

The Role of Domain-Specific Knowledge in Promoting Generative Reasoning in Genetics

Ravit Golan Duncan, Rutgers University, Graduate School of Education 10 Seminary Place New Brunswick NJ 08901, rgduncan@rci.rutgers.edu

Abstract: Promoting the ability to reason generatively about novel phenomena and problems students may encounter in their everyday lives is a major goal of science education. This goal proves to be a formidable challenge in domains, such as molecular genetics, for which the accumulated scientific understandings are daunting in both amount and complexity. To develop effective instruction that fosters generative reasoning we need to have a sound understanding of the types of knowledge in the domain that are critical for such reasoning. In this study I examined the ensemble of knowledge undergraduate students employed in explaining genetic phenomena and solving problems about them. I characterized two key knowledge types: *domain-specific heuristics* and *domain-specific explanatory schemas* that are crucial for understanding the dynamics and entities central to molecular genetics phenomena. I then developed a cognitive model that highlights the role of these powerful conceptual understandings in promoting generative reasoning in genetics.

Introduction

One of the main challenges of science education is helping students become scientifically literate (American Association for Advancement of Science, 1993; National Research Council, 1996). It is clear that a crucial aspect of such literacy is that students' understandings need to be *generative*. That is, we want students to be able to reason not only about scientific phenomena currently taught in class, but also about problems and phenomena they are likely to encounter outside of the classroom and throughout their lives. The challenge of fostering generative understandings is particularly acute for domains in which scientific advances are rapid, phenomena are complicated, and the amount of information accumulated is daunting. One compelling example of such a domain is molecular genetics. Over the past several decades there have been numerous advances, both scientific and technological, in this domain (sequencing of the human genome, genetic screening, and stem cell research, to name a few) and the knowledge base has grown tremendously. Given that we cannot possibly teach students everything, we need to make decisions regarding what to include in the curriculum in order to foster understandings that are generative. Furthermore, it is imperative that these decisions be informed by a thorough understanding of the types of knowledge that are important for such generative reasoning in genetics.

The study I describe herein examined the ensemble of knowledge students employed in reasoning about genetic phenomena and the roles that different knowledge types played in promoting generative understandings in this domain. Towards this end I conducted several clinical interviews with nine undergraduate students enrolled in an introductory molecular genetics course in a university in the Midwest. The rich data I obtained of students' reasoning about and explanations of the mechanisms underlying molecular genetic phenomena allowed me to carry out a fine-grained analysis of the knowledge types and reasoning patterns involved. Specifically, three questions guided my analysis:

1. What is the nature, in terms of defining form and content where possible, of the different knowledge types students used to reason in the domain?
2. What are typical interactions between these domain-specific knowledge types in episodes of reasoning and problem solving in genetics?
3. What is the role and contribution of these knowledge types to generative reasoning?

The third question, which builds upon the first two, strives towards important instructional implications. If we can identify the knowledge elements that are most productive for generative reasoning in this domain we can attempt to design instruction to foster the construction of that knowledge.

Understanding Molecular Genetics

Molecular geneticists strive to explain how genes bring about their effects. While often the effects of genes can be observed at the organ and organism level, molecular geneticists are primarily concerned with the underlying

causal mechanisms at the cellular and molecular level. For the purpose of such explanations, it is critical to understand that the genetic code does not directly specify the observable feature rather it specifies the first step in a chain of events that results in the observable feature. More specifically, the information embodied in genes specifies the composition of a central molecule in molecular biology: the protein. Simply put, each gene specifies the structure of one protein; the protein in turn affects the structure and function of the cell, which affects the function of the tissue, etc, resulting in the observable effect (i.e., symptoms of a genetic disease). Understanding this chain of events is at the crux of explanations in molecular genetics.

In contrast to the blossoming of scientific knowledge in genetics noted earlier, the state of affairs in terms of students' understanding of genetics is rather bleak (Golan & Reiser, 2003; Marbach-Ad & Stavy, 2000; Marbach-Ad, 2001; Venville & Treagust, 1998). Students are not prepared to understand and benefit from technological advances in genetics that they may encounter in their everyday lives, such as genetic counseling and screening (Lewis & Wood-Robinson, 2000); nor are students informed enough to understand and participate in current debates involving genetic issues, such as genetically modified foods, stem cell research, and gene therapy (Garton, 1992; Kindfield, 1992).

