Chapter 3
The Nature of Scientific Meta-Knowledge

Barbara Y. White, Allan Collins, and John R. Frederiksen

Introduction

We argue that science education should focus on enabling students to develop meta-knowledge about science so that students come to understand how different aspects of the scientific enterprise work together to create and test scientific theories. Furthermore, we advocate that teaching such meta-knowledge should begin in early elementary school and continue through college and graduate school and that it should be taught for all types of science, including the biological, physical, and social sciences.

In this chapter we outline a theory of scientific meta-knowledge, which is built around four critical aspects of science. We refer to the four types of meta-knowledge that are needed as (1) meta-theoretic knowledge, (2) meta-questioning knowledge, (3) meta-investigation knowledge, and (4) meta-analytic knowledge. While numerous investigators, working in the philosophy of science, cognitive science, and science education, have developed theories about various aspects of the nature of science, there have been few attempts to provide detailed models that explicate all four of these components, along with an elaboration of how they work together in scientific inquiry. This has resulted in science curricula that provide students and teachers with impoverished views of the nature of science.

Most K-12 inquiry-oriented science curricula, for example, underemphasize the role of theory, particularly competing theories. They also fail to adequately explain the relationship between theory and evidence, nor do they adequately portray the inquiry processes that are involved in the interplay between the two. Even college-level textbooks on research methods fall short. Most focus on investigation and analysis and do not adequately discuss the different forms that scientific models and theories can take or the types of evidence and arguments that can be used to support or refute them.

B.Y. White (✉)
University of California at Berkeley, Berkeley, CA 94720-8099, USA
e-mail: bywhite@berkeley.edu

M.S. Khine, I.M. Saleh (eds.), *Models and Modeling*, Models and Modeling in Science Education 6, DOI 10.1007/978-94-007-0449-7_3,

This chapter is our attempt to provide a more comprehensive and integrated overview of what science curricula should aspire to teach about the nature and processes of scientific inquiry and modeling. Our Inquiry Island and Web of Inquiry learning environments and science curricula, which engage young students in theory-based empirical research projects, provide first steps toward putting this vision into practice.

Perspectives on the Nature of Science

The layperson's view treats science as made up of facts and theories that systematic observation and experimentation have established over time. This is a somewhat static perspective, which leads science education to emphasize the coverage of important scientific facts, concepts, and theories. In contrast, the predominant view among researchers in science education is that understanding the nature of science is important and that scientific inquiry and investigation are at the heart of the enterprise (AAAS; 1990; National Research Council, 1996, 2007). Those who hold this view have developed inquiry curricula that teach students how to form and test hypotheses, carry out careful observations and measurements, analyze the data they collect in their investigations, and report on their findings (see Anderson, 2002).

Some leading researchers in science education have emphasized the importance of argumentation between competing theories as the core activity of science (e.g., Driver, Newton, & Osborne, 2000; Duschl & Osborne, 2002; Duschl, 2007; Kuhn, 1993; Osborne, 2005; Smith, Maclin, Houghton, & Hennessey, 2000). They therefore design curricula that engage students in considering competing theories and in understanding how evidence can be developed to support or refute those theories (e.g., Bell & Linn, 2000; Sandoval & Reiser, 2004; Suthers & Weiner, 1995).

In another vein, some view modeling and theory construction as the central goals of science, where theories are coherent bodies of concepts, laws, and models, which account for a wide range of observations and enable humans to predict, control, and explain what happens as events occur (e.g., Collins & Ferguson, 1993; Gilbert, 1991; Halloun, 2004; Hestenes, 1987; Mellar, Bliss, Boohan, Ogborn, & Tompsett, 1994; Perkins & Grotzer, 2005; Slotta & Chi, 2006; White, 1993; Windschitl, Thompson, & Braaten, 2008). This view leads to science curricula that help students learn about the nature of scientific models and the process of constructing and testing theoretical models (e.g., Grotzer, 2003; Lehrer & Schauble, 2000, 2005; Schwarz & White, 2005; Smith, Snir, & Grosslight, 1992; Stewart, Cartier, & Passmore, 2005; White & Frederiksen, 1998).

Other researchers have emphasized the collaborative nature of science (e.g., Dunbar, 1999, 2000). Indeed, the scientific enterprise can be viewed as a form of collaborative learning that enables society to develop and test theories about the world. This perspective leads to instructional approaches that employ a "community of learners" approach (Bielaczyc & Collins, 2000; Brown & Campione, 1996), which emphasizes collaborative inquiry and knowledge building, to develop students' understanding of the scientific enterprise (e.g., Borge, 2007; Herrenkohl, Palinscar, Dewater, & Kawasaki, 1999; Hogan, 1999; Metz, 2000; Scardamalia & Bereiter, 1994).

We argue that all of these perspectives capture essential components of the scientific enterprise. Scientific inquiry can be viewed as a process of oscillating between theory and evidence, in a practice of competitive argumentation that leads teams of researchers to develop and test alternative scientific models and theories. The ultimate goal is to create theories and develop arguments, which employ explanations and evidence to support or refute those theories, and thereby convince other researchers of the merits of your team's "current best theory" (cf., Carey & Smith, 1993; Driver, Leach, Millar, & Scott, 1996; Duschl & Osborne, 2002; Duschl, 2007; Giere, 1992; Hammer, Russ, Mileska, & Scherr, 2008; Klahr & Simon, 1999; Krathwohl, 1998; Kuhn, Black, Keselman, & Kaplan, 2000; National Research Council, 1996, 2007).

The transition from making theories to seeking evidence, through an investigation, is one where the generation of questions and hypotheses derived from theory is crucial. The transition from carrying out an investigation to the refinement of a theory is one in which data analyses and syntheses are central. This view leads to a basic model of scientific inquiry that has four primary processes: (1) theorizing, (2) questioning and hypothesizing, (3) investigating, and (4) analyzing and synthesizing. Associated with each of these primary processes is a regulatory process that monitors how well the process is being carried out and whether another process should be invoked to deal with issues that arise (such regulatory processes, though important, are beyond the scope of this chapter and are addressed in White, Frederiksen, & Collins 2009).

In our earlier work on teaching scientific inquiry to young learners (White & Frederiksen, 1998), we portrayed such a model as an inquiry cycle, which provides a scaffold for inquiry in the form of a series of steps that one undertakes in a never-ending cyclical process of generating, testing, and elaborating scientific principles and models, with the ultimate goal of developing a widely useful, accurate, and comprehensive theory for a given domain. This is, of course, a simplified view: Mature scientific inquiry does not necessarily proceed in this stepwise fashion. For one thing, it is possible to start anywhere in the sequence. So, for example, one might start with vague questions that are not based on a particular theory or one might start with an investigation or with existing data to generate theoretical ideas. Furthermore, one does not necessarily proceed through these "steps" in order. For instance, analyzing data can lead to the need to do further investigation. So the critical components in the scientific enterprise are closely intertwined, and any view of science education that underplays one of these components fundamentally misleads students as to the nature of science (Chinn & Malhotra, 2002). Nonetheless, for pedagogical purposes, presenting students with an inquiry cycle, in which one starts with theorizing and questioning, is an effective initial model that can enable students to develop capabilities for inquiry, as well as an understanding of its constituent processes (Frederiksen, White, Li, Herrenkohl, & Shimoda, 2008; White & Frederiksen, 1998, 2005).

This model of scientific inquiry reflects the way most sciences include two camps: the theoreticians and the empiricists. Theory and empirical investigation form the two poles of science. Research questions form a bridge between these two poles, in which competing theories generate alternative hypotheses about the answer to a question, which then are tested through empirical investigation. Analysis and synthesis form the other bridge between the poles by providing ways to represent and interpret data from the investigation to bear on the theories in competition and synthesize a new "current best theory."

Scientific knowledge in a field such as physics is usually thought to include basic theories (e.g., Newton's Laws) and concepts (e.g., acceleration), as well as problem solving, investigation, and data analysis methods. We argue that there is meta-knowledge sitting above this basic scientific knowledge that characterizes the different kinds of models, research questions, investigation methods, and data analysis techniques and the relations between these different aspects of science. This kind of meta-knowledge provides scientists with a toolkit of representations and techniques that enables them to be more productive thinkers and better researchers. We think that making scientific meta-knowledge explicit should lead to improving science education, as well as to improving the methods and practices of science.

The inquiry cycle underlying our analysis of scientific meta-knowledge is in fact embedded in the standard form of empirical articles: i.e., introduction, methods, results, and discussion. The introduction relates the investigation to existing theory and derives the research questions and hypotheses that the investigation addresses. The methods section describes how the investigation was carried out. The results section describes the data analyses and the findings from those analyses. The discussion section then brings the analyses back to existing theory and how it should be modified based on the findings. Hence the inquiry cycle we utilize is deeply embedded in the culture of science.

Most scientists tend to focus their efforts in one area or another. For example, some scientists are strong in theory, some in designing investigations, and some in data analysis, while others are more balanced in their approach. It is not necessary that a given scientist be an expert in all aspects of scientific inquiry, but the field must encompass all these different components. We would argue, however, that science education should emphasize an understanding of all the components so that learners come to appreciate what is entailed in constructing theories, generating research questions and hypotheses, designing investigations, and analyzing and synthesizing interpretations of the data to generate arguments that support or refute particular theories. Such an overview is critical to understanding the nature of science (Lederman, 2007).

We became sensitive to different kinds of scientific meta-knowledge in developing advisory systems, Inquiry Island and the Web of Inquiry, which guide students as they do science projects that require them to engage in theory-based empirical research (Eslinger, White, Frederiksen, & Brobst, 2008; Shimoda, White, & Frederiksen, 2002; White & Frederiksen, 2005; White et al., 2003; White, Shimoda, & Frederiksen, 1999). These systems incorporate much of the scientific meta-knowledge we outline below in their various advisors. They include

Quentin Questioner, Hugo Hypothesizer, Ivy Investigator, AnnLi Analyzer, Sydney Synthesizer, and Morton Modeler. These advisors, who live on Inquiry Island, offer guidance to students when they seek help or encounter problems. Each advisor presents goals, purposes, plans, and strategies, as well as definitions and examples, with respect to their particular component of the scientific inquiry process. We think of Inquiry Island as one way to make different facets of scientific meta-knowledge available to students and teachers.

