Do Radical Discoveries Require Ontological Shifts?

Michelene T.H. Chi
chi@pitt.edu

Robert G.M. Hausmann
bobhaus@pitt.edu

University of Pittsburgh
Learning Research and Development Center
3939 O'Hara Street
Pittsburgh, PA 15260
(412)624-7456

To appear in the *International Handbook on Innovation* (Vol. 3), L. V. Shavinina & R. Sternberg (Eds.), New York: Elsevier Science.

The theoretical stance explicated in this chapter assumes that scientific discoveries often require that the problem solver (either the scientist or the inventor) re-conceptualizes the problem in a way that crosses ontological categories. Examples of the highest level of ontological categories are entities, processes, and mental states. Discoveries might be explained as the outcome of the process of switching the problem representation to a different ontological category. Examples from contemporary and the history of science will be presented to support this radical ontological change hypothesis.

Key Words: Scientific discovery, ontological shift, emergent systems, contemporary science, history of science, and representation.

Introduction

This chapter entertains a simple claim. The claim is that many great revolutionary discoveries in science may have occurred because the scientists have undertaken an ontological shift. That is, the scientists had re-conceptualized or re-represented the problem (i.e., the phenomenon to which she/he is seeking an explanation) from the perspective of one ontology or ontological category to another ontology. The remaining chapter explains what this re-representation or shifting across ontological categories entails, and why it is unusual to undertake, thereby explaining the low frequency of revolutionary scientific discoveries. Examples from contemporary and the history of science are cited to exemplify this claim.

The Nature of Ontological Categories

All concepts and ideas (or concepts within ideas and theories) belong to a category. Psychologists define concepts in the context of their category membership. The term concept is used here broadly to refer to a category instance (such as a cat, as an instance of the category Animal, a thunderstorm as an instance of a PROCESS category, and an idea as an instance of a MENTAL STATES category[1]). To assume that all concepts belong to a category is quite standard. For instance, White (1975) also assumed that a concept signifies a way of classifying something. The advantage of categorization is that it allows people to assign the same label to a new instance of the category, and to make inductive and deductive inferences about a new category member (Chi,

---

[1] Trees will be referred to in capital letters, and ontological categories will have its first letter capitalized.

Hutchinson, & Robin, 1989; Collins & Quillian, 1969; Medin & Smith, 1984). Thus, there is clearly a cognitive advantage of having a categorical representation.

What structure does a categorical representation take? Psychologists have by-and-large addressed the hierarchical (subset-superset) nature of categorical representation. Questions of interests have been: Within such a hierarchy, which categories are the basic level (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976); or how are super-ordinate categories coalesced and subordinate categories differentiated (Smith, Carey, & Wiser, 1985)? But most of the attention has been paid to the nature of the features that coheres category members, such as whether they are defining features, characteristic features, core features, explanation-based or theory-based features. In our work, we are concerned primarily with lateral (rather than hierarchical) relationship among the categories. We define the lateral relationship to be an ontological one (Chi, 1997).

Ontology refers to the categorical structure of reality. It has long been assumed, since the times of Aristotle, that things belong to *fundamentally different* ontological categories. What is in dispute, to some extent, is the structure of the ontological categories. Categorical structure is typically hierarchical, with “being” as the topmost category, subsuming everything that exists (Lowe, 1995). Deciding on the structure of ontological categories is a metaphysical question. Our concern, on the other hand, is to propose that *psychologically*, people represent categories with ontological boundaries as well. We are not focusing, at the moment, on the exact structure of such psychological ontological categories, but merely to propose the influence of such ontological categories on discoveries.

We can begin by assuming a few major ontological categories as psychologically real and distinct, as our working definition. Consistent with theories provided by Sommers (1971) and Keil (1979), there is likely no dispute that ENTITIES (or material kinds), PROCESES, and MENTAL STATES are three basic ontological categories, both metaphysically and psychologically. These major categories (or sometimes we will refer to them as trees), with subcategories that they subsume, are shown in Figure 1. Thus, Figure 1 shows a plausible ontological categorical structure. Other major ontological categories that are not shown may be TIME, SPACE, and so forth. Another issue that we are not dealing with is whether metaphysical ontological categories map directly onto psychological ontological categories. We suspect the mapping is not perfect; however, we will sidestep this issue for now.

Insert Fig. 1 here

The psychological reality of ontological categories can be determined by their ontological attributes, which is a set of constraints or properties governing the behavior of members of a given ontological category. For example, objects in the ENTITIES category must have a certain set of constraints that dictate their behavior and the kinds of properties they can have. ENTITIES (such as sand, paint, or men) can have ontological attributes such as being containable, storable, have volume and mass, can be colored, and so forth. In contrast, Events such as war (belonging to the PROCESS ontological tree), do not have these ontological attributes and obey a different set of constraints. Therefore, members of two ontological categories that occupy different trees have mutually exclusive ontological attributes.

Notice that the definition of an ontological attribute is quite different from all the other kinds of categorical features that psychologists have studied. It is not a defining feature, which is an essential feature that a member of a category must have; nor is it a characteristic feature, in which the members of a category frequently or typically have (Chi, 1997). Instead, an ontological attribute is defined by Sommers (1971) as a feature that an entity has the potential to have, such as being colored, even though it does not necessarily have it. War or pain, for instance, do not have the ontological attribute of "being red," whereas squirrels can have such an ontological attribute even though squirrels are commonly not red. Thus, ontological attributes are distinct from other kinds of categorical features in that they are constraints that govern a member of an ontological category, without that member necessarily having that feature.

The psychological reality of ontological categories can be determined by a linguistic test commonly used by philosophers--the predicate sensibility test. As illustrated in Figure 1, each of the major trees (ENTITIES, PROCESSES and MENTAL STATES) generates a hierarchy of subcategories. A predicate (indicated in quotes), which modifies one concept, will sensibly modify all other concepts below it on the same (branch of a) tree (commonly known as "dominates," e.g. Sommers, 1963; or "spans" Wall, 1972), even if it is false. For example, “hungry” can be applied to the category of Animals and all of its subcategories, such as “humans” and “dogs.” Therefore, Animals and Humans are not ontologically distinct. Thus, a bee, an Animal on the ENTITIES tree, has the potential to be "heavy" even though it is false, whereas a bee cannot be "an hour long," a predicate of the PROCESS tree. Conversely, an event can be "an hour long" but not "skinny." The point is that predicates on the same tree can modify concepts below it sensibly even if it is false because it is plausible that a specific concept possesses that attribute, and moreover, the truth or falsity of the sentence can be checked.

When an attribute of one category cannot span members of another category, then the two categories are ontologically distinct, and they will be referred to here as "parallel" categories. For example, "is broken" (an attribute of the Artifact category) cannot be used sensibly to modify an entity from a parallel category such as a dog (a member of the Animal category). That is, "dogs" and "Artifacts" do not involve a direct hierarchal relationship. Similarly, it makes no sense to say, "the thunderstorm is broken." Such statements are called category mistakes. "The thunderstorm is broken" is not merely a falsehood, for otherwise "the thunderstorm is unbroken" would be true. "The thunderstorm is broken" is a category mistake because "broken" is a predicate used to modify physical objects made by humans, whereas a thunderstorm is a type of PROCESS of the Emergent Systems kind. By this definition then, categories on parallel branches of the same tree (such as Natural Kind and Artifacts, see Fig., 1), as well as on branches of different trees (such as Natural Kinds in the ENTITIES tree and Events in the PROCESS tree), are ontologically distinct. However, parallel categories on the same tree do share some attributes. For example, both an Artifact such as a teapot and a Natural Kind, such as a dog, can have color. Nevertheless, they will be considered to be ontologically distinct, as the evidence to be cited below shows, because they have mutually exclusive sets of ontological attributes. In contrast, the attributes of the trees are totally, mutually exclusive.