Why is genetics reasoning so challenging? Prior research attributes students' difficulties to the complicated structure of genetic phenomena, namely, the multiplicity of levels involved and the invisible nature of the entities and process within them (Horwitz, 1996; Marbach-Ad & Stavy, 2000; Stewart & Van Kirk, 1990). Phenomena in genetics transverse multiple organization levels: genes, proteins, cells, tissues, organs, etc. These levels are organized hierarchically such that elements at one organization level constitute the components of progressively higher organization levels (for example, cells constitute tissues that constitute organs). In prior research I have argued that it is not only the number, organizational structure and scale of the levels that contribute to students reasoning difficulties but also the nature of the levels and the relationship between them. In my work I characterize the levels of genetic phenomena as containing entities that are from very different conceptual ontologies. In a sense there are two ontologically distinct levels in genetic phenomena: an information level containing the genetic information (our genes), and a hierarchically organized biophysical level containing proteins, cells, tissues etc. Therefore, genetic phenomena are especially complicated in that they are driven by an information level (the genetic code) but are manifested through a myriad of biophysical entities. This characterization of genetic phenomena as composed of two ontologically distinct levels provides an analytical lens that highlights aspects of the genetic system that are likely to be conceptually critical for generative reasoning. Namely, understanding how the information level brings about effects of a physical nature. Moreover, it is precisely this understanding of the mechanisms through which genes bring about their effects that is at the crux of scientific literacy in this domain (AAAS, 1993; NRC, 1996).

Methods

I used clinical interview methodology to elicit student thinking about molecular genetics phenomena and to analyze the knowledge resources employed in such thinking. I conducted the study in the context of an introductory biology course at the undergraduate level. Instruction in this course was typical of college level teaching: predominantly lecture format using PowerPoint presentations as visual aids. The study participants, nine undergraduate students were mostly sophomores who were taking the course as a requirement for pre-medical education or biology major. The students varied in the extent to which they have studied molecular genetics in the past. These students participated in weekly hour-long interviews throughout the course of relevant instruction (five weeks). The interviews were composed of three kinds of tasks: 1) Questions about genetic phenomena in terms of components and dynamics (these concepts were taught in class)- for example, I asked students: What genes code for? What are proteins and what do they do? 2) Hypothetical tasks in which I gave students a description of a genetic phenomenon at the observable level and asked them to reason about the underlying molecular mechanisms and to speculate about how the information encoded in genes brings about the observable phenomenon. I chose phenomena that students did not study in the class in order to see their ability to reason about genetic phenomena at-large as opposed to their ability to recall a taught example; 3) The remaining tasks were taken from examples taught in class. I showed students the PowerPoint slides used during lecture and asked the to explain the mechanism discussed.

My data corpus consisted of video transcripts of 21 interview tasks that I analyzed in two complementary and sequential steps. The first was a content analysis distinguishing between topics of the students' comments. In this analysis, I segmented the transcripts based on the level of specificity (specific to molecular genetics, specific to biology, or general and applicable across domains) and the content of ideas expressed (ideas about the genetic

information, proteins, etc). The content analysis set the stage for the next analytical step that was based on DiSessa (1996) and Sherin's (2001) versions of knowledge analysis. In this analysis I identified and characterized the sorts of knowledge representations that seemed to underlie students' reasoning behavior, particularly in content analysis segments classified as being specific to molecular genetics. I examined the frequency of use of the different knowledge types identified (Chi, 1997) and their role in promoting generative reasoning about novel and unfamiliar phenomena (hypothetical tasks). This analytical process was iterative in that the coding categories were defined in part by existing knowledge forms described in the literature (mental models, schemas, cases) and in part emerged from the data and/or were refined to capture the different aspects of students' knowledge evident in the data (for example, the domain-specific explanatory schemas, and the domain-specific heuristics). I then constructed a cognitive model of reasoning in molecular genetics showing the different knowledge types associated with understanding the mapping between the information and biophysical levels and the mechanisms involved in mediating genetic effects.