In order to illustrate our views about scientific meta-knowledge, we will use examples from the domain of social psychology. Working with Inquiry Island and the Web of Inquiry, we have begun to develop different models of friendship that students can manipulate and refine. Our goal is to make scientific models readily accessible and modifiable by students so that they begin to develop an appreciation of the power of models for understanding and refining theories. Rather than providing students with models that reflect the "received truth of science," such as Newton's laws, we want to provide students with models, such as models of friendship, which they can critique. The students then are able to test and refine the models or reject them altogether and create new models that reflect their understanding of friendship. In this way students are acting like scientists, searching for weaknesses in models, running models under different conditions, and engaging in empirical research in order to test and refine the models. They are learning how theory development and refinement generate empirical questions and investigations. We will use these friendship models and possible investigations to illustrate the different kinds of scientific meta-knowledge that we think are important.

In the next four sections, we describe meta-scientific knowledge in terms of its four components: meta-theoretic knowledge, meta-questioning knowledge, meta-investigation knowledge, and meta-analysis knowledge (note that we do not mean "meta-analysis" in the statistical sense). We should emphasize though that meta-questioning knowledge includes meta-knowledge about forming both research questions and hypotheses, while meta-analysis knowledge includes meta-knowledge about data analysis and synthesis. For each of the four sections, describing the four types of scientific meta-knowledge, there are five subsections that explicate the different aspects of the meta-knowledge: (1) the different types, (2) the purposes, (3) the creation process, (4) the criteria for evaluation, and (5) the synthesis of the different types. In a final discussion section, we summarize the framework and discuss the issues it raises.

Meta-theoretic Knowledge

Different Types of Models

Meta-theoretic knowledge includes knowledge about the nature of scientific models and theories. In our work on epistemic forms and games (Collins & Ferguson, 1993), we characterized three types of theoretical models (or epistemic forms) that researchers use to guide their inquiry: structural, causal (or functional), and dynamic process (or mechanistic) models. The different forms of structural models include primitive elements (e.g., chemical elements), stage models, cross product tables (e.g., the periodic table), hierarchies, and comparison tables. Similarly there are different types of causal models, such as causal chains, form–function analysis (e.g., Hmelo-Silver & Pfeffer, 2004), and multifactor models (as in medicine). Grotzer and Perkins (Grotzer, 2003; Perkins & Grotzer, 2005) present a more detailed taxonomy of causal model types. Finally there are different process model types, such as system-dynamics models (e.g., Mandinach & Cline, 1994), production systems (Newell & Simon, 1972), and agent models (e.g., Wilensky & Resnick, 1999). All of these representational forms have epistemic games (i.e., rules and strategies) associated with them, which are practices scientists use as they construct models to characterize and theorize about different phenomena.

We can illustrate the different model types with examples of friendship models for each of the three types: a stage model of friendship (structural), a multifactor model of friendship (causal), and an agent model of friendship (dynamic process). We want to reiterate that these models are not meant to be correct models, but rather models with enough intuitive grounding for young students to test and refine.

The stage model of friendship shows how the five top-level factors in the multifactor model of friendship (shown in Table 3.1) might change over time among pairs of friends. There might be variations on this stage model that reflect a friendship that lasts a very long time (delete stages 4 and 5) or a friendship that breaks up suddenly over some event (delete stage 3).

The multifactor model of friendship, shown in Fig. 3.1, postulates a possible set of factors that affect the strength of friendship between pairs of people. This model has five top-level factors (the same as in the stage model) and a set of factors that contribute to each of the five top-level factors.

We have developed some preliminary prototypes for a number of different agent models, where different types of people interact by calling one another on the phone. The information they communicate to each other (e.g., “so-and-so lied about you”) leads the recipients to alter the values of parameters that reflect their feelings for each of the other people in the network, based on rules of the model (e.g., “if I

Table 3.1 A stage model of friendship that characterizes the relationship between two people

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
	Getting to know	Close friends	Drifting apart	Just broken up	Not friends
Trust	Increasing	Very high	Decreasing	None	Low
Common interests	Increasing	Many	Decreasing	Rejecting	Not applicable
Proximity	Seek out	Seek out	Seek sometimes	Avoid	Neither
Effort	Very high	High	Decreasing	None	Low
Communication	Difficult	Easy	Difficult	None	Easy

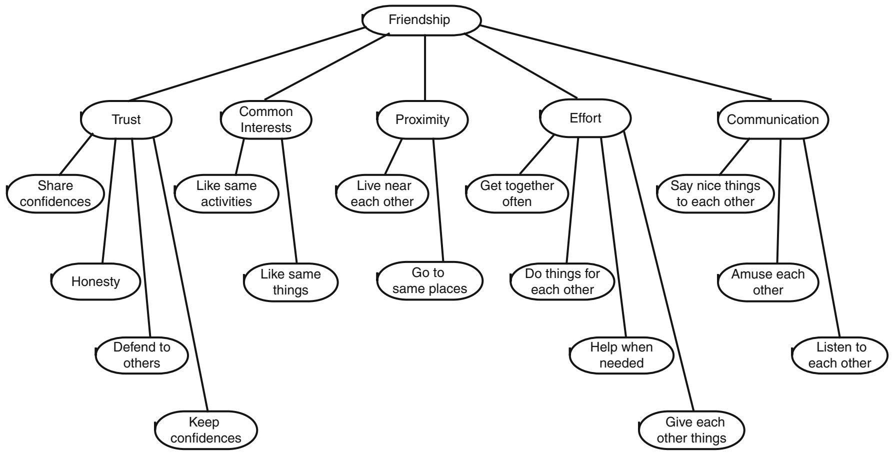

Fig. 3.1 A multifactor model of friendship

learn a person has lied about me, I will decrease my liking of that person"). These are simple agent models and lack graphic representations of the behavior of the agents, such as those in models like the Sims. However, they are both runnable and modifiable, so young students can see how the models behave under different input conditions and rules, which they then can modify. We will give a brief description of one of the more complex agent models.

An Agent Model of Social Interaction (Community Model)

This model embodies a personality for each person in a community, where the personality is determined by values on six different personality characteristics. People vary as to how appealing, amusing, talkative, honest, sweet, and shrewd they are. In this world, people call up people they like and tell them facts about people in the community and hear facts from the people they call. A person does not call the same person a second time until they have new facts to communicate.

People each start with 10 facts, assigned randomly, and acquire new facts that are generated throughout. They also remember any facts they are told. Each fact is about a person in the world and varies in truth and amusement value. People do not know about the truth of a fact, unless it applies to them or they made it up. Facts are remembered for a long time so that when new members join the community, the old members have lots of facts to tell them.

People tend to like people who are appealing and tell them amusing facts. Talkative people share more confidences and call more people. Dishonest people make up facts, and sweet people only say nice things. Shrewd people tell you nice things about people you like and bad things about people you do not like.

Each person has a degree of liking for each of the other people, which varies widely. Liking depends on many things, such as how appealing and amusing the other person is and whether you have discovered them to be dishonest. Depending on how much they like the person they are talking to, they will end the conversation sooner or later. Either party can end the conversation. The system can track how much each person likes the others over time to see how relationships in the community change over time.

Purposes of Different Models

Different model types serve different purposes. Structural models highlight the relationships between different elements in the models. Causal models depict the causal and functional dependencies between elements in the models. Dynamic process models allow one to run models of processes to see the consequences of different assumptions in the models. These runnable models can unpack mechanisms that explain the causal relationships depicted in static causal models, like the multifactor causal model shown above. Each different type depicts different relationships and properties. We will exemplify the purposes of different model types in terms of the three examples introduced above.

Stage models are valuable because they show how processes evolve over time. They are common in historical analysis, psychological analysis, and analysis of any process that can be characterized by a series of states. One of the most famous stage models is the cultural progression from hunter–gatherer to agricultural to industrial societies. Stage models help people to understand changes that occur in the world. For example, the stage model of friendship we presented in Table 3.1 might help children understand the way others are behaving toward them and even the way they are behaving themselves.

Multifactor models are a common way to analyze causality in systems. They are particularly pervasive in psychology and medicine, but are common in many other disciplines where events are caused by multiple factors. In well-specified multifactor models, variables (called factors or independent variables) are linked together in a tree structure. The branches of the tree are AND-ed together if the factors are all necessary to produce the desired value on the dependent variable. They are OR-ed together if any of the factors are sufficient to produce the desired value of the dependent variable. Often the factors are neither necessary nor sufficient, as in the multifactor friendship model we showed, where we identified a number of factors that taken together tended to facilitate friendship between pairs of people. Multifactor models are useful because they specify the different factors that produce a particular outcome and show how they are interrelated.

Agent models vary in the degree of intentionality the agents embody (Russell & Norvig, 1995). The friendship model we described has agents that have character traits, but no specific intentions. Other agent models we have been developing have different strategies for pursuing goals, such as getting people to like them. By running agent models, it is possible to see the consequences of different initial conditions and rules in the model. Some of the kinds of research questions that one might investigate with the model we described are the following: Do dishonest people start out being liked and end up being disliked? Are people with a few friends liked more intensely than people with a lot of friends? What happens when you introduce a liar into a community of honest people? Do people tend to make friends with people who have the same level of appeal? The power of agent models derives from the ability to study these kinds of questions and manipulate different rules and conditions to see the effects.

Creating Models

The different epistemic forms or model types are tools for making sense of the world. The structural forms are often the first forms people use to try to create a model. The simplest structural form is just a list so that if you are trying to account for the variables that affect friendship, you might start by creating a list of the causes or important factors that lead to a good friendship. As we explained in Collins and Ferguson (1993), a good list must satisfy a number of constraints: it should have multiple elements but not too many, the elements should all be of similar types, the elements should cover all the possibilities, and they should be mutually exclusive. If the list gets too long, or the elements are not all of the same type, then it might be that the list should be turned into a hierarchy or table, or one of the causal models, such as a multifactor model (see below). Trying to satisfy all these constraints helps the sense maker create the list.