Gelman (1988) and Keil (1979; 1989) tested the psychological reality of a few ontological categories depicted in Figure 1. Besides the sensibility judgment task, Keil (1989) has also used a physical transformation task such as surgical operations to show that even young 5-year-old children deny the possibility that a toy bird could ever be made into a real bird. Thus, they adamantly honor the distinction between Natural Kinds and Artifacts even at such a young age.

In general, categories are ontologically distinct if one is not super-ordinate of the other. Thus, branches on the same tree can presumably form distinct ontological categories as well, as in the case of the distinction between Natural Kinds and Artifacts. For purposes of our discussion, we will consider the categories on the three major ontological trees in Figure l to be ontologically distinct, as well as "branches" on the same tree that do not occupy a subordinate/super-ordinate relationship. Whether or not there are degrees of ontological distinctiveness (e.g., are humans and dogs less ontologically distinct than Natural Kinds and Artifacts?) remains an epistemological and psychological issue that is not addressed here.

Are Processes of Ontological Shifts Difficult?

What is an ontological shift? By our definition, an ontological shift is merely the re-assignment or re-categorizing of an instance from one ontological category to another. Suppose we acknowledge that Fish and Mammals are distinct ontological categories. Children often mistake “whales” to be a kind of Fish rather than Mammals. However, when children are told that a whale is a mammal, they do not seem to have difficulty reassigning whales to the Mammals category. So it seems that the process of shifting itself may be straightforward, and perhaps it is analogous to a routine learning process of linking or integrating new ideas with old, provided that the old ideas already exist. That is, children already know about the category of Mammals, so that transferring the category membership of whales is not difficult.

The apparent ease of shifting across ontological categories, as demonstrated with the whales example, can be further seen in the ease with which adults shift across ontological boundaries when they use and understand metaphors. That is, metaphors often take the form of combining a predicate from one ontology to modify an entity from another ontology. Take Lakoff's example of anger, which is a MENTAL STATE. In metaphors, anger is conceptualized as a mass Substance, and takes the grammar of mass nouns, as opposed to count nouns. Thus, one would say *How much anger has he got in him?* but not *How many angers does he have in him?* Because anger is treated as a mass Substance, it can occupy space, can be contained, and can come out, such as *He "let out" his anger* or *I could barely "contain" my rage.* This anger Substance is sometimes conceived of as a kind of heated Fluid, such as *She got all* "*steamed up"* or *His pent-up anger "welled up" inside him.* The quoted terms can all be understood as predicates from the Substance category.

Thus, it seems as if shifting across ontological categories is fluid and easily undertaken. On the other hand, there is a class of situations in which an ontological shift is necessary, but people find it extremely difficult to do. This is the case when students misunderstand concepts such as heat transfer, electricity, natural selection (Chi, 1997; Ferrari & Chi, 1998). Thousands of studies have documented the misconceptions students hold and the failure of innovative instruction to remove their misconceptions.

This difficulty in correcting their misconception may have arisen, we proposed, from the possibility that correcting their misconceptions requires an ontological shift (Chi, in press; Chi & Roscoe, 2002). For example, students tend to conceive of concepts such as heat transfer as a kind of Causal Event, whereas they are in fact a kind of Emergent System (see Fig. 1). Causal Events are processes with activities that occur with temporal and spatial contiguity, often with an identifiable agent that directly or indirectly produces an effect. Emergent Systems, on the other hand, are similar to complex dynamic systems (Casti, 1994; Holland, 1998), in which numerous individual agents or elements independently interact in some uniform way to give rise to complex behavior at the macro level that is not exhibited at the level of the individual elements. For example, each element can follow a simple rule simultaneously and independently along with all the other elements, without any explicit goal to achieve. Yet, a macro pattern will emerge that looks as if the system is intending to achieve a certain goal, and that the process of achieving this goal is orchestrated by some director/agent in some sequential way.

Besides this special case of misconceptions in the context of learning and understanding, shifting ontologically may be difficult (if not directly told to do so) for three basic reasons. The first is a lack of the alternative category. The second is the lack of awareness of the need to shift. Finally, the third is that it is resource-intensive. We discuss the first reason briefly since it does not apply to scientific discoveries and elaborate on the latter two reasons more extensively.

Lack of the Alternative Categories

This reason may not be a relevant one for discovery, but in learning science concepts, students often fail to successfully re-conceptualize a phenomenon or concept with respect to the right ontological category because they may lack knowledge about the target ontological category. In particular, as mentioned above, we have identified a class of science concepts that entail an Emergent System kind rather than a Causal Event kind of PROCESS (see Fig. 1 under PROCESSES). Students are unfamiliar with this Emergent kind of PROCESSES; therefore, one could assume that they cannot shift their representation of processes from a Causal to an Emergent kind when that category does not yet exist in their knowledge (Chi, in press).

Lack of Awareness

Even if people do have the alternative ontological categories, they may still not shift spontaneously because they may not be aware that they have to shift their representation of an entity or phenomenon from one ontological category to another. This lack of awareness may arise from four sources. First, it is a low frequency phenomenon. That is, we do not routinely need to re-classify a phenomenon from being of one kind to another kind. We generally categorize an entity correctly on the basis of its outward appearance. How we represent an everyday entity or phenomenon usually corresponds to its true identity. For instance, when we identify an animal that is furry with a wagging tail as a Dog, our categorization is usually accurate. Thus, it is seldom the case that the core features of an entity or phenomenon are in fact different from its surface features, so that we have misclassified it and have to shift. There are a few exceptions. For example, we classify whales as a kind of Fish because it has many of the characteristic features of a fish, such as swimming in water, shaped like a fish and so on. This is one of the few instances in which the surface features belie the true core feature. Occasionally, we must shift in the context of reading stories and watching films. For example, in the popular children's novel *Indian in the Cupboard*, the central theme of the book is that a toy Indian comes alive. A great deal of suspense usually surrounds such a kind of conversion and the main character is clearly surprised. A similar ploy is used in *Velveteen Rabbit*. Similarly, in the film *The Crying Game*, a male character is disguised as a female. The viewers are surprised when the disguise is revealed. Notice that this kind of revelations in stories and movies are much more dramatic than mistaken identity. Misidentifying one person (e.g., an undercover agent) for another (a preacher), as often occurs in mysteries, is suspenseful, but not as dramatic because no ontological categories had to be crossed: Both an undercover agent and a preacher are a kind of Man. Whereas, discovering that a character is really a Man rather than a Woman requires shifting that person's categorical membership. The drama, in the case of crossing ontological categories, arises because the viewers or readers must revisit and revise all the previous assumptions and implications about the character when he was a male because a new set of properties and features have to be inherited when an instance changes membership. Thus, since we are not accustomed to re-represent or re-classify a phenomenon or entity from one ontological category to another, we are not aware that such a shift is sometimes necessary.

However, mistaken classification based on the perceptual features is precisely the source of misconceptions in science (Chi & Roscoe, 2002; Chi, Slotta, & de Leeuw, 1994). For example, we perceive and feel heat flow from one location to another.

Therefore, when heat flow is conceptualized as a Causal Event, it appears as if the hot molecules move from one location to another location, when in fact the flow sensation derives from the collective random bombardment of all the moving molecules moving at different speeds (Chi, in press).

A second reason for a lack of awareness for the need of ontological shift is that our mis-representation is often adequate for everyday functioning. This is especially true in the case of our science misconceptions. It's perfectly adequate for us to close the window in order not to let the heat escape, thereby treating heat as a kind of Substance that can flow from one space to another space, when in fact heat is a kind of an Emergent Systems PROCESS. Thus, our mistaken classification provides an adequate explanation for the phenomena that we encounter everyday, so that we are not aware of our misconceptions.

