CHAPTER EIGHT

The Ubiquitous Patterns of Incorrect Answers to Science Questions: The Role of Automatic, Bottom-up Processes

Andrew F. Heckler

Contents

Psychology of Learning and Motivation, Volume 55

ISSN 0079-7421, DOI 10.1016/B978-0-12-387691-1.00008-9

Abstract

Nonexperts often exhibit regular and persistent patterns of errors when answering questions about science concepts. Typically, these patterns are considered to be due to high-level mental structures such as concepts or mental models that are different from the relevant expert concepts. Here, I consider the systematic influence of automatic, bottom-up processes on answering patterns to science questions. General evidence of the existence of top-down and bottom-up processes is surveyed from a variety of areas in cognitive science. Specifically, it is found that patterns of incorrect answering are a significant empirical driving force behind many investigations in learning and performance, and many of these areas invoke the need for bottom-up mechanisms to explain observations. The application of some of these mechanisms to the area of student answering of science questions is discussed. In particular, it is hypothesized that patterns of incorrect answering on a broad class of science questions are strongly influenced by the phenomenon of *competition* between relevant and irrelevant information in the questions. I investigate the particular cases in which the outcomes of this competition are mediated by the relative *processing times* and allocation of *attention* to relevant and irrelevant information in questions. These mechanisms result in predictable patterns of response choice, response time, and eye gaze fixations, and I discuss some studies suggesting that these mechanisms are at work when students answer specific physics questions. If, as suggested, automatic, bottom-up processes play a role in performance on science tasks, then this has important implications for models of understanding and learning science.

1. Introduction

This chapter revolves around what could be regarded as the most important empirical finding of science education to date, namely, that people often answer simple scientific questions incorrectly, yet in regular, patterned ways. More specifically, following Piaget's numerous demonstrations that children often answer ostensibly simple questions incorrectly, thousands of empirical studies have established that when conceptual questions about simple natural phenomena are posed to students, their answers are often contrary to scientists' answers, remarkably similar to those of other students, and resistant to traditional instruction (for lists, see Kind, 2004; McDermott & Redish, 1999; Pfundt & Duit, 2000). For example, students often believe, even after traditional instruction, that an upward traveling ball must have a net upward force acting on it (Clement, 1982).

Thus, we find ourselves in a fortuitous situation: we have numerous replicable empirical observations of how students respond to specific science questions, sometimes in great detail. If patterns in the responses are found, one can consider two general ways in which these findings can be useful.[1] First, information about students' answering patterns can help to inform instruction. An example of this is the several-decades demonstrated success in physics curriculum design and implementation done at the University of Washington, in which students' incorrect answering patterns have become a fundamental starting point for instructional methods (e.g., McDermott, 2001).

Second, patterns in the empirical data can be used to help build models of hypothesized mechanisms that *cause* the response patterns and perhaps student responses more generally. Ideally, these models can help us to make *predictions* of answering patterns to novel sets of questions. Furthermore, these models of causal mechanisms may also make predictions of how students would respond to specific types of instruction, and as such the models may also prove useful for designing instruction to help students answer difficult questions correctly.

In this chapter, I will concentrate more on the second approach, namely, investigating models of basic mechanisms that can not only help to explain *why* there are patterns of incorrect answers to science questions but can also *predict* answer patterns. While most existing explanations of answering patterns involve higher level mental structures such as *misconceptions*, I will consider the possibility that a number of bottom-up, automatic mechanisms can play a significant role in the generation of answering patterns.

The general idea that both bottom-up and top-down mechanisms are at work in learning and answering questions related to physical phenomena is hardly new. Some researchers have investigated and discussed this topic, even going back to Piaget. Nonetheless, the investigation of the potentially important role of bottom-up mechanisms in student answering patterns has been relatively ignored (especially in the science education arena) and is consequently an underexplored topic ripe for rigorous investigation. Therefore, in this chapter I will explore some of the past work on the influence of bottom-up processes on answering patterns and I will focus on the particular phenomenon of *competition*.

Specifically, I propose that answering patterns are often strongly influenced by competition between relevant and irrelevant information present in a science question. I will examine how competition manifests itself

[1] In addition to the two uses mentioned here, it may also be useful to build phenomenological models that reproduce observed response patterns for given questions, with minimal assumptions about the causes of the patterns.

in two interrelated ways. First, in most cases, the relevant variables in science are not easily observable (e.g., density determines floating) and, as a result, they are less likely to automatically engage *attention* than some of the irrelevant variables (e.g., size does not determine floating). In addition, the relevant information in science is often more difficult to process than irrelevant information and, as a result, more relevant information is *processed slower*. For example, there are data (described below) suggesting that students' well-known preference for utilizing *height* rather than *slope* on a graph is strongly influenced by the fact that in typical contexts height is inherently processed *faster* than slope.

The outcome of the competition mediated by these mechanisms may not only influence and thus help predict response choices but they may also imply patterns in other response metrics, such as processing time (e.g., response times) and attention (e.g., eye gaze) to specific features of a posed question. Thus, the hypothesized role of many of these mechanisms has the virtue of being testable by a number of different measurement modalities.

This investigation of the role of basic, automatic mechanisms in answering science questions stands in contrast to most existing explanations in science education that focus on higher level structures or processes, such as *concepts* or *explicit reasoning*, as causes of incorrect answering patterns. Nonetheless, the more bottom-up mechanisms proposed here are likely to complement higher level explanations.

2. The General Structure of Answering Patterns and the Critical Issue of Similarity

Because the central theme of this chapter is about patterns in student answering to science questions, it is worth considering the often ignored yet important issue of how one comes to claim or establish the existence of a *pattern in answering*. The empirical data of student responses to a set of questions itself are in a sense "raw" data. The question of whether there are any patterns in these raw data is, strictly speaking, a judgment based on an arbitrary (though perhaps reasonable) definition of *pattern*. Such a definition inevitably involves assumptions about the *similarity* of responses and of questions. Therefore, in this section, I will discuss the necessity of including explicitly constructed and acknowledged assumptions with any claims of patterns. The intention of the discussion is to reveal that the issue of *patterns* is fundamental to building a consistent, predictive theory of student responses, is far from resolved, and is certainly a fertile area for further empirical and theoretical investigation beyond the scope of this paper.

I will consider two main categories of patterns in answering: between-student and within-student answering, since these two kinds of answering patterns require fundamentally different explanations[2] (see also Siegler (1981) and discussions within about Piaget's view on this).

2.1. Between-student answering patterns

As is commonly defined, a between-subject pattern is the phenomenon of many subjects exhibiting similar performance on the same task. In science education, this phenomenon occurs when a specific question is posed to a number of students and many of them often answer incorrectly in ways that are judged to be similar (see the next section for a discussion about the similarity of responses). For example, when asked what is inside the bubbles formed in boiling water, a significant number of students answer that the bubbles are filled with air, when in fact the correct answer is that they are filled with water vapor (Osborne & Cosgrove, 1983).

Between-student answering patterns can be explained in a general way (though somewhat vaguely) by the fact that students are biologically similar, namely, they have similar cognitive processes and perhaps even similar "innate knowledge" (e.g., see Carey, 2009; Carey & Spelke, 1996), and students have similar everyday experiences, including experiences of the natural world and social experiences (e.g., Driver, Asoko, Leach, Mortimer, & Scott, 1994; Gelman, 2009), which shape their actions.

2.2. Within-student answering patterns

Within-student answering patterns require a different kind of explanation than between-student patterns. A within-student answering pattern of interest occurs when a specific set of questions, *judged to be similar* in some important way, is posed to a student, and the student provides answers that are *judged to be similar*. Therefore, determining within-student patterns is not straightforward, since it necessitates a *judgment of similarity* of both questions and responses. Since similarity is always a judgment based on a (presumably reasonable) choice of criteria, there is no one "correct" measure of similarity of questions and of responses, but there are certainly some measures that are more useful than others, depending on the task at hand. It is especially important to distinguish between a judgment of similarity of questions and responses on the basis of expert knowledge rather than on the basis of the student (i.e., answerer) point of view.

Judging the similarity of questions and responses based on an expert point of view is often necessary from an instructional point of view, since

[2]Note there are other ways to search for patterns using a purely psychometric approach (e.g., C. Reiner, Proffit, & Salthouse, 2005).

the goal of instruction is for students to recognize similarity and apply consistency as experts do. In fact, the assessment of a particular concept or skill could be seen as the practice of constructing questions that are similar from an expert's point of view in that they test knowledge of that particular concept or skill. In this case, instructors often look for only one kind of pattern: the pattern of correct answering. That is, the pattern that matches the expert point of view. However, not only is there useful information in patterns of incorrect answers, but students often do not use the same bases for judging similarity between questions that experts do (cf. Chi, Feltovich, & Glaser, 1981). Therefore, the interpretation of a pattern or lack of a pattern in answering from an expert scientist's point of view may be misleading, and even instructionally counterproductive. Instead, examining why *students* judge the similarity between questions can be helpful information for instruction (e.g., see Driver & Easley, 1978; Elby, 2001; Hammer, 1996a, 1996b).

Furthermore, since we are investigating the origin of within-student answering patterns, examining *student* judgment of similarity of questions (rather than expert similarity judgment) is warranted. In particular, we are interested in causes of answering patterns (any proposed pattern that has no cause could be regarded as arbitrary and not scientifically useful), and students are presumed to be the cause. Therefore, I will make the general assumption that a set of questions is answered in a similar manner by a particular student because the questions are for some reason being treated by the *student* in a similar manner. That is, within-student patterns occur because the questions are *judged to be similar by the student*, either implicitly or explicitly. For example, a student could perceive two questions as being about the same thing (e.g., force and motion) and thus apply a coherent impetus theory (i.e., misconception) to both questions. On the other hand, two questions could also be treated as similar because some automatic cognitive mechanism (of which the student in not necessarily consciously aware of) is processing both questions in a similar manner. Specific examples of such mechanisms will be discussed in Section 7.