Stage models are a simple type of time-structured list, but they invoke additional constraints. The simplest stage model is a list constructed with the constraint that the stages follow each other sequentially without overlap. Figure 3.2 shows a more complicated version of a stage model. Each stage might be characterized by multiple characteristics, and furthermore these characteristics may be arranged on a set of dimensions (e.g., the boy was angry and tired before his nap, but happy and energetic afterward). In a more complicated stage model, which is both structural and causal,

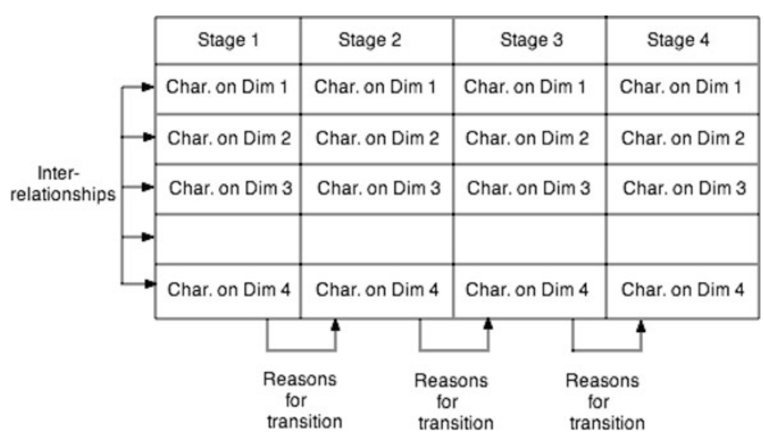

Fig. 3.2 A characterization of stage models

the interrelationship between the variables might be specified (e.g., energy state determines mood) and the reason for the change from one stage to the next specified (e.g., a nap leads to an increase in energy state). The last four constraints (i.e., multiple characteristics, specified dimensions, specified interrelationships, and reasons for transition) are all optional constraints that a person might or might not use in constructing a stage model.

We created the stage model of friendship shown earlier by first identifying a top-level list of factors (i.e., trust, common interests, proximity, effort, and communication) and treating these as dimensions that vary over time. Then we decided on a set of stages that friendships often follow and assigned values to each dimension for each of the stages we identified. As we mentioned earlier, one could also have a stage model for lasting friendships, or friendships that dissolve because of some incident, such as a deep disagreement, a betrayal of trust, or competition for another person.

Multifactor models are also created by starting with a list of factors that are conjectured to affect a dependent variable. Then the goal is to create causal chains that link the different factors to the dependent variable. Multifactor models can be expanded into system-dynamics models by adding feedback loops between the different variables. In the simple multifactor model of friendship we created, the different factors were assumed to affect one of the five top-level factors, which in turn affected the dependent variable (i.e., the strength of friendship).

Creating agent models is more difficult. The modeler first has to identify all the different kinds of agents that will live in the community and the kind of background world in which the agents interact. Then each agent needs to be given a set of states they can have (such as holding beliefs), actions they can take (such as talking to other agents), and reactions they can have (such as changing their beliefs), which can be specified in the form of rules for interacting with other agents in the world. If the agents possess higher order intelligence, one can also specify things like goals to pursue (such as getting people to like them), strategies for achieving the goals (such as flattering people), and even intentions or purposes to motivate them (such as the pursuit of happiness) (see Russell & Norvig, 1995). An important aspect of such a simulation model is how the behavior of the agents will be represented to the user, whether in dialog, tables, movement of agents around the world, or graphs of some kind. Many simulations of agent models provide multiple displays, which the user can manipulate in order to study different aspects of the behavior of the system.

Characteristics of a Good Model

There are a number of characteristics that a well-designed model should have. We have tried to characterize them in a way that is general to all of the different types of models:

- *Accuracy*: The model should accurately reflect some aspect of the way a system behaves or is structured. No model completely characterizes any system, but it should be accurate within the scope of its boundary conditions.

- *Generality*: The model should account for as wide a variety of phenomena in the world as possible. The model should cover all the phenomena within its scope.
- *Parsimony*: The model should be as simple as possible, but no simpler (to paraphrase Einstein). Parsimony helps to make the model clear and transparent to users.
- *Useful*: The model should have potential application to understanding and predicting the behavior of the modeled systems. While many models are constructed without regard to their application, the best models turn out to have useful applications.
- *Coherence*: The model should fit with everything that is known about the domain. In particular, as explained in the next section, it should cohere with other models to form an integrated theory of the domain.

How Different Models Fit Together

Scientific theories, in our view, are made up of a number of linked models. In chemistry, for example, the primitive elements (hydrogen, helium, etc.) are arranged in a cross product table (i.e., the periodic table). There is an underlying atomic structural model of protons, neutrons, and electrons arranged in shells that accounts for the structure of the periodic table. There are also constraints that determine how different elements combine into molecules, based on their atomic structure. Hence the standard theory in chemistry is made up of different types of models linked together in systematic ways.

In our prior work (Frederiksen & White, 2002; Frederiksen, White, & Gutwill, 1999; White & Frederiksen, 1990), we illustrated how various models of electrical circuits can be linked together to create a comprehensive theory of circuit behavior. Models of different types can be used to embody different perspectives on the behavior of a system. For instance, functional models can be used to show the purpose of a circuit and how subsystems within the circuit interact to achieve that purpose. Constraint models can be used to portray circuit behavior at the macroscopic level. They can reason about circuit behavior through the application of a set of laws that govern the distribution of voltages and currents within the circuit. Process models can be used to represent the behavior of circuits at a more microscopic level. They can show, for instance, how electrical forces within a circuit cause mobile charged particles to be redistributed when, for example, a switch is closed.

Such models can be linked in different ways. They can be linked derivationally when the behavior of one model type is emergent from the behavior of another model type (Wilensky & Resnick, 1999). For example, Frederiksen and White (2002) describe how their macro-level model of circuit behavior (which employs circuit laws) can be derived from the local-flow model of electricity they present, and how the local-flow model can be derived from their particle model of electricity (the micro-level model).

In addition, models can be linked developmentally when there is a “progression of models” (White & Frederiksen, 1990) such that a higher order model is created from a lower order model by adding new rules or entities. For example, a simple local-flow model of circuit behavior allows people to solve problems about serial circuits, and an elaboration of the model allows people to solve problems about serial, parallel, and hybrid circuits. Such model progressions, which lead to theory development, can involve the addition, modification, differentiation, or generalization of model components or even the construction of new models. Model progressions can be used to represent the evolution of scientific theories, as well as representing possible paths to understanding a domain (i.e., from low-level to high-level understanding).

The models of friendship we have been developing are linked together in various ways. We mentioned how the top-level factors in the multifactor model form the dimensions of change in the stage model and how emergent properties of the agent model appear in macro models such as the multifactor model of friendship. We have also developed a number of other models that are linked to the models shown: a flow model of trust, a system-dynamics model of popularity that links different variables in feedback loops, and an aggregate model of clique formation that reflects emergent behavior of the agent model described earlier. In addition, we have been creating a variety of agent models, one of which has agents that form friendships based on common interests and one where the agents pursue friendships using strategies derived from the multifactor model of friendship. The goal is for young learners to see how models at different grain sizes and addressing different phenomena are linked together to form a theory of a domain and how such a theory can evolve as research progresses.

Given the growing importance of modeling in science, we think it is critical that students learn about the different forms that models can take and how different models can be linked together to form a coherent and powerful theory. Hence, the essential meta-theoretic knowledge that people need to learn is how theories and models are created, refined, and extended. Our work on epistemic forms and games (Collins & Ferguson, 1993) is a preliminary attempt to outline how models are created. The work on linking models (Frederiksen & White, 2002; White & Frederiksen, 1990) is an attempt to characterize how different models can be integrated to form the theory of a domain, which has implications for how models are refined and extended. We think these various pieces fit together to form the basis for the development of meta-theoretic knowledge in science.

Meta-questioning Knowledge

Different Types of Research Questions

In order to evaluate models and theories, it is necessary to turn elements of the theories into research questions that can be directly investigated. Sometimes research questions are quite vague (e.g., What are the precursors to heart disease?) and sometimes the questions are specific (e.g., Does taking a particular drug reduce one's cholesterol level?). The hypotheses in any study are the different possible answers to a research question, which can be based on alternative theoretical positions. For example, in the Framingham Heart Study, which tried to identify precursors to heart disease, researchers identified some 200 plus possible precursors to measure. They then followed the people in Framingham over time to see if they developed heart disease. These 200 plus variables were their alternative hypotheses as to what might lead to heart disease.

The different epistemic forms (i.e., types of models) generate different types of research questions. Table 3.2 gives examples of the types of questions that arise in constructing structural models, causal models, and process models. These are not an exhaustive set of questions needed to construct the different types of models, but they are some of the most common research questions that arise as scientists create models to express and develop their theoretical ideas.

Table 3.2 Types of research questions generated by different epistemic forms

Form	Examples of questions
Structural	What are all the different types of X?
	What are the characteristics of X?
	What stages does X go through as it evolves?
	What are the components of X and how are they related?
Causal	Does Y cause X?
	What effect does Y have on X?
	What causes X to happen?
	What are all the factors that affect X?
Process	What process produces X?
	What are good strategies for accomplishing X?
	What are the rules of interaction between X and Y?
	What mechanism enables X to be achieved?

If possible, it is always best to construct research questions that differentiate between alternative theories or models. Any of the questions shown in Table 3.2 can be turned into a question to differentiate between two models. For example, the first structural question can be transformed into "Are the different types of X from set Y or set Z?" The first causal question can be transformed into "Does Y or Z cause X?" The first process question can be transformed into "Is X produced by process Y or Z?" Research questions that differentiate between alternative models force researchers to generate alternative hypotheses as answers to their question.

We can illustrate how the different types of research questions are tied to particular models in terms of the three models of friendship described in the previous section. These examples do not exhaust the possible research questions one can ask about each of the friendship models, but they do give the flavor of the kinds of questions each model provokes.