Finally, we have a tendency to pursue endeavors in a rigid and routine manner, analogous to functional fixedness (Ohlsson, 1992). For example, consider Duncker's famous candle problem. The participant is presented with three objects on a table: a book of matches, a box of tacks, and a candle. The goal is to attach the candle to the wall so it will burn normally. The solution requires the solver to reclassify the box, not as a *container*, but as a *platform*. Subjects solve the problem correctly 86-100% of the time when the problem is presented with the box of tacks empty, versus a solution rate of 41-43% when the box is full (Weisberg & Suls, 1973). That is, there is no reason for us to think about re-classifying the box when it is full since we can continue to think of it as a container. Thus, even reclassifying within the same ontology is uncommon.

Resource-Intensiveness

Although we have assumed in a preceding section that shifting itself may not be difficult in the context of using and understanding metaphors, shifting across ontological categories may be resource-intensive in other contexts. That is, the shifting itself may not be effortful, but once undertaken, efforts and cognitive resources have to be allocated for the process of re-inheriting a new set of features. Although there is no direct evidence on this point, there is some suggestive evidence that even shifting perspectives takes extra effort. Black, Turner, & Bower (1979) have found, for example, that people prefer to maintain the same point of view when processing a narrative text. (This implies that changing perspective or one's representation, may require extra cognitive resources, even though changing perspectives is not the same as shifting across ontological categories. See discussion below.)

In sum, our assumption is that shifting per se is not difficult. The ease of shifting per se can be supported by two kinds of anecdotal evidence mentioned above. In the first case, the ease of shifting can be seen in children when they are told that whales are a kind of mammal. Similarly, in novels and movies, children and adults have no difficulty in shifting their representation of a stuffed toy to a live person, or from a man to a woman, and so forth. Another example of ease of crossing ontological categories comes from adults' comfortable use and understanding of metaphors. However, these kinds of shifting occur when we are explicitly told to shift, or we are presented with a metaphorical usage that includes a combination of two ontological categories. Shifting on our own, without being told, may be more cumbersome, largely because we are not aware of the necessity to shift. Finally, the shifting itself may not be the bottleneck (assuming one knows about the category to which one is shifting), what is difficult may be the resource-intensiveness of re-inhering all the attributes of the alternative category. For example, it is difficult for some cognitive psychologists to consider cognition not as knowledge (a Substance) in the head but as a PROCESS of interaction, the difficulty arises not because we cannot undertake such a shift, but because we cannot easily understand all the ramifications for our views on memory, learning, and problem solving (see the debate between Anderson, Reder, & Simon, 1996; Clancey, 1996).

Re-representation and Ontological Shifts

An ontological shift is a kind of re-representation. However, it is distinctly different from other kinds of re-representation. We discuss five other kinds of re-representation. But before discussing them, a few remarks should be made about general issues regarding representation. First, there is no dispute that certain representations of a problem may make a problem easier or harder to solve. For example, in order to give commands to the programming tool "turtle graphics" (Papert, 1980) for the turtle to draw lines across the graphics screen, some problems are easier to solve by taking the turtle's point of view, and other problems are easier to solve by taking a more global point of view. Similarly, Hayes and Simon (1977) found differences in solution times for isomorphic representations of the classic Tower of Hanoi problem. Solution times were twice as fast if the problem is represented as Monsters exchanging Globes, rather than Monsters changing the size of the Globes. One of the reasons that a problem can be more easily solved in one representation than another has to do with the constraints of the problem. Sometimes the constraints become more obvious in one representation versus another. At other times the constraints become an integral part of the representation (being built-in in some sense). And yet at other times the constraints can be released. Thus, whether a specific representation is more or less beneficial to arriving at the problem solution is a different issue from the issue that concerns us here: namely, whether or not shifting between representations (or re-representing) is beneficial, and if so, why it is not undertaken more frequently, as may be necessary in scientific discoveries.

Changing Perspectives

The most often discussed notion of changing representation is in the context of development. Piaget and Inhelder (1956), for example, discuss the way younger children cannot (but older children can) see things from another physical perspective besides their own. Similarly, younger children cannot (but older children can) acknowledge the listener's point of view. For example, Shatz and Gelman (1973) showed that young 2-year-olds did not adjust their speech to the age (and knowledge) of the listener, whereas 4-yar-olds did adjust their speech depending on whether they were speaking to another peer or an adult. Thus these developmental studies implicates that children are capable of shifting their perspectives as they get older.

In the adult literature, besides Papert's (1980) work cited above, Hutchins and Levine (1981) have shown that problem solvers do change perspectives as they solve the Missionaries and Cannibals problem. They used deictic verbs such as "come," "go," "take," "send," "bring," and place adverbs such as "here," "there," "across" to determine the solver's point of view, such as viewing the river that the Missionaries and Cannibals have to cross, from either the left bank or the right bank. One of their interesting findings was that when solvers were “blocked” in their solution, in the sense that they have made two non-progressive moves out of a problem solving state (in the problem space), then they were successful in becoming unblocked when they changed their point of view. So clearly, being able to shift one’s point of view is beneficial for problem solving, and it is a form of re-representation.

Re-representing at Different Levels

Instead of changing perspectives, representational shifts can also be considered in the context of levels. There are four ways to think about levels.

Shallow and deep levels. One way to think about levels is in the context of shallow surface features (such as the entities mentioned in a problem statement) and deeper semantic features (such the principles that govern the solution procedure). So for example, Chi, Feltovich and Glaser (1981) showed that expert physicists represented routine physics problems at a conceptually deeper level (in terms of the principles that guide the solution), whereas physics novices (those who have taken one course in college with an A grade), tended to represent the same problems according to their literal components (such as the pulleys and the inclined planes). Obviously the underlying principles of the problems determine their problem solution and not the literal components. The implication of this work is that as one acquires expertise, one’s representation changes so that it’s organized according to the principles of physics rather than the physical components.

Subset-superset hierarchical levels. Besides re-representation between levels in the sense of from a shallow more surface-oriented level to a deeper more conceptually-oriented level, a second form of re-representation occurs between hierarchical (subset-superset) levels of categories. Suppose we ask students to solve the following two "insight" problems:

1. A man who lived in a small town married twenty different women in that town. All are still living and he never divorced a single one of them. Yet, he broke no laws. How can you explain this?
2. Two strings hang from a ceiling. They are hung far enough apart that a person cannot reach both strings at the same time. The goal is to tie the strings together. Lying on a nearby table are a hammer and a saw.

These problems are typically considered to be difficult (thereby called "insight" problems). To solve them requires that one re-represents an entity within the problem. In the first problem, the solution is to re-represent the man not as a “bachelor,” but as a “clergyman” (a specific type of a “bachelor”). Similarly, in order to solve the second problem, one must re-represent hammer not as a "tool,” but as a "heavy tool.” This heavy tool can then act as a weight to be tied to one of the strings to create a pendulum. One simply swings the pendulum, grabs the first string, and then catches the other string on its return. In both cases, the entities in the problems are re-represented as an instance of a subcategory. This is an example of re-representation within hierarchical levels. Hierarchical levels satisfy the relationship of “kind of.” That is, a clergyman is a kind of a “bachelor,” and a “hammer” is a kind of a “heavy tool.”

