COGNITION AND INSTRUCTION, *19*(4), 479–541


# How Students Understand Physics Equations

Bruce L. Sherin
*School of Education and Social Policy*
*Northwestern University*

What does it mean to understand a physics equation? The use of formal expressions in physics is not just a matter of the rigorous and routinized application of principles, followed by the formal manipulation of expressions to obtain an answer. Rather, successful students learn to understand what equations say in a fundamental sense; they have a feel for expressions, and this guides their work. More specifically, students learn to understand physics equations in terms of a vocabulary of elements that I call *symbolic forms*. Each symbolic form associates a simple conceptual schema with a pattern of symbols in an equation. This hypothesis has implications for how we should understand what must be taught and learned in physics classrooms. From the point of view of improving instruction, it is absolutely critical to acknowledge that physics expertise involves this more flexible and generative understanding of equations, and our instruction should be geared toward helping students to acquire this understanding. The work described here is based on an analysis of a corpus of videotapes in which university students solve physics problems.

With the cognitive revolution in educational studies has come a call for teaching for understanding. We now believe that it is not enough for students to be able to solve problems or answer certain questions; in addition, we want students to have something called understanding that lies behind and undergirds such abilities. In domains in which formal mathematical expressions are prevalent, these issues come into particularly sharp focus. We do not want meaningless symbol manipulation; if students use symbolic expressions, we want them to use the symbols with understanding.

---

Requests for reprints should be sent to Bruce L. Sherin, School of Education and Social Policy, Northwestern University, 2115 N. Campus Drive, Evanston, IL 60208–2610. E-mail: bsherin@northwestern.edu

In many respects, this viewpoint seems eminently sensible. The notion of meaningless symbol use is clear, at least intuitively, and seems undesirable. However, if we adopt this position, we are nonetheless faced with some important and difficult questions. Unquestionably, understanding involves more than the ability to write literal expressions from memory and to perform rote manipulations. But what, exactly?

I attempt to answer this question within the domain of physics. Mathematical expressions are part of the very language of physics. Equations are used to contain and convey fundamental aspects of content; physicists read (and presumably understand) equations in written texts, and they compose equations to express physical notions.

This work is fundamentally concerned with how physics equations are understood. I want to be able to answer questions such as these: How do physicists (and physics students) take a conceptual understanding of some physical situation and express that idea in an equation? In what conceptual terms are physics equations understood?

It is possible that the understanding of physics equations is of a limited sort. For example, consider the equation $v = v_o + at$. In what sense is this equation understood by people with some physics expertise? One possibility is that an understanding of this equation, even for an expert physicist, does not extend beyond knowing the conditions under which the equation can be used and how it can be used. In that case, an expert may have an extensive knowledge of typical circumstances of use, perhaps associated with problem types (Chi, Feltovich, & Glaser, 1981), as well as a knowledge of how this equation, along with other equations, is employed in those circumstances to obtain a result. If this portrayal accurately describes expertise, our hypothetical expert would understand that $v = v_o + at$ gives the velocity ($v$) for the case of constant acceleration ($a$). However, the fact that velocity is exactly given by $v_o + at$ —two terms, separated by a plus sign—does not express any particular meaning for the expert. The details of the equation—the exact arrangement of the symbols that it comprises—are not taken as expressing a meaning.

In this article, I argue that the understanding of physics equations can greatly surpass this limited type of understanding. At least for experts and some students, the particular arrangement of symbols in an equation expresses a meaning that can be understood. Later, I present numerous episodes and the results of a detailed analysis to support this view. By way of introduction, I first present a brief example to give a flavor of what I have observed.

Part of my research involved asking pairs of university physics students to solve a selection of problems. One problem was the air resistance problem, which is about a ball dropped under the influences of gravity and air resistance. When the ball is first dropped, it initially accelerates—it goes faster and faster. Eventually, because of the opposition of air resistance, the speed of the ball approaches a constant value—its terminal velocity. The problem is: How does the terminal velocity differ for two objects of identical size and shape, one of which is twice as massive as the other?

One way to understand this physical situation is that there are two forces acting on the ball: a force down from gravity and a force from air resistance acting upward. Initially the ball is not moving, and the force from air resistance is zero. Gradually the ball speeds up, and the force from air resistance grows until, at terminal velocity, it is exactly equal to the force from gravity (see Figure 1).

Later, I look in detail at how students' wrote equations to describe this physical situation and how they solved the problem. My basic argument is that, at least in some cases, the students built equations from a sense of what they wanted the equations to express. For example, some students wrote equations of the form $F_{up} = F_{down}$ to describe the situation shown in Figure 1d, in which a constant (terminal) velocity has been reached. They understood these equations, I maintain, as expressing the simple notion that there are two influences in balance acting on the object. They did not recall this expression and recognize that it was appropriate for this circumstance, and they did not derive it from a physical principle. Instead, they constructed an equation to express a particular understanding of the physical situation: Two opposing influences balance each other.

Also, at times, students wrote equations to describe the more general circumstance, shown in Figure 1b and 1c, in which the ball is gradually speeding up. The expressions they wrote were of the form (outcome) = (downward influence) – (upward influence). For example, some students wrote expressions of the form

$$a = \frac{F_{down}}{m} - \frac{F_{up}}{m}$$

I contend that these students understood these equations as two opposing influences competing to produce a result.

There is a rich story to be told about the conceptual content of equations. The meaning of physics equations is understood in terms of a specific vocabulary, including notions such as balancing and competing influences, and I lay out that vocabu-

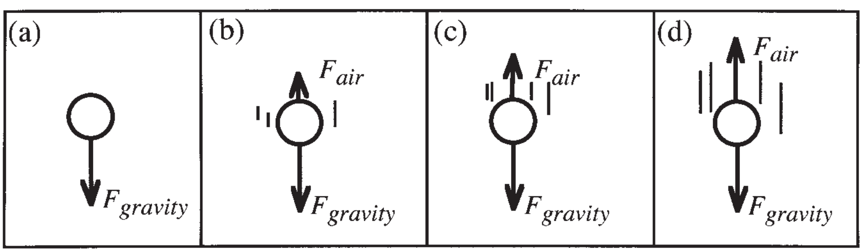


FIGURE 1 (a) An object is dropped. See (b) and (c). As it speeds up, the opposing force of air resistance increases. (d) At terminal velocity, the forces of air resistance and gravity are equal.

lary in what follows. Stated informally, my hypothesis is that successful physics students learn to express a moderately large vocabulary of simple ideas in equations and to read these same ideas out of equations. I call the elements of this vocabulary *symbolic forms*. Each symbolic form associates a simple conceptual schema with an arrangement of symbols in an equation. Because they possess these symbolic forms, students can take a conceptual understanding of some physical situation and express that understanding in an equation. Furthermore, they can look at an equation and understand it as a particular description of a physical system.

This hypothesis has many implications for how we should understand the nature of thinking in physics as well as in other domains involving extensive use of mathematical expressions. If I am right, the use of formal expressions in physics is not just a matter of the rigorous and routinized application of principles, followed by the formal manipulation of expressions to obtain an answer. Rather, successful students learn to understand what equations say in a fundamental sense; they have a feel for expressions, and this understanding guides their work. This portrayal is a more flexible and creative image of what is possible with physics equations than has thus far been presented in the research literature (e.g., Larkin, McDermott, Simon, & Simon, 1980a, 1980b).

These observations, in turn, have implications for what should be taught and learned in physics classrooms. From the point of view of improving instruction, it is absolutely critical to acknowledge that physics expertise involves this type of flexible and generative understanding of equations. We do students a disservice by treating conceptual understanding as separate from the use of mathematical notations.

This article has three major parts. First, I explain and argue for symbolic forms as an important component of physics knowledge. This part of the article establishes that symbolic forms exist and, more generally, that the particular variety of understanding associated with symbolic forms is necessary to understand fully the use of equations in physics. The second major section elaborates on the symbolic forms framework, describing the knowledge system that comprises symbolic forms and providing a list of the symbolic forms necessary to account for the data corpus. In essence, I present a partial dictionary of the vocabulary in terms of which physics equations are understood. I conclude with a discussion of implications of the framework for education and connections to research in mathematics education.

## SYMBOLIC FORMS: INTRODUCTION AND ARGUMENT

I begin this section by discussing some relevant components of physics education research. Then, following this literature review, I describe briefly the data corpus and method for the study. Next, I introduce symbolic forms by looking in moderate detail at example episodes. Finally, I compare my explanation of the example episodes with competing hypotheses from the literature.

## Equation Understanding in Physics Education Research

In this section, I describe two major branches of the research on physics learning, with an emphasis on the relevance of each to equation understanding. First, I discuss prior research on physics problem solving. This body of work is clearly relevant to my analysis because it involves studying individuals working with physics equations. I then briefly discuss research on naive physics knowledge, the knowledge of the physical world that individuals gain prior to formal instruction. This research is relevant because ultimately I argue that naive physics knowledge provides part of the conceptual basis in terms of which equations are understood.

A central point of this review is that there has been a sizable gap between these two areas of research. Although research in physics problem solving has made substantial progress in modeling the problem solving process, it has missed some fundamental varieties of equation understanding. In contrast, work in naive physics has emphasized qualitative types of understanding—it is concerned with conceptual understanding—but has not told us how this conceptual understanding relates to problem solving. Thus, one purpose of this research is to bridge the gap between these two areas of research.

In my research I have also drawn extensively on mathematics education research that bears directly on the question of how equations are understood. For example, there is a long history of work on the solving of mathematics word problems (Reed, 1998). In addition, researchers have contrasted two broad stances that students are said to adopt toward equations: the process and object stances (Herscovics & Kieran, 1980; Kieran, 1992; Sfard, 1987, 1991). I discuss this research later, after presenting my analysis.

*Research on physics problem solving.* Research on problem solving in physics has focused on the process by which students—and, to a lesser extent, expert physicists—solve textbook physics problems (Bhaskar & Simon, 1977; Chi et al., 1981; Larkin et al., 1980a, 1980b). In its earliest incarnations, these studies of physics problem solving were strongly influenced by research into human problem solving conceived more generally (e.g., Simon, 1975). The notion was that moves in the space of equations written on a sheet of paper could be treated much like moves in the solution of a puzzle, such as the Tower of Hanoi.

More specifically, physics problem solving was seen to work such as this: The student or expert reads the problem and takes note of the quantities given in the problem and the quantities that are needed. Then the problem solver writes, from memory, equations that contain these various quantities as variables. Once the equations are written, the search can begin; the equations are manipulated until a route from the given quantities to what is needed is obtained. Note that the search metaphor underlying these models does not leave much room for understanding of any sort. Certainly, these models do not include the sort of equation understanding that I described previously.

Models of physics problem solving that involve something more similar to understanding on the part of the solver do exist. One important variety is *schema-guided forward inference* models (Priest & Lindsay, 1992). Briefly, these models are based on the hypothesis that experts possess a number of schemata directly associated with physical principles, such as Newton's second law or the conservation of momentum (Chi et al., 1991; Larkin, 1983). Within these models, the problem is first recognized as being of a certain type, and then a schema that guides the solution is invoked. These schemata contain equations and techniques for instantiating the equations in particular circumstances.

With regard to the understanding of equations, these models go beyond the search models. They are not limited to heuristic search in an equation space; rather, the solver knows the path that the solution will follow and how to use equations to follow that path. The solver may thus be said to understand equations in the sense of knowing where and how to use them. The problem solver does not, however, understand why individual expressions have their particular makeup.

Strictly speaking, I am not proposing an alternative to these problem-solving models. Problem-solving models are concerned with giving an account of the steps taken by a solver in reaching a solution. They are often concerned with the equations that are written and in what order. Therefore, many of the associated research questions have to do with issues of control, such as whether novices forward chain or backward chain. In contrast, I focus on modeling the understanding of physics equations.

*Naive physics knowledge and using equations.* Contemporaneous with the early research into physics problem solving, researchers came to realize that students enter physics instruction with quite a lot of knowledge about the physical world, and that this knowledge has a strong impact on their learning of formal physics. The study of this prior knowledge—often called intuitive physics or naive physics knowledge—has become an endeavor of its own. Since the initiation of naive physics research, a substantial body of research has accumulated, working from a variety of perspectives.[1] This research includes straightforward listings of difficulties and misconceptions (Clement, 1983; McDermott, 1984), attempts to argue that students possess their own theories of physics (McCloskey, 1983), as well as cognitive analyses that attempt to break naive physics knowledge down into smaller constituents of understanding (diSessa, 1993).

[1]See the extensive bibliography compiled by Pfundt and Duit (1994) for a dramatic illustration of the quantity of this work.

It is not immediately obvious that naive physics research can help, in any way, with the program of describing how physics equations are understood. In general, research into naive physics does not address solving textbook physics problems. The main focus of naive physics research—and naive science research, generally—is on how students understand the physical world prior to instruction as well as on what parts of this qualitative naive understanding remain after instruction. Little research has elucidated the manner in which naive physics or other conceptual knowledge is employed during the use of equations to solve problems.

Thus, these two major programs of research—physics problem solving and naive physics—have remained largely disjointed. In contrast, I attempt to show an absolutely fundamental connection, and I claim that naive physics provides an important part of the conceptual basis in terms of which equations are understood.

Some researchers have attempted to bridge the gap between naive physics and physics problem-solving research. Clement (1994), looking at the reasoning of experts, argued that certain varieties of nonformal reasoning—what he called intuitions and imagistic simulations—play a crucial role in expert thinking. However, Clement did not describe in any detailed manner how these varieties of nonformal reasoning connect to the use and understanding of equations. In other important work, Ploetzner (Ploetzner & Spada, 1993; Ploetzner & Van Lehn, 1997) looked at the role of qualitative reasoning (cf. deKleer & Brown, 1984; Forbus, 1984) in physics problem solving. Similar to Clement, Ploetzner filled in somewhat different pieces of the research puzzle than I address here. Most important, the qualitative knowledge identified by Ploetzner consisted primarily of qualitative versions of expert knowledge and was tied to specific expert domains (e.g., motion with friction). In contrast, I argue that equations can be understood in terms of more basic and generic intuitions that cut across expert domains.

## The Data Corpus, In Brief

I present here a brief account of the methods and data corpus for this study. A more thorough description of methods may be found in Appendix B.

*Participants.* Participants in the study were university students enrolled in a third-semester physics course for engineers. This group was selected because they were at an intermediate level of expertise; they were not experts, but they had substantially more experience than students enrolled in their first physics course. The analysis presented here was based on observations of five pairs of students, selected randomly from a pool of volunteers. Four pairs consisted of two men, and the final pair consisted of one man and one woman.

*Tasks.* I observed each pair of students as they stood at a white board and solved physics problems. Each pair participated in four to six sessions, with a session lasting between 1 and 1.5 hr. The total time of all observations was approximately 27 hr (see Table 1).

It was essential that the tasks were neither too difficult nor too easy for students. If the tasks were too difficult, students would have had little notion of how to proceed, and they may well have been reduced to manipulating symbols in hope of discovering an answer. If the tasks were too easy, students may have been able to produce a solution quickly without having to think about what to do. The tasks were primarily standard textbook problems but included a few more unusual tasks. The analysis reported here focused on a subset of the tasks, corresponding to 11 of the total 27 hr of videotape. I refer to this portion of the data corpus as the focus corpus, and the great majority of the examples are drawn from these data. The tasks included in the focus corpus are listed in Table 2.

*Analysis.* The analysis on which my framework and arguments are based is primarily qualitative. I repeatedly viewed transcripts of the focus corpus and iteratively adapted the framework to the corpus. During each iteration, I paid particular attention to episodes in which equations were used with understanding. Analysis procedures are described more fully in Appendix B.