- The stage model suggests the following questions: What are the critical stages through which friendships evolve? Are there different trajectories for different

kinds of friendships? What are the critical characteristics on which the stages differ? What causes the transition from one stage to another?
- The multifactor model suggests the following questions: What are all the factors that affect whether two people form a friendship? Which factors have the largest effect? How do the factors combine to affect the degree of friendship? What causal chains link each factor to the degree of friendship?
- The agent model suggests the following questions: What strategies do people follow to get others to like them? What is the process of clique formation and do cliques change over time? Which kind of statements have the most effect on increasing or decreasing one's liking for another person?

Another type of research question addresses the issue of generalization. When one conducts an investigation, it is carried out in a particular context and, in the social sciences, with particular participants. In order to generalize the results of the study, it is possible to generate research questions of the type "Do the findings hold in other contexts?" or "Do the findings hold with different types of participants?" Establishing the bounds on models and theories is an important part of the scientific enterprise, and so these kinds of generalization questions are pervasive in research.

Purposes for Research Questions

The fundamental reason for constructing research questions is to derive some implication from a model or theory that can be directly investigated. Hence, research questions perform the job of narrowing down or extracting from complex theories a particular issue that can be studied. For example, Einstein's general theory of relativity was a broad and complex theory, which was not easy to test. But specific and surprising implications of the theory could be tested. One implication is derived from the theory's prediction that light waves bend in a gravitational field. Hence, a light source at a great distance, such as a distant quasar, should appear in two different places in the sky, if located behind an object with a strong gravitational field. The research question this generates is "Can we identify a distant object that appears in more than one place in the sky?" This question then led to a specific investigation that could test one implication of a complex theory.

Ideally research questions help to differentiate between possible theories. Finding crucial questions that in fact distinguish between alternative theories is very difficult. And when a crucial question is investigated, the researchers whose theories are not supported by the data usually can come up with some explanation that still preserves their theory, albeit in a modified form. But even in investigations that do not compare alternative theories, researchers have to come up with questions and data that enable them to differentiate their explanations for their findings from obvious alternative explanations that other researchers might generate.

Very often the purpose of a research question is to solve some practical problem or help achieve some worthy goal in the world. Hence the question driving the Framingham Heart Study (i.e., What are the precursors for heart disease?) was a very important research question to investigate, since so many lives depended on the answers that were found. In fact the study identified factors (e.g., smoking, high cholesterol, and high blood pressure) that could be treated, and hence the findings of the study have led to the saving of millions of lives.

Criteria for Good Research Questions

There is an art to generating good research questions. We have identified a number of characteristics that make for a good research question. These represent criteria that researchers should use when they generate possible research questions to investigate and need to choose which question(s) to pursue.

- *Interesting.* A good question addresses an issue other people care about. It is particularly interesting if the model or theory generating the question predicts a surprising answer to the question.
- *Worthwhile.* A question that addresses a salient issue in the world, such as the precursors of heart disease, is clearly worth pursuing.
- *Distinguishing.* A question that leads to answers that would help distinguish between competing theories is an important research question.
- *Accumulative.* If a question builds on previous research, it helps researchers create a more comprehensive theory for the domain being studied.
- *Feasible.* A research question needs to suggest investigations that can clearly be carried out and give results that fairly answer the question.

Generating Research Questions

There are generally three sources of research questions: theory, data, and practical issues or problems in the world. We will elaborate on how each of these sources acts to generate research questions.

When constructing a new theory, there are many questions that arise. As we indicated earlier, each of the epistemic forms or model types gives rise to a number of different questions. Testing the implications of a model similarly leads to a particular set of questions depending on the form of the model. And as we suggested, when there are competing theories, questions that distinguish the theories are particularly worth pursuing. Finally there are questions that come out of the limitations of any theory, such as questions of generalization or of filling in holes in the theory. For example, when Mendeleyev first constructed the periodic table, there were a number of elements that the table predicted should occur in nature, but that had not yet been identified. Hence his theory led to the research question: "Can we find elements with the missing characteristics?" As these examples suggest, models and theories are a very rich source of research questions.

In addition to theory, data or findings from an investigation often lead to new research questions. This happens in two ways. Often, some data found in a study seem to contradict the predictions from a theory. This can lead to revising the theory, but it can also lead to questions about boundary conditions of the theory (i.e., its generalizability). In other cases, patterns of data may not contradict the theory, but may suggest that the theory is incomplete in some way. In most cases, anomalous data suggest revisions to the theory, which in turn lead to new research questions.

As we indicated in our discussion of the Framingham Heart Study, research questions often arise to achieve goals that derive from needs or observations of the world. The history of science is filled with stories of how the world has posed questions for researchers that have led to important investigations in science.

How Research Questions Fit Together

When one research question is answered, it often raises a set of related questions. One way this occurs is when a particular structural pattern is found, as when Mendeleyev discovered the periodic table, it raised the question of why the elements in a single column have similar properties. This question led eventually to the model of the atomic structure of atoms. Similarly, a causal model, such as a multi-factor model, raises questions about the mechanisms that lead each of the factors to have the given effects. These two examples show how answering one research question can lead researchers to generate related questions about underlying processes, structures, and mechanisms, which in turn leads to more comprehensive theories for any given domain.

In summary, the meta-questioning knowledge that we think students need to acquire includes learning about the different types of research questions that can be asked and how each type of question is related to particular epistemic forms. Students also need to develop an understanding of how questions can be created to distinguish between competing theories and how, in the process of creating a deeper, more coherent theory for a domain, one question leads to another.

Meta-investigation Knowledge

Different Types of Investigations

The third component of scientific meta-knowledge is an understanding of the different forms that scientific investigations can take. There are many different investigation methods, but they generally fall into two basic types: (1) exploratory inductive investigations (often referred to as scientific induction) and (2) confirmatory investigations (often referred to as the hypothetico-deductive method). Exploratory inductive investigations are employed when one has broad research questions and some general theoretical ideas, which suggest interesting data sources to study, but which are not specific enough to generate particular hypotheses. The goal is to obtain data that will constrain one's efforts to develop more detailed models and theories. Confirmatory methods are used when one has a well-developed model, or set of competing models, which allows one to develop a set of theory-based hypotheses to test. The goal is to test each of the hypotheses to see if the findings are consistent with its theoretical predictions. This allows one to determine which models are most consistent with the data and which are not suitable for explaining the phenomena that have been investigated.

Exploratory Inductive Investigations

Galileo is famous for developing exploratory inductive methods in science. In his experiments on pendulums and gravity, he systematically varied the elements that he thought might affect the period of the pendulum and the speed of a ball rolling down an incline. From these exploratory investigations, he derived equations for the motion of pendulums and falling bodies. The Framingham Heart Study is a modern variation on his method using natural variation rather than controlled manipulation. The investigators in this study collected data from many people in Framingham Massachusetts on a large number of variables that they thought might influence the likelihood of getting heart disease. They then followed the people over many years to see if they developed heart disease and identified a number of variables that were precursors to heart disease.

The kind of data collected in exploratory investigations has a strong effect on the types of models that can be constructed from the data. Quantitative data support the construction of constraint-equation models, as we see with Galileo, or multifactor models, as we see in the Framingham Heart Study. To construct process models, such as the agent model described earlier, one needs a richer data stream, such as observational, protocol, or discourse studies provide. The goal of exploratory studies is to identify patterns in the data and systematic relationships that allow for the construction of models. These models can then be evaluated using confirmatory methods.

Confirmatory Investigations

Confirmatory investigations, which are designed to test theory-based hypotheses, can take many different forms. The best known is the randomized controlled trial, in which one or more hypotheses are tested by comparing conditions that correspond to each of the hypotheses being tested. Often such a test of a hypothesis contrasts an "experimental" condition, which includes some particular feature, with a control condition that lacks that feature. In such cases, the competing hypothesis is that the feature will have no effect, which is known as the "null hypothesis." In order to ensure the generality of the findings, the participants, or objects being studied, are assigned randomly to the different conditions. Often special efforts are made to control any variables that might affect the results, other than those being deliberately varied. After the experiment has been carried out, the data are analyzed to see if they are consistent with what was predicted by any of the hypotheses.

In one example of such a randomized controlled trial, we tested hypotheses about the impact of self-assessment on students' learning (White & Frederiksen, 1998).

In this study, three middle school teachers taught science lessons in two different ways using our ThinkerTools Inquiry Curriculum in which students construct theories of force and motion. By randomly assigning their classes to the experimental or control conditions, it was possible to hold the teacher variable constant. The experimental group engaged in systematic self-assessment of their work, whereas the control group spent the same amount of time reflecting on what they liked and did not like about the ThinkerTools learning environment. The experimental group showed significant gains as compared to the control group on a number of different assessments of their learning. The findings support the theory that self-assessment can positively impact the learning of science and scientific inquiry. One advantage of collecting multiple measures of performance is that the patterns found can lead to further hypotheses and theory refinement, such as theories about the impact of low-achieving students working on self-assessment in collaboration with high-achieving students.

Examples of Investigations

We can illustrate different confirmatory and exploratory investigation methods in the context of creating and testing models of friendship. In order to construct a multifactor model of likeability, students could break into groups to conduct interviews or surveys of people they know in order to identify critical factors. They might ask: What is it about other people that makes you like them? If different groups of students construct different multifactor models of likeability, these models could be tested by different confirmatory methods.

One kind of confirmatory study might set up a series of small parties, where actors talk to different people, who do not know the nature of the study. The actors would apply different strategies to get people to like them. Each actor would apply a different strategy in each party so that the strategies and actors are properly counterbalanced. For example, one pair of actors could use a flattery strategy, another pair could appeal to common interests, and a third pair could tell amusing stories. After talking to each person, the people participating in the party could rate each of the actors on likeability. The average likeability rating for actors using each different strategy would be a measure of the strategy's effectiveness. This experiment thus could determine which of the strategies is most effective and the relative efficacy of each.