Component levels. A third way to think about levels is in terms of decomposition. For example, if we were asked to explain how the human circulatory works in terms of its function of delivering oxygen to body parts, we would explain it by appealing to the components of the circulatory system, such as the heart, the lungs, blood, and blood vessels, and that it is the contraction of the heart that sends blood to different parts of the body. One can then further ask how does the heart contract. To answer this question, we would have to discuss the components of the heart, such as the role of the rise of ventricular pressure and so on. Basically, each question and its accompanying explanation must reduce each component into its finer and finer constituent parts. Miyake (1986) collected protocol data that illustrated representational shifts in terms of a reduction-decomposition approach to levels. She showed that dyads, in attempting to understand how a sewing machine works, would move to lower and lower levels when they recognized that they had not understood the mechanism. For example, in figuring out how a stitch is made, one can understand it by explaining that the needle pushes a loop of the upper thread through the material to the underside, so that the upper thread loops entirely around the lower thread. However, in order to understand how this looping mechanism works, one has to explain the mechanism at a lower level, of how the bottom thread is able to go through the loop of the upper thread.

Component levels satisfy a “part of” relationship, rather than a “kind of” relationship. In these examples, ventricular pressure is part of the cardiovascular system, and looping is part of the mechanism of making a stitch.

Emergent levels. A fourth way to think about levels is one in which the relationship between them is an “emergent” one (Chi, in press; Resnick & Wilensky, 1998). (See the discussion about Emergent Systems above.) In this kind of level, the (often) observable “macro” level behaves independently of the “micro” level objects.

Moreover, the macro observable level arises from local interactions of the micro level individuals. The most commonsense example is a traffic jam. A traffic jam is a gridlock of cars, so that cars can no longer move at the same speed before the jam. This is the macro-level phenomenon. The behavior of the individual cars in a jam is independent of the jam. Each individual car may be following the same simple rule, which is to accelerate if there was no car in front within a certain distance, and slow down when a car comes within a given distance. The independence of the macro and micro levels can be seen in that the jam itself can move backward even though the individual cars move forward. However, the jam arises or emerges from the interaction of the cars.

For another example, changes in a moth's pigmentation, due to the smoky industrialization in England in the middle of the 19th century, can be understood in terms of emergent levels as well. That is, the emergent pattern, that moths were getting darker over generations, occurred because individual moths were independently being eaten or not eaten by birds. It so happened that lighter moths, which perched on tree trunks that were getting sootier and sootier, became more apparent and visible to hungry birds. Thus, it was more likely that the lighter moths were more visible and thereby were eaten while darker moths tended to survive. Even though one cannot specify on an absolute basis whether a given dark or light moth (on a relative scale) would survive (since it depended on a number of local conditions, such as whether a moth happens to land on a darkly sooted tree trunk, and whether a hungry bird happens to be nearby and can see the moth) or not. Nevertheless, over generations, the probability is such that the darker ones tended to survive and reproduce, thus producing the changing pattern of moths getting darker over generations. Thus, this overall pattern is an emergent one. It is not caused by any specific actions on the part of any specific moth or groups of moths. This is not a case of re-representation from one level to another lower level, as in the example of the sewing machine. Rather, this is a re-representation from a macro level to a representation of the relationship between the micro and the macro level, in order to understand the macro level pattern. Representing the inter-level emergent relationship is extremely difficult for students (Chi, in press).

In sum, two different kinds of re-representation are described above, one involving changing spatial perspectives and one involving four kinds of changing levels. Among these five examples of re-representation, only the last one constitutes an ontological shift. In that case, the shift is between the ontologies of “Causal Events” and “Emergent Systems” (see Fig. 1 again). The fact that only one kind of re-representation involves an ontological shift (or sometimes referred to as radical conceptual change, Chi, 1992), may explain why it occurs infrequently. In the next section, we revisit whether or not we have evidence that ontological shifts are possible.

Evidence of Ontological Shifts

Besides the anecdotal evidence that people can shift across ontological categories readily, as in the case of children re-representing *whale as a mammal* or the case of understanding stories about an *Indian in the Cupboard,* is there evidence that ontological shifts actually occur? There are two kinds of evidence. We have evidence of ontological shifts when we compare novices with experts. In one study, we compared the explanations of ninth grade students with those of graduate physics students on simple conceptual problems, such as *Explain which coffee cup is a better insulator, a styrofoam* *cup or a ceramic cup?* We coded their explanations not for accuracy (which would not be fair to the high school students), but for the kind of predicates they used in their explanations. The predicates they use would indicate which ontological category they represent the concept/phenomenon. Ninth graders (the novices) would explain their choice of the coffee cup with justifications such as *The coffee in the ceramic mug is hotter because the heat in the Styrofoam cup is "gonna escape."* In contrast, the graduate students' (the experts) answers might say something to the effect that *The energy loss in the ceramic cup is through transfer of heat from a hotter source to a cooler source, due to the movement or "motion of electrons."* The basic difference in their explanations is that novices use Substance-based predicates. That is, viewing heat as a kind of Substance that can be contained, stored, or escaped; whereas experts use PROCESS-based predicates such as motion of electrons (Slotta, Chi, & Joram, 1995). Thus, experts appeared to have undertaken an ontological shift whereas novices have not.

A better example that is not contaminated by knowledge of the domain jargon comes from our study with expert swimming coaches. Two experts with a minimum of 12 years as full-time head coach and who have produced from 20 to 100 top national caliber swimmers were compared with two novices coaches with a maximum of two years full-time coaching experience. Their task was to diagnose underwater films of four swimmers performing the freestyle stroke. The general diagnoses rendered by the novices focused on specific body parts, such as *the elbow was bent on extension,* or *the right arm not underneath the body.* In contrast, experts tended to diagnosis more holistically, referring to processes, such as *Little and unequal body roll,* or *stroke unbalanced*. Basically, the more novice coaches diagnosed swimming deficiencies by attending to individual body parts (Concrete ENTITIES), whereas the experts focused on aspects of the swimming PROCESSES, such as body roll (Leas & Chi, 1993). According to the aforementioned studies (Leas & Chi, 1993; Slotta et al., 1995), one undergoes an ontological shift in the process of acquiring expertise.

Does Ontological Shift Underlie Discoveries?

As in the case of resistance to ontological shifts in learning about many science concepts (i.e., the robustness of misconceptions), it may be the case that scientists are also resistant to change and choose to persist in generating explanations or hypotheses within the same ontology. The thesis of this chapter is to explore the claim that major scientific discoveries may have arisen because the scientist underwent an ontological shift, in terms of the nature of the novel, revolutionary explanations. Few empirical studies directly support this claim; however, Hutchins and Levine (1981) found that subjects who were blocked in their problem solving become unblocked when they shifted their perspective.[2] However, as we said earlier, shifting perspective is not exactly the same thing as shifting across ontological categories. Below, we review two domains of scientific discovery and/or scientific revolution: one in contemporary (20th century) science, and the last one in the history of science, to show that these discoveries did require an ontological shift. By scientific revolution, we mean changes that are more extensive than at the level of individual theories, but rather changes that involve the research tradition or paradigm under which specific theories are constructed. We use the term paradigm in the same sense as Kuhn (1996), Lakatos (1970), and Laudan (1977),

---

[2] See also (Andersen, 2002) for historical evidence.

which considered research paradigms as world-views that encompass a set of theories with similar assumptions and similar concepts. "Similar" can be conceptualized, in our opinion, as those belonging to the same ontology. So for instance, "evolutionary theory" really refers to a family of doctrines that all assume that organic species have common lines of descent. All variants of evolutionary theories would implicitly make that assumption. Thus, theories within the same paradigm basically are modifications and extensions of one another, since they adopt the same assumptions and concepts, scrutinize the same set of problems, and use the same sort of methodologies and instruments to do their science.

