By posing a continual stream of pertinent questions, a nonmathematical computer program can prod freshman physics students toward an analytical solution to one-dimensional kinematics problems.

The use of computers in physics pedagogy is almost entirely focused on “parameter-based” software: students focus on a set of initial variables, which result in a subsequent output. For example, we can investigate the effects of launch angle, speed, and air drag (the parameters) on a projectile’s motion¹ (the output) in several ways, including the PHeT Web-based projectile simulator (http://phet.colorado.edu), Physics Academic Software’s Windows-based simulator (http://webassign.net/pas), or, for an example in Mathematica, Wolfram’s Projectile Motion (http://demonstrations.wolfram.com/ProjectileMotion). However, despite vast differences in software form (symbolic, numeric, graphical, and so on), perhaps spanning generations of computer technology, the user experience remains fixed in this parameter-based mode, giving much of the physics pedagogical software the same “feeling.” From a programming perspective, this commonality results from the procedural programming languages used to build the software. That is, the programmer identifies the problem to be solved, chooses a desired implementation scheme, and gives the computer specific step-by-step instructions on how to proceed, subject to some terminal condition(s).

Such software is fundamentally weak, however, where input parameters, data, or desired outcomes are not well defined.²,³ Consider a typical end-of-chapter physics problem, such as problem 1.34 in Randall Knight’s book:⁴

A Porsche at rest accelerates from a stop light at 5 m/s² for 5 seconds, then coasts for 3 more seconds. How far has it traveled?

Or take problem 2.47 from a book by Douglas Giancoli:⁵

A falling stone takes 0.30 s to travel past a window 2.2 m tall. From what height above the top of the window did the stone fall?

Any number of available software packages can numerically, symbolically, or graphically simulate such problems, reinforcing relevant concepts in the procedural programming domain. But there’s another desirable software mode not being addressed. What about software that “understands” the laws of physics to the point where it can generate a textual, step-by-step, analytical solution to such problems? (By “analytical” we mean solutions that are found through direct, often symbolic, manipulation of the relevant physical laws, using algebra, trigonometry, and calculus.) This question motivated our work for two primary reasons.

First, there’s a lack of software available to tutor students in this manner. Developing heuristic software is apparently an untapped research area in computational physics pedagogy. A literature search yielded volumes on “intelligent tutoring systems,”⁶,⁷ but very little on solving physics problems.⁸,⁹ (Mastering Physics is a homework submission system tied to many current physics textbooks; while it appears to offer heuristic-like help with end-of-chapter problems, it’s not a generalized problem solver. See www.masteringphysics.com.)

Second, current physics texts continue to be written in an analytical fashion, with hundreds of problems calling for analytical solutions following hundreds of pages of analytical lessons. A disconnect appears to exist between parameter-based physics pedagogical software and the analytical emphasis of current physics texts.

We’ve developed a software tool that can generate a textual, step-by-step, analytical solution to an end-of-chapter physics problem without any prior knowledge of the problem itself. Our software has successfully generated solutions to one-dimensional kinematics (1DK) problems involving a single object—such as the Porsche and the stone in the problems above. The software works by asking questions about the problem and, based on the user’s responses, offers instructions that lead to a solution—similar...
to how an expert system works.\textsuperscript{10,11} In short, our software simulates, at a simple level, the professor-student interaction that might occur during a one-on-one tutorial session.

**Software Implementation**

Procedural languages themselves stand as an enormous hindrance in developing the type of heuristic software that we propose. Such software is vague in terms of inputs and outputs and raises many questions: What is the student confused about? What is the nature of the problem? How should we present the problem’s facts to the computer? What solution do we need? Attempting to code this software in a rigid, step-by-step procedural language would be unnecessarily labored. Clearly, we need another programming approach.

**Why Prolog?**

A long-ago interest in AI led to our awareness of Prolog and its ability to manipulate symbols while emphasizing a strict logical plan. Prolog enables an entirely different approach to computer programming, using a declarative rather than procedural method. The declarative method doesn’t require the programmer to supply the computer with step-by-step instructions. Instead, we must present rules to satisfy and data to be tested against those rules. The Prolog “engine” then works to find relationships among the data that are consistent with the rules. Already this sounds a lot like how we solve a physics problem.