Still, any claim of the existence of a within-student answer pattern caused by student-judged similarity of a set of questions must be based on an *inference* about the basis upon which the student is judging similarity. This inference is inevitably made by the person who is claiming the existence of a pattern. For example, is an explicit rule or concept (such as impetus theory) used by the student to judge similarity of two questions, or is it some bottom-up perceptual similarity? This is an important point because any claim of a within-student pattern is not solely an empirical observation but necessarily also depends on an assumption about the student's basis for similarity judgment. A typical assumption is that the students base their judgment of similarity of questions on some particular naïve concept. However, if the identification of the concept used by the student is incorrect or the student's responses are based on some other mechanism that does not involve an explicit concept (such as an automatic bottom-up mechanism), then any claimed pattern may be less meaningful. Alternatively, incorrect assumptions about a student's basis for judging similarity may result in a failure to recognize the presence of a within-student pattern of answering.

In short, any claimed within-student pattern in answers practically entails some assumption about the student's implicit or explicit judgment of similarity among the questions. Of course, any claim of within-student answering patterns also depends on the nature of the judged similarity in the *responses.* Such a judgment is usually done by the one who is making claims of answering patterns and is inevitably related to the assumptions of the student's bases for judging similarity of questions. In addition, one can measure and compare not only the content of the responses but also the other factors, such as time to respond and allocation of attention.

Therefore, the task of claiming within-student patterns on a given set of questions must critically include a detailed characterization, via comparative measurements or other analysis, of the bases upon which one is claiming similarity of both the questions and the answers. Ultimately, a careful description of the nature of the similarities will help to provide insight and clarity about the mechanisms underlying these answer patterns.

In practice, the issue of determining the basis of similarity judgments necessary for claiming patterns of answers has been implicit and relatively straightforward. It is common to find that questions and student responses are grouped into a few readily recognizable (by experts and even many students) and robust categories that include the correct response and a couple of prevalent incorrect response types (e.g., Bao & Redish, 2006). For example, Siegler (1976) found that when students are given balance task problems, one category of responses is to choose the side with the larger mass as winning, regardless of the length of the lever arm. In some cases, Chi (2005) points out that responses have been categorized in terms of past scientific theories, such as the impetus theory for force and motion questions. In order to account more for the student perspective, many researchers have carefully studied student responses and constructed reasonable categories of student responses that are specific to the domain. For example, Vosniadou and Brewer (1992) categorized student models of the earth according to various specific student models (flat, hemisphere, round, etc.). Chi (2005) has also pointed out another way to categorize responses in a more domain general manner, by looking for students' tendencies to answer according to ontological categories.

Nonetheless, it is important to keep in mind that these above examples of claims of student answering patterns necessarily make assumptions about how students interpret and answer the questions. We will discuss such assumptions in more detail in Section 3.

2.3. Summary

While between-student and within-student answering patterns tend to empirically occur simultaneously, both require different explanations. For the former, one must explain why *students* are similar; for the latter, one must explain why *questions* are similar. Both must characterize how *responses* are similar, though one can compare more than just the content of a response and look to other metrics such as response time.

From a general viewpoint, between-student patterns are somewhat trivially explained by the fact that students have similar biology and have many similar daily experiences. However, it is still a challenge to explain why many students tend to choose a specific answer to a specific question.

Explaining within-student answering patterns requires an explicit characterization and demonstration of the basis upon which *students* (rather than experts) are perceiving—either implicitly or explicitly—the similarity of questions.

3. Existing Explanations for Incorrect Answer Patterns to Science Questions

In this section, I will briefly review major existing explanations of incorrect answer patterns for science questions. I will focus on explanations of results from students that are typically between 10 and 20 years of age. There is also a significant amount of work on the development of concepts in young children that can be relevant to incorrect answering patterns on science questions (e.g., Carey, 2009; Gelman, 2009), though I will not discuss this here.

3.1. Misconceptions

The most widely assumed explanation for incorrect answer patterns stems from the abductive inference that the patterns are caused by somewhat coherent and generally applied "misconceptions" or "naïve theories" cued by the question and constructed by students from their everyday experience (e.g., Carey, 1985; Driver & Erickson, 1983; McCloskey, 1983; Vosniadou, 1994; Wellman & Gelman, 1992). A student, for example, might (incorrectly) answer that a ball traveling on a curved track would continue to travel in a curve after leaving the track because he/she has developed a coherent theory predicting that when objects are moving in a curved path, they will continue to move in a curved path, even in the absence of external forces (McCloskey, Caramazza, & Green, 1980). Note that this explanation directly addresses the case of within-student answer patterns. Since it is also found that many students exhibit the same consistent answer patterns (i.e., between-student patterns), presumably these students have all formed the same misconception because they have derived it from everyday experiences common to all students.

The term *misconception* is frequently used in the literature, and it is important to note that the term actually describes an *inference* about the cause of patterns of incorrect answering rather than an empirical observation of student answering. Clearly, it is logically valid that *if* students held coherent, incorrect theories (i.e., misconceptions) and *if* they consistently applied these theories, then they would likely answer relevant questions in patterned incorrect ways (following the pattern of the misconception and its consistent application). However, it is not necessary to have a misconception in order to produce patterns of incorrect answers: the pattern may also be due to other causes. Thus, I will sometimes refer to patterns of incorrect answering as *misconception-like* answers.

The misconceptions explanation has been critiqued because it was recognized that the model of student-held coherent yet incorrect theories was not universally valid at least in its strictest interpretation in two ways. First, when students were asked questions about or relevant to their putative theories, the theories themselves were often highly fragmented, incomplete, and logically inconsistent certainly from the point of view of the expert and often even from the perspective of the student (e.g., diSessa, Gillespie, & Esterly, 2004; for a discussion, see Keil, 2010). Second, student answering was shown to often be fairly sensitive to context; thus, within-student patterns of incorrect answers could be disrupted by simply making small changes to the context of the question. For example, on the question concerning objects moving in a curved path, Kaiser, Jonides, and Alexander (1986) found that significantly many students answered that water would come out of a curved hose in a straight line, and significantly less answered that a ball would come out a of curved tube in a straight line.

Of course, this critique of the misconceptions explanation could be at least partially addressed by the fact that the questions judged to be similar by an expert may not be perceived as similar by the student, therefore a lack of a pattern could be expected. Furthermore, there are many examples in which students do consistently answer incorrectly in ways that are consistent with them holding an incorrect concept for a significant set of questions. However, in some cases, the answering is so fragmented even with small changes in question context that it is difficult to imagine that the student holds a robust theory that is applied to many situations.

3.2. Knowledge in pieces or resources

In light of the demonstrated sensitivity of student answering to the context of some questions, others have suggested that rather than cuing coherent theories, questions with different contexts instead cue different (and often incorrect) combinations of "pieces of knowledge." In this way, within-student patterns of incorrect answers could be disrupted. Between-student patterns were still observed for a given question, and this could be explained because students have many of the same experiences; thus, a given question will often cue similar combinations of pieces of knowledge, resulting in student responding with similar incorrect answers. The pieces of knowledge represent basic phenomena such as "force as mover" (diSessa, 1993) or relations such as "more *x* means more *y*" (e.g., Stavy & Tirosh, 2000). A perhaps more general version of the pieces of knowledge explanation claims that questions may cue incorrect "resources" that students use to answer a question (e.g., Hammer, 2000). These resources can be wide ranging and include factual knowledge, basic relations, procedural knowledge, and epistemological beliefs. An especially powerful aspect to the knowledge in pieces model is the notion that student often have "untapped" knowledge and skills that can be used to improve their learning and performance (e.g., Hammer & Elby, 2003). To support this, there is evidence that students sometimes have the correct knowledge available to answer correctly, but this knowledge is often not cued (Hammer, Elby, Scherr, & Redish, 2005; Heckler, 2010; Sabella & Redish, 2007).

Finally, it should be noted that a more comprehensive version of the knowledge in pieces explanation is in fact extended to include loosely bound collections of pieces of knowledge that can form something resembling a coherent concept, thus explaining the presence of within-student patterns in some cases (e.g., diSessa & Sherin, 1998).

3.3. Ontological categories

Somewhat independent of the misconceptions and knowledge in pieces explanations is a third prominent explanation for incorrect patterns in answering. Some researchers (e.g., Chi, 2005; Reiner, Slotta, Chi, & Resnick, 2000) provide arguments and evidence for a "domain general" mechanism for misconception-like answering patterns, as opposed to "theory-specific" or "domain-specific" explanations of misconceptions and knowledge in pieces. The domain general process they investigate is the incorrect categorizing of the ontological nature of certain physical variables or phenomena (e.g., Chi, Slotta, & de Leeuw, 1994). For example, students commonly (incorrectly) believe that force is a substance in the sense that objects "have" a force (Reiner et al., 2000). Like the misconceptions explanation, this explanation can account for within-student answering patterns.

3.4. Summary

The explanations involving student-held concepts, theories, or models tend to naturally explain the presence of within-student patterns, while explanations involving more fragmented knowledge in pieces explanation tend to naturally explain the absence of within-student patterns that might be expected if students held coherent misconceptions. Both kinds of explanations have been modified to explain the presence or lack of answer patterns to some degree, though there is still debate about the validity of each explanation (e.g., see diSessa, Gillespie, & Esterly, 2004).

The existing explanations of coherent yet incorrect concepts or theories, incorrect ontologies, or knowledge in pieces all account for within-student patterns for at least some sets of questions. However, since each explanation is different, they may also identify and explain different patterns of answering. For example, the misconceptions explanation tends to be quite domain specific and will tend to search for and identify patterns within a specific domain, whereas the ontological category explanation is more domain general; consequently, this kind of approach will tend to search and identify patterns that are more domain general. On the other hand, these different explanations all appear to agree on the general reason for between-student patterns, namely, that whatever mechanism is responsible for the within-student patterns is common to all students.

Finally, a caveat: I have focused here on incorrect answer patterns specifically to science questions, yet there is also significant work on pattern of incorrect answers to more general questions, of which science is a subset. In particular, I refer to the field of heuristics and biases and the pioneering work of Tversky and Kahnemann (1974). I will address this in Section 6.1.

4. The Insufficiency of Existing Explanations

While explanations discussed above are useful in the examination of student answering patterns, this section discusses three limitations of these explanations. Section 5 then describes an empirical example highlighting these limitations and the need for the inclusion of a more bottom-up mechanistic explanation.