Some Examples of Ontological Shifts in Contemporary 20th Century Theories

In this section, we briefly describe what may be considered major breakthroughs in the 20th century, drawing on examples from the treatment of diseases as well as theories for natural or scientific phenomena. This treatment is very superficial, and captures merely the highlights of the changes in the scientific explanations, as rendered by other scientists.

Peptic ulcers. Thagard (1998a; 1998b) detailed the historical shifts in the causal explanation for peptic ulcers. The prevailing hypothesis prior to 1979 for the cause of stomach ulcers was an excess acidity in the stomach. The increased production of gastric acid was thought to be caused by elevated levels of stress experienced by an individual. The causal sequence of events that eventually lead to an ulcerated stomach lining were: increase in stress levels, which lead to the increased production of stomach acidity, which then lead to erosion of the stomach lining. There was a general agreement and satisfaction with the stress leading to excess acidity model, although treatment was only successful at reducing acidity (the symptom), but not stress (the cause).

Evidence against the prevailing explanation came from a source not initially designed to explain the etiology of ulcers. In 1979, while Warren (a pathologist) was performing an autopsy on a patient, who was diagnosed with nonulcer dyspepsia and gastritis, he observed an unusually high count of bacteria in his stomach. Warren's observation went against conventional wisdom because the medical community believed the stomach was too hostile of an environment for the survival of microorganisms. Warren began collaborating with Marshall (a gasteroenterologist), and they noted a correlation between gastritis and bacterial infection, as well as duodenal ulcer and bacteria. To strengthen their hypothesis, Warren and Marshall successfully treated 90% of their patients, who were diagnosed with duodenal ulcers, with antibiotics. After several different types of experimental studies, Warren and Marshall finally proposed the hypothesis that the bacteria, *Helicobacter pylori*, were responsible for peptic ulcers. Today, the bacteria hypothesis remains the most viable explanation for peptic ulcers because it succeeds in treating patients at the level of the hypothesized causal mechanism, instead of merely controlling the symptoms. Instead of using antacids to control the increased stomach acidity, physicians now also prescribe a regimen of antibiotics to eradicate the infection (Graham, 1993). Notice that a breakthrough in finding a better causal explanation of peptic ulcer consisted of shifting the ontological categories of the explanations, from one of MENTAL STATES (stress) to one of Animate Concrete ENTITIES (bacteria).

Dinosaur extinction. Clearly, one of the most intriguing scientific issues of the modern era has been to provide a coherent explanation for the massive extinction of dinosaurs, nearly 65 million years ago. Several interesting ideas, ranging from infertility due to rising climatic temperature, ill reactions to newly evolved flowering plants, cataclysmic events (such as massive meteors or massive volcanic activity), all attempted to explain why dinosaurs became extinct (Gould, 1985). However, several of these theories are not viable because they are not consistent with other empirical observations. For example, because the extinction was on a global scale, an acceptable theory must also explain why oceanic life (such as plankton) also died at the same time, and it must be consistent with the fossil record. Therefore, the first two explanations (infertility and increased toxicity in flowering plants) must be ruled out because neither hypothesis can explain why plant and oceanic life also died out (Gould, 1985).

The prevailing dogma, which finds its intellectual roots in the gradualist ideas put forth by geologist Charles Lyell, was that the Earth was scalped and modified over millions of years (W. Alvarez, 1997). The Earth has undergone long periods of climatic changes (the “ice age” of the Quaternary Period, for example), as well as changes in sea level. To explain dinosaur extinction, gradualists assumed that, like every other species, dinosaurs became extinct because they could not adapt to the climatic changes. Scientists of the day argued that an incomplete fossil record is evidence that dinosaurs died out over a protracted period (W. Alvarez, 1997). At this point, iridium did not play a role in the development of a theory of dinosaur extinction. According to the gradualist perspective, if iridium existed on Earth, then it would have had an Earth-bound origin (volcanic activity, for example).

Although the gradualist explanation seemed congruent with the available evidence, Walter and Luis Alvarez proposed a different theory (L. W. Alvarez, Alvarez, Asaro, & Michel, 1980). They proposed that an enormous meteor struck the earth, which created enough dust to block the sun around the planet. They reasoned that, the obstruction of sunlight would cause the process of photosynthesis to stop, thereby causing the entire food chain to suffer. The impact hypothesis satisfies the two aforementioned constraints. First, the impact hypothesis explains why there is global extinction (not just animals, but plants and oceanic life as well). Second, the hypothesis is consistent with the fossil record because the amount of iridium embedded within the Crustaceous-Tertiary boundary is orders of magnitude higher than the amount found on Earth during any other time period. High concentrations of iridium are mainly found on asteroids or comets, as well as the core of the Earth. To rule out the alternative hypothesis that volcanic activity caused the extinction of the dinosaurs, the Alverez team turned to a different strand of evidence. They demonstrated that a specific type of rock, called "shock quartz," was also observed at the same time period. Shock quartz is formed by a violent strike to a layer of quartz. Because of the overlapping evidence, the meteor hypothesis gained popularity.

These two hypotheses have maintained the scientists' interest perhaps because they are ontologically distinct. The impact theory is based on the notion of a one-time discrete Causal Event, involving a Concrete ENTITY such as a meteorite, whereas the gradualist explanation can be considered a continuous PROCESS of climatic and sea-level changes, perhaps likened to an Emergent System.

Coronary heart disease. Discovering the cause for coronary heart disease (such as acute myocardial infarction) represents one of the most significant challenges to the medical community because an estimated 7,500,000 Americans die each year from myocardial infarction (American Heart Association, 2001). The direct explanation for the cause of heart attacks has been atherosclerosis, which is the thickening of the coronary arteries, thereby restricting the flow of blood. The question is what caused such thickening to occur. The prevailing explanation is that arteries are thickened by cholesterol deposits into the artery walls and production of atherosclerotic plaques. The plaques cause a narrowing of the vessels. In the last two decades, Americans have been obsessed with diets that can reduce their cholesterol level to an acceptable healthy level. More recently, however, two alternative theories have been proposed. One is the hypothesis that, instead of cholesterol, it is iron overload that causes heart attacks. This hypothesis is consistent with a number of pieces of evidence. For instance, men who regularly donate blood have a lower risk of heart disease, and pre-menopausal women who regularly lose blood, and thus iron, also have a lower risk of heart disease. Tuomainen et al. (1998), using a case-control study of Finnish men, found that individuals with high iron levels were more likely to have suffered an acute myocardial infarction in the past 6-7 years. Presumably, iron can increase “oxidative stress” on the lining of the blood vessels. That is, iron can somehow interfere with nitric oxide, a chemical that relaxes blood vessel walls, allowing the blood to flow more freely.

A second promising new hypothesis is currently being proposed, that coronary heart disease is caused by inflammation in the bloodstream. This hypothesis is compatible with the finding that many people (in fact, over 50%) with no known risk factors (such as high cholesterol levels) nevertheless do have heart attacks. Moreover, an enzyme called myeloperoxidase (MPO) and a substance called interleukin 6 were both elevated among people who had heart attacks and narrowed coronary arteries (Zhang et al., 2001). Both substances are associated with inflammation. For example, MPO is normally found in infection-fighting white blood cells, so that their elevation indicates the body's attempt to fight inflammation. Using a cross-sectional survey, Meier et al. (1999) found adults who had taken tetracyclines and quinolones (antibiotics) showed a reduced likelihood of experiencing an acute myocardial infarction. The exact mechanism that mediates atherosclerosis through inflammation is not completely clear. One possibility is that bacteria and other infections (such as gum infection) can cause clot-forming cells (or platelets) to clump together, thereby causing arteriole blockage.