Results and Discussion

As noted earlier a crucial aspect of generative reasoning about molecular genetics phenomena is the understanding of what the informational content specifies and how it is mediated by physical elements in the system. The focus of this paper is the nature and interactions of the domain-specific conceptual resources that afford such reasoning. I identified two domain-specific knowledge forms that played a key role in such reasoning: *domain-specific heuristics* and *domain-specific explanatory schemas*. The domain-specific heuristics are a form of principles that are applicable to a wide variety of phenomena in the domain. Their role in reasoning is to constrain and guide reasoning by suggesting domain-appropriate entities and relationships. These heuristics can guide actions in the sense that they direct the construction of the explanation core. For example, invoking the *genes-code-for-proteins* heuristic prompts the reasoner to identify or suggest a relevant protein (or look for an entity that might be a protein), given that a gene is involved in the phenomenon, thereby adding the protein element to the explanation.

The domain-specific explanatory schemas are similar to the heuristics in that they also define important aspects of genetic phenomena (and explanations about them), namely prevalent mechanisms that mediate genetic effects. For example, the explanatory schema *inhibit* embodies the understanding of a very important mechanism in molecular biology: inhibition. The explanatory schemas are more specific, or detailed, in comparison to the heuristics and are constrained by them. Thus, values taken by the various slots in the explanatory schemas must conform to the general principles embodied in the heuristics. In the following sections I shall discuss each of these knowledge forms in turn, provide some illustrative examples of their use and conclude with a cognitive model of reasoning with these knowledge types in molecular genetics.

Domain-Specific Heuristics

I identified three domain-specific heuristics in the data corpus: *genes-code-for-proteins*, *proteins-as-central*, and *effects-through-interaction*. The *genes-code-for-proteins* heuristic links the information and physical levels by denoting the relationship between the genetic information and physical entities in the system. The *proteins-as-central* heuristic emphasizes the central role of proteins in genetic phenomena and guides the reasoner to assume that most molecular phenomena involve proteins. It is often invoked along with the *genes-code-for-proteins* heuristic and seems to have a synergistic effect, strengthening the guidance of the former. The *effects-through-interaction* heuristic defines physical interaction between molecules as the key mechanism by which effects are brought about in molecular genetics (protein-protein, DNA-protein, protein-substrate, etc). That is, in order to function and exert their effects, components in the system (often proteins) need to physically interact with each other. The results of these interactions are defined by the explanatory schemas, yet the notion that physical interaction is necessary is embodied in the *effects-through-interaction* heuristic. Taken together these heuristics: 1) constrain the nature of the genetic information to being about proteins; thus, defining the relationship between the information and physical levels of genetic phenomena, and 2) posit proteins as being the central mediating (physical) agents of genetic effects, and 3) constrain the mechanisms of action in the system to being interaction-based. Due to space limitations I shall only elaborate and illustrate the use of the *genes-code-for-proteins* heuristic.

The *genes-code-for-proteins* knowledge form serves to restrict the informational content of genes by specifying that genes contain information specifically about proteins (as opposed to cells, organs, traits, etc). This allows the reasoner to safely assume that if a gene influences a phenomenon, it is highly likely that the phenomenon is mediated by a protein; even in cases in which the specifics are unknown. The following excerpt illustrates the use of this knowledge form. Here Vera is reasoning about genetic antibiotic resistance in bacteria (antibiotic resistance

was mentioned in class but the mechanism was not explained); initially she provided a very general solution claiming that the gene confers resistance through some unknown protection mechanism. When I pressed her to suggest a specific mechanism for her general solution she invoked her heuristic of *genes-code-for-proteins* and inferred that the “protective” gene likely codes for a protein. With more prompting she suggested that the protein is involved in making a coat that protects the bacteria (this coat presumably prevents the antibiotic from entering the bacteria).

- Interviewer: How does this ampicillin resistance gene give the bacteria resistance, what is it doing?
- Vera: Umm, I have no idea. I don't really know the mechanism for how it protects it; I just know that if it has the gene it protects it.
- Interviewer: Could you speculate what a potential mechanism might be like?
- Vera: Let's see, maybe if you have it, then you create some proteins...
- Interviewer: Ok, so this would be coding for a protein?
- Vera: Yeah, I think so. So maybe you make some proteins that umm..... maybe it makes a protein that, I don't know, makes up the coating for the bacteria, I don't know.