In this article, I describe and argue for the framework primarily by presenting examples from the focus corpus and explaining how the examples can be understood in terms of the framework. In addition, I contrast my framework with alternatives and argue that it can better account for the range of phenomena than existing frameworks do. In this regard, my main point is that there are phenomena that simply have not been addressed in prior work and that are at least partly explained by the existence of symbolic forms.

## Symbolic Forms Explained

*Mike and Karl invent a new equation.* To introduce and define symbolic forms, I begin with a somewhat unusual episode that clearly illustrates some of the key phenomena. In this episode, two students were working on the shoved block problem (see Table 2), in which a block resting on a table is given a hard shove. The

TABLE 1
Time Spent by Students in Experimental Sessions

| | *Mike and Karl* | *Jack and Jim* | *Mark and Roger* | *Jon and Ella* | *Alan and Bob* | *Total* |
|---|---|---|---|---|---|---|
| No. of Sessions | 4 | 5 | 5 | 6 | 6 | 26 |
| Total hours | 5.0 | 4.5 | 6.0 | 6.0 | 5.5 | 27.0 |
| Focus hours | 2.0 | 1.5 | 2.0 | 3.0 | 2.5 | 11.0 |

TABLE 2
Tasks Included in the Focus Corpus

| *Problem Name* | *Problem Text* |
|---|---|
| 1. Shoved block | A person gives a block a shove so that it slides across a table and then comes to rest. Talk about the forces and what's happening. How does the situation differ if the block is heavier? |
| 2. Vertical pitch | (a) Suppose a pitcher throws a baseball straight up at 100 mph. Ignoring air resistance, how high does it go? (b) How long does it take to reach that height? |
| 3. Air resistance | For this problem, imagine that two objects are dropped from a great height. These two objects are identical in size and shape, but one object has twice the mass of the other object. Because of air resistance, both objects eventually reach terminal velocity: (a) Compare the terminal velocities of the two objects. Are their terminal velocities the same? Is the terminal velocity of one object twice as large as the terminal velocity of the other? (hint: keep in mind that a steel ball falls more quickly than an identically shaped paper ball in the presence of air resistance). (b) Suppose that there was a wind blowing straight up when the objects were dropped, how would your answer differ? What if the wind was blowing straight down? |
| 4. Mass on a spring | A mass hangs from a spring attached to the ceiling. How does the equilibrium position of the mass depend upon the spring constant (k) and the mass (m)? |
| 5. Stranded skater | Peggy Fleming (a one-time famous figure skater) is stuck on a patch of frictionless ice. Cleverly, she takes off one of her ice skates and throws it as hard as she can: (a) Roughly, how far does she travel? (b) Roughly, how fast does she travel? |
| 6. Buoyant cube | An ice cube, with edges of length L, is placed in a large container of water. How far below the surface does the cube sink? (?ice = .92 g/cm$^3$; ?water = 1 g/cm$^3$) |
| 7. Running in the rain | Suppose that you need to cross the street during a steady downpour and you don't have an umbrella. Is it better to walk or run across the street? Make a simple computation, assuming that you're shaped like a tall rectangular crate. Also, you can assume that the rain is falling straight down. Would it affect your result if the rain was falling at an angle? |

block slides across the table and eventually comes to rest because friction between the block and the table slows the block down. The problem is this: Suppose that the same experiment is done with a heavier block and a lighter block. Assuming that both blocks are started with the same initial velocity, which block travels farther?

The correct answer is that the heavier and lighter blocks travel precisely the same distance (see Figure 2). I discuss how a pair of students were dissatisfied with this counterintuitive result, leading them to invent their own brand of physics and

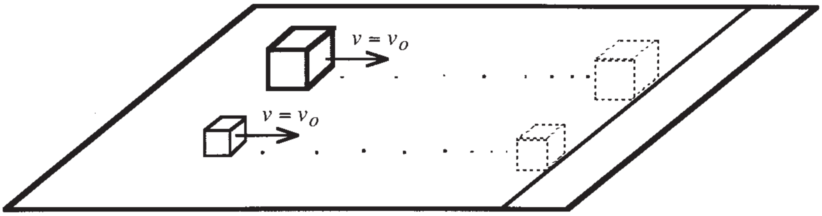


FIGURE 2 A heavier and a lighter block slide to a halt on a table. If they are shoved so as to have the same initial speed, they travel exactly the same distance.

write a completely new equation. This construction is the key feature of this example: Two students begin with a conceptual specification for an equation and then construct a novel expression based on that specification.

Among all the students participating in this study, two competing intuitions were expressed concerning the shoved block. The more common intuition was that the heavier block should travel less far because it presses down harder on the table and thus experiences a stronger frictional force slowing it down. The alternative intuition was that the heavier block should travel farther because it has more momentum and is thus harder to stop. In fact, it is because these two effects precisely cancel that the heavier and lighter block travel the same distance.

The pair of students to be discussed here, Karl and Mike,[2] adopted the latter intuition, that heavier blocks should travel farther, and they clung to this intuition, even after using equations to derive the correct result. This was especially true of Karl, who argued that it was simply common sense that the heavier block should travel farther:

> Karl: Yeah, that's true. But I still say that the heavier object will take the longer distance to stop than a lighter object, just as a matter of common sense.

Mike was somewhat less worried by the counterintuitive result but was willing to stop and discuss what may be changed in their solution. Eventually, Karl came to the conclusion that "in real life" the coefficient of friction may vary with weight in such a way that heavier blocks travel farther.

> Karl: I think that the only thing that it could be is that the coefficient of friction is not constant. And the coefficient of friction actually varies with the weight.

[2]All student names are fictitious.

The coefficient of friction is a parameter that determines the magnitude of the frictional force. Students are taught in introductory physics classes that this parameter can be treated as a constant depending only on the nature of the materials involved, such as whether the objects are made out of wood or steel. Given their counterintuitive result, however, Mike and Karl were willing to reconsider the assumption that this parameter is a constant and set out to compose their own expression for the coefficient of friction. Karl began by expressing their core assumption:

Karl: I guess what we're saying is that the larger the weight, the less the coefficient of friction would be.

To obtain their desired result—that heavier masses travel farther—they needed the coefficient of friction to decrease with increasing weight. In other words, friction must work less on a heavier mass. Over the next 9 min, Karl and Mike laid out some additional properties they wanted their new expression to have. An excerpt from that discussion follows (see Appendix A for a key to the notations used in transcripts).

Karl: Well yeah maybe you could consider the frictional force as having two components. One that goes to zero and the other one that's constant. So that one component would be dependent on the weight. And the other component would be independent of the weight.
Mike: So, do you mean the sliding friction would be dependent on the weight?
Karl: Well I'm talking about the sliding friction would have two components. One component would be fixed based on whatever it's made out of. The other component would be a function of the normal force. The larger the normal force, the smaller that component. So that it would approach a—it would approach a finite limit. It would approach a limit that would never be zero, but the heavier the object, the less the coefficient of friction at the same time. (…)
Mike: I don't remember reading that at all [laughs].
Karl: See, I'm just inventing my own brand of physics here. But, if I had to come up with a way—if I had to come up with a way that would get this equation to match with what I think is experience, then I would have to—that's what I would have to say that the=
Mike: Actually, it wouldn't be hard to=
Karl: =the coefficient of friction has two components. One that's a constant and one that varies inversely as the weight.

Karl outlined what he wanted from the new expression, stating the coefficient of friction should have two components, one independent of weight and the other decreasing with increasing weight. After some fits and starts, Mark and Karl settled on this expression for the coefficient of friction:

$$\mu = \mu_1 + C\frac{\mu_2}{m}$$

Here $m$ is the mass and $\mu_1$, $\mu_2$, and $C$ are constants introduced to flesh out the structure of the expression. This expression has some problems—most notably that the coefficient of friction tends to infinity as the mass goes to zero, but this expression captures much of what Mike and Karl wanted to include: It decreases with increasing mass, and it approaches a constant as the mass becomes infinitely large.

The most important thing to note about this expression is that it is not found in any textbook. Mike and Karl had not learned it previously, and they did not simply derive it by manipulating equations they already knew; instead, they built this expression from an understanding of what they wanted it to do. As Karl said, he was "just inventing [his] own brand of physics."

This behavior poses difficulties for existing models of physics problem solving. There is no obvious way to explain Mike and Karl's ability to write this equation in terms of the manipulation of known equations or even in terms of derivation from given principles. How then did Mike and Karl accomplish this construction? For example, how did they know to write + instead of × between the two terms? How did they know to put the $m$ in the denominator?

*What is a symbolic form?* At the core of my explanation of how Mike and Karl constructed their novel equation—and more generally, how conceptual content and equations can be connected—is the hypothesis that as people develop physics expertise, they acquire knowledge elements I call symbolic forms. A symbolic form involves two components.

1. Symbol template. Each form is associated with a specific symbol template, and forms are knowledge elements that include the ability to see mathematical expressions in terms of this template. For example, two symbol templates that students learn to see in expressions are

$$\square = \square \qquad\qquad \square + \square + \square \ldots$$

The symbol template on the left is a template for an equation in which two expressions are set equal. These two expressions can be of any sort—the □ can be filled in with any expression. Similarly, the symbol template on the right is for an expression in which a string of terms are separated by plus signs.

2. Conceptual schema. Each symbolic form also includes a conceptual schema.[3] The particular schemata associated with forms are relatively simple structures, involving only a few entities and a small number of simple relations among these entities. In this respect, they are similar to diSessa's (1993) p-prims and Johnson's (1987) image schemata, which are also presumed to have relatively simple structures. Although I have drawn on both p-prims and image schemata, the view I describe shares more of the commitments of diSessa's construct.

Roughly speaking, the schema is the idea to be expressed in the equation, and the symbol template specifies how that idea is written in symbols. Thus, in essence, my central hypothesis is that students learn to associate meaning with certain structures in equations.

The nature of symbolic forms can best be illustrated by returning to Mike and Karl's invention of an equation for the coefficient of friction. Mike and Karl's core specification for this equation was that the coefficient of friction should consist of two parts: one constant and one varying inversely with the weight.

Karl: =the coefficient of friction has two components. One that's a constant and one that varies inversely as the weight.

Two forms are evident in this specification, the first of which I call *parts-of-a-whole*. The symbol template for parts-of-a-whole is two or more terms separated by plus (+) signs. The conceptual schema has entities of two types: a whole composed of two or more parts. This schema can be seen behind Karl's statement that the coefficient of friction consists of "two components." Furthermore, the equation has the structure specified in the symbol pattern; it has two terms separated by a plus sign.

To describe various symbol patterns, I employ a shorthand notation. For parts-of-a-whole, I write the symbol pattern as this:

*parts-of-a-whole* [□ + □ + …]

The □ refers to a term or group of symbols, typically a single symbol or a product of two or more factors. The brackets around the whole pattern indicate that the entity corresponding to the entire pattern is an element of the schema. The shorthand notation for symbol patterns is summarized in Appendix B.

The second form in Karl's statement is *proportionality minus* or *prop–*. The schema for prop– involves two entities: One quantity is seen as inversely propor-

[3]I use the phrase *conceptual schema* to distinguish these schemata from problem-solving schemata. In problem-solving schemata, the schema specifies steps in a solution. Here, in contrast, the schema specifies physical entities and relations among those entities.

tional to a second quantity. In Karl's specification for this expression, the second component of the coefficient of friction is inversely proportional to the weight. As he said, one component is constant and one "varies inversely as the weight."

The symbol pattern for prop– is relatively simple. When a student sees prop– in an expression, all they are attending to is the fact that some specific symbol appears in the denominator of an expression. I write it as this:

$$\textit{prop-} \qquad \left[\frac{\cdots}{\ldots x \ldots}\right]$$

*X* usually corresponds to a single symbol. In Mike and Karl's expression, *m* is the symbol of interest and, in line with the prop– form, it appears in the denominator of the second term.

Two additional forms also played a role in Mike and Karl's building of their novel equation. The first has to do with the *C* in front of the second term, which was inserted by Mike and Karl almost as an afterthought at the end of the episode. I take this to be a case in which the coefficient form is in play.

$$\textit{coefficient} \qquad [x\ \square]$$

In coefficient, a single symbol or number, usually written on the left, multiplies a group of factors to produce a result. I describe typical features of the coefficient part of the schema later in this article. One feature is that the coefficient is frequently seen as controlling the size of an effect. This, roughly, was Mike and Karl's reason for inserting the coefficient. As Karl wrote, he commented that "You find out that uh you know the effect is small, or you find out the effect is large or whatever so you might have to screw around with it." For Mike and Karl, then, the factor of *C* was a way to tune their expression.

Finally, the μ= that appears on the left side of Mark and Karl's equation constitutes an application of one of the simplest and most ubiquitous forms: the identity form.

$$\textit{identity} \qquad x = [\ldots]$$

In the identity form, a single symbol, usually written on the left, is separated from an arbitrary expression by an equal sign.

These explanations in terms of symbolic forms provide one kind of answer to the questions I posed earlier: How did these students know to write + instead of × between the two terms? How did they know to put *m* in the denominator? I have argued that, to compose a new equation, these students had to develop a specification for that equation in terms of symbolic forms. Once this was done, the symbol templates that are part of these forms specified, at a certain level of detail, how the expression was to be written.

*Symbolic forms and more typical problems.* In this section, I present and analyze a second problem-solving episode involving activity with physics equations that is more typical for introductory physics courses. The episode involves the air resistance problem, in which a ball is dropped under the influence of air resistance. Previously, I explained that the ball gradually speeds up until it reaches terminal velocity (see Figure 1). Here, I examine how Mike and Karl solved this problem.

Mike and Karl began by agreeing that what they called the “upward” and “downward” accelerations of a dropped object must be equal when it reaches terminal velocity.

Karl: So, we have to figure out (0.5) what, how do you figure out what the terminal velocity is.
Mike: Well, okay, so, you have a terminal velocity, You have terminal velocity, that means it falls at a steady rate. Right, which means the force opposing it
Karl: It means it has an upward acceleration that’s equal to the downward acceleration.
Mike: Ri:::ght.

Following this discussion, Mike drew the diagram reproduced in Figure 3, talking as he drew.

Mike: Air resistance. You have a ball going down with a certain velocity. A for:::ce,, You reach terminal velocity when air resistance, the force from the air resistance is the same as the force from gravity. But, we also know that force of air resistance is a function of velocity. [w. f(v)=] Right?

In this last comment, Mike reiterated the notion that the influences of air and gravity must be equal at terminal velocity but in terms of opposing forces, rather than

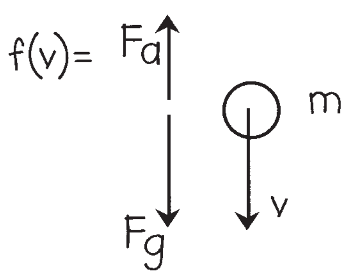


FIGURE 3 Mike and Karl’s air resistance diagram.

opposing accelerations. Finally, in the next passage, Mike and Karl composed an expression for air resistance:

$$a(t) = -g + \frac{f(v)}{m}$$

Mike: So, at least we can agree on this and we can start our problem from this scenario [g. Indicates the diagram with a sweeping gesture]. Right? Okay? So, at any time,, At any time, the acceleration due to gravity is G, and the acceleration due to the resistance force is F of V over M [w. g + f(v)/m]. This is mass M [w. m next to the mass on the diagram].
Karl: Ah:::.
Mike: Okay, // now they're opposing so it's a minus.
Karl: // (So) as your mass goes up,, [g. m in equation]
Mike: So, this is negative G. [g. down, makes a negative sign in front of the g] Positive direction. You have negative G plus F of V over M. That's your acceleration at any time. Right?
Karl: Well, wait. You want them to have,,
Mike: This is the acceleration at any time T. [w. a(t)= next to the expression]

Mike and Karl's conceptualization of the situation took shape as an equation. This is the conceptualization discussed in the introduction; they understood the situation in terms of two competing influences, and they directly embodied this understanding in an equation. My explanation of this construction involves three forms: competing terms, opposition, and identity.

| | |
|---|---|
| *competing terms* | □ ± □ ± … |
| *opposition* | □ – □ |
| *identity* | $x$ = […] |

The identity form also appeared in the previous episode. Competing terms and opposition require brief explanations. Competing terms indicates that each of the competing influences corresponds to a term in the equation. This is the root of Mike and Karl's understanding: There are terms, corresponding to influences, that compete to "have their way." Opposition refers to a situation in which the two influences compete in a specific manner; they are opposing influences. This conceptual opposition is associated with the use of a negative sign in the equation. "Now they're opposing so it's minus."