It seems that there is a rapidly growing body of evidence that inflammation in the bloodstream can cause heart attacks. Moreover, there is a growing consensus among scientists that other disorders, such as colon cancer, Alzheimer's disease, may also be caused by chronic inflammation. Hence, inflammation seems to be the hypothesis that is favored at the moment (in the sense that an increasing number of studies are testing this hypothesis and trying to understand its mechanism). The excitement, we belief, arises from the fact that a hypothesis based on inflammation (a PROCESS) is ontologically distinct from a hypothesis based on both cholesterol deposits and iron overload (both cholesterol and iron are a kind of Substance). Thus, an ontological shift in hypothesizing implies that the new hypothesis has inherited many other ontological attributes that require scientists to further investigate its ramifications.

Some Examples from the History of Science

In this section, a few examples are cited of major scientific discoveries that may be thought of as having undergone ontological shifts.

Epilepsy and other diseases. Between 430 and 330 B.C., the approach to diseases was a religious one. For example, epilepsy was caused by divine visitation (Thagard, 1996). Divine beliefs fit the ontology of a kind of MENTAL STATES. Consequently, treatment of diseases was consistent with the explanation at the time, such as by appealing to the gods for mercy or other magic.

After 330 B.C., Hippocrates rejected the religious approach by postulating that epilepsy is caused by an excess of phlegm, one of the four fluids (or humors) that were thought to constitute the human body. Healthy bodies supposedly had the correct proportion of each of the four fluids (blood, phlegm, yellow bile, and black bile). Imbalances in these fluids were thought to produce diseases, and different kinds of imbalances produced different diseases. Treatment of course, was again, compatible with the prevailing explanation. To redress humeral imbalance, one would change one's diet, or rid the body of excess phlegm by inducing vomiting or evacuation of the bowels, and/or letting blood from the veins.

One can say that Hippocrates' humeral theory is radically different from the divine visitation theory because the explanation of epilepsy had undergone a major ontological shift, from a kind of a MENTAL STATE to a kind of Substance. In the 19th century, another major conceptual shift occurred, this time to postulate that diseases were caused by bacteria and other microorganisms. This breakthrough occurred from, presumably, the analogy that Pasteur made in the late 19th century between fermentation and disease. Since fermentation was caused by yeast and bacteria, so too can diseases be caused by microorganisms. Thus, another ontological shift had occurred, from one of Inanimate Substance (imbalance of phlegm) to Animate ENTITIES (mircro-organisms).

Notice that before Pasteur, Fracastoro had posed another theory, the notion of "contagion," which can be thought of as "seeds" that can be passed from one person to another. Treatment would therefore consist of expelling and destroying these seeds, rather than restoring humeral imbalance. Although the contagion theory and the humeral theory appear to be rather different, they were not ontologically distinct. They are both a kind of Substance. Thus, Fracastoro's theory was not a breakthrough and had little influence after 1650 (Thagard, 1996).

Theory of evolution through natural selection. In the 18th century, Lamarck postulated that environmental forces modified an organism's morphology. The classic example is the giraffe's long neck. To explain why the giraffe has a long neck, Lamarck hypothesized that reaching for leaves on tall trees gave the giraffe a longer neck. These changes in animal physiology could then be passed onto the organism's offspring. The central idea in Lamarck's theory of evolution was that traits could be acquired from intentional use. If there were a change in the animal's environment, the animal would then be able to respond (i.e., adapt) by changing its traits. Thus, the Lamarckian theory proposed that evolution is a kind of a Causal Event, in that the agent (the giraffe in this case) can directly cause and manipulate (by using and stretching his neck, for example) his own traits to suit the environment.

In the 19th century, Darwin rejected the idea that animals have the capacity to actively change their heritable traits (Gould, 1985). Instead, he postulated a radically different mechanism: natural selection. Darwin's theory can be stated in the following way: Basically, each giraffe is different from every other giraffe, from variability in the inherited genes. So some giraffes have longer necks and others shorter necks. The ones with longer necks can reach their food source, so they are more likely to survive and reproduce, although some of the shorter neck giraffes may also survive if some short trees happen to be around. But in general, the taller giraffes have an advantage, given that trees tend to be tall. Because offspring inherit such a trait (of long necks), then the taller giraffes, having had more opportunities to survive and reproduce, tend to produce offspring with longer necks. Over generations, the entire population of giraffes tends to have longer and longer necks.

The shift from intentional modification of heritable traits to probabilistic survival and reproduction due to random variation in a trait represents a radical ontological shift. Evolution via intentional adaptation might be categorized as a kind of a Causal Event, whereas Darwin's theory might be best understood as a kind of an Emergent System. (See Chi, in press, for more details.)

Electricity. Another ontological shift can be found in the theoretical development surrounding the topic of electricity. Among the very first documented cases of static electricity (i.e., ancient Greece) included the observation that amber attracted small pieces of straw after it was rubbed (Meyer, 1971). These early experiments are now considered demonstrations in electrostatics. The explanations generated by early theorists and experimenters were targeted at explaining why two objects are attracted to one-another after being rubbed with a cloth or fur. The first explanation for the electromotive forces began with a type of Material Substance called "effluvium." In

1600, William Gilbert proposed the effluvium stretched out across space to hold two objects together. The larger the space, the thinner the effluvium became, and the weaker the felt attraction. The experimental data confirming the existence of the effluvium as a Substance was the observation that the interposition of any object would break the attraction between two objects (Home, 1981). This was taken as evidence as a Material Substance because one property of material objects is that they are unable to occupy the same space. Home (1981) provides a description of the antiquated construct:

> The effluvia are described at one point as ‘fine,’ at another as ‘subtle;’ however, it is emphasized that they are material, and as such are subject to ordinary physical laws. The presence of moisture and other impurities in the atmosphere can hinder the action for the effluvia in two ways: the moisture will condense on the surface of the glass or other substance to be electrified, and in doing so may block the pores from which the effluvia are normally emitted; alternatively, the mere presence of impurities in the atmosphere should increase its density and hence the resistance it offers to the motion of the effluvia. (p. 5)

Although initially successful, the effluvium theory was unable to account for a few enigmatic observations. For example, the theory required some medium for the action at a distance. It was hypothesized that the effluvium stretched out across the air, from one object to the other. However, experiments conducted after the introduction of the vacuum demonstrated that a medium was unnecessary to observe the same results. Another difficulty was the observation that some charges attract while others repel. Therefore, it became increasing apparent that the effluvium theory was incomplete.

The effluvium theory was eventually replaced by Benjamin Franklin’s single fluid theory of electricity (Meyer, 1971). Objects were considered positive if they contained more fluid, and negative if they had less fluid. The flow of electricity, according to Franklin, was from positive (objects with more fluid) to negative (objects with less fluid). Franklin also made the assertion that the electrostatic demonstrations in the lab also applied to natural phenomena, such as lightning (Holyoak & Thagard, 1995). Franklin did, however, express some doubt as to the fluid theory. He stated, “The electrical matter consists of particles extremely subtle since it can permeate common matter, even the densest, with such freedom and ease as not to receive any appreciable resistance” (quoted in Hecht, 1994). It seems reasonable for Franklin and his contemporaries to conceptualize electricity as a fluid, given the popularity of “fluid-like” explanations for other physical phenomena (e.g., the caloric theory of heat and the impetus theory of motion).

Electricity, however, did not remain the flow of a fluid. First, in a series of experiments, Faraday provided empirical support for the claim that different “forms” of electricity were actually manifestations of a common origin. That is, electricity generated by a rotating magnet was derived from the same cause as a voltaic pile or frictional (i.e., static) electricity (Harre, 1981). Second, Faraday went on to propose a field theory of electricity, which stated electricity was induced in a current carrying wire when moved relative to a magnetic field. The movement of the wire thereby “cut” the lines of forces surrounding the magnet, causing the induction of an electric current (Nersessian, 1992; Tweney, 1989). James Clerk Maxwell, a contemporary of Faraday, went on to formalize these ideas with his famous field equations.