Another kind of study, using observation and interviews, might be helpful for constructing a stage model, a multifactor model, and an agent model of friendship. The idea is you would study how the relationships in a new group of people evolve over time—for example, a book group made up of people who do not know each other beforehand. The book group would be videotaped, and interviews would be conducted with each participant shortly after each meeting. The interviews might ask (1) How do you feel about X and why? (for X = each person), (2) Did you do anything to make others like you?, and (3) Did you do anything to attract or repel X and if so what? The research questions could be the following:

- How do the relationships change over time?
- What factors determine how much each person likes another?
- What strategies do people use to attract or repel others?
- Do cliques form and change over time? What is the process of clique formation?

Purposes of Different Investigations

The goal of exploratory investigations is to create or refine a model. Often different methods are applied to triangulate on the phenomena in order to construct a more robust model. Different inductive methods should be applied depending on the type of model the investigator is trying to create. We described one method that might be used to create a multifactor model, though such self-report data may be systematically biased. If one wanted to create the kind of agent model we outlined in the theory section, you would need to apply more intensive methods. For example, you might carry out an observational study in different groups, with cell phones that record all conversations, in order to identify salient personality types, the nature, duration, and frequency of the messages between different people, and the bases for the cliques that form among the groups. A limitation of such intensive investigations is that the researchers may have systematic biases in what data are collected and in how they are interpreted. That is why confirmatory studies are critical to the development and refinement of models.

Confirmatory investigations are needed to evaluate the accuracy of models and to resolve conflicts between different theories. They are also critical to determining the boundary conditions for different models in order to decide the range of situations to which the model applies. As described above, theories and models generate research questions, which in turn generate hypotheses. These hypotheses are then tested with investigations designed to see whether or not the prediction made by each hypothesis is supported.

Creating Investigations

There are a number of steps to creating an investigation. The first step is usually to decide on your research questions. Some research questions, such as “what are all the factors that affect X,” lead to exploratory investigations. Other questions, such as “does X have a positive effect on Y,” suggest confirmatory studies. In order to address your question, you need to simulate the situation or process you are trying to model. If your question is “what factors affect whether a person likes another person,” you need to create a situation where people are deciding whether they like someone or commenting on why they like or dislike other people. If you want to know whether a particular strategy causes people to like the person employing the strategy, then you need to create a situation where people are judging other people who employ that strategy. If you have complex research questions, such as “How do cliques form in a community?” and “What are the characteristics that determine which people form cliques together?", then you need to carry out an investigation that embodies the entire process of clique formation.

Next you have to decide how to collect a representative sample with respect to the phenomena the model addresses. If you are surveying or interviewing people, you choose a random sample of people who fit the class that the model is supposed to characterize. If you are creating a situation that is supposed to reflect situations in the world, then consider whether there are characteristics of the situation you have created that will bias the results you find. For example, if you are creating a party where actors are trying to make people like them, it is important to select a random sample of people that is large enough to represent the population you are modeling (e.g., high school students). It is also critical to consider ways that the party might be unrepresentative of the ways that people form opinions about others. Ensuring that the subjects are unaware of the nature of the experiment is one way to make the party more representative, but it is still a highly artificial situation that distorts in different ways the basis on which people form opinions about others. To compensate for this artificiality, it is necessary to carry out other investigations that address the same issues.

To carry out an investigation, you need a detailed plan of the steps you will follow, specified in enough detail that other investigators could carry out the investigation themselves. For example, if you are conducting interviews with people to determine the factors that make for likeability, you need to specify the questions you will ask in a survey or interview, how the respondents will be chosen and how the data will be collected. In surveys, you have to specify the layout and order of questions. In interviews, you have to specify the method of recording answers and the order in which the questions will be asked. In the party simulation, you need to specify how different actors will carry out the strategies so that there is replicability in implementing the strategies.

It is wise to check your plan to see how feasible it is to execute and whether it produces the kind of data you expect. The most common way to evaluate a plan is to conduct a pilot study with a few subjects over a short period of time. You do not have to analyze the pilot data carefully, but you should check to see that the data you are getting are reasonable. You should consider whether any changes in your plan are warranted on the basis of the pilot results.

When you carry out the investigation in full, you should look carefully at the results to see if they are plausible and consistent. You also need to determine whether the plan was executed as specified or whether there were problems that arose that caused the investigation to vary from the planned procedure.

Further Issues to Consider When Designing Investigations

Confirmatory studies, designed to confirm or test hypotheses, are best suited for situations where there are a small number of hypotheses and variables. When situations are complex, investigators may only be able to test a few specific predictions of a theory. Many confirmatory studies simplify or standardize the situation and collect a large amount of data, hoping that factors that have not been controlled are contributing randomly to the effects being investigated and will not affect the group averages in any systematic way. Another way to deal with complexity is to use multivariate methods, such as regression or covariance analyses. The intent in these methods is to control complexity by taking into account, through statistical adjustments, factors, other than those specified in the hypotheses, which might have effects on the results. The problem with these multivariate methods is that they depend on correlational evidence and ignore variables that are not quantifiable. Many verification studies, such as those using analysis of variance methods, simplify the situation in order to get enough constraint to establish a strong causal argument.

In exploratory inductive investigations, different methods and data sources are often used to cover the phenomena of interest in order to construct a more robust theory. Dewey (1910, Chapter 7) suggested three principles for regulating the observation and collection of data in forming "explanatory conceptions or theories" that are still wise advice today: (1) Care must be taken in differentiating between what is observed and what is inferred, so as not to jump to hasty conclusions about one's theory. (2) One needs to look for multiple cases to see how general one's conclusion is, but one also needs to look for contrasting cases in order to determine the factors that are critical to the conclusion. (3) One needs to look for exceptions and contrasts that may challenge one's initial conclusions and suggest others (which is similar to having control conditions in a confirmatory investigation). One often learns more from examining anomalous cases that do not have the expected features.

Characteristics of a Good Investigation

There are a number of characteristics that make for a good investigation:

- *Directness*: The investigation should be tightly linked to the phenomena being modeled. The more indirect the measures, the more likely that the data will not accurately reflect the real world (Frederiksen & Collins, 1989).
- *Replicability*: The investigation should be carefully specified so that other researchers can carry out the investigation and ideally find the same results. This is related to the generalizability of the results. The plan for the research should characterize the situations within which the same results can be expected.
- *Transparency*: The procedures should be readily understandable by other researchers so that they can evaluate how well the investigation addresses the claims made. Transparency is also critical so that other researchers can replicate the investigation.
- *Systematic*: The procedures should be carried out thoroughly and systematically so that the results found can be relied upon. This requires representative samples and careful measurement. The accuracy of the findings depends on carrying out the investigation in a systematic way.
- *Distinguishing*: For confirmatory investigations, the experimental design should test the hypotheses so that the findings will support particular theories or models

and rule out other plausible theories and models. Such investigations are labeled "crucial experiments."

- *Feasible*: The investigation should be possible to carry out by other investigators so that the results can be replicated. Ideally the experimental conditions should be easy to create and the instruments and resources needed widely available.

How Different Investigations Fit Together

There are three major ways that different investigations can relate to one another: (1) One investigation can replicate another. (2) One investigation can complement another in order to triangulate on the results. (3) An exploratory investigation can lead to a confirmatory investigation, which can trigger further exploratory investigation. We will briefly discuss each of these relationships.

No investigation is ever an exact replication of another. The situations will be different, the participants will be different, and the procedures will be different. Replications help to determine how general the results are by showing whether they hold up under all these differences. When different results are obtained in a replication, it leads investigators to question whether the original results were spurious or whether some critical difference between the two investigations led to the different results. So replication is critical to the progressive refinement of theory.

Triangulation is an important strategy when investigators are dealing with complex phenomena. The goal is to bring different methods to bear on the same phenomena. For example, an investigation about what factors make someone like another person might collect data from three different sources: surveys with a range of people, data on what strategies are effective at a simulated party, and structured interviews with people about their friendships and feelings toward others. Other kinds of data, such as ethnographic observation, would also be valuable in order to triangulate on the best theory.

We have argued that exploratory inductive studies help in constructing models and theories, whereas confirmatory investigations are used to test hypotheses that represent competing theories; however, the process is really cyclical. Often confirmatory studies, especially if they include collecting rich data, provide clues for further theory development through an embedded inductive investigation of those data. This can lead to theory refinement and suggests new confirmatory investigations that can further refine and test the theory. Thus in scientific inquiry, testing hypotheses deduced from theories and interpreting patterns in data to construct new theories are intertwined. This process is complex and depends on meta-knowledge of the forms of inquiry one is engaging in at a given time and how they are interrelated as one moves from one form of investigation to another. Meta-investigational knowledge also makes one aware of the pitfalls and limitations of the forms of investigation one is using at any particular time.

Meta-knowledge for Data Analysis

Different Data Analysis Methods

Data analyses are systematic procedures for examining the information obtained in an investigation. We classified investigations as either exploratory, confirmatory, or mixtures of both, so we can characterize data analysis methods the same way. In addition, we can classify them as qualitative or quantitative. In fact, most graduate schools in the social sciences offer separate courses for quantitative and qualitative methods, thereby treating them as distinct categories.

Although the methods for analyzing qualitative versus quantitative data may differ, we argue that the primary goals of the analysis remain the same. Thus data analyses can perhaps best be characterized by how they support these common goals: (a) coding and representing data to reveal patterns, (b) testing hypotheses, (c) inducing new theoretical models, and (d) considering the generality of the findings. Exploratory studies, in which the main goal is to develop a model that accounts for the findings, require the testing of hypotheses to induce the model. Confirmatory studies not only test hypotheses derived from existing models but also provide rich data usually that lead to new models or theories (which are often about underlying mechanisms).

In the following, we illustrate some different data analyses by describing how data from the three studies we outlined previously, in the section on investigation, could be analyzed.

