Chi, M.T.H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61-82). Hillsdale, NJ: Erlbaum.

3

Three Types of Conceptual Change: Belief Revision, Mental Model Transformation, and Categorical Shift

Michelene T. H. Chi
University of Pittsburgh

CONCEPTUAL CHANGE KIND OF LEARNING

Learning of complex material, such as concepts encountered in science classrooms, can occur under at least three different conditions of prior knowledge. First, a student may have no prior knowledge of the to-be-learned concepts, although they may have some related knowledge. In this case, prior knowledge is *missing*, and learning consists of *adding* new knowledge. Second, a student may have some correct prior knowledge about the to-be-learned concepts, but that knowledge is *incomplete*. In this incomplete knowledge case, learning can be conceived of as *gap filling*. In both *missing* and *incomplete* knowledge conditions, knowledge acquisition is of the *enriching* kind (Carey, 1991). In a third condition, a student may have acquired ideas, either in school or from everyday experience, that are "in conflict with" the to-be-learned concepts (Vosniadou, 2004). Knowledge acquisition under this third case is of the *conceptual change* kind. It is customary to assume in this case that the prior "in conflict with" knowledge is incorrect or misconceived, and the to-be-learned information is correct, by some normative standard. Thus, learning in this third condition is not *adding* new knowledge or *gap filling* incomplete knowledge; rather, learning is *changing* prior misconceived knowledge to correct knowledge. This chapter focuses on this conceptual change kind of learning.

Although this definition of conceptual change appears straightforward, conceptual change kind of learning entails several complex, non-transparent, and interleaved issues. Some of the key non-transparent ideas are: (a) In what ways is knowledge misconceived? (b) Why is such misconceived knowledge often resistant to change? (c) What constitutes a *change* in prior knowledge? and (d) Ho should instruction be designed to promote conceptual change? The existence of decades of research on conceptual change speaks to the complexity of these issues. This chapter hopes to add clarity to some of these issues by laying out three different grain sizes in which knowledge can be "in conflict with" the to-be-learned materials, postulating for each grain size the processes by which such "in conflict with knowledge" can be changed, and speculating on the kind of instruction that might achieve such change. We start by providing some definitions and assumptions about concepts and categories in conceptual change.

"CONCEPTS" AND "CATEGORIES" IN CONCEPTUAL CHANGE

In this section, we elaborate on (1) the scope of the term "concepts" in conceptual change research, (2) the assumptions about the role of categorization in learning and conceptual change, and (3) the relationships among different levels and kinds of "categories."

Scope of Concepts

Several decades of psychological literature (see Medin & Rips, 2005, for a recent review; and see Jackendoff along with the Forum published in *Mind & Language*, 1989, for a broader view) have dealt with determining how concepts can be identified and defined. That classic literature has typically been devoted to defining isolated and static concepts and categories such as *robins* and *birds*. From that literature, we adopt the common assumptions that a concept has several perceptual features and conceptual attributes, and a concept can be viewed as belonging to some category. For example, a *robin* has a red breast (a perceptual feature), lives in a temperate climate (more of a conceptual attribute), and belongs to the category of *birds.* (Throughout this chapter, we will use the term features to refer to perceptual properties, "attributes" to refer to conceptual properties, and italicize category terms and scientific concepts.)

Although prior conflicting ideas are often referred to as mis*conceptions,* and learning that involves altering such incorrect ideas is referred to as *conceptual* change, the grain size of that prior knowledge does not have to be at the level of a concept, in the traditional sense of static concepts typically studied by psychologists, such as *chairs* and *furniture*. Even though psychologists have begun to expand the notion of a category beyond concrete static types to include explanation-based categories such as *food items for a diet* (e.g., popcorn, diet soda, lean turkey, Barsalou, 1983) or principle-based categories (such as physics problems that share the same principle, Chi, Feltovich, & Glaser, 1981), the kind of misconceived knowledge in subject matter domains taught in schools (especially science domains) are at a much larger grain size, more complex and inter-related. For example, students are expected to learn about systems (such as the *circulatory system*) consisting of many inter-related components (such as *blood, organs*, etc). Students are also expected to learn not only about static concepts, but also about dynamic concepts, such as the processes of *heat transfer* and *natural selection*. In short, the term "concepts" in conceptual change research often refers to a broader scope than isolated and static concepts.

ROLE OF HIERARCHICAL CATEGORIZATION IN LEARNING

Categorizing is the process of identifying or assigning a concept to a category to which it belongs. One of the most important assumptions about categorizing that we also adopt is its role in learning (Bransford, Brown, & Cocking, 1999). Categorization is an important learning mechanism because a concept, once categorized, can "inherit" features and attributes from its category membership. For example, we can infer that *robins lay eggs* even if we were never told that fact, as long as we know that *robins are birds* and *birds lay eggs*. By knowing that *robins are a kind of bird* allows us to infer that *robins* inherit the properties of *birds*. Thus, categorizing, or assigning a concept to a correct category, is powerful because a learner can use knowledge of the category to make many inferences and attributions about a novel concept/phenomenon (Medin & Rips, 2005). Even young children can do this. For example, 4- to 7-year-old dinosaur aficionados can generate many appropriate inferences about an unfamiliar dinosaur once they have categorized it on the basis of surface features (Chi & Koeske, 1983; Gobbo & Chi, 1986).

Besides the common assumption that categorization allows new concepts to inherit categorical properties, we propose two additional assumptions about the role of categorization in learning. The first new assumption is that when learners have *no* obvious basic category to assign a new concept or phenomenon — they will assign it to the next higher level of category that is appropriate. For example, suppose an observer in a museum sees a strange large creature (a *gavial*) with four short legs, scaly skin and a flat bill-like snout. Not knowing that it's a kind of *reptile*, like a crocodile, the observer would categorize it at the next level up, as a kind of *animal*, since it appears to have the properties of *animals,* can move on its own, eat, and so forth. (The second new assumption will be described in the next section.)

As illustrated above, the type of relationships cognitive psychologists have explored about inheritance of properties are hierarchical ones. Hierarchical relationships among categories are primarily inclusive in nature. For example, *living beings* include *animals*, and *animals* include *reptiles* and *birds*, and *birds* include *robins.* (See Figure 3.1, left-most hierarchical tree.) *Living beings,* in turn, can be subsumed under an even higher category, such as *objects*; and *objects* can be subsumed under yet an even higher category such as *Entities*. The classic psychological research that dealt mostly with hierarchical relationships among categories asked questions such as: What level within this hierarchy is the most "basic" and useful? How does correct categorization support reasoning and inferencing? Can priming the correct super-ordinate category enhance recognition?

Little research has focused on incorrect hierarchical categorization, perhaps because it is not wrong but merely too specific or overly general. As in the preceding *gavial* example, the overly general hierarchical categorization of *gavial* as an *animal* is not that damaging, since the observer can still benefit from correct inferences and attributions inherited from the *animal* category. For example, the observer can understand new instruction about *gavials,* such as that *they breathe air through their snouts*. The observer can assimilate this new piece of information because it is compatible with what s/he knows about *animals* in general. Therefore, categorizing a concept at a higher categorical level is not damaging to learning.

Lateral and Ontological Categories

Research in cognitive psychology has paid much less attention to the role of "lateral" (rather than hierarchical) categories. For example, *artifacts* can be considered a lateral category more-or-less "parallel" to *living beings* (see Figure 3.1) *Artifacts* does not include the subcategories of *living beings*, such as *animals, reptiles, birds,* or *robins.* Instead, *artifacts* includes a different set of subcategories, such as *furniture* and *toys,* and *furniture* includes subcategories such as *tables* and *chairs* (see Figure 3.1) In short, *artifacts* and *living beings* can be thought of as occupying different branches of the same hierarchical tree (Thagard, 1990), in this case the *Entities* tree. We will refer to categories on different branches as "lateral" (vs. hierarchical) categories and, when lateral categories occur at about the same level within a tree, we will refer to them as "parallel."

Although *artifacts* and *living beings* can both be subsumed under the higher-level category of *objects* and therefore share higher-level properties of *objects* such as "has shape" and "can be thrown," the properties of *artifacts* and *living beings* tend to be distinct and mutually exclusive. For example, *living beings* "can move" on their own volition, whereas *artifacts* cannot; *living beings* "can reproduce" whereas *artifacts* cannot. (Examples of properties of each category are shown in quotes in Figure 3.1.). Gelman (1988) and Schwartz (1977) might have referred to these categories as different in "kind..