My analysis suggests that even in this relatively typical problem-solving episode, a qualitative conceptual account was partly driving the students' work with equations, with symbolic forms mediating the interaction between conceptual understanding and symbolic expressions. However, my interpretation is not incontrovertible. In particular, it is possible that the students were, in some manner, deriving their expressions by applying Newton's laws or other principles, rather than building the equations from an understanding of what they wanted to express. I take up this issue in the next section.

## An Alternative Model

There are aspects of Mike and Karl's equation that my analysis in terms of symbolic forms does not explain. In particular, I have not said anything about how the students knew to write $f(v)/m$ for the term corresponding to air resistance. Likely this has to do with the equation $F = ma$, an equation that Mike and Karl certainly remember. They stated that the force of air resistance depends on velocity, which led them to write this force as $f(v)$. Then, because they were talking about opposing accelerations, they used $F = ma$ to write the acceleration as $a = f(v)/m$.

At the least, then, forms cannot be the whole story in this construction episode—and they will not be the whole story in most construction episodes. More formal considerations and remembered equations must also play a role. In my rendition of this episode, a remembered equation was used to fill in one slot in a symbol pattern. However, the episode is also open to some overall interpretations that are significantly different than the one I provided. Most notably, one may argue that Mike and Karl simply applied the equation $F_{tot} = ma$. This equation states that the total of all the forces on an object is equal to the mass of the object multiplied by its acceleration.

Mike and Karl's work may be seen as a straightforward application of this equation. One may argue that they simply have instantiated this equation for the situation, writing the sum of the forces and then dividing this sum by the mass to get the acceleration. This alternative suggestion is a version of the schema-guided forward inference models discussed previously.

This sort of model is an important alternative and merits discussion. My intent is not to argue that this alternative view is incorrect or even that it competes with the one I am developing. In fact, I believe that a more complete description of the knowledge involved in expert symbol use would include knowledge of multiple sorts, including knowledge more closely tied to formally recognized principles, as well as symbolic forms. However, I argue that some student behaviors are better explained by symbolic forms and that symbolic forms—and the related behavior—persist into expertise, complementing more formal aspects of knowledge.

A version of this alternative (schema-guided forward inference) view was worked out by Larkin (1983). According to Larkin, experts possess a number of schemata that are directly associated with physical principles. When one of these schemata is invoked, it drives the solution of a problem. Larkin provided two examples, the forces schema and the work-energy schema. These schemata are closely related to physical principles as they would be presented in a physics textbook. For example, Larkin wrote that the forces schema "corresponds to the physical principle that the total force on a system (along a particular direction) is equal to the system's mass times its acceleration (along that direction)" (p. 82). This is essentially Newton's second law, $F_{tot} = ma$. In this view, problem solving involving the forces schema is seen to work as follows:

1. The problem solver identifies the forces acting on an object.
2. These forces are added together to yield an expression for the total force, $F_{tot}$.
3. This expression for $F_{tot}$ is substituted into $F_{tot} = ma$.
4. The equation is manipulated to solve for the desired quantities.

The equation $F_{tot} = ma$ is a sort of template that allows the construction of equations not identical to those seen before. There are particular limits on its range of applicability, however: The entities in the sum must be forces, and they must be set equal to *ma*.

Clearly, as I have already suggested, Larkin's account may be used to describe Mike and Karl's work on the air resistance task. In this view, Mike and Karl first wrote the force on each object; $-mg$ is the force of gravity, which is negative because it acts in the downward direction, and the force of air resistance is simply taken to be $f(v)$. These forces are then summed and set equal to *ma*:

$$ma = -mg + f(v)$$

(Presumably, Mike and Karl do this step mentally because it does not appear as one of the equations they write.) Finally, they divide this equation by the mass, *m*, to produce an equation for *a*.

One crucial difference between Larkin's view and my own is that her schemata are closely tied to physical principles (in addition to some more global differences in orientation). For Larkin, it is crucial that the influences in competition are forces because Newton's laws only involve forces. In contrast, in my view, the situation is generically schematized as competing influences—a notion that cuts across physical laws. Many types of entities, other than forces, may compete.

Models such as Larkin's are implicit and explicit in many studies of physics problem solving. For example, Chi et al. (1981), in their well-known work on the categorization of physics problems, explicitly adopted a similar position. According to these researchers, when an expert looks at a physics problem, a basic set of features cues a problem-solving schema associated with a fundamental physics principle. (These schemata correspond to the problem categories that Chi et al. discovered.) Then, according to Chi et al., the "knowledge—both procedural and declarative—contained in the schema is used to further process the problem in a more or less top-down matter" (p. 150).

Because of its emphasis on principles I refer to Larkin's account as the principle-based schemata view. In addition to the emphasis on principles, there is also an important difference in orientation and goals between these other programs and my own. The schemata in Larkin and Chi's models are problem-solving schemata, in which the structure of the schema corresponds to the steps one would take in solving a problem. In contrast, the schemata associated with symbolic forms are conceptual schemata, in which the structure corresponds to an understanding of conceptual relations or structure in the world.

*Deviations from the principle-based model in the air resistance example.* I now take a closer look at some episodes to argue that they can be better explained by the symbolic forms account. On the air resistance task, Mike and Karl alternated between describing the situation in terms of competing accelerations and describing it in terms of competing forces. This alternation indicates that Mike and Karl were aware that force and acceleration are formally interchangeable via $F = ma$. This alternation also suggests, however, that a view of the situation in terms of opposing influences is a core driving element of these students' work and that the specific nature of these influences is not of utmost importance to them at all stages of the solution attempt. This observation cuts to the heart of the difference between the forms model and the principle-based schemata model. I believe that Mike and Karl's work was driven at least in part by an understanding tied to a notion of opposing influences that cuts across principles, and not by a schema tied solely to forces and Newton's laws.

Furthermore, Karl's use of the terms *upward acceleration* and *downward acceleration* would be considered inappropriate by many physicists. To a physicist, the term *acceleration* is usually reserved for describing actual changes in the speed of objects, and not for virtual changes that may be canceled out by other virtual changes. Thus, Mike and Karl's work departed somewhat from a strict and careful application of traditional methods that follow a universal schema associated with physical principles.

Finally, as they worked, Mike and Karl had interpretations readily available. For example, Mike said that one acceleration is "opposing" the other. That such interpretive comments are readily available suggests that these notions may have played a role in the students' work.

*Other deviations from the principle-based model.* Additional types of deviations from the principle-based model appeared in the work of other students on the air resistance task. Some pairs of participants never wrote an expression for the total force. Instead they drew a diagram and then jumped directly to equating the upward and downward forces acting on the falling object. For example, Mark and Roger drew the simple diagram shown in Figure 4 and then wrote the equation $R = mg$, where $R$ is the force from air resistance and $mg$ is the force from gravity. Similarly, Jon and Ella jumped straight to writing $F_g = Cv^2$, where $F_g$ is the force from gravity and $Cv^2$ is their expression for the force from air resistance. After

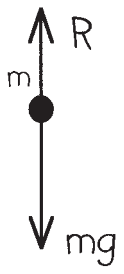


FIGURE 4 Mark and Roger's diagram for the air resistance task.

writing this equation, Ella commented "The terminal velocity is just when the—I guess the kind of frictional force from the air resistance equals the gravitational force?"

I believe that, rather than following the forces schema, these students understood the situation in terms of the balancing form. In this symbolic form, two competing influences are seen as equal and opposite. Furthermore, each of these influences is associated with one side of an equation, as indicated in the symbol template.

*balancing* $\square = \square$

These balancing solutions are not quite aligned with the principle-based schemata model. If the forces schema were appropriate for understanding Jon and Ella's work, they should first find the total force, $F_{tot} = F_{grav} + F_{air}$, and then substitute this total force into $F_{tot} = ma$ to get $F_{grav} + F_{air} = 0$, if we presume that $a = 0$. Instead, by immediately writing an expression of the form $F_{grav} = F_{air}$, Ella seems to have skipped over some steps: The total force does not appear anywhere in the solution, and the presumption that the acceleration is zero is not visible. This skipping of steps was common. In fact, in one task involving balanced forces, the buoyant cube problem, every pair of participants began by writing a balancing expression.

Of course, students may have done the missing steps mentally, without expressing them verbally or in writing. However, the expressions produced by these two techniques differ by a sign; it is not possible to simply manipulate $F_{grav} = F_{air}$ to get $F_{grav} + F_{air} = 0$. This suggests, at the least, that the students were not simply doing manipulations of this sort mentally.

In addition, I saw direct evidence that students were not consciously following principles but leaving them unsaid. In my final session with each pair, I engaged them in open-ended discussion during which I asked them to revisit balancing solutions they had given in previous sessions. I asked questions to investigate what they were leaving unsaid. Few students could explain how balancing expressions of the form $F_{grav} = F_{air}$ followed from basic principles such as $F_{tot} = ma$. Furthermore, when directly presented with such explanations, students seemed surprised. These episodes are described in more detail in Sherin (1996).

*Understanding can play multiple roles in problem solving.* There are other aspects of solutions to the air resistance problem that are difficult to account for within the principle-based framework. As their final result for the air resistance problem, all the participant pairs ultimately derived an expression roughly of the form:

$$v = \frac{mg}{k}$$

Once they derived this equation, every pair of participants made statements about its meaning. For example, after Alan and Bob derived this equation, Alan commented "So now, since the mass is in the numerator, we can just say velocity terminal is proportional to mass and that would explain why the steel ball reaches a faster terminal velocity than the paper ball." Concerning the same relation, Mark said simply "So the terminal velocity of the heavier object would be greater." These interpretations are based on the proportionality (prop+) form; the students recognized that they derived an expression directly proportional to the mass.

The important point is that understanding plays a new role in problem solving. Earlier, I argued that understanding can drive the construction of initial equations, but because symbolic forms allow students to interpret existing equations as well as construct new ones, understanding can also play a role in the intermediate and final stages of problem solving. In these examples, the students drew implications from the final result; they saw that a heavier object will have a higher terminal velocity. In addition, they likely made judgments as to the reasonableness of the expressions they derived. If the final equation were not sensible, they may well have revisited some aspects of their solution.

Principle-based models do not obviously account for this sort of behavior. If there is anything similar to equation understanding in those models, it is localized to setting up the problem. In contrast, in the forms model, understanding can potentially play a number of roles in problem solving. For example, forms-based understanding can allow one to (a) construct expressions, (b) reconstruct partly remembered expressions, (c) judge the reasonableness of a derived expression, and (d) extract implications from a derived expression.

*Students construct novel expressions.* I raise here the most important evidence supporting the forms hypothesis. Previous examples have shown episodes that cannot be explained by any version of the principle-based hypothesis. For example, it would be difficult for Larkin's (1983) schema view to explain Mike and Karl's ability to construct their own equation for the coefficient of friction, $\mu = \mu_1 + C\frac{\mu_2}{m}$. There are no physical principles that would lead to the construction of this equation, and there are no textbook equations from which this expression can be derived.

Thus, at least some students have abilities that are not described by the principle-based model. Furthermore, this observation should inform any account of their thinking on more usual tasks, such as simple systems involving balanced forces. I could, for example, attempt to introduce special-purpose schemata to explain some common special cases, such as situations with equal and opposite forces, but given the existence of form-like knowledge to explain behavior in other contexts, it may be unnecessary to introduce special-purpose schemata. Presuming that the form-like knowledge explains the deviations allows for a more economical model.

In other words, once we accept that students such as Mike and Karl have abilities not captured by the principle-based schemata model, we have to live with that observation in all our accounts of physics problem solving. When we analyze Mike and Karl's work on the more familiar air resistance task, there is no reason to believe that their form-related capabilities will not be employed. Suppose, for a moment, that something such as Larkin's (1983) forces schema is responsible for Mike and Karl's work on the air resistance task. Because Mike and Karl have the ability to see meaningful patterns—forms—they may, at any time, be seeing these patterns in equations that they write.

*Some final caveats.* In concluding this general description of symbolic forms, some tempering of my claims is necessary. In the episodes presented previously, principle-based knowledge was clearly playing some role. For example, even if we accept that some students understand the buoyant block problem in terms of balancing, we still have to explain how they knew it was forces that needed to balance. This competence almost certainly depends, in part, on students' ability to recognize problem types and on knowledge of what principles apply to those problem types.

Thus, knowledge employed during the use of physics equations, in its full and finished expert form, must have components that are tied to physics principles. In fact, I believe that some of this knowledge is similar to principle-based schemata. Thus, I do not mean to suggest we should discard the principle-based schemata view, but I hope that I have shown cracks in this model—some pretty big cracks, in fact—that make room for other important classes of knowledge and that greatly broaden the range of student activity that we can expect and explain.

## SYMBOLIC FORMS: THE KNOWLEDGE SYSTEM

In this section, I paint a broader picture of the knowledge system that comprises symbolic forms. I first situate this work within a larger program called knowledge analysis, which provides many of my beginning assumptions. Then I discuss, in a systematic manner, the full range of symbolic forms uncovered by the analysis.

Throughout this discussion, I work from a large number of short excerpts, rather than from longer example episodes, with the goal of providing a more complete picture of the scope of the system and the range of phenomena covered.

## Introduction to Knowledge Analysis

The knowledge analysis program has been exemplified in such projects as diSessa's (1993) work on p-prims, which I describe here, as well as diSessa and Sherin's (1998) attempts to provide theoretical constructs to replace the notion of concept for describing scientific thinking and learning. Knowledge analysis focuses on describing the nature and development of extensive knowledge systems. It is less concerned than many other endeavors in cognitive science with the articulation of general structures or mechanisms that cut across many domains of knowledge. For example, a program in knowledge analysis would not argue for the existence of schemata as a form of knowledge. Instead, it would focus on listing and describing the contents of the specific schemata that constitute understanding in a particular domain. Stated simply, knowledge analysis is more about the content of knowledge than its form.

Furthermore, the interest in extensive knowledge systems signals an intent to focus on moderately large systems for which the number of elements may be substantial. In fact, a basic presumption underlying this program is a belief in a complex mental ecology consisting of diverse types of elements. According to this viewpoint, much behavior is likely to involve interactions among this complex ecology of elements; thus, understanding behavior requires, to some extent, attending to this complexity.

Knowledge analysis often attempts to analyze knowledge systems in terms of the primitive constituents of a complete cognitive analysis. For example, imagine a researcher conducting a clinical interview asking a student, "What would happen if I were to toss a ball straight up into the air?" and the student answering, "The ball will slow down as it goes up." Many claims could be made about what this statement implies about what the student knows. For example, perhaps the student has the belief that a tossed ball slows down. Or, perhaps, the student has the belief that when a force acts on an object, opposite to the direction in which it moves, it slows down. Either belief could potentially account for the student's statement.

Of course, without further data there is no way to choose between alternatives of this sort. More dramatically, additional data may reveal that the student's statement cannot be explained in terms of a single belief. Indeed, there is no reason to assume a one-to-one correspondence between any question we choose to ask and the knowledge elements possessed by students. The more general assumption is that any single answer given by a student is generated by an ensemble of knowledge elements.