A unified, mechanistic explanation for the various forms of electricity was provided by the discovery of the electron in 1899 (Harre, 1981). Sir Joseph John Thomson, who was credited with the discovery, attempted to measure both the mass and charge on the electron (Hecht, 1994). Although not everyone at the time subscribed to the atomic theory of matter, Thomson made the argument that there were subatomic particles, which could be observed experimentally. Given Thomson's discovery, electricity could now be understood as net movement of free electrons through a closed circuit. Thus, the flow of electricity is a kind of Emergent System.

The history of electricity thus demonstrates a non-radical shift from the progression of the effluvium to the fluid theory, in that both theories involved substances moving from one place to another, as a kind of a Causal Event. The breakthrough did not occur until Thomson's ideas were established, that current flow is the net movement of subatomic particles. This revolutionary idea considered electrical current to be a kind of an Emergent System.

The caloric theory of heat. Fluid-type models of physical phenomena were used to explain not only electricity, but heat transfer as well. The caloric theory of heat, proposed largely by Lavoisier, was the dominant theory of the day (Spielberg & Anderson, 1995). The theory held two basic tenets. First, heat transfer was the result of two bodies, of unequal temperature, coming into contact with one another. The caloric was understood to flow from the relatively "hot" object to the "cold" object. The second tenet stated the caloric was conserved (i.e., it can neither be created, nor destroyed). Although the theory was useful in understanding new developments, such as the steam engine, it became apparent that the theory was not totally correct.

One anomaly, which could not be explained by the two tenets, was the explanation of heat generated during friction (Spielberg & Anderson, 1995). Because the caloric was always conserved, then it was hard to explain why two objects, which start in thermo-equilibrium, could produce heat when rubbed together. This observation violated the second tenet. To investigate this anomaly more systematically, Joule designed a method for precisely measuring the heat generated from friction. His apparatus included a container, with several compartments. The compartments were filled with fluid, and a paddle agitated the fluid contained within. Joule measured the temperature of the liquid with a thermometer. He found that the temperature of the liquid increased slightly when the paddle was turned. Because the system was in thermo-equilibrium at the beginning of the experiment, the only source of the heat generated was the agitated fluid. Therefore, Joule concluded that the caloric theory of heat must be incorrect, and the motion of the fluid led him to postulate a new theory of heat. The kinetic theory of heat dispensed with the concept of a separate fluid; instead, it treated heat as a property of matter. Joule reframed heat as the motion of the molecules that made up the fluid in his apparatus.

The switch from the caloric theory to the kinetic theory of heat is hailed as "among the greatest intellectual achievements of the nineteenth century" (Hecht, 1994: 565). The kinetic theory can be considered a major ontological shift in two ways. First, the caloric was considered a material Substance, which can "flow" from one object to the next. However, the kinetic theory conceived of temperature as the average speed with which molecules move, so that heat energy is not a kind of Substance, rather it refers to a PROCESS, such as speed of molecules. Moreover, heat energy is transferred from one object to the next by changing the velocity of the molecules that comprise the object. Thus, heat transfer is a kind of Emergent Process.

Alternative Mechanisms Leading to Scientific Discoveries and Theory Change

What this chapter proposes is a mechanism for inducing a radically new hypothesis or theory, to explain a set of phenomena or data. The idea is that it may be the case that novel theories are often noticed, focused upon, explored, and sometimes succeeded in explaining the phenomena better, when the novel theories are ontologically distinct from the former theories. We are not suggesting that a novel theory, in order to be successful, must be ontologically distinct. Rather, we are suggesting that when a given theory, and its related family of theories, are inadequate for explaining a phenomenon or a disease for many years, perhaps an ontological shift in hypothesizing can allow a major breakthrough to occur. We illustrated this possibility with a few examples.

What other mechanisms have been proposed for scientific discoveries? The crux of understanding scientific discovery is to understand how a new hypothesis is induced, especially a radically new one. The new hypothesis presumably can explain a pattern of observations that other existing hypotheses have failed to explain successfully. Psychological research attempting to understand how new discoveries are made often focus on the generation of experiments to test the hypothesis, as well as evaluation of the resulting evidence. Far less work has been done on the processes of inducing a new hypothesis. However, Klahr and Dunbar's (1988) research, does addressed the issue of inducing new hypotheses. Their task involved discovering the function of a mystery key, labeled "RPT," on a robotic toy. The participants tested their hypotheses by writing small programs that included the mystery key. Most participants used their semantic knowledge of "repeat" to induce their initial hypothesis. As they received feedback from the device, it became evident that the repeat function had violated some participants' initial hypotheses. However, a new hypothesis, in this task, is not ontologically distinct from earlier hypothesis. So the findings learned from this task cannot generalize to revolutionary discoveries that did involve ontological shifts.

Similarly, computational research on the process of inducing a new hypothesis took a few directions that also differed from those involving ontological shifts. Langley, Simon, Bradshaw, and Zytkow's (1987) work, for example, addressed the problem of discovering laws given a set of empirical observations. For instance, BACON.5 can look for a linear relation among a set of variables by using a standard linear regression technique. If the variables and their values fit a monotonically increasing relation, then the ratio between the two terms are found, given the signs of the values of the two terms are the same. If a monotonically decreasing trend occurs, then a product is defined, and so on. BACON.5's discovery is restricted to finding an appropriate fit of a mathematical expression to the empirical values that correspond to a pre-specified set of variables. This process corresponds to the formulation of a new theory or principle (in the form of a mathematical expression) on the basis of experimental results. This is essentially a curve-fitting process, which seems to be quite different from the process of inducing a radically new hypothesis.

Another computational approach was Thagard's (1992) model, ECHO. ECHO basically describes the mechanism by which a specific hypothesis is accepted or rejected, on the basis of the number of pieces of evidence that supports or contradict the hypothesis. In this sense, ECHO is not a model of hypothesis generation per se, but rather, it is a model that evaluates which of several hypotheses better fit the evidence.

Thagard (1992) mentioned two other mechanisms that characterize scientific discoveries. One is replacement of an old theory by a new theory. Replacement does accurately portray the *result* of ontological shift, but it does not refer to the ontological shift per se. Thagard (1992) and others (e.g., Hampton, 1997) also proposed conceptual combination, in arriving at new ideas and concepts. However, it is difficult to imagine how two concepts from different ontologies can be combined, nor how combining two concepts from the same ontology can derive a radically (i.e., ontologically) different concept. A third mechanism that is often proposed is analogical reasoning (Gentner et al., 1997; Holyoak & Thagard, 1995). However, by our assumption, analogy can only produce similar accounts since by definition, analogies have similar structures. For example, it was mentioned that diseases were caused by bacteria was discovered by analogizing it to the process of fermentation caused by yeast. Thus, yeast and bacteria are analogous. The ontological shift (from phlegm to bacteria) occurred when the scientist noticed the similarity between yeast and bacteria. That was the insight. The analogical mapping between yeast causing fermentation and bacteria causing diseases was straightforward. Thus, we would argue that the analogical mechanism per se, mapping yeast to fermentation and bacteria to disease, did not create the discovery. Perhaps some prior ontological shift (from phlegm to bacteria) was the mechanism that caused the scientist to notice the similarity between yeast and bacteria in the first place.

