation of metaphor, which is still central to many theories of metaphor, was incorporated four years later by Mary Hesse, a philosopher of science, whose "treatise on analogy in science argued that analogies are powerful forces in discovery and conceptual change" (Holyoak, Gentner and Kokinov, 2001). This treatise, *Models and Analogies in Science* (1966), argued that "the deductive model of scientific explanation should be modified and supplemented by a view of theoretical explanation as metaphoric redescription of the domain of the explanandum." Additionally, she coins three new terms in analogical reasoning: the positive, negative and neutral analogies. In positing an analogy, "positive analogies" are known to transfer from the base to the target, "neutral analogies" are possible elements of the base that are present in the target, and "negative analogies" are elements that do not transfer.

Analogies in cognitive science

In the 1980's, concurrent with Lakoff and Johnson's work on metaphor, there was a shift in research from four-term analogies (of the type found in standardized tests – a:b::c:d) to more complex analogies, such as those found in science and language. According to Holyoak, Gentner and Kokinov (2001), who were themselves pivotal in this shift,

> This exploration led to a more general focus on the role of experience in reasoning and the relationships among reasoning, learning, and memory, giving rise to an approach termed " case-based" reasoning (e.g., Kolodner 1993). In contrast to rule-based approaches to reasoning (the approach that was dominant in AI at the time), case-based reasoning emphasized the usefulness of retrieving and adapting cases or analogs stored in long term memory when deriving solutions to novel problems.

This work continued to adopt Black's model of metaphor, most significantly the system of relations in the base of the analogy. Among the best-known models developed at this time is Gentner's theory of structure-mapping (1983). In this model

> the central idea is that an analogy is a mapping of knowledge from one domain (the base) to another (the target) such that a system of relations that holds among the base objects also holds among the target objects. In interpreting an analogy, people seek to put the objects of the base in one-to-one correspondence with the objects of the target so as to obtain the maximal structural match… Thus, an analogy is a way of aligning and focusing on relational commonalities independently of the objects in which those relations are embedded. (Gentner and Jeziorski, 1993.)

From this model a computational model, the Structure Mapping Engine, was developed by Falkenhainer, Forbus and Gentner (1990). Holyoak and Thagard (1989) have a similar (ACME) program for analogies that is instead based on a connectionist network, but the underlying model of analogy is the same. It is important to note that these models will take two systems (for example, the solar system and the Rutherford model of the atom) and, as Falkenhainer, Forbus and Gentner (1990) describe, construct "matching algorithms consistent with [the] theory."

In contrast to these models are those posited by Douglas Hofstadter and the Fluid Analogies Research Group (FARG) at Indiana University, beginning with Melanie Mitchell's (1990) dissertation. Their computational models do not interpret a given analogy, but generate novel analogies. These models, Copycat, LetterSpirit and Metacat, are based on the thesis

> that mental activity consists of many tiny independent events and that the seeming unity of a human mind is merely a consequence of the regularity of the statistics of such large collections of events. Thus the metaphor of the "intelligent ant colony" and the image of "active concepts"…have inspired our models for over two decades. The models all involve the nondeterministic interaction of many tiny events that take place in simulated parallel. The models also feature both long-term and short-term memories, the former of which houses permanent concepts, and the latter of which is like a stage on which temporary mental structures are built, modified, and eventually razed. Events in each memory profoundly influence the multiple tiny parallel processes, and in that way, each memory affects what goes on in the other. (Hofstadter, 2004)

This thesis concerns itself with the creation of analogies, as opposed to their interpretation. I will show in later chapters that structure-mapping cannot account for analogy creation, and argue that the Hofstadter's interpretation of cognition as "many tiny independent events" is consistent with modern views on categorization, which are consistent with observed properties of student-generated analogies in science.

Analogies in linguistics

Concurrent with this shift in cognitive science (and, in particular, artificial intelligence) from rule-based to case-based reasoning, the field of linguistics began a departure from traditional Chomskian emphasis on linguistic competence towards "an increasing concern with linguistic performance and pragmatics" (Ortony, 1993). Representative of this shift is Lakoff and Johnson's *Metaphors We Live By*. In this book the authors argue that our ideas are not only referred to linguistically through metaphors (defined by the authors as "understanding and experiencing one kind of thing in terms of another") but also actually conceptualized in metaphorical terms. The choice of a secondary subject in these metaphors is one that allows the speaker to conceptualize "the nonphysical *in terms of* the physical – that is, we conceptualize the less clearly delineated in terms of the more clearly delineated… the prime candidates for concepts that are

understood directly are the simple spatial concepts, such as UP." This claim is evident in physics, where we speak of high and low potentials and energy and conceptualize atomic forces as wells.