Knowledge analysis goes down to the level of this ensemble, to the level of primitive elements. This account in terms of primitive elements is adequate when it can explain the full range of behaviors of interest. In fact, this is one sense of what it means to say that the elements are primitive; no further decomposition of the knowledge is necessary to account for all cognitive behavior across the relevant range of contexts.

It is possible to perform productive analyses by looking at a more narrow range of behaviors and producing analyses in terms of beliefs. For example, Ploetzner and VanLehn (1997) produced a strong and intriguing analysis of the role of qualitative physics knowledge in problem solving that is self-consciously at this level. In a footnote, they made the following comment:

> Although we use rules in our model, we are not claiming that people have rules in their heads …. Clearly, to say that the student "knows" the rule and "applies" it is a vast simplification. Yet science sometimes needs such simplifications in order to understand nature's regularities. (p. 171)

Although these simplifications can sometimes be productive, analyses at the primitive level are a critical component of educationally oriented cognitive studies for a number of reasons. Many important characteristics of expertise become evident only at the primitive level of analysis. For example, the analysis in this article has helped to bring forward the informal and intuitive nature of physics problem solving.

More important, the constructive nature of learning is most likely to be made evident at the primitive level. Narrowly tuned explanations at the level of beliefs are often brittle; that is, they are not good at capturing gradual change (they do not bend—they break). Suppose, for illustration, that after an instructional intervention, a student gives a different answer to the question about the tossed ball. If this hypothetical student's knowledge is modeled as a single belief, then the only ways to describe this change are to say that one belief has been replaced by another or that the single belief has been modified. In contrast, if the knowledge is modeled as an ensemble of elements, learning may best be captured in terms of small changes in when and how existing elements are employed (Smith, diSessa, & Roschelle, 1993).

This strong constructivist principle is at the heart of knowledge analysis and is part and parcel of the complex systems viewpoint. Taken as a whole, the picture of learning painted here is one of gradual changes in a complex ecosystem of elements. The description of knowledge analysis given in the preceding paragraphs is summarized by the principles listed in Table 3. These are the principles that guided the analysis presented here.

## A Theory of Naive Physics: DiSessa's Model of the Sense-of-Mechanism

Next, I describe diSessa's (1993) research on phenomenological primitives, conducted within the knowledge analysis framework. This work is presented here because it illustrates knowledge analysis research and because it has provided a basis for my own analysis.

TABLE 3
Heuristic Principles for Knowledge Analysis
(Adapted From diSessa, 1993, and Sherin, 1996)

| *Name of Principle* | *Summary* |
|---|---|
| Principle of content over form | It is important to list the specific content of knowledge systems, rather than only articulating general mechanisms and knowledge types. |
| Assumption of extent and diversity | Expert knowledge involves a large system of knowledge elements of diverse types. Attending to this complexity is essential for understanding behavior. |
| The search for primitives | It is important to provide accounts in terms of a core set of knowledge elements that can explain behavior across a wide range of types of cases. |
| Strong principle of constructivism | Learning involves gradual changes to the system of elements. |

diSessa (1993) set out to describe a portion of intuitive physics knowledge that he called the *sense-of-mechanism.* He contended that elements of the sense-of-mechanism form the base level of intuitive explanations of physical phenomena. As an example, diSessa asked what happens when a hand is placed over the nozzle of a vacuum cleaner, resulting in an increase in the pitch of the vacuum cleaner. According to diSessa, people explain this phenomenon by saying that, the vacuum cleaner has to work much harder because the hand is getting in the way of its work. This explanation relies on a certain primitive notion: Things have to work harder in the presence of increased resistance to produce the same result.

diSessa's (1993) program involved the identification of these primitive elements of knowledge, which he called *phenomenological primitives,* or p-prims (the p-prim used to explain the vacuum cleaner phenomenon is known as Ohm's p-prim). The word *phenomenological* appears in the name of this knowledge element in part because p-prims are presumed to be abstracted from experience. Ohm's p-prim, for example, is abstracted from the many experiences the physical world in which things have to work harder in the presence of increased resistance. diSessa listed a variety of p-prims, including p-prims pertaining to force and motion (e.g., force as mover, dying away) and p-prims pertaining to constraint phenomena (e.g., guiding, abstract balance).

According to diSessa (1993), the sense-of-mechanism consists of a number of these elements. He listed 29 individual p-prims and suggested the existence of many more. Thus, diSessa described a complex system involving a moderately large number of primitive elements. These are features of a typical knowledge analysis program.

DiSessa also discussed the cognitive mechanisms governing the behavior of the sense-of-mechanism. In a given situation, individual p-prims are activated by a relatively simple mechanism: They are simply recognized. A person looks at a physical situation and sees it in terms of a particular p-prim, such as Ohm's p-prim. In addition, p-prims are connected in a network so that the activation of one p-prim may make the activation of others more or less likely. With learning, the factors that are associated with activation may change, leading to changes in the circumstances in which particular p-prims are applied.

## Features of the Symbolic Forms Knowledge System

In a number of respects, diSessa's (1993) model of the sense-of-mechanism provided a starting point for the model described here. Most important, the model I am proposing is intended to be structurally similar to diSessa's. In this section, I lay out additional details concerning my model of the knowledge system that consists of symbolic forms.

*A moderately large system of elements.* The symbolic forms knowledge system consists of a moderately large number of elements. I discuss about 20 symbolic forms in the comprehensive list presented in the next section; a full account would include many more. Furthermore, there is no simple structure to this collection. The organization is not hierarchical, and it does not align in any simple manner with physical principles.

*Intermediate level of abstraction.* A single symbol pattern can be associated with more than one symbolic form. For example, expressions of the form $A + B$ can be understood in terms of either parts-of-a-whole or competing terms. Thus, in some respects, symbolic forms are more specific than the symbolic expressions that they interpret; a symbolic form adds semantics to an equation. Nonetheless, symbolic forms represent a significantly abstracted understanding of the world. To understand the blooming buzzing confusion of the world in terms of symbolic forms is to abstract or strip away some details of the situation. Thus, the elements in the symbolic forms system constitute an intermediate level of abstraction—more abstract than a full, rich understanding of the world but less abstract than equations.

*Mechanism.* With regard to mechanism, the key question concerns when and how a symbolic form is cued to an active state. I describe only simple mechanisms, and those only in broad strokes. First, a form can be activated by being recognized in an equation. A person looks at an equation and sees the symbolic form there. Alternatively, an understanding of the physical situation to be described may activate a form. The situation is understood, and then this understanding in some manner activates a particular symbolic form.

In addition to being activated, forms—through their schema component—support specific inferences. Generic capacities act on the conceptual schema to produce inferences. For example, prop– permits some simple inferences: If the denominator varies, the whole varies. If the denominator increases, the whole decreases. The reader is encouraged to read the example episodes provided with this sort of inference in mind.

In describing these simple mechanisms, I am not proposing a full model of the problem-solving process or of other extended behaviors involving physics equations. Rather, the behaviors I model are on a smaller time scale, the time scale at which a person looks at an equation and understands it.

*Genesis.* Where do symbolic forms come from? When and how do they develop? The genesis of symbolic forms is largely a topic for other articles, but I make several brief points here. Some symbolic forms are generated from experiences working with physics equations. Students solve many physics problems, and they watch instructors solve problems. From these experiences, students may somehow abstract symbolic forms. However, symbolic forms are not developed entirely during physics instruction. I later present evidence that some symbolic forms originate in early mathematical experiences. In fact, working explicitly within the symbolic form framework, Izsák (1999, 2000) traced in detail the development of particular symbolic forms in the context of mathematical work. Furthermore, some symbolic forms (e.g., balancing) have similar conceptual content to specific p-prims. Thus, some symbolic forms may develop from associated p-prims.[4]

## Forms: A Semi-Exhaustive List

In this section, I present a semi-exhaustive list of symbolic forms, along with a number of brief examples to illustrate them. Because the collection of forms is not organized according to the principles of physics—and because it apparently has no other simple structure—there are no shortcuts for describing the breadth of forms. The broad scope provided by this overview is essential, with important and subtle points along the way. Because this list accounts only for the corpus employed in this study, I describe it as semi-exhaustive. A different set of tasks or a different population of students would have resulted in a different list of symbolic forms.

Although there is no simple structure to the forms knowledge system, I nonetheless organize forms in clusters (diSessa, 1993). This organization is primarily for rhetorical purposes—not to reflect any psychological grouping of the ele-

[4]For more detail on the relation between forms and p-prims, see Sherin (1996).

ments. However, within a given cluster, the various schemata tend to have entities of the same or similar ontological type. For example, forms in the competing terms cluster are primarily concerned with influences. In addition, the forms in a cluster tend to parse equations at the same level of detail. For example, some forms deal with equations at the level of terms, whereas others involve individual symbols.

A list of the forms uncovered by the analysis of the data corpus is provided in Table 4. I discuss a sampling of items from this table. A brief description of every item is given in Appendix B.

*Competing terms.* To a physicist or moderately advanced student, an equation is sometimes seen as an arrangement of terms that conflict and support

TABLE 4
Symbolic Forms by Cluster

| *Cluster* | *Symbolic form* | *Symbol pattern* |
|---|---|---|
| Competing terms cluster | Competing terms | □ ± □ ±□ … |
| | Opposition | □ – □ |
| | Balancing | □ = □ |
| | Canceling | 0 = □ – □ |
| Terms are amounts cluster | Parts-of-a-whole | [□ + □ + □ …] |
| | Base ± change | [□ ± Δ] |
| | Whole – part | [□ – ▫] |
| | Same amount | □ = □ |
| Dependence cluster | Dependence | $[\ldots x \ldots]$ |
| | No dependence | $[\ldots]$ |
| | Sole dependence | $[\ldots x \ldots]$ |
| Coefficient cluster | Coefficient | [$x$ □] |
| | Scaling | [$n$ □] |
| Multiplication cluster | Intensive–extensive | x × y |
| | Extensive–extensive | x × y |
| Proportionality cluster | Prop+ | $\left[\frac{\ldots x \ldots}{\ldots}\right]$ |
| | Prop– | $\left[\frac{\ldots}{\ldots x \ldots}\right]$ |
| | Ratio | $\left[\frac{x}{y}\right]$ |
| | Canceling (b) | $\left[\frac{\ldots x \ldots}{\ldots x \ldots}\right]$ |
| Other | Identity | x = … |
| | Dying away | $\left[e^{-x\ldots}\right]$ |

or that oppose and balance. The competing terms cluster contains the forms related to seeing equations in this manner, as terms associated with influences in competition. Frequently, although not always, these influences are forces in the technical sense.

A number of forms from this cluster have already been discussed. During Mike and Karl's work on the air resistance task, they composed this equation:

$$a(t) = -g + \frac{f(v)}{m}$$

Their construction of this equation involved the competing terms and opposition forms. Mike and Karl schematized the situation as involving two opposing influences, sometimes understood as forces and sometimes as accelerations. The opposition form was associated with Mike's insertion of a minus sign: "Now they're opposing so it's a minus." In addition, we encountered some episodes involving the balancing form. During their work on the air resistance task, some students saw the forces of air and gravity as balancing after terminal velocity was reached.

In another episode involving the balancing form, which I have not yet presented, a symbolic form was applied inappropriately. This example concerned the shoved block problem, which involves two objects shoved so that they both start moving with the same initial speed. The question is, given that one block is twice as heavy as the other, which block travels farther? As discussed previously, the correct answer is that the two blocks travel the same distance. The derivation of this conclusion can be quite short (refer to Figure 5). After the shoved block is released, the only force acting on it is the force of friction, which is $F_f = \mu mg$. As the only force, it alone is responsible for any acceleration the block undergoes as it slides along the table. For this reason, this force can simply be substituted for $F$ in the equation $F_{tot} = ma$ to give the equation $\mu mg = ma$. Then the $m$ that appears on each side of the equation can be canceled to produce the final expression $\mu g = a$. This final expression for acceleration does not involve the mass; regardless of the mass, the shoved block always travels the same distance.

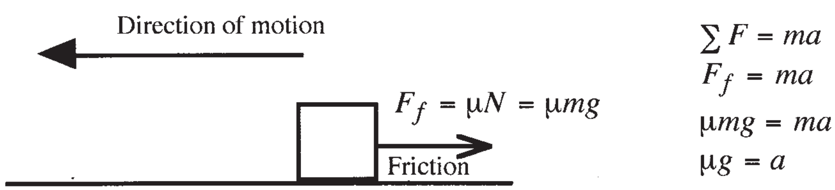


FIGURE 5 A solution to the shoved block problem.

In the episode of interest, Alan and Bob began by producing a solution similar to the one presented in Figure 5. However, after they were done, Alan pointed to the equation $\mu mg = ma$ and added this brief addendum to their explanation: "What Bob didn't mention is that this is when the two forces are equal and the object is no longer moving because the force frictional is equal to the force it received."

In this instance, Alan understands this expression in terms of the balancing form. When Alan looked at the equation $\mu mg = ma$, he saw two equal and opposing influences: the "force frictional" and the "force it received." Each influence was associated with one side of an equation. More specifically, Alan seemed to see this equation as specifying a condition for an end state, in which the block comes to rest because two influences balance. However, once the block is released, only one force acts on it: the force due to friction. Thus, Alan's interpretation of this equation in terms of balancing is inappropriate.

The existence of this sort of error supports the importance of my analysis. Without a framework focused on equation understanding, errors of this sort may not even be noticed. The forms framework helps draw attention to difficulties in equation understanding and provides a way to talk about these difficulties. For a brief discussion of the last form in this cluster, canceling, see Appendix B.

*Proportionality.* When a physicist or physics student looks at an equation, the line that divides the numerator of a ratio from its denominator is a major landmark. Forms in the proportionality cluster involve seeing individual symbols as either above or below this landmark. One of these forms, prop–, was involved in Mike and Karl's construction of an expression for the coefficient of friction:

$$\mu = \mu_1 + C\frac{\mu_2}{m}$$

Karl: The coefficient of friction has two components. One that's a constant and one that varies inversely as the weight.

Karl's statement specified that the second component of their expression should be inversely proportional to the mass. Thus, the prop– form was invoked, and the *m* symbol was written in the denominator of the second term.

Some additional examples of prop+ and prop– can be found in Jack and Jim's work on Focus Task 1, in which a mass hangs motionless at the end of a spring. The work of these students is reproduced in Figure 6. They seemed to find the problem to be relatively straightforward and solved it in about 2½ min, obtaining the following result:

$$x = \frac{mg}{k}$$

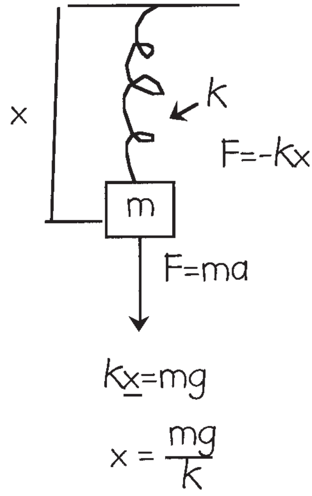


FIGURE 6 Jack and Jim's solution to the spring problem.

In this expression, *m* is the mass, *g* is the acceleration due to gravity, and *k* is the spring constant. Jim then commented that this is a completely sensible result:

> Okay, and this makes sort of sense because you figure that, as you have a more massive block hanging from the spring, then your position *x* is gonna increase, which is what this is showing. [g.~ *m* then *k*] And that if you have a stiffer spring, then your position *x* is gonna decrease. [g. uses fingers to indicate the gap between the mass and ceiling] That why it's in the denominator. So, the answer makes sense.

In the first part of this interpretation, Jim saw the expression in terms of the prop+ form, with *x* seen as directly proportional to the mass: "As you have a more massive block hanging from the spring, then your position *x* is gonna increase." In the second part of the interpretation, the expression is seen through the lens of prop–: "If you have a stiffer spring, then your position *x* is gonna decrease."