Having mutually exclusive properties means that it does not make sense to talk about a concept of one category as having a property from a lateral category. Conversely, a concept can be

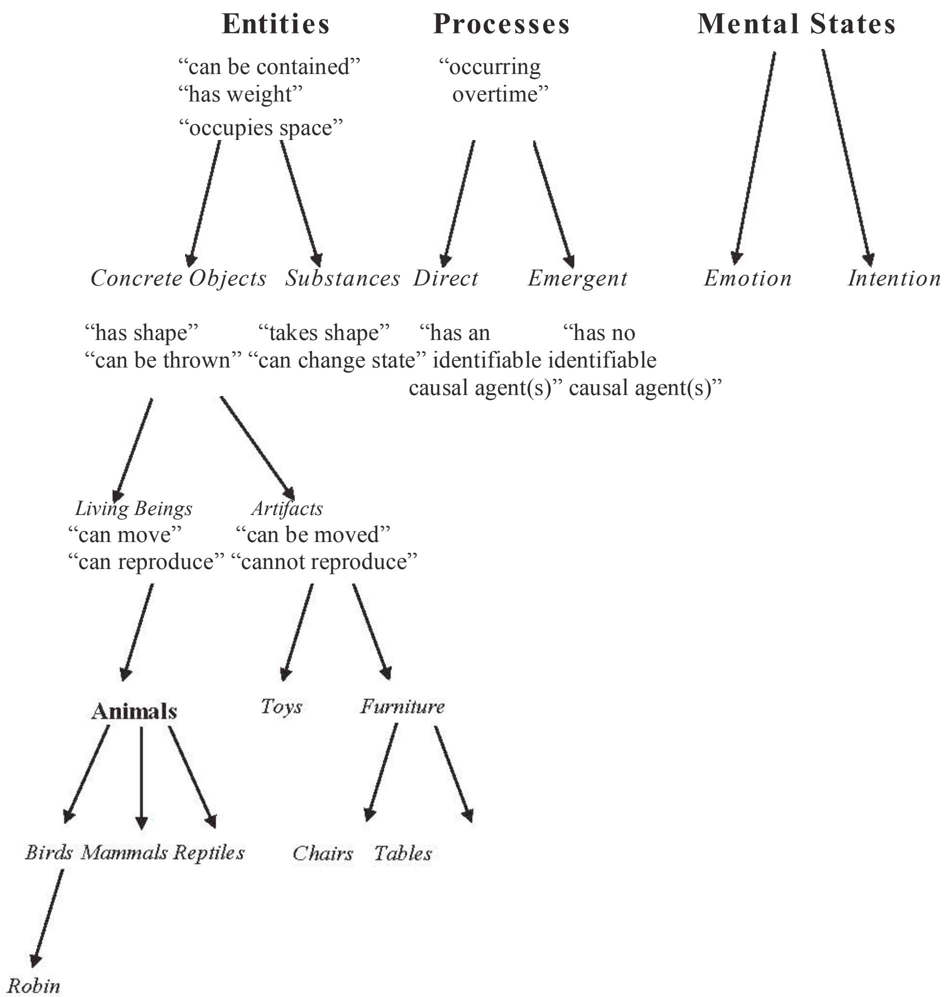

FIGURE 3.1 Distinct ontological trees: hierarchical and lateral categories within a tree and between trees.

described as having a property of its own category whether or not it is true. For example, *living beings* can reproduce whereas *artifacts* cannot. This means that Fido (a dog), being a *living being*, has the potential to reproduce even though Fido (a specific dog), having been neutered, cannot. On the other hand, a *toy dog* (an instance of an *artifact*) does not have this potential. Thus, it makes sense to say that Fido will have grey puppies even though Fido cannot have puppies, but it does not make sense to say that the mechanical toy dog will have puppies. Thus, a property of a category can be applied to members of that category or its subcategories, whether or not it is true, whereas it cannot be applied to a member of a lateral category. Thus, having mutually exclusive properties means that it does not make sense to talk about a concept as having a property of a lateral category, whereas it does make sense to talk about a concept as having a property of its own category even if it is false.

To take another example, a *object* such as a piece of clear glass, being an *Entity*, can have the property of "color," even though a specific piece of glass is colorless. That is, it is acceptable and sensible to say "the glass is green" even though it is not, whereas it makes no sense to say "the baseball game is green." This is because a baseball game, being a *direct* process, which is a category on an alternative tree, cannot take on the property of "color," so that it does not make sense to say that "the baseball game is green." Thus, one way to determine that two categories are laterally distinct (either within the same tree or between trees) is to use such a sensibility judgment task (Keil, 1979). Although in the past, we and others have called such lateral categories *ontological* (Chi, 1997; Keil, 1981), we reserve the term *ontologies* to refer to categories between different trees (as shown in Figure 3.1), since categories on different trees never share any properties, given that they do not share any super-ordinate level categories. For example, *Entities* have properties such as "can be contained" and "has volume," whereas *Processes* have properties such as "occurring over time." Thus, no process, whether it's an event such as a baseball game, a procedure such as baking a cake, or a state change such as melting, can have the property of "has volume," "has color," or "can be contained," whereas no entity, such as a cake or a ball, can have the property of "lasting two hours." Thus, each tree might be considered an "ontology," (and its name will be capitalized) (Chi, 1997, 2005), in that the trees refer to a system of taxonomic categories for certain existences in the world, as defined by philosophers (Sommers, 1971). Thus, in this chapter, we will refer to categories that occupy different trees as different "ontologically", and categories that occupy parallel branches within a tree as different laterally or in "kind."

The goal of our research is not to lay out the exact ordering and structure of hierarchical and lateral categories and trees, nor to decide which categories deserve the name ontology, or how many kinds or ontologies there are. The nature of categorical structure is an epistemological issue. Our goal instead is to focus on the role of lateral and ontological categories in conceptual change kind of learning. Thus, Figure 3.1 is offered merely as an example of a crude and intuitive rendition of categorical structures. It is by no means the absolute or the correct one.

Our second additional assumption about categorization and learning is that, when an observer or learner cannot classify a concept or phenomenon, instead of assigning it more generally to a higher-level category (as mentioned in the first additional assumption above), the observer may instead assign it to a lateral category. Using the gavial example again, the observer might categorize it as a *mammal* rather than an *animal*.

The central question to pose about lateral and ontological categories is the cost of category mistakes. We define a category mistake as the case when a concept has been assigned inappropriately to a lateral or alternative ontological category. In contrast to incorrect hierarchical categorization, category mistakes are damaging in that categorical inferences and attributions will be erroneous, creating a barrier to correct learning with deep understanding. The central thesis of our explanation is that such category mistakes account for the existence of robust misconceptions and their resistance to change. This explanation will be addressed in detail later in this chapter.

KNOWLEDGE MISCONCEIVED AT THREE GRAIN SIZES

Superficially, the notion of misconceived knowledge seems easy to define objectively, in that it is incorrect and to-be-learned material is correct. However, this contrast between correct and incorrect knowledge is too simplistic because it cannot address the issue of why such incorrect knowledge is often resistant to change. In order to understand the difference between incorrect knowledge and misconceived/conflicting knowledge, we need to consider the representation of knowledge at three different grain sizes: individual beliefs, mental models, and categories. Although our framework does not necessarily commit to any notions of hierarchy in these grain sizes, one could presume that they occur at different "levels," with individual beliefs at the lowest level and categories at the highest. What is critical, however, is our proposal that the grain size at which one considers misconceived knowledge determines the level at which instruction should target conceptual change. More specifically, how conflict is defined (between misconceived knowledge and to-be-learned material) determines how instruction should be designed.

In this section, we will focus on each of the three grain sizes of conflicting knowledge. In particular, we will examine how students' ideas conflict with to-be-learned information, the kind of conceptual change that occurs, and the type of instruction or confrontation that might trigger conceptual change. In the discussion below, our examples will be drawn primarily from science domains for three reasons. First, it is relatively easy to agree on what is considered correct or normative scientific information, and thus to contrast it with misconceived knowledge, which, by definition, implies prior knowledge that is incorrect as compared to some normative or scientifically-based information. Second, misconceptions historically were recognized largely in science domains. Third, we draw our examples from science domains for which we have some data, primarily taken from concepts such as the *human circulatory system* and *heat transfer*. For the headings of the three subsections below, the first segment serves as a label for how knowledge is misconceived, the second segment describes the kind of conceptual change that can occur, and the third segment refers to the kind of confrontation and/or instruction that may produce conceptual change.