Luckily, in the limited domain of freshman-level 1DK problems, the text of each problem always contains sufficient logical connections to enable a step-by-step analytical solution. This approach doesn’t perform any numeric or mathematical computation, nor does it deliver a specific result; it’s decidedly nonparameter-based. Instead, our approach focuses on three key elements for analytically solving a physics problem:

- understanding the proper inter-relationships between kinematic variables,
- eliciting all possible knowledge from the problem text, and
- organizing knowledge learned along the way.

Figure 1 shows a session with our software, in which a student solves the “Porsche” problem above (the student’s input is shown in bold).

We wrote the Prolog program to contain the rules, which are essentially the 1DK equations.\textsuperscript{4} The student provides the data that the rules are applied to when prompted by the software; the data itself emerges from facts and other clues found in the problem text. The software never requires specific numbers. Instead, it works with either yes/no questions (such as, “Do you know the acceleration of the Porsche?”) or short-answer questions (such as, “Where is the Porsche when it starts?” for which the answer might be “at the traffic light”). This data set can grow internally as the program learns things along the way. Prolog is uniquely suited to these tasks, as it’s rule-based, with built-in database capabilities. The rule-based character provides a natural structure for translating the laws of physics (the rules) into the computer, while the database element permits careful knowledge organization as the solution unfolds. The entire program—which can lead to a solution for any freshman-type, single-object, 1DK problem—consists of approximately 100 lines of Prolog code.

**How Prolog Works**

Prolog reached its most widespread adoption in the late 1980s as the language of “artificial intelligence,” but it never went mainstream. It has largely become an academic computer language, although several commercial implementations exist including Visual Prolog (www.visual-prolog.com), Quintus Prolog (www.sics.se/quintus), Logic Programming Associates Prolog (www.lpa.co.uk), and Arity Prolog (www arity.com). There are two open-source Prologs: Gnu Prolog (www.gprolog.org) and Swi Prolog (www.swi-prolog.org), which we used in our work. Here, we provide a brief synopsis of the language; interested readers can refer to the books *Programming in Prolog\textsuperscript{12} or Prolog Programming for Artificial Intelligence\textsuperscript{13} for more details.

Prolog is a remarkably simple language with almost no formal constructs. At its core, Prolog has a string, called an *atom*, that can’t be further processed. An atom is a sequence of lowercase letters, such as *position*, *acceleration*, or *car*. From atoms, developers can build predicates. For example, if we know a car’s position in a physics problem, we might declare a *know* predicate which, when connected with the car’s position, would be *know(car,position)*. This predicate is a valid Prolog statement, ending in a period, and represents the most basic representation of knowledge about a system (that is, “we know the position of the car.”).

From predicates, we can define rules, as in

\begin{verbatim}
 can_find(X,v):-know(X,v0),know(X,a),know(X,duration_a_is_applied).
\end{verbatim}

This is a valid Prolog rule; it states that we “can find” \( v \) (the speed) of object \( X \) if (\( \cdot \) means “if”) we know both \( X \)’s \( v_0 \) (initial speed) and \( a \) (acceleration), as well as the time over which \( a \) was applied to the object.

\begin{verbatim}
know(car,position). know(car,acceleration). know(car,acceleration).
\end{verbatim}
Do you know about a time interval that involves the porsche? [y/n]: 

What is the porsche doing at the start of this time interval? at_rest. 

What is the porsche doing at the end of this time interval? going_fast. 

Do you know the acceleration of the porsche during the time interval from at_rest to going_fast? [y/n]: 

Do you know the speed of the porsche when at_rest? [y/n]: 

Do you know the speed of the porsche when going_fast? [y/n]: 

Do you know the position of the porsche when at_rest? [y/n]: 

Do you know the position of the porsche when going_fast? [y/n]: 

Is it possible that an event in this list is closely connected to (or is actually the same as) another event? [at_rest, going_fast] [y/n]: 

Using just a few words, can you say anything else that the porsche does in this problem? [y/n]: 

Figure 1. The software’s input and output for the “Porsche” problem. The student’s input is shown in bold.