> As our choice of metaphor will reflect our conceptualization of the phenomena, our metaphorical choice may change depending on the context of the problem. Consider the following statements, "Light consists of particles" which apparently contradicts "light consists of waves," but both are taken as true by physicists relative to which aspects of light are picked out by different experiments. (Lakoff and Johnson, 1980)

There is enormous utility in being able to conceive of light in different ways – to categorize light as a particle or as a wave. As Lakoff and Johnson state (1980),

> understanding our experiences in terms of objects and substances allows us to pick out parts of our experience and treat them as discrete entities or substances of a uniform kind. Once we can identify our experiences as entities or substances, we can refer to them, categorize them, group them, and quantify them – and, by this means, reason about them. When things are not clearly discrete or bounded, we still categorize them as such, e.g. mountains, street corners, hedges, etc.

That is, there is a relationship between these categories and metaphorical language: our language, which Lakoff and Johnson demonstrate is profoundly metaphorical, allows us to categorize phenomena.

Analogies in general, and metaphors in particular, are a topic of widespread interest and debate in linguistics at this time. Some of these debates will be further reviewed in later chapters. For a more complete review, see *The Analogical Mind* (Gentner, Holyoak and Kokinov, 2001).

> *Analogies in science and science education*
>
> …I think it would also be practical to design a curriculum based on an inquiry into the use of metaphor. Unless I am sorely mistaken, metaphor is at present rarely approached in school except by English teachers during lessons in poetry. This strikes me as absurd, since I do not see how it is possible for a subject to be understood in the absence of any insight into the metaphors on which it is constructed. All subjects are based on powerful metaphors that direct and organize the way we will do our thinking.
>
> -Neil Postman, *Conscientious Objections: Stirring up Trouble about Language, Technology, and Education*

A significant feature of the existing research on the role of analogies in instruction is that the focus has been on analogies drawn by the teacher and explained to the students. Research paradigms for constructing and testing models of analogical

reasoning have similarly focused on the comprehension of analogies that have been created by the researcher. Very little work has looked at generated analogies. This paradigm is indicative of a tradition of science education in which the students are interpreting instruction from the teacher, and not discussing and debating their own views of science. As reforms in science education call for greater attention to student ideas and student reasoning, the existing theories on analogy interpretation are of less use and there is a greater demand for understanding analogies as they are generated. Here I provide a brief overview on the research on the role of analogies in the classroom, the role of analogies in science, and the disconnect between what scientists do and what is attended to in classrooms.

In a pioneering work on scientific analogies in instruction, *Flowing Waters or Teeming Crowds: Mental Models of Electricity* (Gentner and Gentner, 1983), students were given instruction in circuits using one of two different analogies: one visualized the flow of current as a crowd of electrons, one a flowing water-like substance. (Experts use both models to represent different features of current.) Tests given post-instruction were designed to test whether analogy is an important source of insight (the Generative Analogy hypothesis) by "asking whether truly different inferences in a given target domain are engendered by different analogies" (Gentner and Gentner, 1983). The answers that students gave were overwhelmingly representative of the analogical model they had been taught.

Clement's work in physics education has focused extensively on analogical reasoning in the physics classroom, with attention primarily on explicit analogies generated by teachers and interpreted by the students. In Clement (1992), he describes "bridging analogies." These analogies are instantiated by the teacher and are used to motivate the students to see a likeness between two phenomena (say, the force provided by a spring supporting a book and a similar force from a table supporting a book) by providing intermediate analogies (a tight spring, a flexible board). The idea that a student will recognize "A is like B" and "B is like C" but not recognize that "A is like C" without explicit instruction supports a model of categorization in which "family resemblance" to the category prototype defines the membership of the category. In this way, prior research on analogical reasoning can facilitate the link between analogies and categorization. In the following section I will provide an overview on research into categorization.

More recent research on the role of analogies in physics continue in the paradigm of instruction by analogy:

Taber, in *When the Analogy Breaks Down: Modeling the Atom on the Solar System* (2001), notes that: "Analogy is one of the most potent tools in a teacher's repertoire and has been recognized as a common feature of quality science teaching." He voices concerns regarding the utility of a solar-system model of the electron and the students' lack of knowledge regarding the solar system in drawing such an analogy.

Mould (1998) is one of many instructional analogies that are frequently introduced in *Physics Education.* In this article, he suggests the use of an analogy to resolve the puzzle of the lost energy when two capacitors are joined together by comparing this scenario to the concept of water flowing between two tanks.

Harrison and Treagust (1999) have explored the relationship between the analogies we tell our students and the models students construct from these analogies, arguing that "students do not interpret scientific analogical models in the way we intended" and investigate potential factors in the consistency of analogy use in models of atoms.