FALSE BELIEFS: BELIEF REVISION FROM REFUTATION

Students' prior knowledge can be represented at the grain size of a single idea, corresponding more-or-less to information specified in a single sentence or statement. We will refer to single ideas as "beliefs." As described earlier, students' prior beliefs can be *missing* or *incomplete*, but learning under these two conditions of prior knowledge would not constitute conceptual change, since missing beliefs can simply be added and gaps in incomplete beliefs can be filled. For example, a student might not know that *a human heart has four chambers*, and telling the student this piece of information would be *adding* to her prior beliefs. Similarly, a student might know that the upper chambers are called *atria,* but not that the lower chambers are called *ventricles.* Telling the student the name of the lower chambers can be thought of as *filling* a gap in her knowledge about the names of the chambers. We had a priori considered processes of adding and gap-filling as an enrichment kind of learning.

For conceptual change to occur, prior knowledge must conflict with new information. When prior knowledge conflicts with new information at the grain size of a single idea, we can refer to that idea as a *false belief*, as in incorrectly thinking that "the *heart* is responsible for reoxygenating blood" or that "*all* blood vessels have valves." Such false beliefs conflict with correct text sentences that describe the *lungs* as being responsible for oxygenating blood or only *veins* but not arteries as having valves (Chi, de Leeuw, Chiu, & LaVancher, 1994; Chi & Roscoe, 2002). Thus, false beliefs and correct information are in conflict in the sense that they *contradict* each other. For example, it is the *lungs* and not the *heart* that oxygenate blood.

If false beliefs and correct information conflict in the contradictory sense, then one would expect that designing instruction that is targeted at refuting false beliefs might succeed at correcting them, resulting in *belief revision*. It appears that this is true (Broughton, Sinatra, & Reynolds, 2007; Guzetti, Snyder, Glass, & Gamas, 1993). That is, false beliefs for some subject matter domains can be corrected when learners are explicitly confronted with the correct information by contradiction and refutation. In Chi and Roscoe (2002), we reported results by de Leeuw (1993) that focused on eighth-grade students' understanding of the circulatory system. In a pretest assessment, 12 students exhibited a total of 31 "stable" and unique false beliefs of the type exemplified above. As suggested by many others (Engel, Clough, & Driver, 1986; Licht & Thijs, 1990), by "stable," we mean false beliefs that were manifested consistently in students' explanations and answers on more than one occasion, so that they were not simply generated on-the-fly in the context of answering questions. We found that 71% of the 31 prior false beliefs were correctly revised if the text students subsequently read included sentences that directly refuted the false beliefs, such as mentioning that "only veins but not arteries have valves."

Moreover, the false beliefs exemplified above could be revised even when the text sentences did not refute them explicitly, as in directly denying the false beliefs. For example, the text did not explicitly say that "the heart does not oxygenate blood, only lungs do." Instead, the text merely provided the correct information that "the lungs oxygenate blood." Thus, even with such indirect implicit refutation, false beliefs about the circulatory system could be revised, even though they were "in conflict with" the correct information. Thus, we might conclude that conceptual change can sometimes be readily achieved, and it might be described as belief revision through explicit or implicit refutation of prior false beliefs. But such belief revision can be achieved only when misconceived knowledge conflicts in the *contradictory* sense.

There are many false beliefs in other domains that are not so readily revised by refutation at the grain size of a single idea. Consider, for example, false beliefs such as "a thrown object acquires or contains some internal force" or "coldness from the ice flows into the water, making the water colder." Although students can readily learn by *adding* new beliefs about "internal force," such as the equation for its relation to mass and acceleration, the definition of acceleration, and so on, these newly added beliefs cannot correct a student's false belief that "a thrown object acquires or contains some internal force." Moreover, such false beliefs cannot be easily denied or corrected by contradiction. For example, saying that "a thrown object does not acquire or contain internal forces" will not succeed in achieving correct understanding. This is because, as we will propose later, misconceptions about force and temperature/heat do not conflict with normative correct ideas in a *contradictory* sense. Before addressing how these misconceptions about force and heat are misconceived, we should consider misconceived knowledge at the next grain size.

Flawed Mental Models: Mental Model Transformation from Accumulation of Belief Revisions

An organized collection of individual beliefs can be viewed as forming a mental model. A mental model is an internal representation of a concept (such as the *earth*), or an inter-related system of concepts (such as the *circulatory system*) that corresponds in some way to the external structure that it represents (Gentner & Stevens, 1983). For example, mental models can be "run" mentally, much like an animated simulation, to depict changes and generate predictions and outcomes, such as the direction of blood flow. As with beliefs, a mental model can be "in conflict with" the correct scientific model to varying degrees, such as a *missing* or non-existing mental model, or an *incomplete* mental model. For example, some students' prior conceptions of the human circulatory system may be so sparse and disconnected that it is difficult to capture what, if any, structure their mental models have (Chi et al., 1994), so that we could not say whether or not their mental models are "in conflict with" the correct scientific model. In these cases of spare and incomplete mental models, learning would begin by *adding* and *filling* in missing components. Adding and gap-filling a mental model would not constitute conceptual change, by our definition.

In what other ways can mental models be "misconceived," so that learning is the conceptual change kind and not merely the enriching kind? Many of us have proposed that a learner's mental model conflicts with the correct scientific model when it is *fl awed.* By flawed, we mean that it is coherent but incorrect (Chi, 2000; Chi et al., 1994; Vosniadou & Brewer, 1994), in the sense that the student can use the mental model to offer similar and consistently incorrect explanations and predictions in response to a variety of questions. The pattern and consistency of the generated explanations allow us to capture the structure of the flawed mental model (Chi et al., 1994; Vosniadou & Brewer, 1992, 1994). We can then validate the accuracy of the flawed mental model by predicting and testing how the student will respond to additional questions. For example, about half of the participants in our studies had an initial 'single-loop' model of the human circulatory system. According to this flawed model, blood goes to the heart to be oxygenated, then it is pumped to the rest of the body, then back to the heart. (In contrast, the correct "double-loop" model has two paths. One path leads from the heart to the lungs, where blood is oxygenated before returning to the heart. The second path leads from the heart to the rest of the body and back to the heart.) In order to confirm that our assessment of the flawed "single-loop" model is accurate, we can design additional questions to see if students will respond as expected, on the basis of the "single-loop" model.

In what way does a flawed "single-loop" model conflict with the correct "double-loop" model? The flawed "single-loop" model conflicts with the correct "double-loop" model in that it results in *different predictions* about where blood goes after it leaves the heart, *different explanations* with respect to where blood is oxygenated, and *different elements* in terms of whether or not lungs play an important role in oxygenation. Thus, we could say that two models are "in conflict with" each other if they make different predictions, generate different explanations, include different elements, and so forth. (Notice that these criteria of conflict—different predictions, different explanations, and different elements—are similar to the ones mentioned by Carey (1985) as compatible with the notion of "incommensurate" from the philosophy of science. In our framework here, we propose that these two conflicting models are not incommensurate and we would reserve the term "incommensurate" for knowledge that is "in conflict" either laterally or ontologically, to be discussed in the next grain size.)

Likewise, Vosniadou and Brewer (1992) have shown that young children have fl awed mental models of the earth, including a flattened disk and a hollow sphere. Students with a flattened-disk model consistently say that the shape of the earth is round, that one should look down to see the earth, and that there is an edge from which people can potentially fall off. In short, fl awed mental models are coherent in that students retrieve and use them consistently to answer questions and make predictions, allowing researchers to capture the structure of their mental models by analyzing the systematicity in the pattern of their responses (see also McCloskey, 1983; Samarapungavan & Wiers, 1997; Vosniadou & Brewer, 1992; Wiser, 1987). Thus, a flawed mental model "is in conflict" with the correct model in the sense that the two models generate different predictions and explanations, and may contain different elements.

In the previous section, we concluded that refuting false beliefs with correct statements that contradicted those beliefs can lead to belief revision. In this section, we refer to successful modification of a flawed mental model as mental model *transformation.* But how should we design instruction to induce mental model transformation, given that we have defined conflicting models in terms of different predictions, explanations, and elements? Since mental models and correct models conflict at the mental model level (flat earth vs. spherical earth; single-loop vs. double-loop), a holistic confrontation may induce successful model transformation. One way to design a holistic confrontation might be to have students examine a visual depiction (e.g., a diagram) of the flawed mental model, then contrast it with a diagram of the correct model, in terms of the predictions, explanations and elements of each model. We are not aware of any instruction offering this kind of holistic confrontation, and we are conducting a study to address its feasibility.