In its declarative mode, Prolog relentlessly searches for “solutions” to rules based on other rules and any available data. Procedural languages follow a strict path through the given code until terminal conditions are met.

Prolog and 1D Kinematics Problems

To develop a physics-problem solver in Prolog, we first needed a framework.

Kinematic Equations

All 1DK problems in a freshman-level textbook (such as Knight’s) can be solved with two equations:

\[ x = x_0 + v_0 t + \frac{1}{2} a t^2, \]

\[ v = v_0 + a t, \]
and
\[ v = v_0 + a \Delta t, \]
where the usual variable meanings apply.\(^5\) In forming a program design, we need a strict (and ever-obvious) interpretation of these equations as follows.

Because these equations allow for advancing \( x \) and \( v \) forward in time, identification of the time interval \( \Delta t \) is of primary importance. Typical 1DK problems might include several different time intervals, each requiring specific identification and ordering to develop a solution. The program can specify these kinematic variables only after these time intervals are clearly defined. For example, the initial position of an object, \( x_0 \), is strictly the object's position at the time interval's beginning, and \( x \) is its position at the time interval's end. Analogous meanings hold for \( v_0 \) and \( v \). It assumes acceleration, \( a \), is constant and acts over the entire time interval. Unless they're properly connected to some time interval, these kinematic variables can play no logical role in a solution.

Prolog's natural fit with our endeavor is now apparent. We program Equation 1 as a rule into Prolog via the construct

\[
\text{physics\_law}(\text{Object}, \{x(\text{Object}, \text{Tend}), x0(\text{Object}, \text{Tstart}), v0(\text{Object}, \text{Tstart}), a(\text{Object}, \text{Tstart}, \text{Tend}), \Delta t(\text{Object}, \text{Tstart}, \text{Tend}), \text{'}x=x0+v0*\text{\Delta t} + 1/2*a*\text{\Delta t}^2\text{').}
\]

Here \text{physics\_law} is a Prolog rule, containing three parameters, an \text{Object} variable, a list of kinematic variables, and textual advice for the student. In this rule, the kinematic variables are tied to an object \text{Object} and a time interval \((\Delta t = \text{Tend} - \text{Tstart})\) via \text{Tstart} and \text{Tend}. Prolog's built-in pattern-matching capabilities let this rule succeed only if the pattern of kinematic variables—which are linked carefully to a time interval—are matched to known data. To see how this works, consider Prolog's attempt to unify the first term in the \text{physics\_law} rule's list above:

\[ x(\text{Object}, \text{Tend}). \]

In its knowledge base, Prolog might, for example, find \( x(\text{porsche}, \text{top\_speed}) \), meaning "we know the position of the Porsche when it's reached its top speed." This find will instantiate \text{Object} to \text{porsche} and \text{Tend} to \text{top\_speed} for all subsequent terms in the rule. When it comes to \( x0(\text{Object}, \text{Tstart}) \), Object has already been instantiated to \text{porsche}, so it searches for terms matching \( x0(\text{porsche}, \text{Tstart}) \), with \text{Tstart} as yet uninstantiated. It might find \( x0(\text{porsche}, \text{at\_rest}) \) instantiating \text{Tstart} to \text{at\_rest}. Thus, \text{Object}, \text{Tstart}, and \text{Tend} are now instantiated and will remain so for the rest of the terms in the rule. Failure of any subgoal will force backtracking, in which case the program will seek alternative solutions to the \( x0 \) and \( x \) goals.

We again emphasize the organizational structure here. The variables can occur only at the end of a time interval, which is denoted here by the time when the Porsche has attained \text{top\_speed}, according to the student. This end-time boundary must match any other end-time boundaries needed by other kinematic variables if this rule as a whole is to succeed. The same applies for kinematic variables involving start-time boundaries. In traditional "paper and pencil" format, this rule would be written as

\[
\begin{align*}
x_\text{porsche, top\_speed} &= x_\text{porsche, at\_rest} + \\
v_\text{porsche, at\_rest} &= v_\text{porsche, top\_speed, at\_rest} + \\
\frac{1}{2} &\Delta t_\text{top\_speed, at\_rest} = v_\text{porsche, at\_rest} + \frac{1}{2} a_\text{acceleration} \Delta t_\text{top\_speed, at\_rest}
\end{align*}
\]  

(3)

It's true that there's excessive subterm matching capabilities, but it's this specific term-by-term organization that lies at the core of our work. In practice, the \text{physics\_law} rule succeeds if any three of the four variables in its kinematic-variable list are known. The fourth might then be derived (by the student) using the kinematic equation the rule represents (as the software will advise). A similar written rule exists for Equation 2.