Notable exceptions to investigations on teacher-generated, didactic analogies include Duit, Roth, Komorek and Wilbers (2001) who "studied analogy generation and development and analogical reasoning with 25 German tenth graders in a physics class. Results show the advantages and disadvantages of using analogies in promoting conceptual change and as a teaching technique." Their focus, however, was on the utility of analogy in conceptual change and not the generation of analogy as a goal in itself. And Yerrick, Doster, Nugent, Parke and Crawley (2003) who investigated the role of analogies in pre-service teachers' conversations and argue for analogy as part of preservice teachers' conceptual development.

Turning to the role of analogy in science, and not just the science classroom, Dunbar (2000, 2001) has conducted research on "in vivo analogies," going to research groups and listening to the ways in which they use analogies and analogical reasoning in their work and discourse. Analogies, he finds, are "frequent in science and in all aspects of human thinking." They are ubiquitous and crucial to the ways experts reason about science. One goal of science education could be considered to develop habits and skills that scientists employ in scientific reasoning. The focus in the literature on analogies as a means of arriving at correct conceptual understanding ignores the import of developing analogical reasoning as a skill to be mastered in and of itself. As May, Hammer and Roy (2004) noted, "inasmuch as expertise at inquiry supports… students and scientists developing conceptual understanding, young children's development of understanding and abilities for analogical reasoning will serve them better than learning the content knowledge of an expert."

Summary of analogy research

The main theme from this research that I will address is that the majority of research on analogies, and educational analogies in particular, has concerned how people interpret analogies and not on the creation of analogies and its implications on mental models of concepts. If analogical reasoning is an important feature of scientific reasoning, then one goal of science education should entail fostering the use of analogies and developing students' abilities in this realm. As educational researchers, understanding analogical reasoning, what assertions analogies make, what they reveal about the mind, are of importance – *not* because such findings allow us to better convey content, but because they inform us about a skill that is, in itself, one we should foster in our students.

Categorization

Overview

> Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech, Every time we see something as a *kind* of thing, for example, a tree, we are categorizing. Whenever we reason about *kinds* of things – chairs, nations, illnesses, emotions, any kind of thing at all – we are employing categories. Whenever we intentionally perform any *kind* of action, say something as mundane as writing with a pencil, hammering with a hammer, or ironing clothes, we are using categories.
>
> -George Lakoff,
> from *Women, Fire and Dangerous Things*

In 1987, George Lakoff published *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. This work thoroughly summarizes the research on categorization up to that time and interprets the philosophical implications of these findings. Rosch, who tied together existing research on categorization and created an experimental paradigm for investigating categorization, began from the opposite end – she received her undergraduate training in philosophy and brought this to bear on psychology. Her honors thesis was on Wittgenstein's 1953 treatise, *Philosophical Investigations* (Scharmer, 1999), a work that is credited with "the first major crack in the classical theory [of categorization]" (Lakoff, p16).

The classical view of categories held that there were rules of membership and if an item met these rules then it was a member of the category (or, there are properties that define a category and all members must share these properties). In this view categories, therefore, were seen as binary with no internal structure: an item either was or was not a member of that category, and the research paradigm for categorization was to define the rules or properties. Wittgenstein's work addressed the fact that certain categories, such as "game," do not fit this classical description. There is no one property that all games share, but rather there are family resemblances between games, so that "chess and Go both involve long term strategy… chess and poker both involve competition. Poker and old maid are both card games" (Lakoff, 1987 p. 16). Additionally, these categories had no fixed boundary – they could be extended and redefined as new games are introduced, or as previous games shift their context. About this time, a similar recognition of inconsistencies in our conception of categories occurred in mathematics.

Categorization and fuzzy sets

In mathematics, the analog of a linguistic category is the set. In classical set theory an instance either belongs to a set or it does not; there is no middle ground. An item's membership in a set is determined by whether or not it has the properties that define that set, called the class intension. For example, if the set is the class of all objects that are green, the intension of that set is the property green. This clean definition of the set allows for simple rules to be associated with them. The set of objects that are either green or square is the union of the sets of green objects and square objects. The set of all things that are green and square is the intersection of the sets of green objects and square objects. However, a complication associated with set theory occurs at the border between the set and non-set. For an intension that splits the set of all objects in two, such as