Hence, in general, scientific discoveries, in the form of generating or inducing a radically novel hypothesis that explains the observed pattern of findings, have not been explored extensively. What has been explored extensively is scientific thinking more broadly, such as skills of weighing all the available evidence, ways of experimenting that is systematic, such as holding some variables constant while varying others (Klahr & Dunbar, 1988; D. Kuhn, Amsel, & O'Loughlin, 1988). However, little of these research findings bear directly on the mechanism of inducing a radically novel hypothesis. This chapter proposes one such mechanism: the possible role of ontological shifts in representation as required for the generation of a truly new hypothesis, one that can be considered a major scientific discovery.

Acknowledgements

The authors are grateful for funding from the Spencer Foundation Grant No. 200100305. We would also like to thank Ryan D. Tweney for his insightful comments on an earlier version of our chapter.

References

Alvarez, L. W., Alvarez, W., Asaro, F., & Michel, H. V. (1980). Extraterrestrial cause for the cretaceous-tertiary extinction. *Science, 208*(4448), 1095-1108.

Alvarez, W. (1997). *T. rex and the crater of doom*. Princeton, N.J.: Princeton University Press.

American Heart Association. (2001). *2002 Heart and stroke statistical update*. Dallas, Tex.: American Heart Association.

Andersen, H. (2002). The development of scientific taxonomies. In L. Magnani & N. J. Nersessian (Eds.), *Model-based reasoning: Science, technology, values* (pp. 95-112). New York: Kluwer Academic.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior, 18*(2), 187-198.

Casti, J. L. (1994). *Complexification: Explaining a paradoxical world through the science of surprise*. New York: HarperPerennial.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. N. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science* (Vol. XV, pp. 129-186). Minneapolis: University of Minnesota Press.

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 209-234). Washington, DC: American Psychological Association.

Chi, M. T. H. (in press). Emergent systems versus causal events: Schemas for overcoming misunderstandings in science. *Journal of the Learning Sciences*.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M. T. H., Hutchinson, J. E., & Robin, A. F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly, 35*(1), 27-62.

Chi, M. T. H., & Roscoe, R. D. (2002). The process and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reframing the process of conceptual change: Integrating theory and practice*. The Netherlands: Kluwer Academic Publishers.

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27-43.

Clancey, W. J. (1996). Conceptual coordination: Abstraction without description. *International Journal of Educational Research, 27*(1), 5-19.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior. ; 8(2): 240 247, 8*(2), 240-247.

Ferrari, M., & Chi, M. T. H. (1998). Naive evolutionary explanations and radical conceptual change. *International Journal of Science Education, 20*(10), 1231-1256.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*(1), 65-95.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B., Wolff, P., et al. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *The Journal of the Learning Sciences, 6*(1), 3-40.

Gould, S. J. (1985). *The flamingo's smile: Reflections in natural history*. New York, NY: W.W. Norton & Company.

Graham, D. Y. (1993). Treatment of peptic ulcers caused by *Helicobacter pylori*. *New England Journal of Medicine, 328*(5), 349-350.

Hampton, J. A. (1997). Emergent attributes in combined concepts. In T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 83-110). Washington, DC: American Psychological Association Press.

Harre, R. (1981). *Great experiments in science*. Oxford: Phaidon.

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, D. B. Pisoni & G. R. Potts (Eds.), *Cognitive Theory* (Vol. 2, pp. 21-41). New York: Lawrence Erlbaum Associates.

Hecht, E. (1994). *Physics*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Holland, J. H. (1998). *Emergence: From chaos to order*. Reading, Mass.: Addison-Wesley.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps*. Cambridge, MA: The MIT Press.

Home, R. W. (1981). *The effluvial theory of electricity*. New York: Arno Press.

Hutchins, E., & Levine, J. (1981). *Point of view in problem solving* (Technical Report 105). La Jolla, Ca: Center for Human Information Processing.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, Mass: Harvard University Press.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: The MIT Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1-48.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, Fla: Academic Press.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3 ed.). Chicago: The University of Chicago Press.

Lakatos, I. (1970). Fasification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-195). Cambridge: Cambridge University Press.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: The MIT Press.

Lauden, L. (1977). *Progress and its problems*. Berkeley: University of California Press.

Leas, R. R., & Chi, M. T. H. (1993). Analyzing diagnostic expertise of competitive swimming coaches. In J. L. Starkes (Ed.), *Advances in Psychology: Cognitive issues in motor expertise* (Vol. 102, pp. 75-94).

Lowe, E. J. (1995). Ontology. In T. Honderich (Ed.), *The Oxford Companion to Philosophy* (pp. 634-635). Oxford: Oxford University Press.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35*, 113-138.

Meier, C. R., Derby, L. E., Jick, S. S., Vasilakis, C., & Jick, H. (1999). Antibiotics and risk of subsequent first-time acute myocardial infarction. *Journal of the American Medical Association, 281*(5), 427-431.

Meyer, H. W. (1971). *A history of electricity and magnetism*. Cambridge, Mass.,: MIT Press.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10*, 151-177.

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science* (Vol. XV, pp. 3-44). Minneapolis: University of Minnesota Press.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the Psychology of Thinking* (Vol. 1, pp. 1-44). London: Harvester-Wheatsheaf.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideals*. New York: Basic Books.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & K. Paul.

Resnick, M., & Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *Journal of the Learning Sciences, 7*(2), 153-172.

Rosch, E. R., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(382-439).

Shatz, M., & Gelman, S. A. (1973). The development of communication skills: Modifications in the speech of young children as a function of listeners. *Monographs of the Society for Research in Child Development, 38*(5, Serial No. 152).

Slotta, J. D., Chi, M. T. H., & Joram, E. (1995). Assessing students' misclassifications of physics concepts: An ontological basis for conceptual change. *Cognition and Instruction, 13*(3), 373-400.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition, 21*, 177-237.

Sommers, F. (1963). Types and ontology. *Philosophical Review, 72*(3), 327-363.

Sommers, F. (1971). Structural ontology. *Philosophia, 1*, 21-42.

Spielberg, N., & Anderson, B. D. (1995). *Seven ideas that shook the universe* (2 ed.). New York: John Wiley & Sons.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, N.J.: Princeton University Press.

Thagard, P. (1996). The concept of disease: Structure and change. *Communication and Cognition, 29*, 445-478.

Thagard, P. (1998a). Ulcers and bacteria I: Discovery and acceptance. *Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 29*, 107-136.

Thagard, P. (1998b). Ulcers and bacteria II: Instruments, experiments, and social interactions. *Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 29*, 317-342.

Tuomainen, T.-P. P., Kari, Nyyssonen, K., & Salonen, J. T. (1998). Association between body iron stores and the risk of acute myocardial infarction in men. *Circulation, 97*(15), 1461-1466.

Tweney, R. D. (1989). A framework for the cognitive psychology of science. In B. Gholson & W. R. Shadish Jr. (Eds.), *Psychology of science: Contributions to metascience* (pp. 342-366). Cambridge, England UK: Cambridge University Press.

Wall, R. E. (1972). *Introduction to mathematical linguistics*. Englewood Cliffs, N.J.,: Prentice-Hall.

Weisberg, R., & Suls, J. M. (1973). An information-processing model of Duncker's Candle Problem. *Cognitive Psychology, 4*(2), 255-276.

White, A. R. (1975). Conceptual analysis. In C. J. Bontempor & S. J. Odell (Eds.), *The owl of Minerva* (pp. 103-117). New York: McGraw Hill.

Zhang, R. M. D. P., Brennan, M.-L. P., Fu, X. M. S., Aviles, R. J. M. D., Pearce, G. L. M. S., Penn, M. S. M. D. P., et al. (2001). Association between myeloperoxidase levels and risk of coronary artery disease. *Journal of the American Medical Association, 17*, 2136-2142.

Figure 1. A plausible structure of ontological categories.