Surprisingly, this completes our kinematic equations programming into the Prolog program. Prolog's built-in inference engine handles the rest, relentlessly searching all available data as it attempts to find rule versions for which all but one of the variables is known. At that point, the unknown is assumed computable and added to the program's knowledge base.

Time-Interval Identification

In the Porsche problem, there are two time intervals (when the Porsche is accelerating and when it's coasting). The stone problem also has two time intervals (when the stone is falling in front of, and toward, the window from above). Because of the time intervals' importance, the software aggressively attempts to learn about them as soon as possible. It does so by asking the student what the objects "are doing" at the start and end of each such intervals. To answer these questions, the student is forced to examine a given object's actions and devise descriptions of what seems to occur at the beginning and end of each time interval. This effectively defines "time bounds" for the time interval.

In 1DK problems, there are two types of time intervals. The first are explicitly stated, as in the Porsche and stone problems above. The Porsche accelerates for "5 seconds" and the stone

\text{January/February} 2010
falls for “0.3 seconds.” As Figure 1 shows, the Porsche’s explicit time interval is recognized and bounded by times when the Porsche is “at rest” and “going fast.” This tells the software that we know a fact called \( \text{dt}(\text{porsche, at_rest, going_fast}) \) or \( \Delta_{\text{at_rest-going_fast}} \). To the software, this means the student knows an actual value of \( \Delta \) and can use it in a later computation. For the stone, the 0.3 seconds is bounded by expressions such as “top of window” and “bottom of window.”

The second type of time interval is not explicitly stated, but is known to exist via a particular set of actions that an object takes or imposes. For example, in the stone problem, there’s a time interval during which the stone is falling toward the window, but it’s not known how long this interval lasts. However, even without knowledge of the time interval’s magnitude, the student can supply descriptive time interval boundaries if the program asks “if they can say anything else about what the stone does in the problem.”

The student can recognize that, for a part of the problem, the stone “falls toward the window,” which is a time interval bounded by the acts of “being released by a hand” and “reaching the top of the window.” With this knowledge, the software could discover a \( \text{dt(stone, released_by_hand, top_of_window)} \) (or \( \Delta_{\text{released_by_hand-top_of_window}} \), which currently has no known magnitude. In the Prolog code, this is handled by a modified form of the \( \text{physics_law} \) construct that has the rule \( \text{adjacent_times} \) (Object, Description, \( T_{\text{start}}, T_{\text{end}} \)) (as in \( \Delta \) “boundaries”) in place of \( \text{dt} \) (Object, \( T_{\text{start}}, T_{\text{end}} \)). The former notion is always forced to fail as a Prolog fact because the student can’t compute with a time interval of unknown magnitude. The failing \( \text{dt_bounds} \) predicate instead triggers the software into asking specific questions about the time interval, with the goal of finding the time interval’s magnitude from the kinematic equations. Such questions involve asking the student whether \( a \) is known during the time interval, if \( x \) or \( v \) are known at the time interval’s end, and if \( x_0 \) or \( v_0 \) are known at the time interval’s start (alternatively, it could search its internal database for the same information). After gathering such information, the software sees whether a \( \Delta \) is computable from Equation 1 or 2; if so, it’s added to the problem’s knowledge base.

**Time-Interval Sequencing and Connections**

At any given point in the problem, several textual descriptions might accumulate in the database, linked to the time intervals’ beginnings and ends. Connecting and sequencing such expressions is another critical step toward an analytical solution. In the Porsche problem (Figure 1), the acceleration time interval ends with the description \( \text{going_fast} \), while the coasting time interval begins with \( \text{no_acceleration} \). Although different descriptions, these times actually represent the same time instant. The problem’s story line would indicate that the \( \text{no_acceleration} \) just precedes the \( \text{going_fast} \). In other words, the Porsche begins its coasting just as it ends its acceleration.