There were a number of instances in which students interpreted final expressions of this sort by announcing proportionality relations. For example, earlier I discussed students' interpretations of the final result of the air resistance problem:

$$v = \frac{mg}{k}$$

Alan: So now, since the mass is in the numerator, we can just say velocity terminal is proportional to mass and that would explain why the steel ball reaches a faster terminal velocity than the paper ball.

Mark: So the terminal velocity of the heavier object would be greater.

In all these examples, forms are involved in the understanding of existing equations, rather than in the composition process. In short, the students are providing interpretations of the equations that they derive. The mere fact that students can and do interpret equations is worth emphasizing once again. Equations are not just derived and used to obtain numerical results; students have something to say about what some of the equations mean.

*Forms and the limits of qualitative reasoning.* This discussion of the proportionality cluster provides an appropriate context to consider an important issue—what I call the qualitative limits of forms. First, I look again at the air resistance problem. As I have discussed, two forces act on the dropped object in the air resistance task: gravity and a force from air resistance. All students know an expression—a force law—that specifies the force of gravity, $F_g = mg$. However, most students do not know an expression for the force due to air resistance. Although some may have seen such an expression once or twice, it is rare for students to have committed an equation to memory. Thus, the students have to construct their own expression for the force of air resistance.

The accepted answer—the answer you will find in an introductory textbook—is $F_{air} = kv^2$ where $v$ is the current velocity of the object and $k$ is a constant.[5] All the students in this study stated that the force of air resistance should be proportional to velocity and wrote either that expression or $F_{air} = kv$.

$$F_{air} = kv$$

Bob: Okay, and it gets f-, it gets greater as the velocity increases because it's hitting more atoms [0.5] of air.

$$R = \mu v$$

Mark: So this has to depend on velocity [g. $R$]. That's all I'm saying. Your resistance—the resistor force depends on the velocity of the object. The higher the velocity the bigger the resistance.

The important issue here is that the prop+ form does not specify whether the force of air resistance should be $kv$ or $kv^2$. Some students ran headlong into this difficulty, realizing that understanding that the force of air resistance must be proportional to the velocity was not sufficient to determine what relation they should

[5]A more complete expression could include other parameters, such as the cross-sectional area of the dropped object. However, for this task it is sufficient to presume that these other parameters are part of the constant, $k$, and that this constant is the same for both objects.

write. Jack went so far as to list the possibilities. In the following, the symbol $\propto$ means "is proportional to."

$$\begin{aligned} F_u &\propto v \\ &\propto v^2 \\ &\propto v^3 \end{aligned}$$

Jack: Somehow they're related to the velocity but we're not sure what that relationship is. All we can say is that $F_U$ is either proportional to $V$ or to $V$ squared or to $V$ cubed.

This type of ambiguity seems to be a fundamental property of forms; there are limits to the specificity with which forms describe equations. This is true when both reading and writing equations. If an equation is seen in terms of prop+, it is simply not consequential whether the proportionality relation is linear or quadratic.

The students in this study came up against these limits in a number of cases and not only in the context of the air resistance task. For example, when working on the spring task, Mike and Karl could not remember whether the force due to a spring is $F = kx$ or $F = ½\, kx^2$.[6] Although they knew that the expression needed to be consistent with prop+, this was not sufficient to distinguish between these two equations.

$$F = kx \quad ; \quad F = ½\, kx^2$$

Karl: Okay, now, qualitatively, both *kx* and half *kx* squared do come out to be the same answer because as (…) looking at it qualitatively, both half—both half *kx* squared and *kx*, um, you know, increase as *x* increases.

Furthermore, this was no fleeting difficulty. Mike and Karl struggled with this question for about 25 min, but to no avail. Up to the level of their specification, these equations were indistinguishable. As Karl said, both "increase as *x* increases."

Karl expressed the intuition that two expressions are "qualitatively" equivalent. Of course, what it means for two expressions to be qualitatively the same is far from clear. Interestingly, however, there is a relevant body of literature that attempts to make this intuition precise. This literature describes what the researchers call "qualitative reasoning about physical systems" (deKleer & Brown, 1984; Forbus, 1984).

Paradoxically, the focus of qualitative reasoning is what physicists call physical quantities. The literature on qualitative reasoning makes specific claims about how people reason about physical quantities to describe and predict behavior. One hy-

[6]For reference, the equation for the potential energy stored in a spring is $U = ½\, kx^2$.

pothesis concerns the nature of the information about a quantity that people encode. The range of possible values a quantity can attain is seen as divided into a set of discrete regions, with the boundaries at physically interesting places. For qualitative reasoning, only the region the quantity is in and its direction of change are relevant.

Later researchers in this field hypothesized about how people propagate knowledge about changes in one quantity to changes in another. Of particular interest are proportionality relations, which Forbus (1984) called qualitative proportionalities. For example, if two quantities are qualitatively proportional and one quantity is increasing, then the second quantity is also increasing. As in the case of forms, it is not appropriate to ask whether the relation is linear, second order, or otherwise. Thus, the knowledge system formalized by the qualitative reasoning literature appears to possess the same limits as forms. As with prop+, a qualitative proportionality relation between force and velocity says only that force increases with increasing velocity (or decreases with decreasing velocity), and that is the limit of the specification.

*Terms are amounts.* This cluster includes forms, similar to those in the competing terms cluster, that address expressions at the level of terms. However, in the terms are amounts cluster, terms are treated not as influences but as quantities of generic stuff. Thus, whereas signs in the competing terms cluster are commonly associated with directions in physical space, signs in this cluster generally signal adding on or taking away from the total amount of stuff.

The first form in this cluster, parts-of-a-whole, played a role in Mike and Karl's construction of their novel expression for the coefficient of friction. The friction had two components that together constituted the coefficient of friction. However, that episode does not provide the clearest and most illustrative example of the parts-of-a-whole form; in particular, it is hard to see the coefficient of friction as a quantity of stuff. Indeed, Mark and Karl's episode is a marginal instance of parts-of-a-whole. For that reason, I present here an example of a more prototypical instance of the parts-of-a-whole form. The episode involves the running in the rain problem:

> Suppose that you need to cross the street during a steady downpour and you do not have an umbrella. Is it better to walk or run across the street? Make a simple computation, assuming that you're shaped like a tall rectangular crate. Also, you can assume that the rain is falling straight down. Would it affect your result if the rain was falling at an angle?[7]

A diagram from Alan and Bob's work on this task is shown in Figure 7. After a brief discussion, Alan and Bob agreed that rain would strike two surfaces of the

[7]This problem is adapted from *The Flying Circus of Physics* (Walker, 1975).

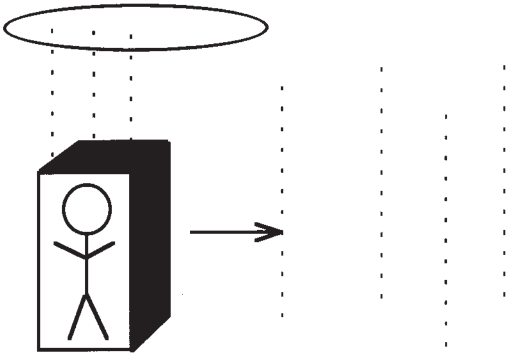

FIGURE 7 Alan and Bob's diagram for the running in the rain problem.

crate—the top and front sides (these two surfaces are shaded in Alan and Bob's diagram). They went on to construct an expression for the total rain that strikes the person. This expression consisted of two parts: one corresponding to what strikes the top of the crate and one corresponding to what strikes the front.

$$\textit{total rain} = \frac{\#\,\textit{raindrops}}{s} + C$$

Bob: There's two—two sources of rain. Okay, rain that you walk into and rain that drops on you. Okay. Walk or run into. Okay, so, the number of rain per second [g. #raindrops/s] is how much rain hits your head. … if you're a crate and you—you move, move into some rain. Since you move from here to here [g. two points on the board], there's—since there's rain coming down at a constant rate, there's always rain in front of you. Okay, so there's raindrops in front of you. So if you walk forward you hit it, hit that rain. Okay. And—and—I mean, you're gonna hit some rain when you walk forward.

Alan and Bob's expression follows from parts-of-a-whole; the total amount of rain is seen as consisting of two parts. The first term corresponds to the rain striking the top of the crate, and the second term corresponds to the rain that strikes the front of the crate. In this case, thinking of the total rain as a total amount of stuff is relatively natural. One can almost imagine a container filling up with all this rain, flowing in through the two possible surfaces.

This expression is not a syntactically correct equation. In the place of the first term, there are just some units—the number of raindrops per second. Furthermore, these are not even the correct units for this term. Given the forms view, such partially complete expressions are to be expected. Because forms such as parts-of-a-whole only describe equations at a certain level of description, it makes sense that expressions may appear in a schematic form during early stages of their construction. This is seen in Alan and Bob's expression; the units written in the position of the first term function as a placeholder for a more complete expression to be written later. In fact, this is exactly how this expression functioned in Alan and Bob's work. Ultimately they elaborated both the first and second term in terms of some useful parameters.

The next two forms in this cluster, base + change and base – change, are flip sides of the same coin. Each form includes a symbol template with two terms. The first term is a base quantity, and the second term is an amount that is taken away or added to this base. The base + change form may appear indistinguishable from parts-of-a-whole, but they are different. In parts-of-a-whole, the two parts are combined to form a new third entity. In contrast, in base + change, no third entity is introduced; instead, the change increases the size of the base quantity.

I present one example in which base + change appears, an example worthy of discussion because it involves an equation that is learned early and used frequently: $v = v_o + at$. At points during the problem-solving sessions, I pressed students about this equation, asking them such things as how they knew it was true and whether they knew it "by heart." During the final session, which involved some more open-ended interviewing, I also each asked pair whether they thought that this equation was "obvious." Mike and Karl responded:

Mike: Well yeah it is obvious because, well velocity will equal to *V* naught if it's not being disturbed. But if it's being acc— If there's a f— acceleration action on it, then uh—and that's constant, you know then velocity will be decreasing as time goes on. Or increasing, whatever it works, I mean whichever it does. So, it's like whatever it is and then plus a correction on the acceleration. So yeah it makes sense. It's obvious, yes it is.

…

Karl: What's obvious to me is that you have the final velocity is obviously going to be equal to the initial velocity plus however much, however faster it gets. That's what's obvious to me. What's not necessarily positively obvious is that the amount, the amount that it gets faster is the acceleration times the time.

Mike began by answering that, yes, $v = v_o + at$ is obvious. It was obvious for Mike because it is consistent with the base + change form; there is a base, $v_o$, plus some "correction." Karl agreed that the equation is obvious but only in a limited sense. In particular, he was not sure that it is obvious that the second term must be precisely *at.* Although he agreed with the need for an initial velocity plus some change that depends on the acceleration, he did not know exactly what that change should look like.

This example again shows the limits of forms. Forms specify the equations to be written but only up to a certain level of detail. The interpretation of $v = v_o + at$ given by Mike and Karl was the one provided by almost all students. They typically saw this equation in terms of base + change and at no finer level of detail. In fact, it was not uncommon for students to write the second term incorrectly. For example, rather than writing $v = v_o + at$, Mark and Roger wrote this:

$$v = v_o + ½ at^2$$

Mark: 'Cause we have initial velocity [w. circles $v_o$ term] plus if you have an acceleration over a certain time [w. circles $½at^2$]. Yeah, I think that's right.

Although Mark's expression is consistent with base + change—it has an initial velocity plus a change involving the acceleration—this expression does not correctly give the dependence of velocity on time for cases of constant acceleration. In fact, the units of the second term are not even correct. It is likely that this particular error is due to a confusion between the equation for velocity and the equation for position: $x = x_o + v_ot + ½ at^2$.

*Dependence.* The forms in the dependence cluster concern the mere presence of a specific symbol in an expression. Most basic of all these forms is no dependence, in which the symbol pattern involves the absence rather than the presence of symbols. Seeing an expression in terms of no dependence entails noting that a particular symbol does not appear in the expression. For example, in the construction of their expression for μ, Mike and Karl wanted to have one component that does not depend on the weight (i.e., mass). This is an application of no dependence:

$$\mu = \mu_1 + C\frac{\mu_2}{m}$$

Karl: Well yeah maybe you could consider the frictional force as having two components. One that goes to zero and the other one that's constant. So that one component would be dependent on the weight. And the other component would be independent of the weight.

A prototypical use of no dependence can be found in students' work on the shoved block problem. The correct answer to this question is that the heavier and lighter block travel the same distance if started with the same initial speed. To arrive at this conclusion, the students needed to derive an equation in which the mass did not appear. Alan and Bob derived such an expression and then commented

$$a = g\mu$$

Alan: Right, so, actually, they should both take the same.=
Bob: =Wait a minute. Oh, they both take the same! [surprised tone] … So, no matter what the mass is, you're gonna get the same, the same acceleration.

Alan and Bob saw their equation in terms of no dependence. They stated that, because no mass appears in this expression, the acceleration does not depend on the mass, so both objects travel the same distance. This really does constitute a specific way of seeing this expression. To see this expression in terms of no dependence, Alan and Bob had to blur certain features into the background and highlight other features for themselves. In this case, the absence of the symbol *m* is the key registered feature. Furthermore, the character of Alan and Bob's reaction shows that no dependence is not subtle; it is striking and obvious to them and elicits a strong reaction.

The flip side of no dependence and the next step up in complexity is dependence, in which the presence of a particular individual symbol is noted. In the construction of expressions, this form is sometimes used in tandem with other forms. For example, in the early specifications of their expression for μ, described previously, Mike and Karl stated that one component should be dependent on the weight. They later elaborated this claim into the statement that this component should be inversely proportional to the weight. Similarly, dependence sometimes played a role in students' steps toward the construction of an expression for the force of air resistance:

$$R = \mu v$$

Mark: So this has to depend on velocity [g. *R*]. That's all I'm saying. Your resistance—the resistor force depends on the velocity of the object. The higher the velocity the bigger the resistance.

Refer to Appendix B for a discussion of the remaining forms in this cluster.

*Coefficient.* In the coefficient form, a product of factors is seen as broken into two parts. One part is the coefficient itself, which often involves only a single symbol and almost always is written on the left. The coefficient form is distinguished from other forms (particularly in the multiplication cluster) because of some properties that are typically possessed by the coefficient. Students sometimes comment, for example, that a coefficient is "just a factor" or "just a number." It is, in fact, the case that important dynamic variables rarely appear in the position of the coefficient. For example, forces do not appear in coefficients, nor do velocities or positions that vary over the time of a motion. Instead, coefficients involve parameters, such as the coefficient of friction, that, in a sense, define the circumstances under which a motion is occurring.

The proportionality constant introduced by students in the air resistance problem is a typical example of a coefficient

$$F = kv^2$$

Jack: Right, all I did was introduce a constant of proportionality [g. *k*]. We have no idea what it is.

Jack saw this expression in terms of the coefficient form—the constant *k* is seen as multiplying the rest of the expression. Furthermore, Jack's comment that "We have no idea what it is" is characteristic of the type of things that people say about symbols that are being treated as coefficients. Jack made this comment in a somewhat flippant manner. He implied that it does not matter that they do not know what this coefficient is. Coefficients can tune the size of an effect or influence and in that way have a quantitative effect on a motion, but they are not generally seen as influencing the overall character of a motion or the result in question.

It is worth noting that the coefficient cluster is in one sense unique among the five clusters discussed to this point; it is the only cluster concerned with structure in the arrangement of a product of factors. Consider an expression of the form *abc*. In terms of what forms can a student see this expression? The terms as amounts forms and competing terms forms could allow the whole of the expression to be seen as an influence or as some "stuff." In addition, we could see this expression in terms of dependence in three ways, as dependent on *a*, *b*, or *c*. Similarly we could see prop+ in three separate ways, as focusing on any of the factors involved. However, only coefficient breaks down this string of factors into a meaningful structure; it splits off the first factor, *a*, and treats it differently.