Although we have described conflicting mental models at the mental-model level (such as a flat earth vs. a spherical earth and a single-loop vs. a double-loop), instruction to confront a flawed mental model typically occurs at the belief level. Typically, when a student reads a text, instruction consists of a description of the correct model, one sentence at a time. When a learner's flawed mental model is confronted with a description of the correct model presented one sentence at a time, each sentence can either refute (explicitly or implicitly) an existing belief or not, as discussed in the preceding section on *Belief Revision*.

From the perspective of a mental model, there are two possible outcomes when instruction is presented sentence-by-sentence. In the first case, information presented in a given sentence or sentences may not refute (explicitly or implicitly) any of the learner's prior beliefs. Instead, the information might be new or more elaborated than what the learner knows. In such a case, the learner can assimilate by embedding or *adding* the new information from the sentences into her existing flawed model, so that her mental model is enriched, but continues to be fl awed. For example, in the case of a "single-loop" flawed model, learners assume that *blood from the heart goes to the rest of the body to deliver oxygen*. Such models lack the idea that blood also goes to the *lungs,* not to deliver oxygen but to receive oxygen. Upon reading a sentence such as "The right side [of the heart] pumps blood to the lungs and the left side pumps blood to other parts of the body," students with a "single-loop" model may not find it to contradict any beliefs in their flawed single-loop model, since they interpret the sentence to mean that the right side pumps blood to the lungs *to deliver oxygen* (rather than *to receive oxygen*), just as it does to the rest of the body. Therefore, even though at the mental model level, the sentence conflicts with the learner's flawed model, at the belief level, the sentence does not directly contradict the learner's prior beliefs. Thus the learner does not perceive a conflict, and the new information is assimilated into the flawed model (Chi, 2000). In short, assimilation of new information occurs when a learner does not perceive a conflict at the belief level, even though from the researcher's perspective, the new information is in conflict with the learner's flawed mental model.

The second possible outcome of sentence-by-sentence instruction is that new information presented does refute a learner's false beliefs and the learner recognizes the contradiction. Under such circumstances, as described in the preceding section, false beliefs that are explicitly or implicitly refuted do predominantly get revised. The relevant question with respect to mental models is: Does the accumulation of numerous belief revisions eventually results in the transformation of a student's flawed mental model to the correct model? The answer is *yes*, by-and-large.

According to our data, by reading and self-explaining a text passage about the human circulatory system, five of eight students (62.5%) with prior flawed "single-loop" model, transformed their flawed models to the correct model. Similarly, in Vosniadou and Brewer's (1992) data, 12 of 20 children (60%) developmentally acquired the correct spherical model of the earth by the fifth grade, suggesting that their flawed mental models had undergone transformation. In short, again, for domains such as the circulatory system and the earth, coherently flawed mental models can be successfully corrected and transformed into the correct model, in over 60% of the population, with either relatively brief instruction from text (in the case of the circulatory system) or from general development and learning in school (in the case of the earth). Thus, conceptual change can be achieved in that "in conflict" flawed mental models can be transformed into the correct model when false beliefs within a flawed model are refuted by instruction and recognized by students as contradictions, so that the students can self-repair their flawed mental models (Chi, 2000)..

Whether or not flawed mental models are successfully transformed into the correct model also depends on whether some critical false beliefs are revised. That is, a flawed mental model is composed of many correct and many false beliefs. The incorrectness of the flawed mental model does not depend on the number of incorrect beliefs, but on the number of critical false beliefs. For example, across the various studies for which we have assessed students' initial mental models of the circulatory system, we found 22 students (about 50%) to the fl awed "single-loop" model prior to instruction. The number of correct beliefs held by these 22 students varied widely, ranging from 5 to 35. Five students held 10–15 correct beliefs, and 4 students held 25–35 correct beliefs, all embedded within the flawed "single-loop" model (see Figure 2 in Chi & Roscoe, 2002). This variability suggests that knowing and learning many correct beliefs does not guarantee successful transformation of a flawed mental model to the correct model. Some critical or important beliefs serve to discriminate a flawed model from a correct model (in terms of generating correct explanations and predictions), and these critical beliefs need to be revised.

To recap, students' knowledge consists of an interrelated system of false beliefs and correct beliefs, forming a coherent but sometimes flawed mental model. A flawed mental model can be said to conflict with a scientific model if it is incorrect but coherent, in the sense that it consistently leads to different predictions and explanations and contains different elements. During instruction, when a specific sentence contradicts a false belief through explicit or implicit refutation, such refutation can cause students to revise their false beliefs when they are aware of the contradictions. Without such awareness, students may assimilate instruction, especially for implicit contradictions. The accumulation of multiple belief revisions can lead eventually to a transformation of a flawed mental model to the correct model for over 60% of the students, either through direct instruction (in the case of the *circulatory system)* or from exposure to everyday experiences (as perhaps in the case of the *earth)*. For students whose flawed mental models were not correctly transformed, this may be due to a lack of awareness of contradictions, especially for critical false beliefs. There may be other ways to design instruction, such as holistic confrontation, that may encourage revision and reduce the likelihood of assimilation or *adding* to a fl awed model, so that successful transformation can be achieved by all students.

Category Mistakes: Categorical Shift from Awareness and Building a New Category

The preceding sections described two grain sizes at which conceptual change is often achieved successfully. At the level of false beliefs, we found that refuting them can lead to belief revision. At the level of flawed mental models, multiple refutations can cause multiple belief revisions, the accumulation of which can result in transformation from a flawed mental model to the correct model for a majority of the students. However, we have also mentioned that there are numerous concepts (such as force and heat/temperature) across a variety of domains for which conceptual change cannot be achieved at the belief level. This section begins with an example of failure to transform a flawed mental model successfully, illustrating succinctly what robust misconceptions mean, in that they are *persistent and resistant* to conceptual change.

Robust Misconceptions: An Example

Law and Ogborne (1988) carried out a study in which students were asked to use Prolog to design and build a computational model of their own understanding of motion. The Prolog programming required students to express their ideas in propositional rule-based statements, which we can consider to be analogous to beliefs. Building and running such a model forced students to externalize and formalize their ideas, making them explicit, explorable and capable of offering explanations. Students assessed their models by running their programs, then made modifications based on program results or feedback from their instructor. Since programs could be run, allowing students to make predictions and observe outcomes, we can consider such a program to be analogous to an externalized mental model.

As with our circulatory system data, only some students had clear structural frameworks based on a core set of hypotheses about various aspects of motion that the researchers could identify.

We can consider these students as having flawed but coherent mental models. Other students had no clear conceptualization, and these students can be deemed to have missing or incomplete models. For students with coherent but flawed mental models, the question is, can they change their flawed mental model? One way to determine whether they change their mental model is to see whether they change their *implicit* core hypotheses or misconceptions, which include the following set of false beliefs:

a. Force is the deciding factor in determining all aspects of motion;
b. Force is *an entity* which can be possessed, transferred, and dissipated (rather than an interaction);
c. All motions need causes;
d. Agents cause and control motion by acting as *sources* that supply force;
e. Sources that supply force can be internal or external, and the supplied force is referred to as an internal or external force;
f. Weight is an intrinsic property of an object (even though gravity is conceptualized as an external factor that pulls harder on heavier objects).

This set of core hypotheses about force and motion are compatible with various other analyses of students' misconceptions about force and motion in the literature.

The advantage of the Prolog programming environment is that it allowed students to explore the consequences of their externalized beliefs or rules. For example, one student who held the core hypothesis d, that there is a *source* that supplies the force for every motion, wrote the following Prolog rules for determining the cause of motion:

1. _object motion-caused-by itself if _object force-supplied-by _object
2. _object motion-caused-by machine if _object force-supplied-by machine
3. _object1 motion-caused-by _object2 if _object1 force-supplied-by _object2
4. _object motion-caused-by gravity if not (_object under-the-influence-of other-external-force).

She then tested her program for the cause of a falling apple, expecting the computer to say that the motion was caused by gravity (her fourth rule). The reason was that in one of her earlier sessions, she included weight as an external supply of force, along with other forces such as friction and air current. The program's outcome can be thought of as providing explicit refutation of her fourth rule.