greenness, items will either belong to the set of green objects, or to the set of non-green objects, but there is an ambiguity for items that falls at this boundary between green and not-green. Applying this simple mathematical structure of sets to human-created categories does not work. There is logical inconsistency in assigning such items to both sets and to neither set. And the commonly accepted solution to this problem is the "law of excluded middle," which simply forbids there being an object that falls exactly at the border of set and not-set. Zadeh (1965) saw this solution as indicative of the fact that classical set theory does not capture the way in which we experience the world. Things come in gradations: some animals are birds, and some animals are mammals, while some, like the platypus, with its bill, webbed feet and fur, seem to fall somewhere in between. Statements are not always either true or false. The standard solution to the incompatibility of set-theory sets and human categories is to claim that mathematics, with its rigid structure and well-defined systems, does not apply to many real-world problems. Zadeh's solution was to allow for gradual sets and "fuzzy logic." The amendment he made to standard sets was to allow membership in a set to be a non-binary concept, and then extend the operations on ordinary sets to account for this allowance. If an element x has membership in set A with value v and in set B with a value w, then the operations on sets are adjusted in the following way:

Intersection: The value of x in $A \cap B$ is the minimum of v and w.

Union: The value of x in $A \cup B$ is the maximum of v and w.

Complement: The value of x in the complement of A is $1-v$.

This theory allows for a category of "P and not P" – for example, the category defined by "an apple that is not an apple." The crabapple and Adam's apple are both judged to be members of this category. Zadeh's fuzzy set theory reflected a change in the conception of categories that had been developing in anthropology.

Rosch and prototype theory

About the same time, anthropologists, influenced by the Whorfian hypothesis (1956), assumed that color categories were arbitrary and different languages could carve the color spectrum in different, arbitrary ways. However by the late 1960's it was found that, though different languages do vary in the number and kinds of colors they name, there is regularity in color categories among different cultures (Berlin and Kay, 1969). Speakers of different languages that disagree on color category boundaries will agree on which colors were good examples of these categories. Rosch, who was conducting research on the Dani in New Guinea, began looking at their color categories. She depicts the prior research on categorization in this way:

> When psychologists did research on concept learning, they used artificial concepts and sets of artificial stimuli that were constructed so that they formed little micro-worlds in which those prevailing beliefs about the nature of categories were already built in. Then they'd do their learning experiments. But what they found out in terms of the nature of categories was already a foregone conclusion because that was what they had already built into it (Scharmer, 1999).

Rosch argued that, because of the way the perceptual system works, certain areas in the color space are more salient than others, and that those salient colors are first noticed, most easily remembered, and become prototypes around which color categories form in cultures. The Dani had only two basic color terms, dark and light, making this culture ideal for testing the hypothesis. By teaching them novel color categories, structured around natural and unnatural color schemes, Rosch found that the Dani remembered the hypothesized "universal prototype" colors better than other colors, and it was much easier to learn categories structured around those colors than categories structured some other way. Further research extended this to shapes and other categories, and supported her thesis that

> categories form around and (or) are mentally represented by salient or information rich or highly imaginable stimuli which become prototypes for the category. Other items are judged in relation to these prototypes; that's the way they form gradients of category membership. There don't need to be any attributes which all category members have in common, no defining attributes, and category boundaries don't need to be definite (Scharmer, 1999).

In the research that followed, Rosch (1975) showed that certain category members were judged to be better examples of the category than others. A robin is judged to be a better example of the set birds than a penguin, although, strictly speaking, both are birds. This psychological rendition of Zadeh's fuzzy set theory established fuzzy logic as an important component of AI research. In addition, Rosch's research established experimental paradigms for investigating categorization, attending to features such as the following (using the category "bird" for illustration):

- Direct rating. (How birdlike is this?)
- Reaction time. (Show a picture and ask: Is this a bird?)
- Producing an example. (Draw a bird.)
- Asymmetry of similarity. (Are ducks like robins? Are robins like ducks?)
- Asymmetry of generalization. (Robins get the flu; do ducks? Ducks get the flu; do robins?)

In this paradigm, a prototype will receive a high rating, low reaction time, and resembles the example produced. From this research, properties of categories were determined. These are detailed in the following section.

Properties of categories

It was readily apparent that the structure of real categories, as researched by Rosch and others, was not consistent with the classical view. Categories, which in the classical view were devoid of any internal structure, were shown to possess both centrality gradience (the idea that some categories have members that, though clearly within the category boundaries, are more or less representative of the category) and membership gradience (the idea that some categories have degrees of membership, so that the distinction between member and non-member, the category boundary, is not clear). Members of many categories were related to one another in a "family

resemblance" manner, so that no one property is common among all members. The structure of categories leads to "a basic psychological asymmetry: the less prototypical category member is conceived of as closer (i.e., more similar) to the more prototypical member than vice versa" (Shen, 1999). People will more readily compare a non-prototype to a prototype and will more likely generalize from the prototype to the non-prototype than vice versa (Rips, 1975). There is a primary level of categories, known as the basic level, that are "primary with respect to the following factors: gestalt perception, image formation, motor movement, knowledge organization, ease of cognitive processing, and ease of linguistic expression" (Lakoff, 1980). The most central members of a category can function as metonyms for that category – a property common to many languages. For example, American Sign Language has no sign for the category *jewelry* and in ASL this category is referred to by listing the prototypical members (Newport and Bellugi, 1978 p 62); the Hopi call all trees "cottonwood," the name of the most common deciduous tree in their habitat (Trager, 1936-9); Shoshoni speakers refer to large birds in general as well as to eagles themselves as eagles (Hage and Miller, 1976).