4.1. Limitation 1: patterns are typically assumed to be caused by high-level mental structures and processes

I would like to emphasize that the critical question addressed in this chapter is "What causes student incorrect answer patterns to science questions?" This is a question about an *empirical observation* that allows a broad range of possible explanations. In contrast, the typical approach to the empirical evidence is to assume that the patterns are caused by "higher level" mental structures such as concepts, schemas, mental models, or loose collections of pieces of knowledge (e.g., Carey, 1985; Driver & Erickson, 1983; McCloskey, 1983; Novak, 2002; J. P. Smith, diSessa, & Roschelle, 1993; Vosniadou, 1994).[3] These approaches tend instead to ask questions such as "What are the student concepts that explain the answer patterns?" or "How are incorrect concepts learned?" These questions are not directly about empirical observations of answering patterns, rather they are questions about *inferences* about the observations. In short, the typical approach to the empirical observation of incorrect answering pattern is to *already assume* the cause of the patterns, namely, that they are a result of some high-order mental structure such as concepts or mental models.

The origin of this assumption may be traced back to Piaget (1952/1936, 1972/1970), who argued that scientific knowledge cannot be learned from sensory information alone, but rather requires explicit higher order thinking and interaction with the world in order to form high-level mental schemas necessary for scientific knowledge (see also Driver et al., 1994; Leach & Scott, 2003; Taber, 2010; Vosniadou, 1996).[4] Therefore, the argument goes: since higher level structures of knowledge are needed to understand science, such mental structures are needed to answer science questions in a correct and consistent manner.

However, the topic of this chapter is not directly about the origins or nature of scientific knowledge, it is about the origins of incorrect answering patterns to science questions. While one might agree that answering science questions consistently correctly may require correct higher level mental structures, answering incorrectly in patterned ways does *not* necessarily require a higher level mental structure. The patterns could be caused or strongly influenced by more basic, bottom-up processes that are implicit and relatively unknown to the answerer.

In other words, even if we assume that consistently correct answering occurs if and only if the answerer holds the correct concept (let us ignore

[3]Some models, such as Vosniadou's (1994) framework theory, include lower level unconscious aspects to the proposed mental structure.

[4]Note that there is also much discussion about the difference between individual cognition and social cognition (e.g., Leach & Scott, 2003), which we will not discuss here.

the possibility of false positives for simplicity), it still does not logically follow that incorrect answering patterns imply an incorrect concept.[5] Rather, incorrect answering only implies the *absence of a correct concept*, which could imply either the presence of an incorrect concept or the *absence* of any concept at all. For example, misconception-like answers could stem from implicit, automatic, and relatively unconscious processes that direct the student toward "undesired" answers in regular ways and may have little to do with consistently applied explicit concepts. One might claim that a pattern in answering requires some regularity in mental structure. This may be true, but it does not require a high-level mental structure—regularity in answering could be due to more basic processes. Therefore, most explanations of patterns of incorrect answers assume only one cause, namely, high-level mental structures; here, I would like to consider another influence (if not an alternative cause), namely, bottom-up processes.

Finally, the lack of a clear operational definition of high-level *mental structures* (e.g., a scientific concept) severely limits scientific progress of the high-level mental structures approach to explaining incorrect answering patterns to science questions (cf. diSessa & Sherin, 1998; diSessa et al., 2004). If one is to argue that mental structures such as concepts or mental models cause answering patterns, it is critical to establish a robust, unambiguous definition of such structures based on empirical observations characterizing the extent to which a student has a particular mental structure. Constructing such a definition will be a challenge. For example, if one cannot decisively claim that high-level mental structures are the sole cause of answering patterns, then one cannot use answering patterns as a sufficiently decisive empirical measure of the existence of high-level mental structures. This is related to the argument discussed earlier that any claim of the existence of patterns practically requires some assumption of the cause of the patterns, which may or may not be due to high-level mental structures.

4.2. Limitation 2: current explanations have limited predictive power

As mentioned in the introduction, one of the main scientific reasons for constructing a causal explanation of an empirical phenomenon is to make specific predictions about other, related empirical phenomena. There is

[5] Perhaps it should not be surprising that such a compelling, complementary, converse idea was also assumed. That is, if correct answering patterns result from correct concepts, then one might also imagine that incorrect answering patterns result from incorrect concepts or misconceptions. This "counterpart concept," so to speak, may be especially compelling given the strong evidence that prior knowledge interfered with learning the correct concept. The symmetric picture is completed with the assumption that this interfering prior knowledge is none other than a misconception.

value in post hoc explanations of existing empirical evidence, though an explanation increases in usefulness (a) the more it can be generalized to predict other situations and (b) the more specific the predictions can be about any given situation. This also implies that the more scientifically useful a model is, the more testable it is.

Current explanations of misconception-like answering patterns do have some predictive power, though the predictions are quite limited. One reason for this is because most current explanations are done *post hoc*. For example, the well-known phenomenon of incorrect answering to force and motion problems was not *predicted*. Rather, it was empirically *discovered* (e.g., Clement, 1982; Viennot, 1979) and then later explained as being due to students having an incorrect impetus theory of force and motion. This explanation could not be generalized to predict student answer patterns for questions about other physics topics such as simple circuits. It might be argued that since it is to be expected that students construct many concepts in the course of everyday life, the misconception model predicts that, in general, incorrect answer patterns are likely to be found for many other topics, though it cannot predict which topics or what those patterns will be.

There are, however, some specific predictions that the misconceptions model does make about students answering questions, specifically about force and motion questions. In its strictest interpretation, the model predicts that a subpopulation of students will answer all force and motion questions with an impetus motion model. Scientifically, there is an advantage to this model: it makes a specific prediction. As it turns out, the model's predictions are marginally successful in that they do predict some observed patterns, but other times the prediction of patterns are incorrect; thus, the model does not hold up to all empirical observations. Nonetheless, this test should be considered scientific progress: the model made a prediction, and the prediction was empirically tested. The picture for the predictive power of the ontological category model is very similar to that of the misconception model, and it too has succeeded in making some specific predictions but has failed at least one test of its strictest interpretation (Gupta, Hammer, & Redish, 2010).

The resources or knowledge in pieces models, which include the cuing of much finer-grained mental structures, have similar limitations of predictiveness as the misconceptions model in that the resources model also employs post hoc explanations for specific answer patterns (due to cuing of "incorrect" resources) rather than specific predictions of answer patterns. However, because the model is so flexible, there has yet to be any specific, testable predictions for this model (to the knowledge of the author), though there have been some preliminary attempts (e.g., diSessa et al., 2004; Elby, 2000). In order to achieve more scientific progress with such a model, more effort must be made to deduce testable predictions from it.[6]

In sum, current high-level mental structure explanations, such as the misconceptions model, do make a limited number of specific predictions about answering patterns to some specific questions. The misconception model predictions have been somewhat accurate in specific domains, though they have very limited predictive power and scientific usefulness in their current state. If more scientific progress is to be made in this area, mental structure theories models need to make significantly more specific, testable predictions about answering patterns that apply to a range of questions. This will likely entail the incorporation of specific mechanisms and quantifiable models, which tend to be missing in current models.

4.3. Limitation 3: current explanations rarely consider response data beyond the response content

The overwhelming majority of studies on student responses to science questions investigate the content of the student responses, such as the correctness of the response, the patterns of answer choices in a multiple choice test, the explanations in an interview, and the solution method in a problem solving task. However, there are a number of other response measurements that can be extremely useful, including response time, eye tracking, gesturing, and measurements of brain activity. Some of these modes will be discussed in Sections 6 and 7. Indeed, this is a growing area of activity that will provide much needed empirical data useful for testing models.

Since these additional response metrics tend to measure rapid, bottom-up processes of which the answerer is unaware, they will allow the testing of models that include bottom-up as well as top-down processes. A challenge for models such as the misconceptions model or the knowledge in pieces model will be to make testable predictions of such measurements.

5. Example: The Case of Competing Relevant and Irrelevant Information

In light of the general limitations discussed in the previous section of existing high-level mental structure explanations of student answering patterns to science questions, I would like to point out a particular

[6]However, the resources model has proved useful in other ways (e.g., Hammer 1996), namely, for orienting strategies for instruction. The interest here is in the scientific value of the model in terms of predicting answering patterns, which is not the same. The scientific liability of current models is to be distinguished from the *instructional* usefulness of models such as the misconception model or the resources model.

limitation. This limitation can be framed in terms of the notion of *competition*.

Explanations employing high-level mental structures typically assume that a specific question activates a specific mental structure in the student, be it a coherent theory, mental model, an ontological frame, or loosely bound pieces of knowledge, and this activated structure in turn leads to a specific response. However, these explanations do not (currently) include specific mechanisms that explain (or predict) *which* specific mental structure will be cued as opposed to another. For example, if a number of concepts and resources are plausibly relevant to the student for a given question posed, then why are only particular ones used in a given case? Since these explanations to not provide a specific mechanism responsible for activating one "plausible" concept over another in a given case, they will not be able to explain why a particular answer was chosen. Furthermore, if more than one concept is activated, then how is an answer choice determined? Does this mechanism for determining the answer choice result in a pattern of answers?

This limitation can also be described in terms of the *information* that is presented by the question and perceived by the student. Presumably, the student will often attend to both relevant and irrelevant information (as judged by an expert). What is the mechanism that determines which information is attended to or used to determine the answer? In other words, if a question presents a variety of information relevant to various competing mental structures, then what determines the outcome of the competition, and ultimately the student response?