When she did not get the result she expected, she modified her fourth rule by excluding gravity as an external force. After this patching, the computer still did not give her the expected answer of gravity as a cause of the apple's fall, since anything placed in air would be affected by air-current, since air current is an external force. She then revised her fourth rule again to read: _object motion-caused by gravity if not (_object motion-caused-by _something). Her problems continued even after various patchings of her other rules.

This example illustrates the point that, despite numerous revisions in response to refutations at the rule or belief level, the revisions and the accumulation of multiple patches did not transform her flawed mental model into a correct model, in that the *implicit* underlying core hypotheses of her program (or misconceptions) were not changed. That is, she still assumed that all motions need causes (hypothesis c), that agents cause and control motion by acting as sources that supply force (hypothesis d), and so forth. As this example illustrates, the student was not resistant to change per se, since she readily revised her rules, but the multiple belief revisions she did undertake did not add up to a correct model transformation since the revisions did not change her underlying core hypotheses.

In short, there are many domains and concepts for which one's initial flawed mental model is not transformed to the correct model, despite repeated corrections or patchings at the individual belief level. This example shows that, even though the student willingly modified individual rules or beliefs as a result of external feedback (or explicit refutation from the program's outcomes), the revised beliefs, cumulatively, did not transform the mental model into the correct model, in that the implicit underlying core hypotheses were still incorrect. Thus, the flawed model was resistant to change. (There are occasions, of course, when students themselves resist making changes by dismissing the feedback or explaining it away. The point here is that, even with the best of intentions and willingness to change, this student could not transform her misconceived view.)

What should we conclude? This suggests that, for robust misconceptions, refutation at the belief or mental-model level is not the right grain size to achieve conceptual change. In such cases, we propose that instruction be designed to target conceptual change at a different grain size, at the categorical level.

CONFLICT BETWEEN LATERAL CATEGORIES

Findings similar to the Law and Ogborne's (1988) study have been documented for several decades, and we can refer to it as the *robust misconception* problem. That is, many misconceptions are not only "in conflict" with the correct scientific conceptions, but moreover, they are robust in that the misconceptions are difficult to revise, so conceptual change is not achieved. The robustness of misconceptions has been demonstrated in literally thousands of studies, about all kinds of science concepts and phenomena, beginning with a book by Novak (1977) and a review by Driver and Easley (1978), both published almost three decades ago. By 2004, there were over 6,000 publications describing students' ideas and instructional attempts to change them (Confrey, 1990; Driver, Squires, Rushworth, & Wood-Robinson, 1994; Duit, 2004; Ram, Nersessian, & Keil, 1997), indicating that conceptual understanding in the presence of misconceptions remains a challenging problem. The daunting task of building conceptual understanding in the presence of robust misconceptions is sometimes referred to as *radical* conceptual change (Carey, 1985). We propose the operational definition that certain misconceptions are robust because they have been mistakenly assigned to an inappropriate lateral category.

Our claim, then, is that some false beliefs and flawed mental models are robustly resistant to change because they have been laterally or ontologically miscategorized. That is, if a misconception belongs to one category and the correct conception belongs to another lateral or ontological category, then they conflict by definition of *kind* and/or *ontology.* This means that conceptual change requires a shift across lateral or ontological categories. In order to support this claim, we have to characterize the nature of misconceptions and the nature of correct information to see whether they in fact belong to two categories that differ either in *kind* or in *ontology*, thereby are "in conflict."

The Lateral Categories to which Misconceptions and Correct Conceptions are Assigned

In order to characterize the nature of misconceptions in terms of the category to which they have been mistakenly assigned, and also to characterize the nature of scientific conceptions in terms of the category to which they have been assigned, we analyzed students' causal explanations of a variety of science concepts, consolidated researchers' findings on misconceptions, and examined the history and philosophy of science literature. In particular, we examined the extent to which robust misconceptions reflect common implicit core hypotheses, as exemplified by the ideas about force and motion listed earlier. From such core hypotheses, we induced the properties of the mistaken category that they characterized. We then determined the lateral category into which correct scientific conceptions fall. We illustrate below two sets of conflicting lateral categories that we have identified: the conflict between two ontological trees — *Entities* (the misconceived view) and *Processes* (*the correct* view), and the conflict between two branches within the *Process* tree, *direct* processes (the misconceived view) and *emergent* processes (the correct view).

Entities Versus Processes

Entities are *objects* or *substances* that have various attributes and behave in various ways (see Figure 3.1, the *Entities* tree). For example, a ball is a physical object with attributes such as mass and volume, and behaviors such as bouncing and rolling. In reviewing students' explanations for four science concepts — force, heat, electricity, and light — we arrived at the commonality that students mistakenly categorize these concepts as *Entities*. On the basis of our analyses across these four concepts, we proposed that misconceptions for some concepts are *Entity*-based (Reiner, Slotta, Chi, & Resnick, 2000). For example, Law and Ogborne's (1988) hypothesis b, described above, indicates that many students view force as a *substance*-kind of *Entity* that can be possessed, transferred and dissipated. Students often explain that a moving object slows down because it has "used up all its force" (McCloskey, 1983), as if force were like a fuel that is consumed. Similarly, students think of heat as physical *objects* such as "hot molecules" or a *material substance* such as "hot stuff" or "hotness" (Wiser & Amin, 2001), as indicated by phrases such as "molecules of heat" or expressions such as "Close the door, you're letting all the heat out." The misconception is that heat can be "contained," as if it were *objects* like marbles or *substances* such as sand or water. In either case, heat is misconceived as a kind of *Entity*.

Not only do such *Entity*-based misconceptions occur for a variety of concepts across a variety of disciplines, but they are held across grade levels, from elementary to college students (Chi, Slotta, & de Leeuw, 1994), as well as across historical periods (Chi, 1992). They may even account for barriers that were only overcome by scientific discoveries (Chi & Hausmann, 2003). In short, robust misconceptions are extremely resistant to change, so that everyday experiences encountered during developmental maturation and formal schooling seem powerless to change them (in contrast to the success with which a majority of flawed mental models can be transformed from everyday experiences or formal schooling, as described earlier).

How are *Entity*-based misconceptions in conflict with scientific conceptions? Our initial conjecture was that scientists view many of these concepts not as *Entities*, but as *Processes*, an ontological tree distinct from *Entities,* verifiable by the predicate test (see Figure 3.1). For example, heat or the sensation of "hotness," is the speed at which molecules jostle: the faster the speed, the "hotter" the molecules feel. Thus, heat is not "hot molecules" or "hot stuff" (an *Entity*), but more accurately, the *speed* of molecules (a *Process*).

The naïve conception of the term "heat" is that it's like "hotness." "Hotness," as we illustrated above, refers to molecular motion, and motion is a *Process*. But the technical term *heat*, although a noun, actually refers not just to the motion of the molecules, but to the transfer of "hotness." That is, *heat* is defined as "the transfer of energy" or *energy in transit from one object or substance to another*, and is therefore a *Process*. The use of a noun to represent a transfer process, and defining heat as the transfer of *energy*, which is also a noun, is unfortunate, because such terminology encourages students to maintain their misconceptions since they can continue to conceive of the term "energy" as a kind of *substance*. In other words, robust misconceptions cannot be easily refuted by merely presenting scientific information in technical terms.

Direct Versus Emergent

Although we were able to explain a good deal of robust misconceptions as category errors involving the ontological trees *Entities* and *Processes* (Chi, 1997), our explanation for the robustness of many misconceptions was incomplete. Whether or not students conceive of heat as an *Entity*, most students nevertheless do recognize that heat transfer is a *Process* because they have experienced the apparent movement of "hotness" from one location to another, for example from a warm cup to cold hands. Thus, characterizing heat misconceptions as *Entity*-based does not adequately explain why students have difficulty understanding heat transfer. To explain the latter kind of misconceptions, we had to propose conflicts between two additional kinds of lateral categories within the *Process* tree, which we have called *direct* and *emergent* (Chi, 2005). Our claim is that *students* misconceive of some processes as *direct* kinds when in fact they are *emergent* kinds. Table 3.1a and Table 3.1b list two sets of mutually exclusive properties for *emergent* and *direct* processes.

Briefly, a *direct* process is one that usually has an identifiable agent that causes some outcome in a sequential and dependent sort of way. We will describe an everyday example, a less familiar example, and a scientific example, highlighting with each example properties of *emergent* and *direct* processes, as listed in Table 3.1a and Table 3.1b.