From this data, initial theories on categorization argued for interpreting category membership and structure as degree of similarity to the category prototype. However, further work showed that categories are not defined solely by family resemblance to a prototype, but have an intellectual and ecological basis. Barsalou's (1983) studies of "ad hoc" categories, categories that cannot be interpreted as fixed cognitive structures, such as "foods not to eat when on a diet," or "things to do at a convention," found that members of these categories retain the graded structure and typicality effects that Rosch found. However, these categories did not necessarily show a family resemblance to the prototype. Instead, Barsalou argues, these categories are goal-oriented; a chocolate cake has little resemblance to peanut brittle, but abstaining from these satisfies the goal of eating as few calories as possible. Similarly, research was beginning to reveal that similarity alone could not account for even the more typical, stable categories. For example, "the claim that something is a *dog* does more than assert some degree of similarity to a prototype; it also appeals to our underlying intuitions and beliefs about the nature of animals. The effect of these beliefs is to make some similarities between objects decisive and others simply irrelevant" (Neisser, 1987 p3). The claim that similarity alone is an explanation, according to Goodman (1972), is "a pretender, an imposter, a quack. [Similarity] has, indeed, its place and its uses, but is more often found where it does not belong, professing powers it does not possess." In Murphy and Medin's (1985) paper, "The Role of Theories in Conceptual Coherence" they argue against similarity arguments, claiming that

> Current ideas, maxims, and theories concerning the structure of concepts are insufficient to provide an account of conceptual coherence. All such accounts rely directly or indirectly on the notion of similarity, and we argue that the notion of similarity relationships is not sufficiently constraining to determine which concepts will be coherent or meaningful. These approaches are inadequate, in part, because they fail to represent intra- and inter-concept relations and more general world knowledge. We propose a different approach in which attention is focused on people's theories about the world.

The argument entails that

> categorization assumes a (folk) theory on the part of the person who is engaged in that particular cognitive process. This theory 'guides' him in selecting the relevant features and the relevant feature correlations; in other words, noticing features and feature correlations is not an 'objective' process based on similarity, but is instead theory-dependent." (Shen, 1992)

The effect of these findings is to make some similarities between objects decisive and others irrelevant. With these and other findings, Rosch eventually

> came to the conclusion that prototype effects, defined operationally by experiment, underdetermined mental representations. The effects constrained the possibilities for what representations might be, but there was no one-to-one correspondence between the effects and mental representations. The effects had 'sources,' but one could not determine the sources given the effects. (Lakoff p 43)

An alternative to prototype theory is described below.

Idealized cognitive models

As a theory to explain the prototypes that Rosch first documented, Lakoff, in *Women, Fire and Dangerous Things* claims that "prototype effects result from the nature of cognitive models, which can be viewed as 'theories' of some subject matter." (p. 45) Lakoff terms these models "idealized cognitive models," or ICMs, and suggests that, to the degree to which the model does not represent reality, these ICMs will lead to categorization and prototype effects. A classic example from linguistic research of categorization, prototype effects and gradience of membership in a category is the term *bachelor* (Fillmore, 1982). While most people will define a bachelor as an unmarried adult male, certain unmarried adult males are not representative members of the category of bachelors. Lakoff (1987) argues that

> *bachelor* is defined with respect to an ICM in which there is a human society with (typically monogamous) marriage, and a typical marriageable age. The ideal model says nothing about the existence of priests, "long-term unmarried couplings," homosexuality, Moslems who are permitted four wives and have only three, etc. With respect to this ICM, a *bachelor* is simply an unmarried adult man.

This idealized model, however, does not fit the world very precisely. It is oversimplified in its background assumptions. The person referred to deviates from prototypical bachelorhood if either the ICM fails to fit the world perfectly or the person referred to deviates from being an unmarried adult male.