*Multiplication.* Given the preceding comments about the coefficient cluster, it may seem that many forms are missing from this semi-exhaustive list. Perhaps it is surprising that more symbolic forms are not associated with meaningful structure in a product of factors. Of course, it may simply be the case that the vocabulary of forms is sparse in this territory, at least for students working on the tasks in this study. To a certain extent, I believe this is the case. However, two multiplicative forms cover some of this territory. I discuss these forms with the qualifications that they appeared rarely in the data and their identification is tentative.

The two forms in the multiplication cluster are intensive–extensive and extensive–extensive. Intensive and extensive quantities (Greeno, 1987; Hall, Kibler, Wenger, & Truxaw, 1989; Schwartz, 1988) are distinguished in roughly the following way: An intensive quantity specifies an amount of something per unit of something else. In contrast, an extensive quantity is a number of units. For example, suppose that an arithmetic word problem gives a number of apples per basket and a number of baskets, and asks how many total apples there are. The number of apples per basket is intensive, and the number of baskets is extensive.

As the names imply, the intensive–extensive form applies when an expression, such as $xy$, is seen as a product of an intensive and an extensive quantity. Similarly, the extensive–extensive form applies when such an expression is seen as a product of extensives. In these data, intensive quantities were almost always densities or rates. Thus, the intensive–extensive form usually involved the product of a rate and a time or a density and a volume. The fact that this form applied in such a narrow range was part of the reason that it was difficult to establish its general validity.

*The identity form.* To conclude this survey of the forms knowledge system, I discuss a symbolic form that does not fit into any of the previous clusters. Few of the forms discussed thus far involve the use of an equal sign (=). The only exceptions are balancing and same amount. An additional form involving an equal sign covers a large percentage of the expressions that are composed and interpreted: the identity form. In identity, a single symbol, usually written on the left, is separated from another expression by an equal sign. I mentioned this form earlier in my account of Mike and Karl's construction of their novel expression for the coefficient of friction. Identity allowed them to write "μ=" on the left side of the expression.

Identity is so common that it is nearly invisible in student utterances. Furthermore, not only is identity involved in the construction of expressions, but students frequently manipulate expressions to align with identity before interpreting them. This was the case with the final expression from the spring task that Jim interpreted:

$$x = \frac{mg}{k}$$

Jim: Okay, and this makes sort of sense because you figure that, as you have a more massive block hanging from the spring, then your position *x* is gonna increase, which is what this is showing [g.~ *m* then *k*]. And that if you have a stiffer spring, then your position *x* is gonna decrease [g. Uses fingers to indicate the gap between the mass and ceiling]. That's why it's in the denominator. So, the answer makes sense.

Identity allows a simple but important type of inference: It allows us to infer that whatever is true of the right hand side of the expression is also true of the left. Such an inference is implicit in Jim's interpretation.

## DISCUSSION

How can the forms framework inform instruction? Where does knowledge of forms originate? These are the issues I take up next.

### Instructional Implications

The facile use of equations in physics includes the ability to understand equations in terms of symbolic forms, but the leap from analyses of this sort to instructional implications is a large one, requiring work beyond the analysis reported thus far. Recognizing the importance of this type of equation understanding does not tell us how it can be engendered (in fact, it does not even tell us whether any change to instruction is required). In this discussion, I address instructional implications and reiterate some major themes from an instructional orientation.

*When students fail to understand equations.* I begin by looking at cases in which equation understanding breaks down or exhibits limitations because these may indicate sites for potential learning. Perhaps the most dramatic example already presented was Alan's balancing interpretation of the equation $\mu mg = ma$, made during Alan and Bob's work on the shoved block task. This interpretation is incorrect because, among other things, it implies the existence of two horizontal forces when there is only one.

$$\mu mg = ma$$

Alan: What Bob didn't mention is that this is when the two forces are equal and the object is no longer moving because the force frictional is equal to the force it received.

Alan's incorrect interpretation was an isolated incident—no other students made the same error. Other difficulties and limitations, however, appeared repeatedly, both within and across students. One such difficulty concerns how students understood (and failed to understand) the equation $v = v_o + at$. This equation is interesting for a number of reasons. First, it is one of the first equations students learn in an introductory physics course. In addition, to virtually all competent physicists, the expression $v = v_o + at$ is simple and straightforward. The first term, $v_o$, is the initial velocity of the object, and the second term adds on the amount that the velocity changes in a time *t*. Because the acceleration is presumed to be constant, the acceleration multiplied by the time provides a measure of how much has been added to the velocity. Another way to understand this is to think of the acceleration as the rate at which the velocity changes.

As we have already seen, this equation is not straightforward for students, even the moderately advanced students in this study. While working on the vertical toss problem, students often struggled to regenerate this expression, sometimes writing expressions that were incorrect. For example, Mark and Roger wrote the following equation and commented

$$v = v_o + ½\, at^2$$

Mark: 'Cause we have initial velocity [w. circles $v_o$ term] plus if you have an acceleration over a certain time [w. circles $½at^2$]. Yeah, I think that's right.

This statement suggests that Mark interpreted this expression up to the level of the base + change form. He knew that a term must correspond to the initial velocity and a term must describe how the velocity changes over time due to the acceleration. However, because his specification was only at the level of base + change (plus some dependence on acceleration and time), he could not state precisely what should be in the second term. Instead, he resorted to trying to recall this expression. In doing so, he incorrectly substituted a term from the expression for the position as a function of time: $x = x_o + v_o t + ½\, at^2$ .

This is not meaningless symbol use, and the use of base + change here is not incorrect. Rather, the problem is that Mark and Roger's understanding is much weaker than it could be. To a physicist, the precise form of the *at* term is obvious, given the definitions of velocity and acceleration. Certainly educators would like students to approach the point at which this expression makes as much sense to them as it does to a physicist. In terms of my framework, this could mean understanding at in terms of the intensive–extensive form.

Mark and Roger's situation was entirely typical. Every participant pair in this study was able to give only base + change interpretations of $v = v_o + at$. Furthermore, when pressed, every student said this expression would have to be remembered or derived from some other remembered equation, such as the expression for how the position varies with time in cases of constant acceleration:

$$x = x_o + v_o t + ½\, at^2$$

Jack: Well you have to remember something to know how to approach the problem, it's just a question of which um equation you remember of motion. Um, given (1.5) Um given, if you remember this one [g. equation] you can pretty much derive them all.

To follow up on these students' inability to see $v = v_o + at$ as obvious, I asked a number of related questions during the final, open-ended session. One question was a rather bald prompt: "Some students say that the equation $v = v_o + at$ is obvious. What do you think?"

Here is a portion of Mike and Karl's response, presented in full earlier:

Mike: Well yeah it is obvious because, well velocity will equal to *V* naught if it's not being disturbed …. So, it's like whatever it is and then plus a correction on the acceleration. So yeah it makes sense. It's obvious, yes it is …

Karl: What's obvious to me is that you have the final velocity is obviously going to be equal to the initial velocity plus however much, however faster it gets …. What's not necessarily positively obvious is that the amount, the amount that it gets faster is the acceleration times the time.

As discussed earlier, Mike and Karl seem to have agreed that $v = v_o + at$ is "obvious," but only in a certain limited sense. Mike said it was obvious because there should be a "*V* naught if it's not being disturbed" plus a "correction" that has to do with the fact that it is accelerating over time. This is an interpretation at the level of base + change. Karl is clear in stating that, even with this interpretation, the precise form of the second term is not entirely obvious: "What's not necessarily positively obvious is that the amount, the amount that it gets faster is the acceleration times the time."

These observations become even more interesting in light of student answers to another question that I asked during the final session:

> Imagine that we've got a pile of sand and that, each second, *R* grams of sand are added to the pile. Initially, the pile has *P* grams in it. Write an expression for the mass of the sand after *t* seconds.

The idea is that there is a pile of sand in which the total mass grows linearly with time at rate *R*, starting with an initial mass of *P*. Thus, the correct answer to this task is the expression $M = P + Rt$. Notice that this expression has the same structure as $v = v_o + at$; both involve an initial value and a term that grows with time at a fixed, linear rate.

Intriguingly, students found the pile of sand question to be utterly trivial. In every case, including that of Mike and Karl, the students read the question, then one picked up a marker and wrote an expression equivalent to $M = P + Rt$, without comment. Furthermore, they generally seemed perplexed that I had even asked such a simple question.

These students have nearly all the resources they need to see $v = v_o + at$ as trivial and obvious, just as it is obvious to a physicist, but somehow these resources are not getting engaged here. Students do not understand $v = v_o + at$ in the same way they understand $M = P + Rt$. For them, there is some mystery left in $v = v_o + at$.

*Reforming physics instruction.* Because students possess many resources needed to understand equations like $v = v_o + at$, there may be some reasonable steps we can take to get these resources engaged more frequently and at the right times. Even if students possess many of the necessary symbolic forms, however, getting these elements engaged in the right places may not be a trivial instructional goal. Students must learn to adopt particular stances to individual physics expressions such as $v = v_o + at$, and this learning is likely to be tied up with other difficult conceptual issues. For example, it may not be easy for students to learn to treat acceleration as a rate. Furthermore, the expression $v = v_o + at$ may be particularly tricky, compared to $M = P + Rt$ because the acceleration and velocity are both intensive quantities (i.e., rates). However, presuming that physics instructors accept the task of getting students to understand expressions in a more useful and appropriate manner, what should they do? How should instruction be changed?

First, the insights emerging from this research can help; just being able to recognize and understand the problems to be addressed is important. Once it is clear that there is a problem with how students understand $v = v_o + at$ and instructors make efforts to improve that understanding, then it is at least possible that there are some relatively straightforward steps to take. For example, although I did not try this in the study, pointing out the parallel between $M = P + Rt$ and $v = v_o + at$ may have been a useful component of instruction designed to help Mike and Karl better understand the latter expression.

In addition, knowing about specific symbolic forms may help instructors. Of course, it almost certainly will not be sufficient to just tell students about symbolic forms; in the long run, the design of new instructional strategies that take my theoretical results into account will require careful thought and experimentation. However, some reasonable and modest first steps are not at the level of "just tell them." One way that students learn is by watching instructors as they model appropriate behavior. For example, physics instructors frequently model how to solve various types of physics problems by standing at a blackboard and working through some problems while students simply watch. In a similar manner—perhaps as part of modeling problem solving—instructors could strive to model the interpretation of expressions. Simply, in addition to writing and manipulating equations while students watch, they can point to equations and say something about what those equations mean. Of course, many instructors probably do this to some extent already, but a little more frequency and emphasis would likely help.

Some of the most important instructional implications of this work may have to do with how students are prepared for instruction in physics, particularly regarding mathematics. Given the observations presented here, it seems reasonable to suggest that, prior to physics instruction, students should learn to invent at least some simple types of mathematical models and to express the content of those models. The underlying assumption is that instruction in mathematical modeling gives students a leg up when they get to physics instruction. For a student with a mathematical modeling background, each physics equation does not present a new and individual challenge. Instead, the student is able to see physics expressions as instances of particular types of models. For example, Mike and Karl may learn to see $v = v_o + at$ as an instance of a model with linear growth in time.

In many respects, this suggestion is consonant with the approach embodied in standards-based mathematics curricula, such as *Connected Mathematics* (Lappan, Fey, Fitzgerald, Friel, & Phillips, 1997) and *College Preparatory Mathematics* (Hoey & Wootton, 1994). Although the assumptions and elements of these curricula vary, across standards-based curricula there is an increased emphasis on the use of mathematics to model situations and phenomena.

The research presented in this article can help instructors better understand what must be taught and learned in these courses. I am not suggesting that the forms described in this article should be explicitly mentioned or even that there is a recognizable one-to-one mapping between particular symbolic forms and course material. However, the work presented here can help to better specify the sort of understanding instructors must engender and the possible difficulties to expect.

From a certain perspective, this is really what physics is all about, building a certain class of models of the physical world. To worry about student misconceptions and related qualitative understanding is essentially to worry about the particulars of certain particular models, albeit some important ones. I am not alone in arguing that learning to build mathematical models of the physical world is a central goal of physics instruction. For example, Hestenes (1987) adopted a similar view, arguing that the structure of the models inherent in physical principles and equations should be made explicit for students. Furthermore, this view can be seen as part of a more general movement that takes models, not necessarily formulated as equations, as the targets of physics instruction (e.g., White, 1993a, 1993b; White & Frederiksen, 1998).

*Some thoughts on conceptual physics.* Instructors committed to fostering an understanding of physics face a choice: either help students to use equations more meaningfully or use methods that do not spend as much time with equations. In emphasizing the need to teach mathematical modeling, I am implicitly choosing the former alternative. However, there are approaches to physics instruction that take the opposite position. These conceptual physics approaches downplay the use of equations, focusing instead on exploring situations and problems to attack the difficult conceptual issues of physics (e.g., Hewitt, 1971).

In some respects, the analysis presented here can be seen as supporting the reasonableness of teaching students to use equations with understanding. To a certain extent, mathematical equations have developed an undeserved, negative reputation among researchers in science education. Teaching for understanding has seemed frequently to involve moving away from symbolic formalisms. Given the analysis presented here, this shift deserves reconsideration. The use of symbolic expressions can involve significant understanding—not just rote manipulation.

Furthermore, a more fundamental shift in attitude toward the relation between symbolic expressions and understanding may be merited. For example, a physicist's sense of balancing—captured in the balancing form—is rooted in experiences equating mathematical expressions as well as experiences of balancing in the physical world. Consequently, to omit symbolic experiences from physics instruction leads to an expectation of a concomitant change in the nature of the understanding that results.

This is not necessarily an undesirable outcome. It is acceptable to be flexible in instructional goals; learning about qualitative features of the world is valuable, as is learning to mathematize the world. Furthermore, conceptual physics instruction likely has substantial value as a way to prepare students for more traditional algebra-based physics instruction. However, instructors need to understand what they are doing if they choose to teach physics in the manner of conceptual physics. I challenge the assumption that in physics or any domain the conceptual and the symbolic elements of a practice can be separated for the purposes of instruction. Removing equations from the mix changes the nature of understanding. This does not imply that physics cannot be taught without equations. However, it does imply that equation-free courses will result in an understanding of physics that is fundamentally different from physics as understood by physicists.

Again, I must make clear that I do not mean to critique conceptual physics courses; rather, I hope to understand better the consequences of omitting equations. In fact, in the tradition of these more conceptual approaches, there is already a history of striking successes and exciting proposals. For example, White and Horwitz (White, 1993b) designed and implemented a middle school curriculum based on computer models. The students in this curriculum performed better on many tasks than students in a traditional high school course. Clement (1987), through his technique of bridging analogies, succeeded in training expert intuition. I suspect that even many working physicists could learn a great deal from Haertel's (1987) impressive reworking of the entire topic of electricity and magnetism in qualitative terms. All this published research is in addition to the many instructors at universities and high schools that are enjoying success in designing and teaching conceptual physics courses. I hope that the work I have reported here encourages better understanding of the character of conceptual physics, so that instructors can understand and better judge the results of such an approach, in clear contrast to alternatives.