To illustrate this point, consider the known difficulties students have with understanding the relation between electric field (E) and electric

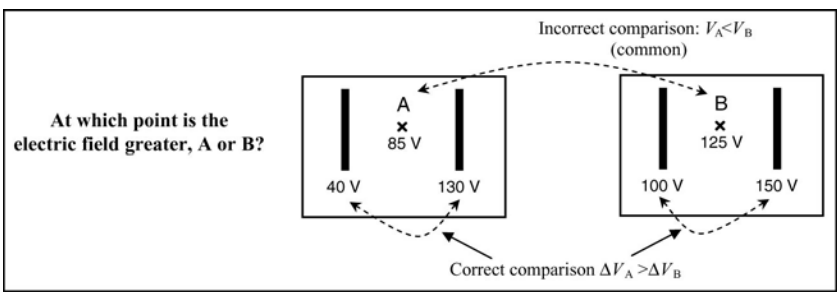

Figure 1 Undergraduate physics students compared the electric fields between the two sets of plates at the indicated voltages. For a series of eight questions, about 50% of the students consistently and incorrectly chose the point with the greater *value* of the voltage at the point between the plates (here, "B"). Less than 40% consistently chose the correct answer, which is found by taking the *difference* between the voltages of the plates. This misconception-like answering pattern could be considered as due to *competition* between relevant and irrelevant information.

potential (*V*), which is measured in *volts* (e.g., Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001). For example, Figure 1 presents some data collected in my lab that are typical of the confusion in the relation between *E* and *V*. Thirty-five undergraduate physics students were shown a diagram of two sets of metal plates held at indicated voltages, and they were asked to determine which point midway between each set had a greater electric field (magnitude). The task was given post relevant instruction, so the students were reasonably familiar with these diagrams. This is a fairly simple and straightforward question in which a student may apply the idea that the magnitude of the electric field is proportional to the gradient of the electric potential, $\mathrm{E} = |\mathrm{dV}/\mathrm{dx}|$. For this task, since the separation between each set of plates is the same, this simplifies to the idea that the point with the greater electric field is the one for which the *difference* in electric potential between the plates is the greatest. Instead, about 50% of students post instruction consistently and incorrectly chose the point with the greatest *value* of electric potential. Why is there often a consistent preference for the value of the potential rather than the *difference* in values between two plates?

Let us consider two possible scenarios that highlight the insufficiency of the misconception and knowledge in pieces explanations mentioned above, thus illustrating the need for a more specific mechanism that can explain and predict an answer preference resulting in a pattern of errors. The first scenario is centered around the notion that the students have learned the (incorrect) concept "the value of electric potential at a point predicts electric field," but have not learned the correct concept "gradient of electric potential predicts electric field," and this explains their patterns in answering. However, the question still remains as to specifically *why* the potential–field association was learned and why the potential gradient–field association was not learned.

The second possible scenario is based on the possibility that many students may have in fact learned both the association of electric potential with electric field and potential gradient with field, but there is nonetheless a preference for one because of the nature of the specific question. In this example (as well as many others), potential and difference in potential *compete*, and the former often wins. From the misconceptions perspective, students hold both the correct and incorrect concepts, but there is no explanation (or prediction) specifying why many students consistently choose the scientifically incorrect concept over the correct one for this question. Likewise, in terms of knowledge in pieces, one might claim that the question cued a basic relation such as "more is more." However, this basic relation could be applied to both the value of the electric potential and the difference in electric potential (more potential is more field, or more difference in potential is more field), and the knowledge in pieces approach also does not specify why one "more is more" was preferred over the other.

In sum, existing explanations of patterns of incorrect answering on science questions do not provide specific predictions or mechanisms that determine why, for a given specific topic, there may be a preference for learning a scientifically incorrect concept relevant to that topic rather than a correct one, or if students have learned both correct and incorrect concepts, why students choose one over the other when responding to a specific question.

It may be possible to modify mental structure models to explain and predict the outcomes of competition. Nonetheless, in the following sections, I describe some specific bottom-up mechanisms that can naturally help to explain and predict the outcome of competition between competing relevant and irrelevant information in at least some questions, namely, that many students tend to base their decision on the dimension that is processed the fastest or garners the most attention, even if it is incorrect. First I will briefly review some previous work that has examined the role of bottom-up mechanisms relevant to science learning and performance.

6. Bottom-Up Versus Top-Down Processing: Evidence from Answer Patterns

The idea that there are two kinds of cognitive systems involved in learning and performance has been discussed in the field of psychology for over 100 years (e.g., James, 1950/1890; Johnson-Laird, 1983; Neisser, 1963; Piaget, 1926; Shiffrin & Schneider, 1977; Vygotsky, 1987/1934). There are a number of recent studies demonstrating that higher and lower order processes interact significantly in decision making and reasoning (Alter, Oppenehimer, Epley, & Eyre, 2007; Evans 2003, 2008; Glöckner & Betsch, 2008; Kahneman & Frederick, 2002; Sloman, 1996), category learning (Kloos & Sloutsky, 2008; Maddox & Ashby, 2004), memory and recall (Poldrack & Packard, 2003), and language learning (e.g., Smith, Jones, & Landau, 1996).

Some of the evidence of two mental processing systems stems from the observation that for many tasks there appears to be two distinct ways to arrive at a response, and in many cases these two paths lead to different responses (cf. the *Criterion S* of Sloman, 1996). One kind of response tends to be fast, implicit, intuitive, automatic, and relatively effortless and is ascribed to being a result of System 1 processes. The other response tends to be slower, explicit, and effortful and is thought to come from a System 2 process (e.g., Evans, 2008; Kahneman & Frederick, 2002; Stanovich & West, 2000).

Figure 2 Examples of tasks that provide evidence of two different systems at work in the course of solution: one automatic and implicit and the other deliberate and explicit. In these cases, the different systems lead to different answers. In particular, the fast, automatic system leads to ''incorrect'' answers, thus implying that one can construct a pattern of incorrect answers with similar tasks. The first task is an optical illusion in which one is to compare the lengths of the horizontal lines. The second task is the well-known Stroop task, in which one is to name the color of the letters. On the third task, constructed and studied by Kahneman and Frederick (2002), most college students answer (incorrectly) ''10 cents.''

An optical illusion is a classic case; an example, as pointed out by Sloman (1996), is the Muller-Lyer illusion (see Figure 2). In this case, the System 1 "perceptual" response is fast and clearly conflicts with the System 2 response that comes from reasoning that would include a concrete measurement. Other examples of two systems at work, some of which are presented in Figure 2, include the Stroop effect (e.g., MacLeod, 1991), belief bias (e.g., Evans, 2003), relapse errors (Betsch, Haberstroh, Molter, & Glöckner, 2004), and perseveration (Brace, Morton, & Munakata, 2006). Of course, these empirical phenomena are not all proposed to be caused by the same mechanism, but all of them have been explained in terms of a dual system similar to System 1 and System 2.

While there are some issues about the ambiguity of the meaning of System 1 and System 2 processes (e.g., see Evans, 2008), there are a number of studies testing predictions resulting from models that assume dual interacting systems at work in learning and performance in reasoning (DeNeys, 2006), relapse errors (Betsch et al., 2004), Stroop effect (Cohen, Dunbar, & McClelland, 1990; Kane & Engle, 2003), and category learning (Sloutsky, Kloos, & Fisher, 2007; Zeithamova & Maddox, 2006). There is further evidence building in neurological findings, showing that different areas of the brain are active during the putative engagement of the two different systems (e.g., Goel, Buchel, Frith, & Dolan, 2000).

A related line of compelling evidence of the existence of nontrivial implicit knowledge and skills is the field of *implicit learning* (e.g., Reber, 1989). In short, humans can unconsciously learn fairly complex rules that are applicable to novel (though somewhat limited) tasks (e.g., Berry & Dienes, 1993; Reber, 1993). For example, people can learn to remember strings of letters better if the strings have fairly complex statistical structure compared to remembering a random sequence, even though the learners are unaware that they are learning any structure. In addition, they can recognize new strings that are similar in structure to the ones they learned, though they are unable to report why they are similar. Evans, Clibbins, Cattani, Harris, & Dennis (2003) provide evidence of the learning in multicue judgment tasks involves both implicit and explicit knowledge, and this may help to explain why experts typically cannot fully explain their knowledge of rules used in tasks, because some of this knowledge is in fact implicit.

Finally, there is an illuminating difference between cognitive science studies on dual systems and science education studies on student answering to science questions. The above empirical studies, such as the Stroop effect, or optical illusions reveal a pattern of incorrect answering, yet the patterns of answering in this case are usually seen as evidence of automatic processes rather than evidence of a high-level mental structure such as a misconception.[7] For example, one would not claim that the Stroop effect is the result of a misconception. In contrast, in science education research, the patterns of incorrect answers to science questions have been taken as evidence of high-level mental structures such as misconceptions.

6.1. Heuristic and biases

The study of judgment and rational choice has a rich history in cognitive science and is related to the topic of incorrect answer patterns to science questions. This is partly because the study of judgment and choice is partially driven by the empirical observation that people often make systematic errors in judgment and choice. For example, in a series of classic studies, Tversky and Kahneman (1974) demonstrated that people tend to make general kinds of systematic errors in questions that require some level of quantitative or probabilistic judgment. For example, people have biases in judgments of the relative sizes of populations due to retrievability from memory. When verbally given a list of male and females, people tend to judge the list has more of one gender if more of names of that gender in the list are famous names.

Based on earlier work by Simon (1955), these patterns have often been explained in terms of *bounded rationality*, namely, that people make rational decisions that automatically include real-world constraints such as limited time and limited access to information. This idea in turn has led to explanations of systematic errors as due to the use of *heuristics*. That is, the hypothesis is that people use fast and efficient heuristics to make

[7]There are exceptions: one might attribute errors in syllogistic reasoning as the result of a mental model (Johnson-Laird, 1983), though this explanation makes assumptions of implicit processes as well (Evans, 2000).

judgments and choices. While in most cases this process is quite successful, in other cases the use of heuristics can lead to biases that cause systematic errors. For example, Goldstein and Gigerenzer (2002) discuss the *recognition heuristic*: if two alternatives are provided and one must choose only one based on some criterion and only one of the alternatives is recognized (familiar), then assume that the recognized one has a higher value of the criterion. The common task used to demonstrate the use of this heuristic is the case in which people are given the names of two cities in the world and asked to choose the one with the higher population. People often choose the city name that they recognize. For reviews of the topic of heurisitics and biases, see, for example, Gigerenzer (2008), Kahneman (2003), and Gilovich and Griffin (2002).

Evidence for the existence of heuristics has typically come from the recognition that, empirically, a given strategy or heuristic is used in many kinds of relevant problems by many people. See Gigerenzer (2008) for a number of examples of heuristics that are empirically well supported. In addition, there has been some progress in establishing testable predictions from the somewhat detailed models of heuristics that bolster the scientific usefulness of the heuristics hypothesis (e.g., see Bergert & Nosofky, 2007; Gigerenzer & Brighton, 2009).

There are two main reasons for bringing up the topic of heuristics. First, since the notion of heuristics was applied to explain patterns of incorrect answering, the pervasive use of heuristics may be an alternative or complementary explanation of misconception-like answers to science questions. Second, the heuristics tends to be regarded as an automatic, bottom-up process rather than an analytic explicit reasoning process (e.g., Evans, 2008; Kahneman, 2003). Therefore, if misconception-like answers to science questions are influenced by bottom-up processes, then heuristics models may be candidates for such processes.