> Under this account bachelor is not a graded category. It is an all-or-none concept relative to the appropriate ICM. The ICM characterizes representative bachelors. One kind of gradience arises from the degree to which to ungraded ICM fits our knowledge (or assumptions) about the world. (Lakoff, 1987)

ICM's can closely match the world, in which case the categories that you developed to create this ICM are robust categories with little gradience. The idea of the ICM has been further parsed and is perhaps best represented by schema theory. Schemas are short "scripts" or stories that we have about the world and the way it works: *event schemas* that are abstracted from our experience of certain events, *image schemas* that provide structure for conceptualizations – "schemas of intermediate abstractions [between mental images in abstract propositions] that are readily imagined" (Palmer, 1996 p. 66) – and *proposition schemas*: abstractions that act as models of thought and behavior and specify "concepts and the relations which hold among them." (Quinn, 1987) It is only within a particular schema that a category is meaningful, and these categories become les meaningful and exhibit a graded structure to the degree that the schema in which they are defined does not apply. A claim of analogies as assertions of categorization then entails analogies as instantiations of particular schemas.

Metonymy, cognitive models and phenomenological-primitives

Phenomenological primitives (p-prims) (diSessa 1993) were developed to address the "preconceptions" of students in problem solving in physics. They are "the intuitive equivalent of physics laws; they may explain other phenomena, but are not themselves explained with the knowledge system." As defined by diSessa, p-prims are "cued to an active state on the basis of perceived configurations, which are themselves previously activated knowledge structures." In this way p-prims are elements within larger models. P-prims "often originate as minimal abstractions of common phenomena," and are "nearly minimal memory elements, evoked as a whole." By way of example, consider one class of p-prims: the "constraint cluster." This class includes bouncing, supporting, guiding, clamping, and carrying. These p-prims are not fundamental for a physicist (all can be explained in terms of forces) but are often elicited in conversations with students as explanations for physical behavior. The p-prims have a "schematization" such as, for the "supporting" p-prim, "'strong' or stable underlying object keeps overlaying and touching object in place." (diSessa, 1993 p. 216)

That a p-prim has a full schematization but is often represented only partially by a particularly salient feature of the schema (e.g., "supporting" as an explanation entails two objects, one of which is strong or stable and underlies another object which touches it) is indicative of the schematization of a p-prim being an idealized cognitive model. This single salient feature of the schema is what Lakoff refers to as a metonym for the idealized cognitive model. Lakoff presents the following example to explain metonymy is the context of ICMs (Lakoff, 1987):

> A linguist who does fieldwork on Ojibwa, a Native American language... asked speakers of Ojibwa who had come to a party how they got there. He got answers like the following (translated into English):
>
> -I started to come.
> -I stepped into a canoe.
> -I got into a car.
>
> He figured out what was going on when he read Schank and Abelson's *Scripts, Plans, Goals, and Understanding* (1977). Going somewhere in a vehicle involves a structured scenario (or, in our terms, and ICM)... In Ojibwa it is conventional to use the embarcation point of an ICM.

That is to say, the embarcation point is a metonym for the entire structured scenario, or ICM. English, too, uses a point of the journey to refer to the whole. Typical English responses to the question of "How did you get here?" may be: "I have a car," or "I biked." Neither comment conveys the entire journey, but chooses one part to represent the whole. This is only possible because we have a model for the journey, and one part can, metonymically, elicit the whole. In the same way, phenomenological primitives as explanations for physical phenomena are only possible because of a larger schema.

The claim I will make, that the p-prim and schema precedes or is in some way more fundamental than the analogy itself, is echoed in diSessa's studies involving the "Montessori bell conundrum." In this problem, students are presented with bells made of the same material, same length, same height, but varying widths. Almost without exception students predict (erroneously) that the thicker bells will have a lower pitch. DiSessa reports:

> Although most subjects were ready with analogies – church bells compared with jingle bells, xylophones, musical instruments of various sizes – I was struck that some initially could not produce any example of the phenomenon they identified to be at the root of the situation. This, along with the rapidity and expressed certainty of responses, heightened my confidence that a p-prim (or several) was at stake rather than analogy.

In following chapters, I will present data that argue that generated analogies stem from a set of schemas or p-prims.

Summary of categorization research

The main points I will be taking from categorization research are the established properties of categories (including graded structure, asymmetry of generalization, prototypes, family resemblance, and metonymy), that categories can be considered *ad hoc* constructions (and still retain the characteristics of a category), that theories, expectations and goals underlie our construction of some categories, and that graded structure in categories is a reflection on the degree to which these theories match the "real" world.