## On the Origins of Symbolic Forms

When are symbolic forms acquired? As I mentioned earlier, some symbolic forms are likely acquired during physics instruction and some much earlier, in the context of mathematics instruction. There is fairly clear evidence that the precursors of symbolic forms are acquired as early as first or second grade, when simple arithmetic word problems are introduced. According to a number of researchers, when elementary school students learn to solve word problems, they come to understand problem situations in terms of a basic set of schemata that correspond to relatively simple semantics. These semantics describe situations in terms of such things as the combining of quantities, changes to quantities, and the comparison of quantities (see, e.g., Carpenter & Moser, 1983; Fuson, 1992; Riley, Greeno, & Heller, 1983; Vergnaud, 1982). Once the situation is understood in terms of these simple semantics, the schemata are mapped directly onto mathematical operations through procedural attachment; that is, there are procedures directly associated with the schemata (Reed, 1998).

For example, Riley et al. (1983) listed four categories of arithmetic word problems: change, equalization, combine, and compare. In change and equalization problems, addition and subtraction are described as actions that lead to increases or decreases in some quantity. More specifically, in change, an amount is simply added onto another quantity. For example, John has four apples and Mary gives him three. In equalization, an amount is added to a base quantity so that it becomes equal to some third quantity: How many apples must Mary give to John so that he has the same number as Fred?

In contrast, combine and compare problems "involve static relations between quantities" (p. 161). Combine problems have two distinct quantities that remain unchanged: How many apples do John and Mary have between them? In compare problems the difference between two unchanging quantities is determined.

In addition to additive types, researchers have also identified a number of multiplicative problem types. This is one place in the literature that the intensive and extensive quantities that I mentioned earlier appear. In some researchers' schemes, for example, there are multiplicative patterns that differ according to whether they involve the products of intensives, extensives, or combinations of intensives and extensives (Greeno, 1987).

Clearly, this whole story bears some strong similarities to my own. The problem types exist at a similar level of abstraction to my symbolic forms, and some specific forms even seem similar to specific problem types. For example, parts-of-a-whole and base + change are similar to Riley's combine and change categories. Thus, it seems plausible that we are seeing knowledge that is at least the precursor of symbolic forms.

However, there are some important differences, both obvious and subtle, between the earlier research and this project. First, the manner in which symbolic expressions are used by students in elementary mathematics is different from how they are used by university physics students. It seems unlikely that the arithmetic schemata possessed by elementary students is strongly bound to symbol templates, as it is for university students. This is one reason I say that we are only seeing the precursors of symbolic forms in elementary mathematics.

More subtly, the research discussed in this article involves a shift in focus from how students solve problems to how students understand equations. This subtle shift has important implications. In the problem type models, the situation activates a schema, and the solution is then guided by the contents of this schema. Thus, as in the schema-guided forward inference models discussed earlier, all the understanding happens at the beginning of a solution. In contrast, if equations can be understood, it is possible that, at any point during a solution, a particular equation may be understood, and this understanding can redirect the course of a solution. A model based on equation understanding can therefore lead to more variety in problem-solving behavior.

With my recasting of this earlier literature, it seems likely that the origins of a subset of symbolic forms are seen in early mathematics problem solving. Roughly speaking, this subset corresponds to a portion of the conceptual vocabulary that involves sets and a few kinds of actions and relations between sets. Carpenter and Moser (1983) characterized the conceptual space that is spanned by their problem types in this way, saying that arithmetic word problems can be characterized according to three dimensions: (a) whether there is "an active or static relationship between sets of objects," (b) whether the problem "involves a set-inclusion or set–subset relationship," and (c) for the case of problems that involve action, whether the action results "in an increase or decrease in the initial given quantity."

Clearly, some important symbolic forms are not covered in this subset. Most prominently, the physical forms, such as competing terms and balancing, cannot be easily captured by the set-like conceptual vocabulary that has been associated with the understanding of arithmetic word problems. Thus, these physical forms may not be acquired in any significant manner until physics instruction. Moreover, if this is correct, then there is an important general moral: A single abstract conceptual vocabulary should not be expected to apply across all domains. Although some forms may have wide applicability, some more idiosyncratic semantics may be specific to individual domains and applications within domains.

Another line of research in mathematics education bears on the development of symbolic forms. Some researchers have contrasted two broad stances that students adopt toward equations: the process stance and the object stance (Herscovics & Kieran, 1980; Kieran, 1992; Sfard, 1987, 1991). In the process stance, the equation is seen as describing a sequence of operations. For instance, the equal sign may be seen as a "do something signal." In the object stance, an equation is treated as an object in its own right, with the equal sign specifying a symmetric, transitive relation. Kieran (1992) argued that learning to adopt the object stance is a develop-mental achievement, with the process stance more natural early in instruction. However, both stances are seen as playing useful roles in expertise.

Because these researchers are focusing on the development of stances toward equations, this research is related to my discussion of symbolic forms. However, some care is warranted in making connections. It is tempting to match my forms identity and balancing with the do-something and symmetric-relation versions of an equal sign. Such a correspondence is not easily made. The equal sign in the identity form does not specify an action to be performed; rather, it states that one quantity can be identified with an expression. In fact, the shift from writing the single symbol on the right to writing it on the left may be significant. In the case of identity, the single symbol is not intended to be the computed result of a computation; instead, it announces the nature of the expression to follow.

## SUMMARY AND CONCLUSIONS

I have tried to show that it is possible, even for third-semester physics students, to understand equations in a relatively deep manner. At least for the students in this study, a particular arrangement of symbols in an equation expresses a particular meaning. Furthermore, I provided a map of this meaningful structure. I argued that physics students learn to understand physics equations in terms of a vocabulary of elements that I call symbolic forms, and I presented a list of symbolic forms that is sufficient to account for the observations in my data corpus. In this way I hoped to paint a picture of a physicist's phenomenal world of symbols—how physicists and physics students experience physics expressions.

In addition, I argued that this work can be understood as bridging the gap that has existed between research on naive physics and research on physics problem solving. Strikingly, the connection is actually quite tight; naive physics provides some of the conceptual vocabulary in terms of which equations are understood.

I used a number of approaches to illustrate and support my viewpoint. Most important, I presented a number of example episodes in varying degrees of detail and argued for my interpretation of those episodes. It is primarily through these examples and supporting arguments that I hope to have built a compelling case. For review, I present a list of some of the more suggestive episodes.

- Mike and Karl constructed their own equation for the coefficient of friction, an equation that does not appear in any textbook and which they had almost certainly never seen before.
- In the context of some more typical textbook problems, students do not always work by carefully and straightforwardly applying principles. Instead, at least sometimes, they construct equations by starting from a more intuitive understanding of the situation.

- While working on the running in the rain problem, Alan and Bob composed an expression with placeholders. This suggests that their construction may be driven by a template-like structure.
- While working on the shoved block problem, Alan gave an inappropriate balancing interpretation of the equation $\mu mg = ma$. A framework that considers equation understanding, the forms framework, allows researchers to talk about this error and can provide a clear characterization of the difficulty.
- Understanding equations is not confined to the first step in a solution. Students interpret equations throughout the solution process. This allows for more flexible problem solving and poses a difficulty for schema-guided forward inference models.
- Students are not always capable of choosing between certain kinds of related expressions. For Mike and Karl, $F = kx$ and $F = ½ kx^2$ were “qualitatively” equivalent.
- Similarly, the equation $v = v_o + at$ was “obvious” for students, but only in a certain way. How can students, in the same breath, say that $v = v_o + at$ is obvious but that they cannot be sure of the *at* term? If students are understanding this equation in terms of the base + change form, then this observation makes sense.

These phenomena, some of which are surprising, are well accounted for by the forms hypothesis. It is part of this hypothesis that students have knowledge for constructing equations that cuts across and is not directly associated with physical principles. Furthermore, it is an essential property of forms that, although there is a sense in which equations are understood, they are only understood at a certain level of detail.

## Future Work

The research presented here points the way to much future work. First, there is work to be done in mapping the learning trajectory from novice to expert. This work could include tracking what happens in physics instruction from a student’s first course through the development of physics expertise. Such an endeavor could answer such questions as what symbolic forms students possess on entry in their first university physics course and when and where experts rely on forms-based reasoning.

In addition, it would be interesting to paint a bigger picture of the development of forms-like knowledge. I would like to track the development of the type of semantics associated with symbolic forms in elementary school through secondary school and on to the college level. As I mentioned earlier, Izsák (1999, 2000) already attempted to track the early development of some symbolic forms. Furthermore, researchers may be able to make some headway simply by reinterpreting existing research, such as the work concerning problem types.

The symbolic forms framework could also potentially be extended to other areas of physics as well as to other domains. If the analysis here is correct, then physics equations are partly understood in ways that may be generic to many disciplines (e.g., parts-of-a-whole) and in ways that are physics-specific (e.g., competing terms). For this reason, looking at other domains would likely be illuminating.

Stretching still farther, it may be possible to generalize the forms framework to other representational forms. For any representation, we could try to map the set of meaningful structures and the meanings associated with those structures. For example, a program of this sort could be fruitfully applied to graphing as it is employed in various domains. In fact, Nemirovsky (1992) hypothesized that people learn a grammar of graphical shapes for graphs—a view closely related to saying that they acquire symbolic forms associated with graphs.

All of these extensions suggest ways that the forms framework can be further studied and applied to other domains and populations, but this framework captures only part of equation understanding. There are almost certainly varieties of understanding that are localized on individual symbols in equations rather than associated with equation-spanning structure (e.g., the *F* in an equation as somehow meaning "force"). In another component of this work, I have identified types of interpretive strategies. These are techniques for orienting oneself to an equation and extracting meaning (Sherin, 1996).

Finally, a central component of future work should be the elaboration and implementation of novel instructional techniques. It is my hope that the work presented here will provide a motivation for revisiting some assumptions concerning instruction in physics as well as other domains that involve the use of symbolic formalisms. I am not advocating that instructors give up the goal of teaching for conceptual understanding; on the contrary, I believe it is a laudable goal. I hope to have shown that we can strive for conceptual understanding while basing instruction on the use of equations.

## ACKNOWLEDGMENTS

This work was supported, in part, by Spencer Foundation Grant B–1393 awarded to A. diSessa.

This work grew out of my interactions with Andrea diSessa, and I am grateful for his help and insight. In addition, many people commented on various drafts of this article, as well as on the dissertation on which this work is based. Thanks especially to Andrea diSessa, Daniel Edelson, Miriam Gamoran, Rogers Hall, and Barbara White. Finally, I thank the editors and my anonymous reviewer for many helpful suggestions.

## REFERENCES

Bhaskar, R., & Simon, H. A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science, 1,* 193–215.

Carpenter, T. P., & Moser, J. M. (1983). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 9–24). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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.

Clement, J. (1983). A conceptual model discussed by Galileo and used by physics students. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 325–329). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Clement, J. (1987). Overcoming students' misconceptions in physics: The role of anchoring intuitions and analogical validity. In J. Novak (Ed.), *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics* (pp. 84–96). Ithaca, NY: Cornell.

Clement, J. (1994). Use of physical intuition and imagistic simulation in expert problem solving. In D. Tirosh (Ed.), *Implicit and explicit knowledge* (pp. 204–244). Norwood, NJ: Ablex.

deKleer, J., & Brown, J. S. (1984). A qualitative physics based on confluences. *Artificial Intelligence, 24,* 7–83.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10,* 165–255.

diSessa, A. A., & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education, 20,* 1155–1191.

Forbus, K. D. (1984). Qualitative process theory. *Artificial Intelligence, 24,* 85–168.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.

Greeno, J. G. (1987). Instructional representations based on research about understanding. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 61–88). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Haertel, H. (1987). *A qualitative approach to electricity* (IRL87–0001). Menlo Park, CA: Institute for Research on Learning.

Hall, R., Kibler, D., Wenger, E., & Truxaw, C. (1989). Exploring the episodic structure of algebra story problem solving. *Cognition and Instruction, 6,* 223–283.

Herscovics, N., & Kieran, C. (1980). Constructing meaning for the concept of equation. *Mathematics Teacher, 73,* 572–580.

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

Hewitt, P. G. (1971). *Conceptual physics.* Boston: Little, Brown, & Company.

Hoey, B., & Wootton, K. (1994). *College preparatory mathematics: Change from within.* Sacramento, CA: CPM Educational Program.

Izsák, A. (1999). *How Amy and Kate used criteria for algebraic expressions to develop knowledge structures for modeling the physical world*. Manuscript submitted for publication.

Izsák, A. (2000). Inscribing the winch: Mechanisms by which students develop knowledge structures for representing the physical world with algebra. *Journal of the Learning Sciences, 9,* 31–74.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). New York: Macmillan.

Lappan, G., Fey, J., Fitzgerald, W., Friel, S., & Phillips, E. (1997). *Connected mathematics.* Palo Alto, CA: Dale Seymour.

Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 75–98). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980a). Expert and novice performance in solving physics problems. *Science, 208,* 1335–1342.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980b). Models of competence in solving physics problems. *Cognitive Science, 4,* 317–345.

McClosky, M. (1983). Naive theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 289–324). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today, 37*(4), 2–10.

Nemirovsky, R. (1992). *On the basic forms of variation.* Unpublished manuscript, Technical Education Research Center, Cambridge, MA.

Pfundt, H., & Duit, R. (2000). *Bibliography: Students' alternative frameworks and science education.* Kiel: Institute for Science Education.

Ploetzner, R., & Spada, H. (1993). Multiple mental representations of information in physics problem solving. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 285–312). Amsterdam: Elsevier.

Ploetzner, R., & VanLehn, K. (1997). The acquisition of qualitative physics knowledge during textbook-based physics training. *Cognition and Instruction, 15,* 169–205.

Priest, A. G., & Lindsay, S. O. (1992). New light on novice–expert differences in physics problem solving. *British Journal of Psychology, 83,* 389–405.

Reed, S. K. (1998). *Word problems: Research and curriculum reform*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsberg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic.

Schwartz, J. L. (1988). Intensive quantity and referent transforming arithmetic operations. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 41–52). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sfard, A. (1987). Two conceptions of mathematical notions: Operational and structural. In J. C. Bergeron, N. Herscovics, & C. Kieran (Eds.), *Proceedings of the Eleventh International Conference for the Psychology of Mathematics Education* (pp. 162–169). Canada: University of Montreal.

Sfard, A. (1991). On the dual nature of mathematical conceptions. *Educational Studies in Mathematics, 22,* 1–36.

Sherin, B. L. (1996). *The symbolic basis of physical intuition: A study of two symbol systems in physics instruction*. Unpublished doctoral dissertation, University of California, Berkeley.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7,* 268–288.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3,* 115–163.

Tipler, P. A. (1976). *Physics*. New York: Worth.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Walker, J. (1975). *The flying circus of physics*. New York: Wiley.

White, B. Y. (1993a). Intermediate causal models: A missing link for successful science education? In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 177–252). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

## APPENDIX A
## Key to Transcripts

| | |
|---|---|
| (… ) | Transcription uncertain |
| // | Overlap with another speaker |
| ::: | Previous sound drawn out |
| ,, | Trail off without finishing statement |
| — | Cut off word |
| = | Not the usual amount of silence between two utterances |
| (0.0) | Pause (approximate time in seconds) |
| [ … ] | Nonlinguistic act |
| [w. … ] | Written on the whiteboard by the speaker |
| [w.mod … ] | Written expression modified |
| [g. … ] | Gesture with description |
| [g.~ … ] | Approximate target of the gesture |

## APPENDIX B
## Symbolic Forms

*Key to symbol patterns.*

| | |
|---|---|
| […] | Whole expression corresponds to an entity in the schema |
| x, y, n | Individual symbols in an expression |
| ☐ | Term or group of terms in an expression |
| … | Omitted portions of an expression that are not consequential or that continue a pattern |

### Competing Terms Cluster

*Competing terms.*

Description: Influences in competition.

Symbol pattern: ☐ ± ☐ ± ☐ …

Identification: Frequently forces but also other directed quantities such as momentum or acceleration. Often used in tandem with free-body diagrams with influences indicated as arrows. The signs of terms are often explicitly associated with directions on the diagram. Utterances often enumerate the influences acting in a circumstance and match to terms in the expression.