Study 1. Qualitative and Quantitative Exploratory Analyses Based on Interviews with Students About Their Views of Friendship

Data for this exploratory study are the tape-recorded answers students have given to the open-ended question: What is it about people that makes you like them? Analysis steps one can take are as follows:

Represent students' responses through data coding. In coding, one seeks to create as short a list of attributes as possible, which still covers all of the ideas that are mentioned by the respondents. What results is a set of "emergent" categories to capture the range of ideas expressed by students in responding to each question. The categorized data are recorded in a data table that indicates, for each student, which attributes were mentioned and which were not. A code book is also produced that provides examples of responses given each code category. The same codes are used for every respondent so that you can make comparisons among them.

Represent data in ways to see patterns. Create bar charts showing the percentage of students mentioning each response category to see what are the most frequent attributes mentioned by students. Also, try to determine which attributes "go together" in the respondents' answers to the two questions.

Use the patterns in the data to build a model. Having identified the relevant attributes, one might construct a multifactor model. By identifying which attributes go together, one can identify factors that interact in the model.

Consider the generality of the findings. The statistical tests of differences among frequencies of categories will give an indication of the likelihood that one would find similar results if the study were repeated with a new, comparable set of respondents. However, the investigation method used (asking questions of respondents) limits one's ability to claim that these results have been shown to apply to real-world friendship formation.

Explore the data to get ideas for follow-on analyses or investigations. As you study the data, you might notice, for instance, that some attributes of friendship seem to be mentioned more often by boys or girls. You could test this hypothesis by looking at frequencies of categories separately for each gender. As you build your multifactor model, you might think of new ways to test your model. For instance, you might decide to ask respondents questions about particular friends, instead of friends in general, to see if the same attributes are important for different people.

Study 2. A Quantitative, Confirmatory Analysis to Test Hypotheses About Strategies for Making Friends

This proposed investigation sought to test elements of the multifactor model using a "party" simulation. Imagine that actors were chosen to adopt different strategies for making others like them, such as (1) flattery, (2) common interests, and (3) amusing stories. For each strategy, there were a number of actors. Each group of actors was chosen to be equally appealing a priori, on average. The people at the party circulated and talked to all of the other people, including the actors (they were not aware that there were any actors at the party). After they talked to each person, they rated the actors on likeability using a five-point scale. In order to get reliable data, the procedure was repeated with three different groups of people, with each actor using a different strategy at the three parties.

Enter data into a data table. For each actor using a particular strategy, his/her rating should be entered into a data table. The data table should use a systematic design, so it will be easy to analyze using a statistics program or spreadsheet.

Represent data in ways to see patterns. Calculate the average likeability for each strategy and actor. Look to see if any one strategy was better than any others and make a bar chart of the means for each strategy and actor.

Test hypotheses. Calculate the mean rating for each strategy. If the pattern of differences among strategies is consistent with any of the hypotheses, you can have increased confidence in the model that generated that hypothesis. If the results are not consistent with any of the hypotheses, think about why. Is the model inadequate? Are there things going on in the study that the model did not anticipate?

Consider the generality of the findings. For each person, calculate the mean likeability for the actors who used each strategy. Then do a statistical comparison of the means for each strategy using analysis of variance or *t*-tests. If there are significant differences among them, there is evidence of generalizable trends for this population.

Explore ideas for follow-on investigations. In order to get a better idea of the reasons why people liked, or did not like, the actors who used particular strategies, one could interview the participants and ask them about their reasons for liking or not liking each actor.

Study 3. Qualitative and Quantitative Exploratory Analyses Based on Observation of the Development of Friendships Within a Book Group

This is a study of how groups of friends evolve over time in an actual social group, here, a book group made up of people who do not know each other beforehand. The goal is to construct an agent model that embodies the interaction patterns. Regular meetings of the book group would be videotaped, and interviews would be conducted with each participant shortly after each meeting. The interviews might ask (1) How do you feel about X and why? (X = each person)? (2) Did you do anything to make others like you? (3) Did you do anything to attract or put off X and if so what? The research questions could be the following:

- What factors determine how much each person likes another?
- What strategies do people use to attract or put off others?
- How do their relationships change over time?
- Do cliques form and change over time? What is the process of clique formation?

Code the data. Code the interview data and enter the results into a data table, as described for study 1. Code the interactions among all participants shown in the videotapes of book group meetings. One way of coding the data is to consider each person and code their interactions with other members of the group (this is the unit of analysis). Things one might code are (a) Who are they addressing/commenting on? (b) Are they interrupting? (c) What is their emotional tone (e.g., smiling, nodding, grimacing, sighing)? (d) Are they agreeing or disagreeing? (e) Are they building on, ignoring, or contesting the other's ideas?

Represent data in ways to see patterns. One goal is to identify closely interacting groups (cliques) based on observations of interactions among individuals. Cluster analyses could be carried out using different kinds of data, such as positive interaction time and negative interaction time, the frequency of positive versus negative interactions, and the positive feelings for each member expressed in the interview. Groups could be identified for each meeting of the book group so that one could look at changes in clique memberships. Another goal is to identify the precursors of clique formation in terms of the kinds of interactions members have in prior meetings. One could also look at the interview response codes for the same interactions, such as the strategies people used to attract others.

Build a model and explore its ramifications with further data analyses. These data analyses will provide a rich array of information about factors that contribute to forming friendships in a new social group. The model is constructed by developing various hypotheses about critical factors and determining whether they apply across different circumstances. If they do, they become part of the model. Considering implications of the model will lead to additional questions and, often, to further hypotheses about factors that might be observable in the data at hand for the study.

For instance, new criteria for video analysis might be suggested, which could then lead to additional data analyses in which these hypotheses are tested.

Consider the generality of the findings. Since the study used only one social group (the book group), data are not available to empirically test the generality of the findings for other kinds of social groups that might be constituted. However, an argument for the plausibility that the results are generalizable can be made based on the typicality of the group used in the study.

Purposes of Data Analysis

The main purpose of data analysis is to support the development of convincing arguments, which show how the findings from an investigation support particular conclusions and have implications for theories. Date analyses examine the information obtained in an investigation in order to meet several objectives in developing and testing models and theories. These include the following:

(1) *Creating representations that will reveal patterns*: One objective is to code and display data in ways that summarize the data and reveal patterns. Achieving this goal requires what diSessa terms meta-representational expertise (see diSessa, 2002a, 2002b, 2004).

(2) *Interpreting how data provide evidence with respect to competing hypotheses*: A second objective is to use patterns found in the data to determine which hypotheses are supported or refuted by the data. These may be hypotheses that were predicted ahead of time, using existing models (confirmatory studies), or that emerge from the data to construct new models (exploratory studies).

(3) *Exploring the data to develop new models, theories, and ideas for further research*: Another objective is to search the various representations of the data for unanticipated phenomena or relationships among variables. Discoveries made through this process may lead to modifying an existing model or to creating new models and theories. This, in turn, leads to further research to investigate the utility and generality of the new model.

(4) *Establishing the generality of the findings*: A final purpose of data analyses is to provide evidence regarding the generality of a theoretical model—the range of circumstances to which it applies. If there are limitations to the conclusions one can draw about the model's generalizability, this can lead to adding boundary conditions to the model or to revising it. Alternatively, it may lead to suggestions for further research in order to more adequately determine the generalizability of the model.

Pursuing these objectives leads to data analyses that enable one to develop and test theoretical models, as well as to improve the range of coverage and explanatory power of a theory.

Creating Analyses

Coding data. In developing ways of coding data, one should think about what features of the data need to be represented to test existing hypotheses and develop new ones. Sometimes these features are not immediately available in the data and need to be inferred or coded. This is most often the case when the data are qualitative, such as when they are based on participants' responses and are coded using categories or abstracted descriptions.

In order to see patterns, data obtained in different situations need to be represented using similar measurements or coded qualitatively using similar categories. This makes it possible to make comparisons of data across different situations. Coding of variables may be based on either predetermined aspects of responses, or they may be "emergent" categories based on examining all responses and creating categories that distinguish them in theoretically interesting ways.

Representing data to reveal patterns. Ways of representing data need to be devised that will (a) indicate whether patterns predicted by a hypothesis or model are present, (b) reveal unexpected patterns that may require modifications to the model, or (c) reveal patterns that lead to the induction of a new model. The goal in each case is to display relational features of the data that show what is "going on" and thereby provide a way of testing the explanatory power of competing theories. Tools used to represent data include statistical ways of describing data (e.g., means or frequencies for different situations) and graphic tools for visualizing relations (e.g., bar charts, scatter plots). Experience and education in data analysis techniques will lead to creating a "library" of forms for displaying data (cf., Giere, 1991). For instance, a graph of average values obtained before and after a treatment, shown for two different treatments, may be a good way to see if there is an interaction between treatment and effect in a confirmatory study.

There are many kinds of patterns that can emerge from data, paralleling the many kinds of relations among variables that are generated by different types of models. In complex, multivariable data sets, the number of pair-wise relations among variables can be great, and the possibility of interactions among variables increases the number of possible patterns even further. Exhaustively searching for all meaningful relationships is often impractical. Thus the particular techniques used in data analysis to reveal patterns are often guided by the epistemic forms of the models one is trying to create and test. The choices of epistemic forms and data analysis methods interact: you see data patterns through analyses that are themselves suggested by theoretical models. In arguing for a particular interpretation of data, you need to be cognizant of how other investigators with different theoretical orientations might interpret the data. One of the most difficult things in data analysis is to be able to "put on the hat" of a different theorist and consider alternative forms of analysis and types of models which they might entail.

Testing hypotheses. If the investigation is a confirmatory study, then the predictions made by the competing hypotheses need to be tested against the data. Similarly in exploratory studies, one generates hypotheses that need to be tested. For each hypothesis, one reviews the summarizations of the data, created in the previous step, to see if the hypothesis appears to be supported or refuted. Investigators often use both descriptive and inferential statistics to do this. In carrying out this step, you must be sure to identify patterns that disconfirm predictions of the underlying theory, or that were not anticipated, in order to see which aspects of the model need to be replaced or improved or if the whole way of thinking about the phenomena needs to be reconsidered.