Analogies as Categorization

Metaphors as category-inclusion statements

Consider the following statements, one typically considered categorization and the second analogy:

1. This ball bearing is a mass.
2. In circuit with an inductor, capacitor and resistor, the inductor is like a mass.

The first is a statement that the ball bearing is a member of the category "mass." The second, however, is not a statement that the inductor is a member of that category, but, I argue, by drawing the analogy we are suggesting that the inductor and that the members of the "mass" category share some categorical grouping in common. One could characterize that category "things which slow the rate of change" (this characterization is referred to as a "ground" in a metaphorical comparison). This category is rarely referred to and therefore is not a stable category with its own name, but more of the *ad hoc* type category that Barsalou (1983) introduced – a spontaneously constructed category that is structured by theories or goals. Glucksberg and Keyser (1990) were the first to identify this relationship and have argued for interpreting metaphorical assertions as categorical assertions, claiming that

> When people use metaphors, they *are* saying exactly what they mean. When, for example, someone says that "Sam is a pig," that is precisely what is meant; that the person designated by the name "Sam" is a member of the superordinate category referred to by the word "pig."

Glucksberg and Keyser argue that the choice of the secondary subject in a metaphor (*pig* in "Sam is a pig") reflects a tendency of languages to have names for basic level objects but not for superordinate categories. Such examples can be seen in the English language, as in the aphorisms "Boys will be boys" or "Cambodia has become Vietnam's Vietnam" (or, as mentioned recently in the 2004 election, "Florida does not want to be the next Florida). These expressions use a "single referring expression… in two distinct ways, to allude to the entity itself and to refer to the category… that this entity has come to exemplify" (Glucksberg and Keyser, 1990 p. 411). (For example, "eagle" in Shoshoni and "cottonwood" in Hopi, as mentioned in a previous section.) When the base of an analogy (termed the vehicle in the context of metaphor) is used as both an exemplar and as an ad hoc name for a category, Glucksberg et al. (1997) call this linguistic move "dual reference." As an example of dual reference,

> the phrase "a responsibility is a shackle" can be used to refer to the concrete, physical device that is made of metal, often has chains, can be locked around someone's arms and legs, and so forth, and it can also be sued to refer to the abstract, general category of constraining entities. We refer to such abstract, general concepts as attributive categories. (Glucksberg et al, 1997 p. 334)

The authors claim that "nouns can be used to make dual reference whenever a superordinate category has not been lexicalized and a category exemplar is available that is prototypical of that category."

Glucksberg and Keysar continue with the assertion that metaphor is not a literal comparison, and must be considered property attributions that extend or create categories. As an example, they contrast the literal comparison with the metaphorical:

> "Copper is like tin"… cannot be paraphrased as category assertions and still make sense, for example, … "Copper is tin."
> Thus the paradox: Two unlike things compared can be paraphrased as a categorical assertion, whereas two like things compared cannot. This paradox may hold the key to a fundamental difference between literal and metaphoric comparisons. We argue that metaphors are not understood as comparisons, but rather as property attributions that either extend old categories or create new ones.

I hold that the difference in these statements is a result of what Roger Brown originally described: "Metaphor differs from other superordinate-subordinate relations in that the superordinate is not given a name of its own. Instead, the name of one subordinate is extended to the other" (Brown, 1958). "Tin" is not a name for the category of which they are members (owing to the fact that this category has a name and that tin is not a prototypical member) and "jail" is (because this category, as it is less stable, has no name and is referred to by its most prototypical member). Furthermore, placing tin and copper in the same category is *expected* and does not violate any previously held ontology or necessitate the construction of a new, ad hoc category, unlike placing jobs and jails into the same category. And, as Barsalou (1983) has shown, ad hoc categories have the same structure and properties as traditional categories. The difference, then, between a metaphorical comparison and a literal one may have cognitive importance in terms of stability of the category and conceptual coherence, but not in terms of the conceptual structure. Shen's (1992) *Metaphor and Categories* makes similar claims as Glucksberg and Keysar. He argues that "in interpreting a metaphorical comparison, an ad hoc category is constructed so that the two metaphorical terms are conceived of as its members," and that the secondary term in a metaphorical statement is typically "a prototypical member of that category." Shen, too, makes a distinction between literal comparisons and metaphorical comparisons, claiming that a literal comparison is indicative of a "common" category, while ad hoc categories are represented by metaphorical comparisons.

Traditionally, metaphors have been interpreted as statements of similitude and not categorization because of the assumption that words (such as "pig" when claim that the person "Sam is a pig") refer to specific taxonomy. Studies on how our *minds* perceive of words and how these perceptions are related to taxonomic definitions versus metaphorical relationships are detailed in the following section.

Idealized cognitive models and lexical networks

Eve Sweetser's 1984 tests of the definition of "lie" (as in "to tell a lie"). She points out that, while most people define a "lie" as "a false statement," in practice:

> A consistent pattern was found: falsity of belief is the most important element of the prototype of *lie*, intended deception the next most important element, and factual falsity is the least important.