*Opposition.*

Description: Two terms, separated by a minus sign, associated with influences that work against each other.

Symbol pattern: ☐ – ☐

Identification: Similar to competing terms but limited to the special case of two influences that oppose. Common words: “oppose” and “opposing.”

*Balancing.*

Description: Two influences, each associated with a side of the equation, in balance so that the system is in equilibrium.

Symbol pattern: □ = □

Identification: Entities should correspond to entire sides of an equation. Influences often associated with directions on a diagram, as in competing terms. Use of the words “balance” and “equilibrium” are strong clues.

*Canceling.*

Description: Two influences that precisely cancel so that there is no net outcome.

Symbol pattern: 0 = □ – □

Identification: Similar to competing terms and opposition, but the outcome is usually mentioned as an entity and appears in the expression. Utterances frequently include the word *cancel* and often mention that the outcome is zero. Distinguished from balancing, the symbol pattern is different, and the schema includes an outcome. In addition, canceling sometimes includes a weak ordering, in time, of the two canceling influences.

## Proportionality Cluster

*Prop+.*

Description: Directly proportional to a quantity, *x*, which appears as an individual symbol in the numerator.

Symbol pattern: $\left[\frac{\ldots x \ldots}{\ldots}\right]$

Identification: One entity under discussion corresponds to an individual symbol that appears in the expression. Often spoken: “As *X* increases, then *Y* also increases.” Often includes the phrase *proportional to*.

*Prop–.*

Description: Indirectly proportional to a quantity, *x*, which appears as an individual symbol in the denominator.

Symbol pattern: $\left[\frac{\ldots}{\ldots x \ldots}\right]$

Identification: Often spoken: “As *X* increases, then *Y* decreases.”

*Ratio.*

Description: Comparison of a quantity in the numerator and denominator.

Symbol pattern: $\left[\frac{x}{y}\right]$

Identification: In most cases, the quantities *x* and *y* have the same units. Utterances involve whether *x* or *y* is greater and whether the ratio is greater than, equal to, or less than one.

*Canceling (b).*

Description: Identical symbols that appear in the numerator and denominator cancel.

Symbol pattern: $\left[\frac{\ldots x \ldots}{\ldots x \ldots}\right]$

Identification: Very common when identical symbols appear in the numerator and denominator of an expression. Use of the word *cancel*.

## Terms Are Amounts Cluster

*Parts-of-a-whole.*

Description: Amounts of generic substance, associated with terms, that contributes to a whole.

Symbol pattern: [ □ + □ + □ …]

Identification: Unlike competing terms, these entities are not influences. Utterances enumerate the parts that contribute to a whole, sometimes in correspondence with a diagram. Also indicated by inferences, such as the observation that if one part increases and the others are held fixed, then the whole increases.

*Base±change.*

Description: Two terms contribute to a whole but play different roles. One is a base value; the other is a change to that base.

Symbol pattern: [□ ± Δ]

Identification: Identification of one term with a base case or initial value. The second term corresponds to a change that occurs through time or across cases. Important words: *gains, correction.*

*Whole–part.*

Description: A new net amount is produced by taking away a piece of an original whole.

Symbol pattern: [□ – ▫]

Identification: Not common in data. Differs from base±change because the part removed is considered to be a portion of the original.

*Same amount.*

Description: Two amounts, each associated with a side, are the same.

Symbol pattern: □ = □

Identification: Not common in data. Differs from balancing because the entities are not influences.

## Dependence Clusters

*Dependence.*

Description: A whole depends on a quantity associated with an individual symbol.

Symbol pattern: [… $x$ …]

Identification: The observation that a particular symbol appears in the expression. Common phrases: *depends on, is a function of.* Also indicated by inferences, such as if $x$ varies, then the whole must vary.

*No dependence.*

Description: A whole does not depend on a quantity associated with an individual symbol.

Symbol pattern: […]

Identification: The observation that a particular symbol does not appear in the expression. Common phrases: *doesn't depend on, isn't a function of.* Also indicated by inferences, such as if the quantity in question varies, then the whole need not vary.

*Sole dependence.*

Description: A whole depends only on one particular quantity associated with an individual symbol.

Symbol pattern: [… $x$ …]

Identification: Similar to dependence but only on a single quantity. Indicated by: *only depends on.*

## Coefficient Cluster

*Coefficient*

Description: A product of factors is broken into two parts and one part is identified with an individual symbol, the coefficient.

Symbol pattern: [$x$ □]

Identification: The symbol treated as the coefficient appears on the left. Utterances include that this symbol is "just a number," "just a factor," or "a constant." Coefficients tune the size of an effect.

*Scaling.*

Description: Similar to coefficient, but the coefficient is unitless. A scaling coefficient is seen as operating on the rest of the factors to produce an entity of the same sort that is larger or smaller than the original.

Symbol pattern: [$n$ □]

Identification: Commonly found in situations where something unitless multiplies other factors. Utterances include the observation that if $n$ is greater than one, then the new quantity produced will be larger than the multiplied factors.

## Multiplication Clusters

*Intensive–extensive.*

Description: A product of an intensive quantity and an extensive quantity. An intensive quantity is an amount of something per unit of something else. An extensive quantity is a number of units.

Symbol pattern: $x \times y$

Identification: Common when the intensive quantity is a density or rate. Utterances involve explicit comments about $x$ and $y$, and frequently the units will be written out.

*Extensive–extensive.*

Description: A product of two quantities, both associated with extensives.

Symbol pattern: $x \times y$

Identification: Not common in data. Representational concomitants may appear, such as diagrams used when computing an area.

## Other Forms

*Identity.*

Description: A single symbol that appears alone on one side of an equation has the same properties as the expression on the other side.

Symbol pattern: $x = \ldots$

Identification: Extremely common but rarely reflected in utterances. The individual symbol $x$ is usually written on the left side of the equation. Allows very quick inferences that anything true of the expression on the right is true of the individual symbol on the left.

*Dying away.*

Description: A whole dies away with some parameter that appears raised to a negative power in the exponent of an exponential.

Symbol pattern: $\left[e^{-x\ldots}\right]$

Identification: Not common in data. Often accompanied by a graph showing exponential decrease. Common with the inference that a large value of the exponent means a small value for the whole.

# APPENDIX C

## Method and Rationale

In this appendix, I fill in some of the methodological details omitted in the body of the article. I describe the design of the study and its rationale. Then, I discuss the methods of analysis—how I got from observations and transcripts to the model presented, including the list of specific symbolic forms. These methodological issues are developed more fully in Sherin (1996).

## The Data Corpus

The overall goal of the study design was to create situations in which to observe students using physics equations with understanding. This goal constrained the nature of tasks as well as the students who could participate. I needed to design tasks that required students to use understanding and to select participants who possessed the requisite understanding.

*Participants.* Because of this desire to see equation understanding at work, I chose to look at university students who were somewhat advanced in the introductory physics sequence. All participants in the study were students at the University of California at Berkeley and were currently enrolled in Physics 7C, a third-semester introductory course in a series entitled *Physics for Scientists and Engineers*. As the title suggests, the course is designed for students who intend to major in a so-called hard science or engineering discipline. In contrast, students headed for medical school or allegedly softer forms of engineering (e.g., architecture) typically enroll in the Physics 8 series.

Two preceding courses in the Physics 7 series are prerequisites for 7C. Physics 7A introduces classical mechanics, and 7B covers topics in electricity and magnetism as well as some thermodynamics. The third semester, Physics 7C, covers optics and electromagnetic waves in addition to some modern physics topics, including special relativity and quantum physics.

Thus, these students were at an intermediate level of expertise. They were not experts, but they had substantially more experience than students enrolled in their first physics courses. This intermediate level was desirable for several reasons. I wanted to have participants who were capable of using equations with understanding, and I wanted to see at least the beginnings of expert behavior. Furthermore, because this work is instructionally oriented, I wanted to be able to understand the difficulties that students encounter and perhaps how they could be overcome. I thus did not want to look at true experts but students who were still at an early enough stage of their studies that I could expect to observe the types of difficulties encountered by students in introductory courses.

The analysis presented here was based on observations of five pairs of students, selected at random from a pool of volunteers. Four pairings consisted of two men, and the final pairing consisted of one man and one woman. The small number of woman participants reflects the percentage of woman volunteers (which is itself reflective of the small number of woman enrolled in the course). Students were paid $5 per 1 hr to participate in the study.

*Tasks* It was essential that the tasks were neither too difficult nor too easy for students. If the tasks were too difficult, then the students would have had little notion of how to proceed, and they may well have been reduced to manipulating symbols in hope of discovering an answer. If the tasks were too easy, students may have been able to quickly produce a solution without having to think about what to do. I believe that some earlier studies have suffered because they looked at situations in which one of these two extremes held. For example, earlier observations that experts forward-chain and novices backward-chain (Larkin et al., 1980a) may be partly an artifact of the problems employed being easy for experts and extremely difficult for novices.

For the most part, I chose problems typical of those found in an introductory mechanics course (such as the course Physics 7A). However, a few more unusual tasks, such as the running in the rain problem, were also included. These tasks were designed to provide contexts in which students would not be able to draw on remembered equations and would thus be forced to construct novel expressions.

During the final session, students were given more open-ended, nontraditional tasks—some of which were designed to try out new instructional techniques. I interrupted students frequently during these exploratory tasks. Moreover, I used part of the final session for an open discussion with students. The analysis for this article focused on a subset of the less exploratory questions, corresponding to 11 hr out of the total 27 hr of videotape. I have referred to this portion of the data corpus as the focus corpus, and the great majority of the examples were drawn from this part of the corpus. The tasks included in the focus corpus are given in Table 2.

*Experimental sessions.* The experimental sessions were conducted in a laboratory setting. Each pair worked standing at a white board, with a single video camera positioned to record the movements of the two students as well as everything they wrote on the board. An attempt was made to maintain an open and relaxed atmosphere and to make the setting as natural as possible. Students work in pairs or larger groups when solving their homework problems; in fact, Berkeley provides course centers for this purpose, which are frequently staffed by teaching assistants.

The students were given the problems one by one, with each problem written on a sheet of paper. For the most part, pairs were allowed to work a problem from beginning to end, without interruption. They were then asked to explain their solutions. I occasionally interrupted the students for various reasons. Sometimes I prompted for clarification. In addition, when students said that they were stuck or they had spent a significant amount of time exploring an unviable avenue, I provided hints or suggestions to start them on a more useful track. For example, Mark and Roger encountered difficulty during the air resistance task. They had the right sort of expressions for the forces of gravity and air resistance and were thus close to a solution. However, they also had a separate expression for the velocity: $v = at$ (this equation for the velocity is not correct here because it presumes that the acceleration is constant). Mark and Roger tried at length to manipulate the various equations they had written to get a result, but the presence of the incorrect equation for velocity confused them, and they floundered. After about 10 min of floundering, I finally interrupted with a hint:

Bruce: You've done a lot of great stuff here, but one thing bothers me. And that's—this relation here [g. $v = at$] … it seems to say that the velocity is getting bigger forever. Like I could put in a bigger $T$ and the velocity would still be bigger.

Mark: Oh yeah, you're right.

Bruce: So that seems a little bit weird to me, it doesn't seem like it's ever getting to a constant velocity.

I pointed out that their equation for the velocity was inconsistent with the assumption that the dropped object reaches a terminal velocity. This is a fairly substantial hint because it points the students directly to the area of difficulty in their solution efforts.

Sessions were typically between 1 and 1.5 hr in length, with pairs participating in four to six sessions (see Table 1). The total time spent by all students working on tasks was approximately 27 hr.

## Analysis

Describing the ensemble of knowledge that underlies thinking in a domain such as physics is a formidable task. Given this level of difficulty, there are a number of ways to simplify the task to make progress. One way to simplify is the approach taken by Ploetzner and VanLehn (1997). These researchers chose to describe thinking in terms of higher level abstractions such as beliefs. Within some limited regime of behavior, people behave as if they have certain beliefs. Researchers such as Ploetzner and VanLehn have attempted to describe these emergent beliefs rather than the knowledge representations that underlie behavior across circumstances. The power of this approach is that, within its regime of applicability, the result is a fully runnable model.

An alternative approach—and the one taken by knowledge analysis—is to focus on the primitive level but only to attempt to describe a portion of the knowledge that underlies any behavior. This approach has both benefits and drawbacks: It gives us a window into fundamental aspects of the development of expertise and makes it easier to build fully constructivist accounts of learning. The major drawback is that the resulting analysis is not a complete description of the knowledge necessary to explain any particular sequence of behaviors. Because behavior depends, in a complex manner, on the entire system, there is no possibility of producing a runnable model. Instead, other components of the system must be filled in by the intuitions of the cognitive modeler.

Even if we adopt this tactic, there are a number of reasons that producing descriptions at the primitive level still involves serious methodological difficulties.[8] The statement, "the student has the belief that a tossed ball slows down," and other accounts in terms of beliefs of this sort are relatively close to the directly observable behavior. In contrast, the leap from observable behavior to primitive elements may be a large one. Furthermore, the nature of these primitive elements may, initially, be difficult to discern. This is particularly true because there is no reason that these elements should divide up the world in the same manner as a textbook, or even that words that line up with elements of knowledge should be available.

For all these reasons, knowledge analysis can be methodologically challenging. I am not able to produce a runnable model that can be checked against behavior. Furthermore, because the leap from observable behavior to primitive elements is large, it is difficult to make it in a way that is not highly heuristic and interpretive. Nonetheless, I have tried to perform an analysis in a systematic and rigorous manner.

I began with 27 hr of videotape of students solving problems. I viewed this videotape once and transcribed it. Then, beginning with a preliminary version of the framework, I attempted to apply the framework to the observations in a coarse manner. This led to the development of a somewhat refined, preliminary framework that included the notion of symbolic forms as well as a preliminary list of forms.

Then the systematic analysis began in earnest. I narrowed my attention to the focus corpus, and within the focus corpus I identified a set of events that were of particular interest: events in which students constructed an equation from an understanding of what they wanted to express or students interpreted an equation.

---

[8]In fact, diSessa (1993) devoted several pages solely to discussing the methodological difficulties associated with his investigation of p-prims.

The episode in which Mike and Karl constructed a novel expression for the coefficient of friction is an example of the construction of an equation. Interpretation events were episodes in which students made comments concerning an equation that went beyond a literal reading of the symbols involved. Jim's understanding of the result derived for the mass on a spring task is an example. In this example, Jim went beyond a literal reading of this equation and told something about what the conceptual content of the equation was and why it was sensible.

$$x = \frac{mg}{k}$$

Jim: Okay, and this makes sort of sense because you figure that, as you have a more massive block hanging from the spring, then your position *x* is gonna increase, which is what this is showing. [g.~ m then k] And that if you have a stiffer spring, then your position *x* is gonna decrease. [g. uses fingers to indicate the gap between the mass and ceiling] That why it's in the denominator. So, the answer makes sense.

In total, 75 construction events and 144 interpretation events were identified. These 219 events were the focus throughout the next stages of the analysis (for more detail, see Sherin, 1996).

Next, I iteratively viewed and reviewed these 219 events, with the goal of describing the understanding evolved in every event in terms of the set of symbolic forms. There were three such iterations, and the set of forms was refined after each iteration. In addition, in parallel with this analysis, I gradually refined a set of heuristics for recognizing each symbolic form in specific events. These heuristics are included in Appendix B.

I emphasize that my purpose is not, at this stage, to present a rigorous coding scheme that can be applied by other researchers. Instead, I want to give a feel for the type of analysis involved in a knowledge analysis program of this sort and, in particular, to show that it is possible to do more than simply select and present some compelling examples. Through the sort of procedure described here, it is possible to take some steps toward ensuring that a framework is faithful to the range of phenomena in a data corpus.