In Section 7, I briefly mention how this may be applied to a specific example, but clearly the application of the hypothesis of general use of heuristics to answering science questions has potential to be a rich area for study in more detail.

6.2. Studies on bottom-up processes in science learning and performance

While the overwhelming majority of studies on student responses to science questions have focused on higher level mental structures, there have been a small number of studies investigating evidence of more implicit lower level processes taking place when students answer questions about natural phenomena.

The phenomenon of *representational momentum* is an example. If a student observes an image of an object undergoing implied or apparent motion and the object then suddenly disappears, the immediate memory of the last position of the object is shifted forward from the actual last position, as if to imply a continuing motion of the object (Freyd & Finke, 1984). This phenomenon is called representational momentum because most have interpreted the results as evidence that the perceptual system internalizes physical principles of motion and creates a representation of the motion that manifests itself, for example, in distorted memories (Freyd, 1987, 1992; Hubbard, 1995, 1998). The effect is small and short-lived but reliable, and the observers are not aware of the distortion; thus, it could be considered as implicit knowledge.

Interestingly, the implicitly projected paths do not always follow Newtonian motion. For example, Freyd and Jones (1994) found that for a ball exiting a circular tube, the perceptually preferred paths were in a continuing spiral rather than a straight line. They argued that this may help to explain why some students explicitly choose the incorrect spiral path when the question is posed explicitly. That is, there may be some influence of the implicit knowledge on the explicit answering. Similarly, Kozhevnikov and Hegarty (2001) found that even experts' implicit knowledge as measured by representational momentum is non-Newtonian, even though their explicit answers are Newtonian. Although it would appear that this implicit knowledge is difficult to change (however, see Courtney & Hubbard, 2008), they propose that it may still affect answering even of experts under certain constraints such as time limitations.

There are other kinds of studies demonstrating students "saying one thing, but doing another" on science-related tasks that would suggest that there are implicit and explicit systems separately influencing performance. For example, Piaget (1976), as pointed out by Oberle, McBeath, Madigan, & Sugar (2006), found that children could hit targets by appropriately letting go a string attached to an object that they were twirling in circles above their heads, but when asked in a paper and pencil task when the ball should be released, they answered incorrectly (i.e., they answered when the string was aligned with the target). Likewise, Oberle et al. asked students to compare the times it would take to two objects of either the same mass and different size or same size and different mass to fall the same (fairly large) height in the realistic scenario when air resistance is explicitly included. They found that students often answered that the objects would fall at the same rate. However, when the students were asked to physically drop two balls such that they would land at the same time, they found that students would drop the balls at *different* times, contrary to their explicit answers, They attributed this difference in answering to two different systems, namely, a perceptual system based on everyday perceptual experience and a higher level conceptual system.

There are also a number of studies investigating the student responses to the motions of objects, demonstrating that although many students may answer incorrectly on questions about motion represented by static displays, they often answer very accurately when the motions are animated (see Rohrer, 2003). For example, when given a choice of trajectories of a ball leaving a curved tube, students will answer correctly more often when given an animation compared to a static diagram (Kaiser, Proffitt, & Anderson, 1985). However, the benefit of animation decreases with increased complexity of motion (Kaiser, Proffitt, Whelan, & Hecht, 1992). This difference in responding has been interpreted as due to the static diagrams cuing explicit (incorrect) reasoning knowledge based on, for example, impetus models and the animated format cuing implicit perceptual knowledge that is based on common experience.

Finally, there are a number of studies on learning that indicate the existence and importance of low-level implicit automatic processes relevant to mathematics concepts. For example, in a study on 8–10-year-old children learning to solve simple addition problems, Siegler and Stern (1998) found that the time to solution was a reliable measure of the solvers' implicit use of a shortcut strategy. They found a bimodal distribution of times to solution with the solvers using the shortcut strategy solving the problem faster. Over a period of weeks, students became better at solving the problems and, perhaps most interestingly, many students started to use the shortcut strategy (as measured by time to solution) before they were explicitly aware of it as verbally reported by the solver. This suggests that there is a process of unconscious strategy discovery.

Furthermore, some researchers have investigated how math learning may be influenced by the phenomenon of perceptual learning, which is lower level, unconscious learning that results in an increase in the ability to extract information simply through experience (no explicit feedback is required). Kellman, Massey, and Son (2010) have found that simple perceptual learning tasks improve performance on higher level math tasks, for example, by increasing the learner's ability to focus on relevant rather than irrelevant dimensions. Goldstone, Landy, and Son (2010) provide evidence for the argument that the low-level perceptual system can adapt (i.e., learn) to achieve specific purposes, such as automatic recognition of symbols or diagrams in math and science, and this learned automaticity at least partially explains continual success on math and science tasks. Both of these examples highlight the educational possibilities of tapping into low-level processes and making required tasks automatic in order to improve math and science performance.

In sum, there are a small number of studies that provide evidence for the idea that automatic bottom-up processes can influence student learning and answering on science and math questions. Some of these studies used other modes of measurement such as response time and nonverbal responses that can help to support the claim that automatic processes are involved. In the next section, I will describe how a large class of science questions involve competing dimensions, and automatic bottom-up processes may at least partially cause the known misconception-like answering patterns to these questions.

7. The Phenomenon of Competition in Science Questions

Section 5 described an example in which competing relevant and irrelevant information (from the perspective of an expert) was present in a science question, and many students consistently based their answer on the irrelevant information. In this section, I will discuss in more detail the phenomenon of *competition* between relevant and irrelevant information in science questions, and the outcomes of this competition as mediated by the low-level mechanisms of relative processing time and allocation of attention. The phenomenon of competition in science questions and its role in misconception-like answering patterns is described in three points:

First, it is assumed that students may consider—either implicitly or explicitly—a number of dimensions (e.g., variables or features) when answering science questions. I would like to emphasize that the dimensions considered by a novice are not always the same dimensions considered by an expert. Novice students may utilize dimensions not scientifically valid according to experts because the students may nonetheless perceive these dimensions as relevant. For example, when determining the period of a pendulum, many students may consider both the mass and length of the pendulum (see Figure 3), yet only the length is scientifically relevant.

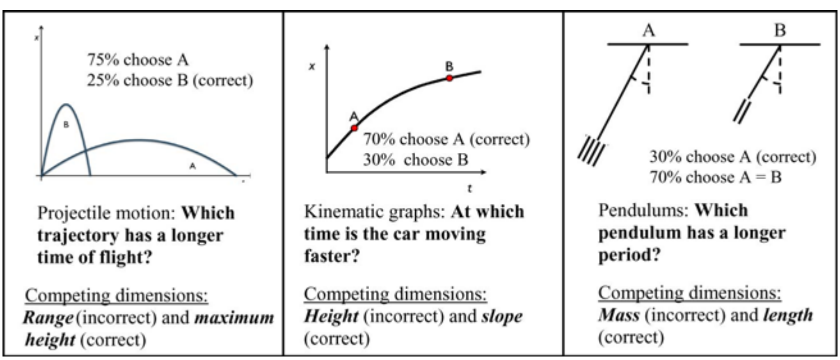

Figure 3 Examples of physics questions with competing dimensions. The indicated student response percentages were collected in pilot studies, with $N > 40$ in for each question.

Second, I propose that for a significant number of science questions, *competition* between relevant and irrelevant dimensions in a question plays a significant role in incorrect answering patterns. To illustrate the pervasiveness of this phenomenon, Figures 1 and 2 as well as Table 1 present a few examples of competing relevant and irrelevant dimensions for questions that have well-known misconception-like answering patterns. It would not be difficult to provide dozens of such examples. Note that collaborators and I have conducted interviews with some students answering these questions to support the validity of the questions.

Third, I propose that well-known mechanisms may at least in part predict and explain the outcomes of competition and the resulting patterns of student answers. These mechanisms are discussed in the next section. It is important to note that we are not explaining the cause of competition itself. Rather, we will help to explain and predict the *outcome* of two competing dimensions.

7.1. Relative processing times of relevant and irrelevant dimensions

I would like to consider the hypothesis that, when there are competing plausible dimensions upon which to base an answer for a given science question, students tend to choose the dimension that is *processed the fastest*. This hypothesis is somewhat similar to the fast heuristic model of "take the best," which chooses the first discriminating attribute to make a decision (e.g., Gigerenzer & Goldstein, 1996; see also Bergert & Nosofsky, 2007). Note that the dimensions of interest need only to be plausible from the perspective of the student; thus, both relevant and irrelevant dimensions may compete.

This hypothesis stems from evidence that if there is competition between relevant and irrelevant information in a question, then the outcome can be influenced by the relative time to process the relevant and irrelevant dimensions. Perhaps the best known method demonstrating this phenomenon is the Stroop effect (e.g., MacLeod, 1991), though the story likely also involves the more general concept of automaticity of processes (Cohen, Dunbar, & McClelland, 1990; Macleod & Dunbar, 1988). The Stroop effect occurs when a well-learned cue that is technically irrelevant to a task nonetheless competes with the relevant cues and interferes in task performance. The classic example is the color-word task, for example, spelling out the word "blue" in red-colored letters and asking participants to name the color of the letters (see Figure 4). Accuracy is typically lower and response times higher when the color of the letters conflicts with the word compared to when the color matches the word. Furthermore, the interfering dimension (word) is typically

Table 1 Examples of Science Concepts with Competing Relevant and Irrelevant Dimensions

Target dimension	Correct predictive dimension/relation	Incorrect/incomplete competing dimension	Common incorrect answer	Correct relationship
Force	Acceleration	Velocity	Greater velocity means greater net force	Greater acceleration means greater net force
Time of flight	Maximum height of projectile	Range of projectile	Both the range and height determine time of flight	Only height determines time of flight
Speed	Slope on x versus t graph	Height on x versus t graph	Higher point on graph means higher speed	Slope on graph determines speed
Period	Length of pendulum	Mass of pendulum	Mass and length determine period	Length only determines period
Electric field	Electric potential gradient	Electric potential	Higher potential means higher electric field	Higher potential gradient means higher electric field
Sliding distance	Initial velocity	Mass	Both initial velocity and mass determine sliding distance	Only initial velocity determines sliding distance
Torque (balance)	(Mass × distance)	Mass	Greater mass (force) means greater torque	Force and position determine torque
Mass	(Volume × density)	Volume	Bigger objects have more mass	Volume and density determine mass
Amount of thermal energy	(Temperature × mass)	Temperature	Higher temperature objects have more thermal energy	Temperature and mass determine total thermal energy
Energy dissipation	(I^2R)	Resistance	Higher resistance means higher energy dissipation	Resistance and current determine energy dissipation

processed with similar or shorter times than the relevant dimension (color).