Patterns can thus be found in data that provide evidence for whether a hypothesis is confirmed or disconfirmed. Confirming a hypothesis increases confidence in the theory's accuracy, but does not confirm the theory itself, because other theories might be constructed that lead to the same prediction and hypothesis. However, disconfirming a hypothesis can support an argument for rejecting the associated theory. Popper (1963) argues that strong theories are subject to refutation when tested, but can never be fully confirmed. Theories that are not fully specified are hard to disconfirm, because they can be augmented to account for factors that had been left out. Having such meta-knowledge about the relationship between theory and evidence should help investigators make appropriate inferences from their data about the status of their theories. It should also lead them to create theoretical models that are increasingly well specified as their research progresses.

Creating new models. If the investigation is an exploratory study, then one is using the data to develop new theoretical ideas. Also confirmatory studies often produce findings that were not predicted by any of the hypotheses and that require a new theoretical model. In creating a new model, you need to incorporate concepts and relationships that predict and account for the patterns found in the data analysis. Choosing the type of model, or epistemic form, will determine more specifically what is needed from the data analysis, such as what type of patterns need to be identified (Collins & Ferguson, 1993).

Finding evidence of generalizability. Theoretical models are expressions of relationships that have general applicability across a range of situations. In scientific inquiry, establishing the generality of a model is important. Commonly, one obtains data from a sample of different situations, or individuals, to provide evidence for the consistency of the results that are predicted by the model across the range of circumstances to which it purportedly applies. However, a theoretical argument for the generalizability of a model to other situations can also be made. This can take the form of specifying the conditions that are necessary for a model to apply, with the implicit suggestion that other factors not mentioned are irrelevant. Such theoretical arguments lead to further research.

When interesting patterns are found during data analysis (e.g., that support or refute a hypothesis or that suggest a new model), there is a need to determine that the patterns are not "flukes," but are reproducible over a range of instances that the model purports to cover. Often this involves studying a sample of similar situations and determining statistically that the patterns have occurred too regularly to have happened by chance.

Emerging ideas for follow-on investigations. Exploratory studies lead to the creation of theoretical models that require further investigation. In confirmatory studies, a careful and deep data analysis often leads to many new questions and hypotheses for future investigations. In addition, the data analysis will suggest many ways in which the investigation could be improved. For example, the investigator may discover factors that were not controlled or may think of different situations that could have been studied.

Characteristics of a Good Data Analysis

In carrying out a data analysis, the desirable characteristics that should drive the analysis include the following:

- *Perspicuous*: The representations of the data should organize and summarize the data in productive ways, so one can see the patterns.
- *Complete*: The analyses should test all reasonable hypotheses in order to be thorough and complete.
- *Systematic*: The analysis procedures should be carried out carefully and systematically.
- *Accurate*: The analysis needs to avoid miscalculations and violation of the assumptions inherent in the methods used.
- *Coherent*: The interpretation should fit all of the data in an integrated way and be consistent with other known sources of data that are pertinent.
- *Transparent*: The analysis procedure should be easily understood and replicable by other researchers.

How Different Analyses Fit Together

As we illustrated in this section, analyzing rich sets of data can allow researchers to test specific hypotheses and develop new theoretical models. This applies, as we have argued, whether the study is primarily exploratory or confirmatory. Often this means collecting both quantitative and qualitative data about a phenomenon. The best analyses involve displaying the data in different forms so that different patterns can be seen. The different patterns then can be tested for statistical significance or for consistency across cases, which can lead to supporting or revising an existing model or to developing new theoretical models. Ideally researchers should be able to synthesize findings from all of their analyses to produce a coherent interpretation that supports a particular theory, one that provides a better account of the data than other competing theories they considered. This can involve, for example, testing causal models, perhaps using inferential statistics, while also analyzing the data to develop process models of the underlying mechanisms. The resulting theory would thus incorporate models that provide both causal and mechanistic accounts of the phenomena being studied.

Discussion

Summary of the Meta-knowledge Framework

We have proposed a framework for scientific meta-knowledge in terms of four primary processes and various supporting processes. The four primary processes are (1) theorizing, (2) questioning and hypothesizing, (3) investigating, and

(4) analyzing and synthesizing the findings. The supporting processes include cognitive, social, and metacognitive processes (see White & Frederiksen, 2005), including regulatory processes that guide scientific inquiry (see White et al., 2009). We developed a standard form for characterizing each of the four primary processes in terms of five elements: (1) the different types, (2) the purposes, (3) the creation process, (4) the criteria for evaluation, and (5) the synthesis of the different types. This is the basic structure of the framework.

With respect to theorizing, we discussed three different epistemic forms or model types: (1) structural models, such as a stage model, (2) causal models, such as a multifactor model, and (3) process models, such as an agent model. We illustrated each of these types in terms of simplified models of friendship. Stage models show how events unfold over time. Multifactor models specify all the factors that influence a particular dependent variable. Agent models allow one to simulate the process of interaction among a set of different actors. Different types of models can serve to embody different parts of an integrated theory, such as a theory of friendship.

Generating research questions is the process that bridges from theory to designing an investigation. Different types of models generate different kinds of research questions. For example, stage models ask how many different phases or stages there are, what are the characteristics of each stage, and what leads to the transition from each stage to the next. Multifactor models raise questions about what factors affect the dependent variable, how the factors are causally connected to the dependent variable, and how the factors combine to affect the dependent variable. Hence, research questions are linked together by the type of model being developed and investigated. Hypotheses constitute the possible answers to the research questions posed and may be derived from competing models and theories.

Investigations can either be exploratory or confirmatory. There are many ways to conduct exploratory studies, such as by carrying out discourse or protocol studies or studies using observational, interview, and survey methods. There are also different forms that confirmatory investigations take, such as randomized controlled trials, where hypotheses are tested by assigning subjects randomly to different conditions and then comparing how the subjects perform. The kind of data collected has a strong effect on the types of models that can be constructed from the data. For example, quantitative data support construction of multifactor models, as in the Framingham Heart Study. To construct process models, such as agent models, one needs a richer data stream, such as observational, protocol, or discourse studies provide. These models can then be evaluated using confirmatory methods.

Analysis and synthesis use data from an investigation to arbitrate between competing models and to develop new models. There are four aspects to this process: (1) coding and representing data to reveal patterns, (2) interpreting the patterns with respect to competing hypotheses, (3) exploring the data to induce new models, and (4) determining the generality of the findings. With qualitative data, a coding scheme helps to create useful data representations. In order to see patterns in the data, there are a variety of representational tools, such as graphs, charts, and computer-based visualization tools. Statistical tests can be applied to help determine relationships in the data, as well as to determine the generalizability of the findings. Often surprising patterns may be found in the data, which suggest revisions to models or lead to new theoretical models, further investigations, and more data analyses.

Teaching Scientific Meta-knowledge

To develop and demonstrate an understanding of the four kinds of scientific meta-knowledge that we argue are needed, students would need to engage in an extensive inquiry curriculum, one aimed at developing and testing models of different types, which would fit together to create a scientific theory for a given domain. For each study they undertake, students would need to consider which type of model they want to develop, which type of question they should ask, which type of investigation they need to carry out, and which data analysis techniques they should employ. Being able to make such decisions presupposes that they know about different model types, question types, investigation methods, and data analysis techniques and, further, that they know how these work together to create and test scientific theories.

While our Inquiry Island and Web of Inquiry software embody much of the meta-knowledge we have described in this chapter, our own research program has provided limited opportunities for determining how best to teach such meta-level expertise to young students (i.e., upper elementary and middle school students). In one attempt to develop students' meta-modeling knowledge, for example, we designed a version of our ThinkerTools Inquiry Curriculum that focuses on enabling students to learn about different types of scientific models and their utility (Schwarz & White, 2005). The results were encouraging, though the impact of the curriculum was strongest on the higher achieving students.

In most of our recent work (Frederiksen et al., 2008; White & Frederiksen, 2005), the teachers we collaborate with only commit to having their students engage in one or two major research projects per year. Given this limited time commitment, we have focused on enabling students to develop a particular type of theory, which usually consists of a causal model (often multifactor) linked with some preliminary models of underlying mechanisms. To evaluate their theories, students design and carry out a particular type of investigation, usually a controlled comparison designed to test their competing hypotheses. They then undertake a limited range of data analyses, typically centering around a comparison of averages and frequency distributions of data from the various experimental conditions. Thus, although the students learn a great deal about scientific inquiry and produce some impressive research projects (Frederiksen et al., 2008), the restricted instructional time and number of investigations students undertake limit the development of their meta-level expertise.

Creating curricula that adequately teach scientific meta-knowledge would require a much more extensive curricular sequence that would give students opportunities to learn more about different types of models, investigations, and analyses, as well as how all of these components of inquiry work together, throughout a series of investigations, to create comprehensive scientific theories. Engaging in such a curriculum would thus necessitate a serious time commitment on the part of schools and teachers. Teachers, along with those who develop curricular standards and accountability measures for science education, need to be convinced of the importance of scientific meta-knowledge.

Concluding Thoughts About the Utility of the Framework

A major benefit of the meta-knowledge framework described in this chapter is to increase the awareness of students, teachers, and scientists as to different types of models, questions, investigations, and analyses. If they have a toolkit that includes a wide variety of model types, they may construct richer theories by building multiple models of different types and linking them together. Similarly if researchers learn how to use different investigation methods, they can combine these methods to triangulate on the phenomena they are trying to understand and hence produce more robust results. The framework provides guidance to make informed decisions as to what to do at different points in the inquiry process. Our basic argument is that scientific meta-awareness provides insights for scientists and students that will enhance their ability to learn and understand scientific inquiry.

One long-term goal of our work has been to develop computer-based tools that support scientists and students in their work. For example, we think a modeling tool could help scientists construct many different kinds of models, just as Stella (High Performance Systems, 2000) helps students construct system-dynamics models and Net Logo (Wilensky, 1999) and Agent Sheets (Repenning, Ioannidou, & Zola, 2000) help students construct agent models. Such a tool could help scientists and students consider different types of models they might construct and provide a structure, which guides the development of each type of model.