Direct Example 1. In the familiar process of a baseball game, the final outcome might be explained as being due to the excellent work of the pitcher, thus attributing the outcome to a single agent (*Direct* property #1) that has special status (*Direct* property #2). Moreover, the behavior of local events within the game corresponds to or aligns with the global outcome. For example, a team with many home runs in a game is more likely to win. Thus, home runs are positive local events and they align with the positive global outcome of winning the game (*Direct* property #3).

TABLE 3.1A Five Inter-level Properties Characterizing the Relationship Between the Agent (micro) Level and the Pattern (macro) Level.

Emergent Causal Explanations	*Direct Causal Explanations*
1. The entire collection or **all** the agents together "cause" the observable global pattern	1. A **single** agent or a **subgroup** of agents can "cause" the global observable pattern
2. All agents have **equal status** with respect to the pattern	2. One or more agents have **special status** with respect to the pattern
3. Local events and the global pattern can behave in **disjoint** non-matching ways	3. Local events and the global pattern behave in a **corresponding** matched way
4. Agents interact to intentionally achieve **local goals**; ignorant of the global pattern	4. Some agents interact to intentionally achieve the **global goal** and direct their interactions at producing the global pattern
5. Mechanism producing the global pattern: **proportional change** (collective summing across time)	5. Mechanism producing the global pattern: **incremental change** (additive summing across time)

TABLE 3.1B Five Micro-level Properties Characterizing the Relationship Among Agents' Interaction in an Emergent and a Direct Process.

Interactions among Agents in an Emergent Process	*Interactions among Agents in a Direct Process*
6. All agents behave in more-or-less the same **uniform** way	6. Agents behave in **distinct** ways
7. All agents interact **randomly** with other agents	7. Agents can interact with predetermined or **restricted** others
8. All agents interact **simultaneously**	8. Agents interact **sequentially**
9. All agents interact **independently** of one another	9. Agents' interactions **depend** on other agents' interactions
10. Interactions among agents are **continuous**	10. Agents' interactions **terminate** when the pattern-level behavior stops

Direct Example 2. A slightly less familiar example is seeing multiple airplanes flying in a V-formation. This V-pattern is intentional, created by the lead pilot telling the other pilots where to fly in order to achieve the global goal *(Direct* property #4).

Direct Example 3. A direct process from biology is cell division, which proceeds through a sequence of three stages. The first, interphase, is a period of cell growth. This is followed by mitosis, the division of the cell nucleus, and then cytokinesis, the division of the cytoplasm of a parent cell into two daughter cells. In each phase, the cells behave in distinct ways, either growing or dividing (*Direct* property #6). Such a process has a definite sequence, in which some events cannot occur until others are completed (*Direct* properties #8 and #9).

In contrast, *emergent* processes have neither an identifiable causal agent or agents nor an identifiable sequence of stages. Rather, the outcome results from the collective and simultaneous interactions of all agents. Let's consider three examples here as well.

Emergent Example 1. The process of a crowd forming a bottleneck, as when the school bell rings and students hurry to get through the narrow classroom door, is an everyday example of an *emergent* process. Although there is an external trigger (the school bell), the global outcome of forming the bottleneck cannot be attributed to any single agent or group of agents, and the process is not sequential. Instead, all the students (*Emergent* property #1) simultaneously (*Emergent* property #8) rush toward the door at about the same speed (*Emergent* property #6), shoving and bumping randomly into whichever student happens to be in the way (*Emergent* property #7).

Emergent Example 2. A slightly less familiar example is migrating geese flying in a V-formation. In contrast to the airplane example, the V-pattern is not caused by the leader goose telling other geese where to fly. Instead, all the geese are doing the same thing, flying slightly behind another goose because instinctually they seek the area of least resistance. Thus, they are pursuing the local goal of flying with minimal effort (*Emergent* property #4), ignorant of the pattern they form. When all the geese do the same thing at the same time, collectively, a V-pattern emerges (*Emergent* properties #1, #2, #6, and #8).

Emergent Example 3. An emergent process from biology is the diffusion of oxygen from the lungs to the blood vessels. This process is caused by all the oxygen and carbon dioxide molecules moving and colliding randomly with and independently of each other (*Emergent* properties #6, #7, #8, & #9). From such random collisions, a greater number of oxygen molecules are likely to move from the lungs to the blood than from the blood to the lungs, simply because there are a greater number of them in the lungs than in the blood. The reverse is true for carbon dioxide molecules. Since all molecules move and collide randomly, both kinds of molecules move in both directions, so that some oxygen molecules do move from the blood to the lungs, and some carbon dioxide molecules do move from the lungs to the blood Thus, the local movements of individual molecules may not match the direction of the movement of the majority of the molecules, thus causing the observed pattern (*Emergent* property #3). Nevertheless, despite local variations, the majority of oxygen molecules move from the lungs to the blood, and the majority of carbon dioxide molecules move in the opposite direction, without any specific intention to move in that global direction (*Emergent* property #4).

To return to our heat example, the technical term "heat" actually refers to the "transfer of hotness." The sensation of hotness moving from one area to another area is understood correctly by students as a *Process*. However, this process is not a *direct* process in that the sensation of hotness moving is not caused by hot molecules moving from one location to another. Rather, the transfer is caused by the collisions of faster jostling "hotter" molecules into slower-moving molecules. That is, when faster-moving molecules bombard slower-moving molecules, it causes the faster-moving molecules to slow down (thus decreasing their hotness) and the slower-moving molecules to move faster (thus increasing their hotness). This is how hotness is transferred. Thus, heat transfer is an *emergent* process. These two sets of examples illustrate general differences between *direct* and *emergent* processes. A more detailed description is provided in Chi (2005).

BEYOND REFUTATION

If misconceptions occur as the result of category mistakes, then instruction needs to focus at the categorical level. When students' misconceived ideas conflict with correct ideas at the lateral category level, then refutation at the belief level will not promote conceptual change, as was shown in the Law and Ogborne study. This is because refutation at the belief level can only cause local revisions at the belief level and not categorical shift. Consider the false belief that "coldness from the ice flows into the water, making the water colder." Essentially, this misconception assumes that ice contains some "cold substance" like tiny cold molecules (the reverse of hot objects, which are often misconceived as containing "hot molecules"), and that this "cold substance" can flow into the surrounding water, which then makes the water colder. To refute this misconception at the belief level, we might point out that ice does not contain a cold substance, that coldness does not flow, and that water does not get colder because it gains coldness. Refutation at the belief level only works when a false belief and the correct conception contradict each other. Moreover, belief level refutation tends to be partial in that many elements of the false beliefs are maintained. For example, it was straightforward to entertain the alternative belief that *lungs* are the source of oxygenation and not the *heart,* because only one element of the false belief had to be changed while many other elements (the concept of oxygenation, etc.) were maintained. But how can a false belief like "ice contains cold substances" be changed? Should a student expect ice to contain an alternative kind of substance if not a "cold" substance? The revision that a student must make has to do with the property "contain," not the feature "coldness" or any other kind of substance. To confront the property "contain" means to confront students at the ontological/categorical level, since "contain" is a property of *Entities*, and not *Processes.*

What about at the mental model level? Suppose we think of the ice and water as a system. The misconception is that coldness flows into the water as a *direct* process. Again, where do we begin in terms of either multiple belief-level refutations or holistic confrontation? We will likely achieve only local patchings, as in the Prolog example, because the core hypotheses underlying the mental model are not addressed. The model transformations that were obtained in connection with the circulatory system and the shape of the earth were transformations that occurred within the same ontology. We propose that, in order to achieve radical conceptual change, we need students to make a category shift by reassigning a concept to another lateral category. To do so, we need to confront students at the categorical level.

How Can We Achieve Shifts Across Lateral or Ontological Categories?

Shifting across lateral categories, in the sense of reassigning a concept from one lateral/ontological category to another, can in principle be straightforward and easily achieved for certain misconceptions. Let's take an everyday example to illustrate the ease of such a shift. Suppose a young child sees a *whale* in the ocean and believes it to be a kind of *fish,* since *whales* possess many perceptual features of a *fish*, such as look like sharks and swim in water. Based on that mistaken categorization, the child will likely assume that *whales*, like other *fish*, breathe through gills through osmosis (a conceptual attribute). To promote conceptual change, we might provide belief-level refutation, pointing out that *whales do not breathe through gills, but through a blowhole*. The child may accept this instruction and revise her false belief about gills, but still continue to implicitly assume that whales are *fish* (rather than *mammals*). Assuming that *fish* and *mammals* are lateral categories that differ in *kind*, maintaining that whales are *fish* will cause the child to have difficulty understanding subsequent instruction or answering questions such as: "Why do sharks suffocate when you take them out of the water, but whales do not?" This example illustrates that, when a category mistake is refuted at the belief level, the belief revision results in superficial or shallow understanding only, since the conception is still fundamentally wrong at the categorical level. Any deep explanations offered in response to more complicated questions will continue to be wrong.