Some physics questions can be considered similar to the Stroop task in that they have two competing dimensions that in some cases lead to conflicting answers and in other cases lead to the same answer. For example, consider the well-known difficulties students have in answering questions about graphs (Beichner, 1994; McDermott, Rosenquist, & van Zee, 1987; Mokros & Tinker, 1987). Students often interpret graph as a physical picture and there is a general confusion about the meaning of height and slope of a graph. In particular, when students are presented with a position versus time graphs for an object (see Figure 4) and asked, "At which point does the object have a higher speed?" many incorrectly answer according to the higher point (incorrect) rather than the greater slope (correct) (McDermott et al.). The graph questions in Figure 4 ask students to compare the speeds (i.e., slopes) at two points on a graph. For this question, the relevant dimension is *slope* and the irrelevant dimension is *height*. One may construct graphs in which the higher point has the higher slope (aligned) or when the higher point has the lower slope (conflicting). Students will often consistently choose the higher point in both cases, basing their answers on the irrelevant dimension of height rather than slope. Consequently, one finds that many students answer the aligned questions correctly and the conflicting question incorrectly (Heckler, Scaife, & Sayre, 2010).

Let us now consider the previously mentioned hypothesis that, among competing plausible dimensions, students tend to choose the dimension

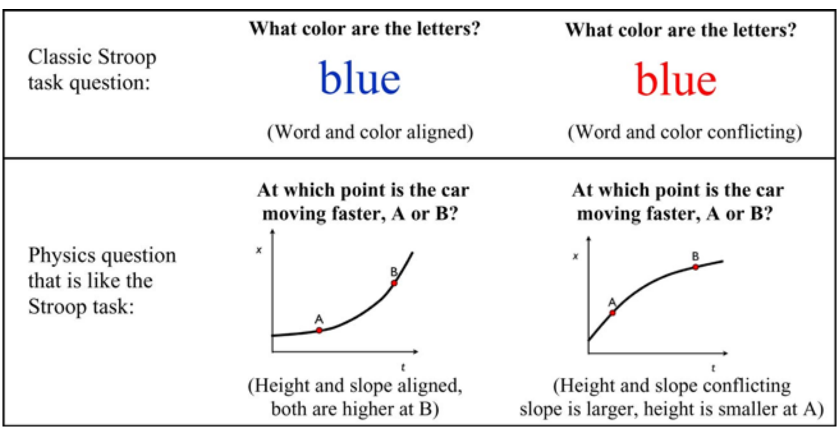

Figure 4 Analogy between a physics question and the Stroop task. Both involve competing dimensions (word vs. color, or height vs. slope), with faster times or higher accuracy for the aligned case.

that is *processedthefastest*. For the example of the graph question above, the hypothesis would predict that since height is often preferred over slope, height is processed inherently faster than slope.

In a recent study, collaborators and I confirmed the predictions of this hypothesis (Heckler & Scaife, 2010; Heckler et al., 2010). In this study (see Figure 5), we used response time as a proxy for processing time, and in speeded comparison task, we found that students could compare the heights of two points significantly faster than the slopes of two points. Furthermore, we found students, as expected, often consistently choose the point with the higher value than the point with greater slope. Perhaps most interestingly, we found that when a short (3 s) delay is imposed on answering, long enough for student to process both dimensions, the students' accuracy significantly improved. Thus, the students were capable of answering correctly, but instead they tend to answer *quickly*, and it may be this preference for answering quickly that drives students to choose the dimension that is processed the fastest.

It is worth noting that the above study also found that students answering incorrectly also answered faster. Thus, there is more than just patterns to the response content; however, there are also patterns to the response *times*. Response times on questions have been investigated in the past to eliminate the effect of guessing, thus improving the accuracy of the tests (Bridgeman et. al, 2004; Schnipke & Scrams,

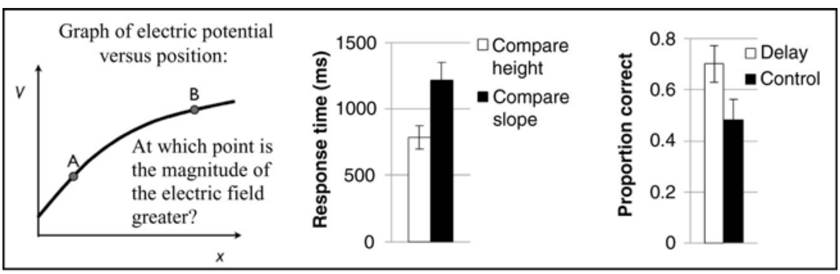

Figure 5 Competition via different processing times of relevant and irrelevant dimensions. Similar to the kinematics graph in Figure 3, the question above (figure left) elicits slope–height confusion in students. For the question above, the electric field is proportional to the slope of the line. Nonetheless, 55% of students consistently chose the higher point (incorrect) rather than the higher slope (correct). A separate speeded comparison experiment demonstrated that students inherently compare heights faster than slopes, supporting the idea that students might simply be choosing the faster processed dimension. In addition, we found that by imposing a 3-s delay on answering, time enough to process both height and slope, the proportion of correct responses increased (figure right). This supports the idea that students can answer correctly, but instead they tend to answer *quickly* (Heckler, Scaife, & Sayre, 2010).

1997; van der Linden, 2008), or to detect cheating (van der Linden & van Krimpen-Stoop, 2003). However, for the questions used in this study, students answered in patterned ways, and thus they are not guessing (e.g., see Heckler et al., 2010).

In sum, some patterns in responses to science questions may arise from lower level implicit decision criteria (e.g., answer quickly) rather than from some higher level conceptual understanding. The influence of this lower level process can be significant enough to mask a student's overall ability to determine the correct answer. The hypothesis that students will tend to base their answer on the plausible information that is processed the fastest makes testable predictions about response *times* as well as response choices. Not only is there some existing evidence to support this hypothesis, but there are also many possibilities of testing it further by designing experiments that include the capture of response time data as well as response choice data.

7.2. High salience of irrelevant cues: attentional learning

Competition can also be manifested in terms of allocation of attention. When two or more cues are present, it is often the case that one of them captures most of the attention. This phenomenon of *cue competition* is fundamental to a wide range of learning and behavior. For example, decades of studies in category learning have identified two major factors that determine which cues are learned and which are ignored among a multitude of competing cues in the environment: learners tend to learn cues that are relatively salient, predictive, or both (e.g., see Edgell, Bright, Ng, Noonan, & Ford, 1992; Hall, 1991; Trabasso & Bower, 1968). There are a number of successful models that can explain the trade-off between the salience and predictiveness of a dimension in terms of *learned attention* (e.g., Kruschke, 2001; Mackintosh, 1975). In our recent work, collaborators and I have provided evidence that when low-salient cues repeatedly compete with high-salient cues, the low-salient cues are learned to be ignored, even if they are more predictive than the high-salient cues (Heckler, Kaminski, & Sloutsky, 2008; 2011). This *learned inattention* to low-salient yet predictive information may contribute to the students' difficulties in correctly answering science questions and learning science concepts.

How can attentional learning lead to incorrect answering on science questions? Science concepts involve highly predictive cues, but these predictive cues can be of relatively low salience. For example, the acceleration of an object uniquely predicts the net force on that object, yet acceleration is often less salient than velocity (e.g., Schmerler, 1976), and students often infer the net force on an object from the velocity of the object rather than its acceleration (e.g., Clement, 1982; Halloun &

Hestenes, 1985). Thus, people's natural preference for attending to more salient cues can be problematic in science learning and performance, because these more salient cues may prevent attention to more predictive but less salient cues.

From this perspective, it is reasonable to expect that answering patterns to science questions may be strongly influenced by the format (i.e., surface features) of the question itself. This is reminiscent of a study by Chi et al. (1981), who found that novices tend to be distracted by surface features of questions rather than the underlying structure.

Therefore, I would like to consider the hypothesis that many students may simply base their response on the most salient and plausibly relevant features of a science question, even if these salient features may in fact be unrelated or contrary to the relevant scientific concept. With several competing features, the most salient one tends to automatically capture attention, with little opportunity for alternative less salient features to be considered.

For example, Figure 6 presents two questions that are based on the slope–height confusion on graph questions mentioned earlier. After relevant instruction, introductory undergraduate physics students were shown the above position versus time graphs of two cars and asked, "When are the speeds of the cars the same?" The speeds are the same at the time(s) when the slopes are the same. The score for the graph with the parallel lines is near perfect, presumably because the sameness of the slopes of the lines captured the attention. In this case then, attention was given to the relevant dimension of *slope*. However, for the crossed-lines graph, many students chose the time at which the lines intersected, presumably because this point captures attention more than the time at which the lines

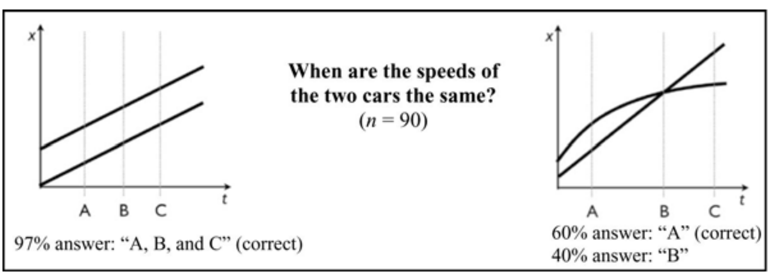

Figure 6 Hypothesized manipulation of attention on kinematics graph questions. For these questions, only the slopes of the lines at any given point are relevant, the relative heights (i.e., values) of the point on lines are irrelevant. In the parallel line graphs, almost all students answered "correctly" presumably allocating most attention to the fact that the *slopes* of the lines are equal. However, in the crossed line graph, many students presumably allocated most attention to the point at which the lines cross (*values* are equal).

have the same slope. Therefore, this is consistent with the hypothesis that students answer according to the features of the question that capture the most attention. In the case of the crossed lines, the irrelevant dimension of the *values* of the lines captured attention and led many students to the incorrect response.