Inquiry Island and the Web of Inquiry provide another kind of computer-based tool that can support scientists and students in their research. Inquiry Island has extensive advice about the issues people should address as they construct models and theories, formulate research questions and hypotheses, carry out investigations, and analyze and synthesize the data they collect. This advice is available when people ask for it or when they appear to need it. It can be modified, in the Web of Inquiry, for particular types of inquiries or for particular groups of people. The goal is to refine Inquiry Island and the Web of Inquiry as general-purpose tools that can support scientific inquiry at many different levels.

Clearly more work remains to be done. This chapter only presents a top-level view of the kind of meta-knowledge we think is central to scientific inquiry. For example, there are many different kinds of model types or epistemic forms (Collins & Ferguson, 1993) that scientists and students might learn to use. In fact, artificial intelligence has led to a proliferation of different model types in its short history, such as production-system models, agent models, constraint-satisfaction models, semantic-network models, frame-system models, qualitative-process models. This proliferation of model types provides new power for making sense of diverse phenomena in the world. Understanding the affordances and constraints in building models of these different types should become a learning goal for future scientists.

Such work on meta-scientific knowledge should thus have benefits that go beyond science education. We think that our framework provides a structure that will enable science to refine its practices and products. Making the meta-knowledge underlying scientific processes explicit fosters systematic reflection by scientists and the field as a whole. This reflection enhances the possibility that a self-improving system will take hold to develop better tools, models, and representational forms. Hence we argue that the development of an explicit understanding of the nature of scientific meta-knowledge will lead to a more productive scientific enterprise.

Acknowledgments This research was funded in part by the National Science Foundation (grants MDR-9154433, REC-0087583, and REC-0337753). We thank the members of our research team for their contributions to this work, as well as the students and teachers who participated in our studies. The views expressed in this chapter are those of the authors and do not necessarily reflect those of the National Science Foundation.

References

American Association for the Advancement of Science . (1990). *Science for all Americans: Project 2016*. New York: Oxford University Press.

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, *13*(1), 1–12.

Bell, P. & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, *22*, 797–817.

Bielaczyc, K. & Collins, A. (2000). Learning communities in classrooms: A reconceptualization of educational practice. In C. M. Reigeluth, (Ed.), *Instructional design theories and models, Vol. II*. Mahwah, NJ: Erlbaum.

Borge, M. (2007). *Regulating social interactions: Developing a functional theory of collaboration*. Doctoral dissertation, Berkeley, CA : University of California at Berkeley.

Brown, A. & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, *28*(3), 235–251.

Chinn, C. A. & Malhotra, B. A. (2002). Epistemologically authentic reasoning in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175–218.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies for guiding inquiry. *Educational Psychologist*, *28*(1), 25–42.

Dewey, J. (1910). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: D. C. Heath.

diSessa, A. A. (2002a). *Changing minds: Computers, learning, and literacy*. Cambridge, MA: MIT Press.

diSessa, A. A. (2002b). Students' criteria for representational adequacy. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel, (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 105–129). Dordrecht: Kluwer.

diSessa, A. A. (2004). Metarepresentation: Naïve competence and targets for instruction. *Cognition and Instruction*, *22*(3), 293–331.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Dunbar, K. (1999). How scientists build models: InVivo science as a window on the scientific mind. In L. Magnani, N. Nersessian, & P. Thagard, (Eds.), *Model-based reasoning in scientific discovery* (pp. 85–99). New York: Kluwer.

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology*, *21*(1), 49–58.

Duschl, R. (2007). Quality argumentation and epistemic criteria. In M. Jiménez-Aleixandre & S. Erduran, (Eds), *Argumentation in science education: Perspectives from classroom-based research*. Netherlands: Springer.

Duschl, R. & Osborne, J. (2002). Supporting and promoting argumentation discourse. *Studies in Science Education*, *38*, 39–72.

Eslinger, E., White, B., Frederiksen, J., & Brobst, J. (2008). Supporting inquiry processes with an interactive learning environment: Inquiry Island. *Journal of Science Education and Technology*, *17*, 6.

Frederiksen, J. R. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27–32.

Frederiksen, J. R. & White, B. Y. (2002). Conceptualizing and constructing linked models: Creating coherence in complex knowledge systems. In P. Brna, M. Baker, K. Stenning, & A. Tiberghien (Eds.), *The role of communication in learning to model* (pp. 69–96). Mahwah, NJ: Erlbaum.

Frederiksen, J. R., White, B. Y., & Gutwill, J. (1999). Dynamic mental models in learning science: The importance of constructing derivational linkages among models. *Journal of Research in Science Teaching*, *36*(7), 806–836.

Frederiksen, J. R., White, B. Y., Li, M., Herrenkohl, L. R., & Shimoda, T. (2008). *Classroom formative assessment: Investigating models for evaluating the learning of scientific inquiry* (Final report to the National Science Foundation). University of Washington: available from the authors.

Giere, R. (1991). *Understanding scientific reasoning* (3rd ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

R. Giere, (ed.). (1992). *Cognitive models of science*. Minneapolis, MN: University of Minnesota Press.

Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, *28*(1), 73–79.

Grotzer, T. A. (2003). Learning to understand the forms of causality implicit in scientific explanations. *Studies in Science Education*, *39*, 1–74.

Halloun, I. A. (2004). *Modeling theory in science education*. Netherlands: Springer.

Hammer, D., Russ, R., Mileska, J., & Scherr, R. (2008). Identifying inquiry and conceptualizing students' abilities. In R. Duschl & R. Grandy (Eds.) , *Establishing a consensus agenda for K-12 science inquiry*. Rotterdam, Netherlands: Sense.

Herrenkohl, L. R., Palinscar, A., Dewater, L., & Kawasaki, K. (1999). Developing scientific communities in classrooms: A sociocognitive approach. *Journal of the Learning Sciences*, *8*(3&4), 451–493.

Hestenes, D. (1987). Toward a modeling theory of physics knowledge. *American Journal of Physics*, *55*(5), 440–454.

High Performance Systems. (2000). *Stella 6.* [software retrievable from http://www.iseesystems.com/. Accessed 5 January, 2011]. Hanover, NH: High Performance Systems.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*(1), 127–138.

Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching*, *36*(10), 1085–1109.

Klahr, D. & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, *125*(5), 524–543.

Krathwohl, D. (1998). *Educational and social science research: An integrated approach* (2nd ed.*)*. New York: Longman.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*, 319–337.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495–523.

Lederman, N. G. (2007). Nature of science: past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–880). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R. & Schauble, L. (2005). Developing modeling and argument in elementary grades. In T. A. Romberg, T. P. Carpenter, & F. Dremock, (Eds.), *Understanding mathematics and science matters*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R. & Shauble, L. (2000). Modeling in mathematics and science. In R. Glaser, (Ed.), *Advances in instructional psychology, Vol. 5. Educational design and cognitive science (*pp. 101–159). Mahwah, NJ: Erlbaum.

Mandinach, E. B., & Cline, H. F. (1994). *Classroom dynamics: Implementing a technology-based learning environment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mellar, H., Bliss, J., Boohan, R., Ogborn, J., & Tompsett, C. (Eds.). (1994). *Learning with artificial worlds: Computer based modeling in the curriculum*. Washington, DC: The Falmer Press.

Metz, K. E. (2000). Young children's inquiry in biology: Building the knowledge bases to empower independent inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 371–404). Washington, DC: American Association for the Advancement of Science.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council (2007). In R. Duschl, A. Schweingruber, & A. Shouse (Eds.), *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academy Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Osborne, J. (2005). The role of argument in science education. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof, (Eds.), *Research and the quality of science education*. Netherlands: Springer.

Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of causal understanding: The role of complex causal models in students' understanding of science. *Studies in Science Education*, *41*(1), 117–165.

Popper, K. (1963). *Conjectures and refutations*. London: Routledge and Kegan Paul.

Repenning, A., Ioannidou, A., & Zola, J. (2000). AgentSheets: End-user programmable simulations. *Journal of Artificial Societies and Social Simulation*, *3*, 3.

Russell, S. J. & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice-Hall.

Sandoval, W. A. & Reiser, B. J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*(3), 345–372.

Scardamalia, M. & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265–283.

Schwarz, C. & White, B. (2005). Meta-modeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, *23*(2), 165–205.

Shimoda, T., White, B., & Frederiksen, J. (2002). Student reflective inquiry: Increasing levels of agency through modifiable software advisors. *Science Education*, *86*, 244–263.

Slotta, J., & Chi, M. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, *24*(2), 261–289.

Smith, C., Maclin, D., Houghton, C., & Hennessey, M. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction*, *18*(3), 349–422.

Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction*, *9*, 221–283.

Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. In M. S. Donovan, & J. D. Bransford, (Eds.), *How students learn* (pp. 515–565). Washington, DC: National Research Council.

Suthers, D., & Weiner, A. (1995). Groupware for developing critical discussion skills. In J. L. Schnase, & E. L. Cunnius, (Eds.), *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning* (pp. 341–348). Mahwah, NJ: Lawrence Erlbaum Associates.

White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, *10*(1), 1–100.

White, B. Y., & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, *42*, 99–157.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118.

White, B. Y. & Frederiksen, J. R. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, *40*(4), 211–223.

White, B., Frederiksen, J., & Collins, A. (2009). The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. Hacker, J. Dunlosky, & A. Graesser, (Eds.), *Handbook of metacognition in education*. New York: Routledge.

White, B., Frederiksen, J., Frederiksen, T., Eslinger, E., Loper, S., & Collins, A. (2003). Inquiry Island: Affordances of a multi-agent environment for scientific inquiry and reflective learning. In P. Bell, R. Stevens, & T. Satwicz, (Eds.), *Proceedings of the Fifth International Conference of the Learning Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

White, B., Shimoda, T., & Frederiksen, J. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: Computer support for metacognitive development. *International Journal of Artificial Intelligence in Education*, *10*(2), 151–182.

Wilensky, U. (1999). NetLogo. [software retrievable from http://ccl.northwestern.edu/netlogo/. Accessed 5 January, 2011]. Center for connected learning and computer' based modeling. Northwestern University.

Wilensky, U. & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3–19.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, *92*(5), 941–967.