These findings are consistent with the idea that humans conceive of the world using idealized cognitive models and that the imperfection of these models results in a graded structure of categories. In the "lie" example the ICM being employed is a model of communication, as studied by Grice (1975). Grice removed the study of language from its Chomskian position of mathematical clarity and tied it to the study of communication arguing that, in order to understand the way language works, one must understand expectations that exist in communication. These expectations (the idealized cognitive model of communication) entail clarity, truth, information, and relevance on behalf of the speaker and influence our expectations for the definitions of words.

If words, such as "lie" are not used in a manner consistent with the definition one would find in a dictionary, how are words represented in the mind? Recent psycholinguistic theory has suggested that the mental lexicon, instead of being organized in a dictionary-style, is far more like a thesaurus. That is to say, the way our minds perceive of words is not so much as obeying rigid definitions with propositional structure, but rather the meaning of one word is tied to a network of related words – words that have appeared in similar contexts, words that have appeared in context with that word, and words that have related meanings. Computationally generated lexical networks have been developed to represent the lexical network of the English language as expressed in dictionaries (one example is the well-known Wordnet (Fellbaum, 1998)). These thesaurus-like structures link words in definitions into a network using various algorithms. Gaume et al. (2002), building on categorization research that they summarize as establishing the "conceptual flexibility" as opposed to "rigid and discontinuous categories," argue that words *themselves* constitute categories and contend that these lexical networks weave a "mental lexicon distributed around metaphoric poles." In this regard, dictionary definitions of terms such as "lie," "pig," and "boys," especially in the contexts noted above, can be expected to fall short of the full meaning of these terms as used in regular language. When these terms are expressed in a lexical-network sense, they can be viewed as representing categories of characteristics or qualities, rendering the metaphorical statements, such as "he's a pig" and "boys will be boys," as assertions of class-inclusion as Glucksberg and Keysar (1992) have argued.

Reconciling categorization and structure-mapping views of analogies

To address the ideas raised by Glucksberg and Keysar and reconcile these with Gentner's structure-mapping theory of analogy, Bowdle and Gentner describe a "path in figurative language comprehension" that claims that there is a shift in the method of comparison in figurative language. Novel metaphors, they claim, are interpreted via a structure-mapping mechanism, while conventionalized metaphors – words that have an original meaning that is different from an often-used meaning (e.g., "roadblock" or "goldmine") – are interpreted as categorization. The issue of *comprehension* of metaphor is not at the heart of this thesis: what someone means when they make a novel analogical statement has to do with the creation of an analogy. It is reasonable, however, to expect that understanding a conventional metaphor might be a similar process to creating your own metaphor: in both cases the categorical commonalities (the ground of the category) is known, while interpreting a novel metaphor would require a "search" of possible meanings of the secondary subject that are being implied.

Conclusion

Words, such as "lie" and "pig," have definitions that exist in a dictionary – stable definitions that people will readily agree on as sensible. These definitions possess a rigid propositional structure, so that "pig" is thought to mean "pink animal with a snout" and "lie" is defined as "an untrue statement." But these definitions are only valid when you accept a certain model about the ways in which we communicate. When you take into account that communication is not so straightforward as our idealized cognitive models of language assume, pinning down the exact description or definition of a word, such as "lie" or "pig," is a much more difficult endeavor – more context dependent and slippery. Lexical maps, which link related terms into a network, more closely approximate the lexical structure (or dictionary) of our minds. Accepting this about language has implications for the interpretation of analogical statements: traditionally, an analogical statement is not literal because the base of the analogy is not intended to be interpreted literally. Instead, I argue that analogical statements are assertions of categorization, and the difference between analogy and traditional categorization is that analogies violate an expected ontology and may even necessitate the construction of a new, ad hoc category. To say that "Sam is a pig" demands that you consider the nodes in the "pig" lexical network that could possibly relate to Sam (including muddy, slovenly, lazy, messy) – just as when one says "I eat pig" (typically viewed as a non-metaphorical statement, but one which calls up the pork chop, ham, and sausage aspects of the pig lexical network). The difference lies in that "pig" is defined in the cognitive model that allows for this animal as distinct from others and "Sam is a pig" requires you to "turn off" some parts of the network.

In the following chapters I will argue that generated analogies in science are statements of category-inclusion that violate the expected categorization of the target. When a student asserts that a cup of water is like a cat in a basket, she is constructing an ad hoc category, this category is intimately tied to a theory and cognitive model, is derivative of this model, and that the secondary item in the analogy is a prototype of this ad hoc category. Furthermore, these categories and their structure are indicative of an underlying idealized cognitive model – ones that are often metonymized by

phenomenological primitives. This interpretation of analogies is responsive to findings from cognitive science, linguistics and education that argue for a manifold ontology of mind.