One advantage of the hypothesis that bottom-up attentional allocation can play a role in incorrect answering patterns is that it can potentially be measured and tested independent of student response choices. For example, one can operationally define overt eye gaze (as measured by an eye tracker) as a measure of attention (cf. Rehder & Hoffman, 2005). Specifically, the dimension that results in the first and longest fixations is considered as the one capturing the most attention (and is the most salient). In the example in Figure 6, one would expect, for example, that attention would be fixed on the intersection of the lines on the second graph.

Note that the term *salience* is used in many contexts, and it is important to use the term consistently. Informally speaking, the salience of a cue or dimension is usually defined as the quality of standing out or being more noticeable compared to other cooccurring dimensions. Salience is often more formally regarded empirically as a quality of a cue or dimension that, separate from relative predictiveness, affects attention to (e.g., Kim & Cave, 1999; Lamy, Tsal, & Egeth, 2003) and the learning of (e.g., Edgell et al., 1992; Hall et al., 1977; Kruschke & Johansen, 1999; Trabasso & Bower, 1968) a cue relative to other present cues. Therefore, one may operationally define salience in a number of ways. For example, one may define the salient dimension as the one that attracts the most attention, as measured by eye tracking.

It is also important to keep in mind that the attention to a cue or dimension depends on the context. For example, the relative attention to two given cues can depend on the presence or absence of other cues; thus, changing the perceptual or conceptual format of the context may change the relative attention to two cues. Furthermore, attention (or salience) depends on bottom-up mechanisms operating at the level of perceptual features as well as on top-down mechanisms operating at the level of cognitive strategies, for example, controlling a search task (e.g., Egeth & Yantis, 1997). Therefore, our measures of relative attention to specific dimensions should be regarded as specific to particular questions and tasks. Nonetheless, the mechanism of attention to salient dimensions and its possible effect on student answering is general.

In a number of recent studies on enhancing multimedia learning, eye tracking results have shown that participants are distracted by irrelevant features and tend to look at more relevant areas of diagrams after instruction (Canham & Hegarty, 2010) and at relevant areas of animations if they are highlighted (Boucheix & Lowe, 2010), or if they are experts (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010). Nonetheless, more studies are needed that specifically focus on expert and novice attention to relevant and irrelevant features in physics problems known to elicit misconception-like responses. That is, the irrelevant features are more than just randomly distracting. There may be previously learned attention to incorrect dimensions that must eventually be overcome. An example of a related study is a study on spatial visualization ability and physics problem solving ability by Kozhevnikov, Motes, and Hegarty (2007) who found that students who made "graph-as-picture" misconception-like descriptions tended to look at the axes less than those students who accurately describe kinematics graphs questions; however, the eye tracking data were more ambiguous about differences between students looking at the lines in the graphs.

In sum, salient yet scientifically irrelevant features of a question compete for attention with less salient yet relevant features, and this may play an important role in incorrect student answering patterns. One may be able to observe the potential role of allocation of attention in the answering of science questions by measuring attention via eye tracking during the course of responding to question tasks. Since allocation of attention is controlled by both bottom-up and top-down processes, the challenge will be designing experiments to separate out these two kinds of processes in order to determine the extent to which automatic attentional mechanisms may be influencing response choices. Furthermore, there are a number of models of attention and attentional learning that may be applicable to misconception-like answering and may offer ways to test such models.

8. Summary and General Discussion

Why do people often answer simple scientific questions incorrectly in regular, patterned ways? This simply posed, powerful question is the driving force behind this chapter. It is one of those simple yet deeply important questions found in science like "what makes stars shine?" or "what causes cancer?" It is a question that can lead to a deeper understanding of how people think and learn and how they interact with the world.

While there is general agreement on the existence of patterns of incorrect answering to science concept questions, there is less agreement about the causes of such patterns. The phenomenon of answering patterns is certainly not monolithic and likely arises from a number of mechanisms. As discussed in Section 2, a critical issue about incorrect answering patterns is that a cause for the patterns must first be assumed in order to identify a practically relevant pattern, and then it must be based on similarities in questions perceived by the *answerer* rather than the expert who poses them.

In the field of science education, prevalent explanations for incorrect answering patterns have focused on high-level mental structures, such as misconceptions and explicit thinking. These explanations have been useful for instruction, but they have very limited predictive power. This lack of predictive power is largely due to the lack of specific models and mechanisms. Furthermore, while these explanations do not exclude the possible influence of automatic, bottom-up processes, they rarely explicitly include them.

On the other hand, in the field of cognitive science, patterns of incorrect answering on a variety of tasks both inside and outside the domain of science have also been a major empirical driving force for investigations in areas such as category learning, language learning, reasoning, decision making, and judgment. However, in these cases, models explaining the patterns of answers often include either (or both) implicit or explicit processes, and this approach has yielded some specific predictive models that have demonstrated some success.

Therefore, in this chapter, the potential role of bottom-up processes in incorrect answering patterns to science questions was explored. In particular, the phenomenon of *competition* was investigated as it relates to the answering of science questions because this phenomenon highlights the limitations of high-level mental structure models and the need for bottom-up mechanisms to explain patterns of incorrect answering.

Two examples of bottom-up mechanisms that can predict the outcome of competing dimensions were examined: relative processing time and allocation of attention to relevant and irrelevant dimensions. First, it is hypothesized that students tend to choose the dimension that is processed the fastest. Second, it is hypothesized that students tend to choose the dimension that captures the most attention (and is plausibly relevant). While specific examples of each mechanism were discussed, it still remains an open question as to how these two mechanisms may be related or interact. Data on response choices supported the predictions of the two mechanisms. For the processing time mechanism, patterns in data of the response times also supported the hypothesized mechanism. Therefore, one advantage to this proposed mechanism is that it makes testable predictions on response measures in addition to the response choice.

8.1. The relevance to science education

The multiple choice questions discussed here, as well as many of the questions used in research on science misconceptions, are similar to science questions commonly found in textbooks and classroom tests. The fact that responses to these questions may be strongly influenced by automatic bottom-up processes in many students has double-edged implications. First, it calls into question the presumed validity of these questions, since they were meant to test the extent to which students have explicit understanding of a particular scientific concept. However, these and similar questions, such as those used on well-vetted concept inventories, (e.g., Ding, Chabay, Sherwood, & Beichner, 2006; Hestenes, Wells, & Swackhammer, 1992), *have* often been validated through student interviews to ensure that the large majority of students can explicitly explain their answer choice. That is to say, these multiple choice questions do often reflect students' explicit understanding as interpreted by their explanations.

Therefore, the second implication of the influence of bottom-up processes on answering patterns is to call into question what is meant by *understanding* of a concept. *Any* claim about "student understanding" or "what a student is thinking" can only be operationally defined by or inferred from student performance on a task, be it the response to an informal question in class, to a multiple choice question on a test, or the success on a semester-long group project. If, as is suggested in this chapter and in the work on dual systems discussed in Section 6, the performance on these tasks is inevitably influenced by unconscious, automatic, bottom-up processes, then our understanding of *understanding a science concept* must include both explicit reasoning and automatic, bottom-up processes. One might say that both "System 1" and "System 2" are a necessary part of what we operationally mean by understanding a science concept, as they both may influence performance on any task relevant to the science concept. Indeed, a significant portion of expert science knowledge may be implicit (cf. Evans, Clibbins, Cattani, Harris, & Dennis, 2003).

If bottom-up processes do play an important role in understanding of a science concept, then this suggests that one should utilize methods of instruction that align these process with goals of explicit reasoning (cf. Brace et al., 2006; Goldstone et al., 2010; Kellman et al., 2010). For example, students may be better able to understand the meaning of tangent slopes on a graph if they can process them as quickly as positions on a graph. Or if one is to reason that velocity is not in the direction of force, this may be facilitated if such examples were highly available in memory due to repeated practice examples.

The goal of this chapter was to investigate the potential role of automatic, bottom-up processes in the well-known phenomenon of patterns of incorrect answering to science concept questions. It seems clear that bottom-up processes can play an important role in student answering, and disregarding such processes risks ignoring a plausible opportunity to improve our understanding of learning and understanding of scientific concepts.

ACKNOWLEDGMENTS

This research is supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305H050125.

REFERENCES

Alter, A. L., Oppenehimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576.

Bao, L., & Redish, E. F. (2006). Model analysis: Representing and assessing the dynamics of student learning. *Physical Review Special Topics: Physics Education Research*, *2*, 010103-1–010103-16.

Beichner, R. J. (1994). Testing student interpretation of graphs. *American Journal of Physics*, *62*, 750–762.

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 107–129.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning.* Hove, UK: Erlbaum.

Betsch, T., Haberstroh, S., Molter, B., & Glöckner, A. (2004). Oops – I did it again: When prior knowledge overrules intentions. *Organizational Behavior and Human Decision Processes*, *93*, 62–74.

Boucheix, J. -M., & Lowe, R. K. (2010). An eye-tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, *20*, 123–135.

Brace, J. J., Morton, J. B., & Munakata, Y. (2006). When actions speak louder than words: Improving children's flexibility in a card-sorting task. *Psychological Science*, *17*, 665–669.

Bridgeman, B., Cline, F., & Hessinger, J. (2004). Effect of extra time on verbal and quantitative GRE scores. *Applied measurement in Education*, *17*, 25–37.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*, 155–166.

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

Carey, S. (2009). *The origin of concepts.* New York: Oxford University Press.

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515–533.

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

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

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

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50*, 66–71.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Bulletin*, *97*, 332–361.