However, confronting misconceptions at the categorical level seems straightforward enough for simple concepts such as *whales*. If a child is simply told explicitly that whales are *mammals* rather than *fish,* she might then be able to explain (or at least understand) why whales do not suffocate on land. The fact that most children eventually learn that *whales* are *mammals* (thereby "*whales are fish"* is not a robust misconception) suggests that lateral categorical shifts can occur readily for certain concepts, perhaps even without explicit refutation. But why is categorical shift not easily achieved for robust misconceptions of the *heat* kind?

Although, to our knowledge, no research has investigated confrontation at the categorical level, a closer examination of the relative ease of categorical shift for the *whale* example, suggests that two instructional steps are needed in order to overcome barriers to conceptual change for robust misconceptions. First, students have to be aware that they have made a category mistake, which amounts to confronting their ideas at the categorical level; and second, students must be knowledgeable about the category to which a concept actually belongs. We briefly discuss these two steps below.

Awareness

Shifting across lateral categories per se is not a difficult learning mechanism from a computational perspective or from everyday evidence as illustrated by the *whale* example and by the ease with which people can understand metaphors. For example, metaphors often invoke a predicate from one category and a concept from a lateral category, often from different ontologi cal trees. For instance, *anger* (a *Mental State*) is often treated as a *substance* (an *Entity*) that can be contained, as in "He let out his anger" or "I can barely contain my rage" (Lakoff, 1987). We propose that part of the difficulty of shifting categories for many science concepts has to do with lack of awareness, in that students do not realize that they have to shift their assignment of a concept to a different category. This is because reassigning a phenomenon or concept from one kind to another kind is a low frequency occurrence in everyday life. That is, students do not routinely need to re-categorize, such as shifting a whale from *fish* to *mammal.* This is because, in our everyday environment, our initial categorizations are mostly correct, since they are based on outward perceptual features. For example, when we identify a furry object with a wagging tail that responds to our commands as a real live dog (thus an *animal)*, we are almost never wrong, in the sense that it is actually a stuffed dog (thus an *artifact).* The fact that these category mistakes rarely occur in real life makes it difficult for learners to recognize that their understanding or lack of understanding of new concepts may originate from a category error at the lateral level. As with metaphors, the rarity of category mistakes is a ploy that is sometimes used in stories and films, to produce interest, drama and suspense, such as in the children's novel *Velveteen Rabbit*. Moreover, if people do make category mistakes, especially across ontological trees, such as confusing reality (either *Entities* or *Processes*) with imagination *(Mental States)*, it is considered bizarre and perhaps a sign of psychological illness.

The rarity of category mistakes in real life also reinforces the strength of commitment to the original category to which a concept is assigned, as well as to the boundary between lateral categories. For example, even four-year-olds treat *living beings* as fundamentally different from *artifacts*, in that they rarely associate *artifacts* with *animal* or human properties (Carey, 1985; Chi, 1988). The commitment to a particular category occurs even as early as age five. Once a concept is categorized, young children are extremely reluctant to change the category to which it is assigned. Keil's work (1989) has shown that, no matter what physical alterations are made to an object (such as a real dog), such as shaving off its fur, replacing its tail, and so on, five-year-olds will not accept such changes as capable of transforming a real dog to a toy dog (thus crossing the boundary between lateral categories *animals* and *artifacts*). However, they will agree that, with appropriate alterations such as replacing black fur with brown fur, one can transform a skunk into a raccoon. This is because skunks and raccoons belong to the same *mammal* category. Thus, once assigned, even five-year-olds honor the boundary between *kinds* and remain committed to the category to which they have assigned a concept.

Even though miscategorization is rare in everyday life, our proposal is that it is the fundamental source of robust misconceptions in science. That is, many phenomena in science look and act like they belong to one category rather than another. For example, geese flying in a V-formation (*Emergent* Example 2) looks like airplanes flying in a V-formation (*Direct* Example 2), heat fl owing into a cool room feels like water flowing down a stream. However, the causal explanations for the similar patterns are distinctly different. Thus, learners can be misled by perceptual similarities and treat such pairs of phenomena as having the same causal explanations, resulting in miscategorization of one but not the other. Therefore, students must be made aware of their miscategorization and must learn to discriminate between the two kinds of phenomena. In short, the lack of awareness of the need to shift categories laterally is due to the low frequency of such shifts in the real world and to superficial similarities among many phenomena. Instruction aimed at promoting such shifts must begin by making students aware that they have committed category mistakes.

Building a New Lateral Category

In our hypothetical *whale* example, it seemed relatively easy for children to shift categories simply by being told that whales are *mammals*. Why is this category shift so easy to implement?

Would science students find it easy to shift categories if we simply told them that heat transfer is an *emergent* rather than a *direct* process? The answer is no, obviously, because students are ignorant of the *emergent* category. That is, we assume that an *emergent process* category is not familiar and available to students and therefore they cannot assimilate novel concepts into it. This missing category situation is tractable and suggests an instructional approach of building such a category. Thus, instruction to promote categorical shift must also include instruction about *emergence*. Our prediction is that, to achieve successful conceptual change for robustly misconceived concepts, we need to first teach students the properties of such a category, which is uniquely distinct from the lateral category with which they are familiar and to which they have mistakenly assigned concepts, such as *direct* processes. Once students have successfully built such a lateral category with its distinct set of properties (as shown in Table 3.1a & Table 3.1b), they can begin to assimilate new instruction (for example, about heat transfer) into the category. Preliminary successes using this method have been shown in Slotta, Chi and Joram (1995), and Slotta and Chi (2006). Descriptions of our current successful attempt is forthcoming (Chi, Roscoe, Slotta, Roy, & Chase, submitted).

CAVEATS ABOUT THE ROLE OF REFUTATION IN HIERARCHICAL AND LATERAL CATEGORIES

This chapter addresses the problem of learning for which prior knowledge conflicts with tobe-learned information. This kind of learning is considered the conceptual change kind rather than the enrichment kind. We propose that prior knowledge can be in conflict with to-be-learned information in three ways. First, at a belief level, prior knowledge can be incorrect or false and conflict with correct information in the contradictory sense. In such cases, conceptual change can be achieved by refuting (implicitly or explicitly) the false beliefs, and this can lead to belief revision in some domains of science. Second, at a mental-model level, prior knowledge can be incorrect and conflict with correct information in the coherent-but-flawed sense. In such cases, conceptual change can be achieved by refuting multiple false beliefs within a fl awed mental model, especially the critical ones. The cumulative effect of many such belief revisions will transform a flawed mental model into the correct model.

However, three caveats need to be noted about the success of these types of refutations for belief revisions and mental model transformations. First and foremost, the success of these two types of conceptual change hinges on the assumption that the misconception and the correct conception are assigned into the same category or hierarchical categories, as shown in the left branch of Figure 3.2. An example might be mistaking the heart as the source of oxygenation rather than the lungs. Moreover, this kind of refutation can be effective whether presented implicitly or explicitly. Finally, the refutation maintains many of the elements of the misconceptions, whether false beliefs or fl awed mental models. For example, in contradicting that the *heart* is the source of oxygenation, we maintained many elements of the false belief and the flawed mental model, such as that *blood is the medium of transporting oxygen and carbon dioxide*, that *oxygen needs to be replenished in the blood,* and so on. In short, the success of belief revision and mental model transformation is domain- and concept-specific. Thus, for concepts such as *heart,* versus *lungs* as sources of oxygenation with systems such as *blood circulation*, misconceptions are *within the same hierarchical branch of an ontological tree* as the correct conceptions.