Her solution was based on the understanding that genes code for proteins and thus a protein must be involved in the protection mechanism. This is a significant step in reasoning about molecular genetics phenomena. In previous work I conducted with high school students, none of the students provided protein-based mechanisms (despite learning that genes code for proteins) to explain genetic phenomena. The idea that the coat is protein-based (either made of proteins or made by proteins) is not trivial or obvious. Rather, it indicates an understanding of genetic phenomena as being mediated by proteins.

Note that Vera does not know exactly how the resistance mechanism works or which protein her speculative bacterial coat is really made of; however, in invoking the *genes-code-for-proteins* heuristic, she is able to conjecture that a protein must be involved. This suggests that domain-specific heuristics are not necessarily a product of detailed understandings of the elements and mechanisms they refer to. Thus, the strength of this and other heuristics lie in their ability to constrain the space of reasoning in a productive and powerful way independently of detailed knowledge of the phenomenon in question. This is an important point because it exemplifies the pivotal role of these knowledge forms in generative reasoning- these heuristics afford reasoning about novel contexts in genetics.

Domain-Specific Explanatory Schemas

The domain-specific heuristics discussed above restrict the space of reasoning or problem solving in a very powerful yet general manner. They do not dictate a mechanism; they merely restrict the realm of possible mechanisms to those involving proteins. I argue that further specification is provided by another prominent and important conceptual resource: the domain-specific explanatory schemas. This knowledge is of schema form (Rumelhart, 1980) and provides the reasoner with an answer to the question of *how*: how do proteins confer resistance to antibiotics, how do proteins cause disease, how do proteins regulate cell growth, and how is gene expression regulated? Therefore, it is the schemas that provide the majority of domain-appropriate mechanisms and therefore play a large role in the construction of domain-appropriate explanations. Moreover, it seems likely that the acquisition of more schemas and perhaps refinement and elaboration of existing schemas is an important process in the development of expertise. I identified ten explanatory schemas in the data: *inhibit*, *activate*, *allosteric regulation*, *translation*, *regulation-of-gene-expression*, *catalyze*, *transport*, *sense/detect*, *build*, and *structure-function*. Due to space limitations I shall elaborate on and illustrate only the *inhibit* explanatory schema.

This schema specifies a mechanism by which one biological entity prevents another entity from carrying out its function. The *inhibit* schema has three main slots: the inhibited element, the inhibiting element, and the mechanism of inhibition. The domain-specific heuristic of *proteins-as-central* constrains the type of variables that can fill these slots. Thus, at least one of these slots is a protein, usually the inhibited element (often an enzyme protein). The *effects-through-interaction* heuristic constrains the mechanism of inhibition to one that involves the binding of the inhibitor to a specific site on the inhibited element (often the active site of an enzyme; a type of inhibition termed competitive). I illustrate the use of this knowledge form with an example taken from a task in which I asked students to explain how virulence genes (found in disease-causing bacteria) cause a disease in the host organism (a mouse in this case). In this example, Kim speculated that the virulence gene codes for an inhibitor that

inhibits a critical biological process in the mouse. She invoked the *inhibit* schema and has rather elaborate knowledge of the mechanisms of inhibition (types of inhibition). In her schema, the inhibited slot is a protein (enzyme), and the inhibitor slot contains a substance that she does not define. We do not know if the inhibitor is a protein or not, albeit since Kim has the *genes-code-for proteins* heuristic (as evidenced by her responses to other interview tasks), it is likely that she is assuming the inhibitor is a protein as well.

- Interviewer: What are these genes?
 Kim: Well, genes that just kill the mouse.
 Interviewer: What would it code for that kills the mouse?
 Kim: Umm, well, it could probably code for some kind of inhibitor in its own metabolic cycle, maybe it could inhibit one of the enzymes that creates critical metabolite in the mouse.
 Interviewer: What is inhibition about?
 Kim: Well, inhibition is where you don't let it perform the way it performs. And, umm, there are different kinds of inhibitors though. It could be like competitively binding to it or binding somewhere else allosterically.
 Interviewer: What do you mean by somewhere else?
 Kim: Well, umm, enzymes have active sites and you could bind to the active site of the enzyme and not let substrate that is supposed to bind there bind, or you could bind to different location like an allosteric site, which is like on the other side of the enzyme and changing its conformation so it doesn't let anything else bind in active site.