Courtney, J. R., & Hubbard, T. L. (2008). Spatial memory and explicit knowledge: An effect of instruction on representational momentum. *The Quarterly Journal of Experimental Psychology*, *61*, 1778–1784.
Deneys, W. (2006). Dual processing in reasoning: Two systems but one reasoned. *Psychological Science*, *17*, 428–433.
Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics: Physics Education Research*, *2*, 010105-1–010105-7.
diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10* (2–3), 105–225.
diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*, 843–900.
diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change. *International Journal of Science Education*, *20*, 1155–1191.
Driver, R., & Easley, J. (1978). Pupils and paradigms: a review of literature related to concept development in adolescent science students. *Studies in Science Education*, *5*, 61–84.
Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. H. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, *23*, 5–12.
Driver, R., & Erickson, G. (1983). Theories-in-action: Some theoretical and empirical issues in the study of students' conceptual frameworks in science. *Studies in Science Education*, *10*, 37–60.
Edgell, S. E., Bright, R. D., Ng, P. C., Noonan, T. K., & Ford, L. A. (1992). The effects of representation on the processing of probabilistic information. In B. Burns (Ed.), New York: Elsevier Science.
Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.
Elby, A. (2000). What students' learning of representations tells us about constructivism. *Journal of Mathematical Behavior*, *19*, 481–502.
Elby, A. (2001). Helping physics students learn how to learn. *American Journal of Physics*, *69*, S54–S64.
Evans J. St. B. T. (2000). In J.A. Garcia- Madruga et al. (Eds.), *Mental models in reasoning* (pp. 41-56). Madrid: UNED.
Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 454–459.
Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.
Evans, J. St. B. T., Clibbins, J., Cattani, A., Harris, A., & Dennis, I. (2003). Explicit and implicit processes in multicue judgment. *Memory & Cognition*, *31*, 608–618.
Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, *94*, 427–438.
Freyd, J. J. (1992). Dynamic representations guiding adaptive behavior. In D. Meyer, and S. Kornblum, (Eds.), (pp. 309–323). Dordrecht: Kluwer.
Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 126–132.
Freyd, J. J., & Jones, K. T. (1994). Representational momentum for a spiral path. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 968–976.
Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, *60*, 115–140.
Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20–29.
Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
Gilovich, T., & Griffin, D. W. (2002). Heuristics and biases then and now. In T. Gilovich., D. W. Griffin, and D. Kahneman, (Eds.), (pp. 1–18). Cambridge, UK: Cambridge University Press.
Glöckner, A., & Betsch, T. (2008). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, *3*, 215–228.
Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage*, *12*, 504–514.
Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, *2*, 265–284.
Gupta, A., Hammer, D., & Redish, E. F. (2010). The case for dynamic models of learners' ontologies in physics. *The Journal of the Learning Sciences*, *19*, 285–321.
Hall, G. (1991). *Perceptual and associative learning.* New York: Oxford University Press.
Hall, G., Mackintosh, N. J., Goodall, G., & Dal Martello, M. (1977). Loss of control by a less valid or less salient stimulus compounded with a better predictor of reinforcement. *Learning and Motivation*, *8*, 145–158.
Halloun, I. A., & Hestenes, D. (1985). Common-sense concepts about motion. *American Journal of Physics*, *53*, 1056–1065.
Hammer, D. (1996 a). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *The Journal of the Learning Sciences*, *5*, 97–127.
Hammer, D. (1996b). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics*, *64*, 1316–1325.
Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, *68*, S59.
Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences*, *12*, 53–90.
Hammer, D., Elby, A., Scherr, R., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), (pp. 89–119). Charlotte, NC: Information Age.
Heckler, A. F. (2010). Some consequences of prompting novice students to construct force diagrams. *International Journal of Science Education*, *32*, 1829–1851.
Heckler, A. F., Kaminski, J.A., & Sloutsky, V.M. (2008). Learning associations that run counter to biases in learning: Overcoming overshadowing and learned inattention. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 511–516). Austin, TX: Cognitive Science Society.
Heckler, A. F., Kaminski, J.A., & Sloutsky, V.M. (2011). Salience, learned inattention and recovery from overshadowing (in preparation).
Heckler, A. F., & Sayre, E. C. (2010). What happens between pre- and post-tests: Multiple measurements of student understanding during an introductory physics course. *American Journal of Physics*, *78*, 768–777.
Heckler, A. F., & Scaife, T.M. (2011). Processing time and competition between relevant and irrelevant dimensions on science questions.(in preparation).
Heckler, A. F., Scaife, T. M., & Sayre, E. C. (2010). Response times and misconception-like responses to science questions. In S. Ohlsson, and R. Catrambone, (Eds.), (pp. 139–144). Austin, TX: Cognitive Science Society.
Hestenes, D., Wells, M., & Swackhammer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141–158.

Hubbard, T. L. (1995). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, *2*, 322–338.
Hubbard, T. L. (1998). Representational momentum and other displacement in memory as evidence for nonconscious knowledge of physical principles. In S. Hameroff., A. Kaszniak, and A. Scott, (Eds.), (pp. 505–512). Cambridge, MA: MIT Press.
James, W. (1950). *The principles of psychology.* New York: Dover Original published 1890.
Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146–154.
Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.
Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.
Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich., D. Griffin, and D. Kahneman, (Eds.), (pp. 49–81). Cambridge, UK: Cambridge University Press.
Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, *14*, 308–312.
Kaiser, M. K., Proffitt, D. R., & Anderson, K. A. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 795–803.
Kaiser, M. K., Proffitt, D. R., Whelan, S. M., & Hecht, H. (1992). Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 669–689.
Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47–70.
Keil, F. C. (2010). The feasibility of folk science. *Cognitive Science*, *34*, 825–862.
Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, *2*, 285–305.
Kim, M. -S., & Cave, K. R. (1999). Top-down and bottom-up attentional control: On the nature of the interference of a salient distractor. *Perception and Psychophysics*, *61*, 1009–1023.
Kind, V. (2004). Beyond appearances: Students' misconceptions about basic chemical ideas: A report prepared for the Royal Society of Chemistry (2nd ed.). London: Education Division, Royal Society of Chemistry. Retrieved December 8, 2009 from http://www.rsc.org/education/teachers/learnnet/pdf/LearnNet/rsc/miscon.pdf.
Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*, 52–72.
Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, *8*, 439–453.
Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*, 549–579.
Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.
Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*, 1083.
Lamy, D., Tsal, Y., & Egeth, H. E. (2003). Does a salient distractor capture attention early in processing? *Psychonomic Bulletin and Review*, *10*, 621–629.

Leach, J., & Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science and Education*, *12*, 91–113.
Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.
MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.
Macleod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 126–135.
Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*, 309–332.
Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*, S12–S23.
McCloskey, M. (1983). Naive theories of motion. In D. Gentner, and A. L. Stevens, (Eds.), (pp. 299–324). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science*, *210*, 1139–1141.
McDermott, L. C. (2001). Physics education research: The key to student learning. *American Journal of Physics*, *69*, 1127–1137.
McDermott, L. C., & Redish, E. F. (1999). RL-PER-1: Resource letter on physics education research. *American Journal of Physics*, *67*, 755–767.
Mcdermott, L. C., Rosenquist, M. L., & van Zee, E. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics*, *55*, 503.
Mokros, J. R., & Tinker, R. F. (1987). The impact of microcomputer-based labs on children's ability to interpret graphs. *Journal of Research in Science Teaching*, *24*, 369–383.
Neisser, U. (1963). The multiplicity of thought. *British Journal of Psychology*, *54*, 1–14.
Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, *86*, 548–571.
Oberle, C. D., McBeath, M. K., Madigan, S. C., & Sugar, T. G. (2006). The Galileo bias: A naive conceptual belief that influences people's perceptions and performance in a ball-dropping task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 643–653.
Osborne, R., & Cosgrove, M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*, 825–838.
Piaget, J. (1926). *The language and thought of the child.* London: Routledge & Kegan Paul.
Piaget, J. (1952). *The origins of intelligence in children.* New York: International University Press Original published 1936.
Piaget, J. (1972). *Psychology and epistemology: Towards a theory of knowledge* (A. Wells, Trans.). Harmondsworth: Penguin.(Original published 1970).
Piaget, J. (1976). *The grasp of consciousness* (S. Wedgwood, Trans.). Cambridge, MA: Harvard University Press (Original published 1974).
Pfundt, H., & Duit, R. (2000). *Bibliography: Students' alternative frameworks and science education, 5th ed. Kiel, Germany: Institute for Education Science.*
Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*, 245–251.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.
Reber, A. S. (1993). *Implicit learning and tacit knowledge.* Oxford: Oxford University Press.

Rehder, B., & Hoffman, A. B. (2005). Eye tracking and selective attention in category learning. *Cognitive Psychology, 51*, 1–41.
Reiner, C., Proffit, D. R., & Salthouse, T. (2005). A psychometric approach to intuitive physics. *Psychonomic Bulletin and Review, 12*, 740–745.
Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naïve physics reasoning: A commitment to substance-based conceptions. *Cognition and Instruction*, *18*, 1–34.
Rohrer, D. (2003). The natural appearance of unnatural incline speed. *Memory & Cognition*, *31*, 816–826.
Sabella, M. S., & Redish, E. F. (2007). Knowledge organization and activation in physics problem solving. *American Journal of Physics*, *75*, 1017–1029.
Schmerler, J. (1976). The visual perception of accelerated motion. *Perception*, *5*, 167–185.
Schnipke, D. L., & Scrams, D. J. (1997). Modeling item-response times with a two-state mixture model. *Journal of Educational Measurement*, *34*, 213–232.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.
Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520.
Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, *46*, 1–84.
Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.
Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, *18*, 179–185.
Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*, 115–163.
Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in children: A dumb attentional mechanism? *Cognition*, *60*, 143–171.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–665.
Stavy, R., & Tirosh, D. (2000). *How students (mis-) understand science and mathematics: Intuitive rules.* New York: Teachers College Press.
Taber, K. S. (2010). Constructivism and direct instruction as competing instructional paradigms: An essay review of Tobias and Duffy's Constructivist instruction: Success or failure? Education Review, 13(8) Pages 1–45. Retrieved September 20, 2010 from http://www.edrev.info/essays/v13n8index.html.
Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research.* New York: Wiley.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, *33*, 5–20.
van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant response patterns in computerized adaptive testing. *Psychometrika*, *68*, 251–265.
Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, *1*, 205–221.

Vosniadou, S. (1994). Capturing and modeling the process of change. *Learning and Instruction*, *4*, 45–69.

Vosniadou, S. (1996). Towards a revised cognitive psychology for new advances in learning and instruction. *Learning and Instruction*, *6*, 95–109.

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

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky, Vol. 1: Problems of general psychology.* New York: Plenum Press.(Original published 1934).

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*, 3878–398.