For misconceptions that are categorically misassigned between lateral branches or ontological trees (as shown in the right branch of Figure 3.2), we assume that this happens because learners cannot tell what category the to-be-assigned concept belongs or because they do not have the category to which the new concept should be assigned. Either way, conceptual change requires a

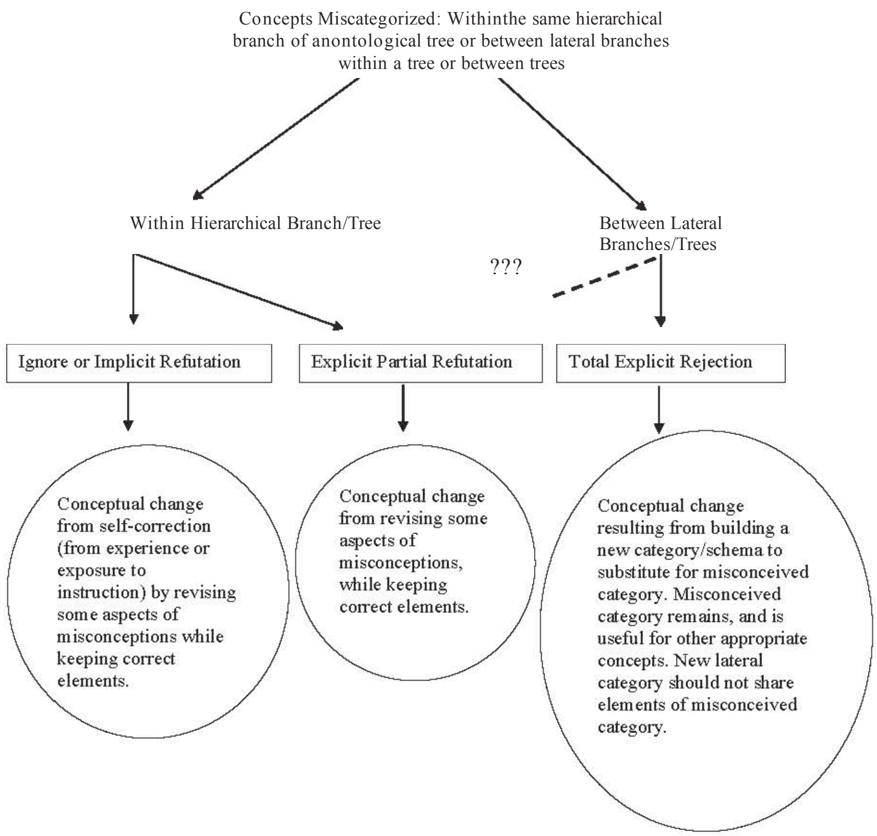

FIGURE 3.2 The type of instruction needed (rectangles), and the processes and types of conceptual change achieved (ovals), depends on how misconceptions are miscategorized.

categorical shift. Such a shift necessitates that the learner is aware that the shift is needed and that the correct category is available. For many robust misconceptions in science, the lateral category to which misconceptions have to be reassigned, *emergent processes,* does not exist in students' knowledge base, so instruction has to build a new category. Because *emergent* and *direct* processes are different in *kind*, with mutually exclusive properties, confrontation needs to reject the misassigned category and build the alternative *emergence* category, perhaps through direct instruction using contrasting cases (see Chi, Roscoe, Slotta, Roy, & Chase, submitted). Of course, the original *direct process* category can remain, as it is important for understanding other *direct processes*. In short, it is difficult to imagine how robust misconceptions can be corrected at a deep level if one maintains many elements of the misconceptions, as in approaches that call for integrating ideas or elements of misconceptions with correct science conceptions. This integrating approach is somewhat analogous to our "explicit partial refutation" approach, shown in Figure 3.2. We denote our skepticism by using a dotted line and question marks to link robust misconceptions with this type of instructional approach to achieve conceptual change, as shown in Figure 3.2.

ACKNOWLEDGMENTS

The author is grateful for funding and support provided by the Spencer Foundation (Grant 200100305). The author will be a professor at Arizona State University, effective August, 2008.

REFERENCES

Barsalou, L. W. (1983). Ad-hoc categories. *Memory and Cognition, 11*, 211–227.

Bransford, J., Brown, A., & Cocking, R. (Eds.) (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Broughton, S. H., Sinatra, G. M., & Reynolds, R. E. (2007). The refutation text effect: Infl uence on learning and attention. Paper presented at the Annual Meeting of the Americal Educational Researchers Association, Chicago, Illinois.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S.Carey & R. Gelman (Eds.), *The epigenesis of mind.* (pp. 257–291). Hillsdale: NJ: Erlbaum.

Chi, M. T. H. (1988). Children's lack of access and knowledge reorganization: An example from the concept of animism. In F. Weinert & M. Perlmutter (Eds.), *Memory Development: Universal Changes and Individual Differences* (pp. 169–194). Hillsdale, NJ: Erlbaum.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Ed.), *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science*, (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.), *Conceptual Structures and Processes: Emergence, Discovery and Change.* (pp. 209–234) Washington, DC: American Psychological Association.

Chi, M. T. H. (2000). Cognitive understanding levels. In A.E . Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 2, pp. 146–151. Washington, DC: American Psychological Association.

Chi, M. T. H. (2005). Common sense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*, 161–199.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

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., & Hausmann, R. G. M. (2003). Do radical discoveries require ontological shifts? In L. V. Shavinina (Ed.) *International Handbook on Innovation* (pp. 430–444). Oxford: Pergamon.

Chi, M. T. H., & Koeske, R. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19,* 29–39.

Chi, M. T. H., & Roscoe, R. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reframing the process of conceptual change: Integrating theory and practice* (pp. 3–27). Dordrecht, The Netherlands: Kluwer.

Chi, M.T.H., Roscoe, R., Slotta, J.D., Roy, M., & Chase, C. (Submitted). Domain-general qualitative understanding of emergence transfers to learning of specific science concepts, submitted to *The Journal of the Learning Sciences.*

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.

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science and programming. In C. B. Cazden (Ed.), *Review of research in education*. Washington, DC: American Educational Research Association.

de Leeuw, N. (1993). Students' beliefs about the circulatory system: Are misconceptions universal? In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 389–393). Hillsdale, NJ: Erlbaum.

Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education, 5*, 61–84.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science.* London: Routledge.

Duit, R. (2004). *Bibliography: Students' and teachers' conceptions and science education database.* Kiel, Germany: University of Kiel.

Engel Clough, E., & Driver, R. (1986). A study of consistency in the use of students' conceptual framework across different task contexts. *Science Eduction, 70*, 473–493.

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

Gentner, D., & Stevens, A. L. (Eds.) (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Gobbo, C., & Chi, M. T. H. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development, 1,* 221–237.

Guzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Meta-analysis of instructional interventions from reading education and science education to promote conceptual change in science. *Reading Research Quarterly, 28*, 116–161.

Jackendoff, R. (1989). What is a concept, that a person may grasp it? *Mind and Language, 4*, 68–102.

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

Keil, F. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88*, 197–227.

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

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Law, N., & Ogborne, J. (1988). Students as expert system developers: A means of eliciting and understanding commonsense reasoning. *Journal of Research on Computing in Education, 26*, 497–514.

Licht, P., & Thijs, G. D. (1990). Method to trace coherence and consistency of preconceptions. *International Journal of Science Education, 12*, 403–416.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models*, (pp. 299–324). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Rips, L. J. (2005) Concepts and categories: Memory, meaning, and metaphysics. In K. J. Holyoak & R. G. Morrison, *The Cambridge handbook of thinking and reasoning* (pp. 37–72). New York: Cambridge University Press.

Novak, J. (1977). *A theory of education*. Ithaca, NY: Cornell University Press.

Ram, A., Nersessian, N. J., & Keil, F. C. (1997). Special issue: Conceptual change. *The Journal of the Learning Sciences*, *6*, 1–91.

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naive physics reasoning: A Commitment to substance-based conceptions. *Cognition and Instruction, 18*, 1–34

Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science*, *21*, 147–177.

Schwartz, S. P. (1977). *Naming, necessity, and natural kinds.* Ithaca, NY: Cornell University Press.

Slotta, J. D., & Chi, M.T.H. (2006). The impact of ontology training on conceptual change: Helping students understand the challenging topics in science. *Cognition and Instruction*, *24*, 261–289.

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*, 373–400.

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

Thagard, P. (1990). Concepts and conceptual change. *Syntheses*, *82*, 255–274.

Vosniadou, S. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction, 14*, 445–451.

Vosniadou, S., & Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.

Vosniadou, S., & Brewer, W. (1994). Mental models of the day/night cycle. *Cognitive Science, 18*, 123– 183.

Wiser, M., (1987). The differentiation of heat and temperature: History of science and novice-expert shift. In S. Strauss (Ed.), *Ontogeny, phylogeny, and historical development* (p. 28–48). Norwood, NJ: Ablex.

Wiser, M., & Amin, T. (2001). "Is heat hot?" inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction*, *11*, 331–353.