Kim's *inhibit* schema is rather elaborate; she has several potential methods of inhibition (competitive or allosteric) and seems to know quite a bit about how those types of inhibitory mechanisms operate. In many cases when students invoked a particular explanatory schema there was ancillary knowledge associated with it; in a sense, the schema was annotated with additional information. Mostly, this knowledge was in the form of propositions (declarative knowledge) and mental models (Gentner & Stevens, 1984). This knowledge helped in the elaboration of explanations and often made them more accurate (or at least more detailed). However, this ancillary knowledge is not an essential part of the schema since some students did not seem to have it and yet were able apply the inhibit schema successfully. It seems that knowing that the mechanism of inhibition is mediated by direct binding of the inhibitor to the inhibited is the critical part of the schema. Any elaboration on this mechanism allows the student to provide a more detailed explanation but does not change the essence of it. Thus, overall the explanatory schemas provide the majority of domain-specific mechanisms needed for explaining how genes bring about their effects.

Before I present the cognitive model I would like sharpen somewhat the distinction between the heuristics and the explanatory schemas. If we assume that knowledge representations can differ in three respects— form, content, and function—then these knowledge forms differ primarily with respect to function and somewhat with respect to form. In terms of form, the heuristics are principle-like abstractions. It may be the case that they are very simple schemas. The explanatory schemas, on the other hand, are more elaborate schemas, and they can be associated with mental model or propositional knowledge that include details of particular instantiations of the schema. In terms of function, the heuristics are most similar to production rules (Anderson, 1987), because they help identify important entities and the relationships between them. The schemas are most similar to scripts (Schank, 1982) that define the event of a particular mechanism in action. Thus, the heuristics function as broad-stroke guiding principles, whereas the schemas are more detailed instantiations of domain-specific mechanisms.

Cognitive Model

My goal here is to present a conceptual model of reasoning in genetics at an abstract enough level for it to be informative for other domains. By discussing some of the general features of the knowledge types and reasoning I observed in the data, I hope to give the reader a sense of what comparable knowledge types in other domains might look like. The model I present, illustrated in Figure 1, depicts the types of knowledge forms that played a key role in students' reasoning, their function in the knowledge system, and the relationships between them. I also highlight the level of specificity at which the knowledge forms exist in order to illustrate the relative contribution of more and less contextualized knowledge.

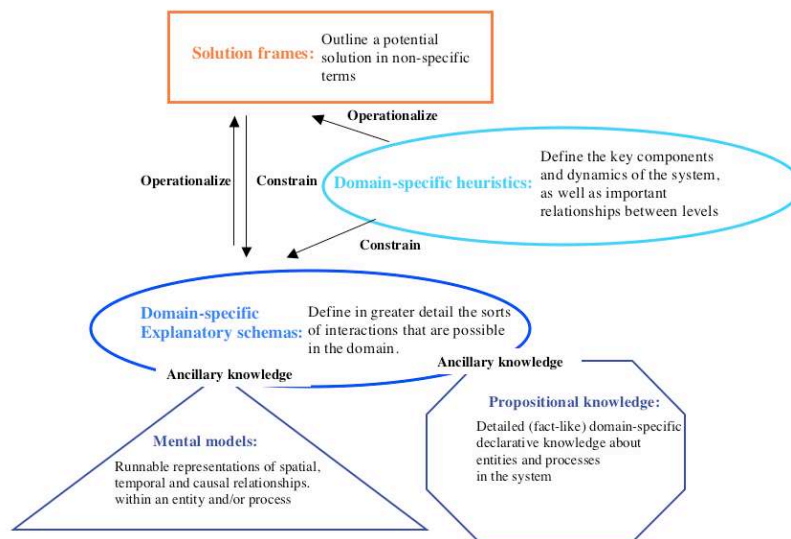


Figure 1. Conceptual Model of Reasoning About Genetic Phenomena.

A prominent reasoning pattern I observed in the data (illustrated in Vera’s reasoning) was a progression from the general to the specific when constructing an explanation. Students often began by offering a general outline of a potential solution to the problem- a solution frame. These frames served as placeholders to be filled with domain-specific knowledge. The domain-specific heuristics and explanatory schemas provided that knowledge. The heuristics are principle-like knowledge forms that define key entities and relations in the domain, such as proteins, genes, and the coding relationship between them; thus constraining the space of reasoning. The schemas in turn define more specific mechanisms and are themselves constrained by the heuristics. That is, the specific variables that the schemas can take on are consonant with entities and relationships defined by the heuristics (for example different types of interactions between proteins). Often, these schemas are annotated with ancillary knowledge (either mental model or fact-like) that adds accuracy or case-specificity to the schema.

Together the heuristics and schemas populate the space of reasoning with domain-appropriate entities and mechanisms. Their power and utility stems from their generalizability *within* the domain and their relative independence of context specific details. These forms can be used to reason about a myriad of phenomena and problems in molecular genetics and are in essence a tool kit for generative reasoning. It is plausible that analogous knowledge forms may be critical for generative understandings in related domains that feature complicated systems and the dynamics within them- the domain of ecology with the focus on ecosystems would make a good candidate for such a search.

Conclusions and Implications

In the research presented herein, I have tried to characterize the knowledge resources used in reasoning about molecular genetics. I have found that while both general and domain-specific knowledge forms play a role in such reasoning, it is a crucial type of domain-specific knowledge that seems to drive reasoning by constraining the problem-space and defining key components and mechanisms in the domain. These knowledge forms support generative reasoning in that they are widely applicable within the domain and embody key understandings in it. Moreover, there appears to be a rather small set of domain-specific heuristics and explanatory schemas that comprise this critical body of knowledge. From an instructional perspective this is very encouraging as it suggests that teaching for generative understandings may be an attainable goal despite the time constraints of the classroom. While this cognitive model is not a prescription for instruction it nonetheless highlights what students need to have in their conceptual toolkit in order to reason generatively in genetics and it suggests a focus for genetics instruction.

References

- American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy*, New York: Oxford University Press.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-methods problem solutions. *Psychological Review*, 94(2), 192-219.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the learning sciences*, 6(3), 271-315.
- diSessa, A. A. (1996). What do “just plain folk” know about physics? In D. R. Olson & N. Torrance (Eds.), *The handbook of education and human development: New models of learning, teaching, and schooling*. Cambridge, MA: Blackwell.
- Garton, G. L. (1992). Teaching genetics in the high school classroom. In M.U. Smith & P.E. Simmons (Eds.), *Teaching genetics: Recommendations and research proceedings of a national conference* (pp. 20-23). Cambridge, MA: National Science Foundation.
- Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Golan, R., & Reiser, B. J. (2003). Students’ reasoning about phenomena generated by complex systems: The case of molecular genetics. In Zandvliet, D.B., and Anderson, C. (Eds.), *Proceedings of the 2003 Annual General Meeting of the National Association for Research in Science Teaching*. Philadelphia, PA.
- Horwitz, P. (1996). Teaching science at multiple space time scales. *Communications of the ACM*, 39(8), 100-102.
- Kindfield, A. C. H. (1992). Teaching genetics: recommendations and research. In M. U. Smith, and P. Simmons (Eds.), *Teaching genetics: recommendations and research. Proceedings of a National Conference* (pp. 39-43). Cambridge.
- Lewis, J., & Wood-Robinson, C. (2000). Genes, chromosomes, cell division and inheritance -- do students see any relationship? *International Journal of Science Education*, 22, 177 - 195.
- Marbach-Ad, G. (2001). Attempting to Break the Code in Student Comprehension of Genetic Concepts. *Journal of Biological Education*, 35(4), 183-189.
- Marbach-Ad, G., & Stavy, R. (2000). Students cellular and molecular explanations of genetic phenomena. *Journal of Biological Education*, 34(4), 200-210.
- National Committee on Science Education Standards and Assessment (1996) *National Education Standards*, Washington, DC: National Academy Press.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce and W. F. Brewer (Ed.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum.
- Schank, R.C. (1982). *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge University Press.
- Sherin, B. L. (2001). How students understand physics equations. *Cognition and Instruction*, 19(4), 479-541.
- Stewart, J., & Van Kirk, J. (1990). Understanding and problem-solving in classical genetics. *International Journal of Science Education*, 12, 575-588.
- Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in genetics using a multidimensional interpretive framework. *Journal of Research in Science Teaching*, 35 (9), 1031-1055.

Acknowledgements

Research reported in this paper was partially supported by a Graduate Research Grant from the University Research Grants Committee at Northwestern University. This research was also funded in part by a graduate fellowship from the Center for Curriculum Materials in Science, funded by the National Science Foundation under Grant ESI-0227557.