Likewise, the stone is seen to be “reaching the top of the window” (after being released) just as its position is known to be at the “top of window.” The software constantly trolls its database for such descriptions and asks the student if any of them actually represent the same physical time instant.

If it knows two time expressions are equivalent—for example \( t_1 \) and \( t_2 \), with \( t_1 \) just preceding \( t_2 \) in the storyline)—then the software can immediately connect the two respective time intervals via kinematic variables, as in

\[
x_0(t_2) = x(t_1)
\]

and

\[
v_0(t_2) = v(t_1)
\]

These equations state that initial parameters in \( x \) or \( v \) at the beginning of the later time interval can be found from the final like-parameters at the previous time interval’s end. We represent Equation 4 as

\[
\text{physics_law(Object,}
\{x'(Object,'/',Tstart)]},
\{x('Object','/',Tend,'})
\text{=} x0 ('Object','/'},
\text{Tstart,'})']
\text{:-}
\text{adjacent_times(Object,Tend,}
\text{Tstart)}
\]

where the rule \( \text{adjacent_times} \) succeeds if two times (\( T_{\text{start}} \) and \( T_{\text{end}} \)) are found in the database and known to be related. Such work is left up to the Prolog engine: the code specifies only a required pattern between any \( x \) and \( x_0 \) (or \( v \) and \( v_0 \)). If one of the two terms is known, the other can be found and added to the database, with instructions given to the student. Knowledge of \( \text{adjacent_times} \) facts comes directly from the student.

**Kinematic Variables**

With clearly identified time intervals in hand, the software can now ask about specific kinematic variables in their proper context. For example, if it
knows the time interval magnitude—either directly or in a derivable sense—between events \( T_{\text{end}} \) and \( T_{\text{start}} \), it can pose a question about \( v \), a kinematic variable that can exist only at the end of a given time interval. The Prolog code

\[
\text{question} \{ \text{Object}, v \text{ (Object, } T_{\text{end}}) \} : - \\
\text{maybe_dt} \{ \text{Object}, \_, T_{\text{end}} \}, \\
\text{not} \{ v \text{ (Object, } T_{\text{end}}) \}, \\
\text{affirm} \{ '\text{Do you know the speed of the} \\
\text{Object, ' when } 'T_{\text{end}}, ' ? ' \}
\]

will handle such a question. The \text{maybe_dt} rule succeeds if the database contains either an explicit or derivable time interval magnitude. The underbar (\_) in this line is the Prolog “anonymous” or “don’t care” variable. Because the code is asking about \( v \), we’re not concerned with the time interval’s beginning, hence the placement of only \( T_{\text{end}} \) in the code. If for example, \text{Object} instantiates to \text{stone} and \( T_{\text{end}} \) to \text{reaching_top_of_window}, the question will read “Do you know the speed of the stone when reaching_top_of_window?” (We fine the broken, grammatically incorrect yet understandable sentences a charming aspect of AI applications.) The \text{affirm} predicate used to pose the question succeeds only if the student answers “yes” to the question.

Such questions are also posed for \( v_0, x_0, x, \) and \( a \). If any of these rules succeed, then the program inserts this kinematic variable knowledge—which is carefully tied to an object and time interval—into the Prolog database as newly gained knowledge.

Other Components
Several lines of code form the software’s core engine. These lines continually drive the software to ask questions, make database assertions, and display any instructions to the student. The software is started with a rule called \text{solve}, which takes a single object name as its parameter—as in \text{solve(porsche)} or \text{solve(stone)}. This rule starts by extracting questions to ask from any available question predicates and updates the database for questions with affirmative answers or those that might generate new time intervals data.

The rule \text{check_laws} is called by \text{solve} after the program asks appropriate questions. It sequences through all \text{physics_laws} rules attempting to find one that succeeds (with all but one of its required variables known). A subgoal of \text{check_laws}, called \text{check_terms}, actually checks each kinematic variable’s availability against the database and returns the one that is unknown, which might now be a known (derivable) quantity.

In practice, the software might terminate without the student finding an answer. If so, it’s run again (and again) as necessary, forcing the student to rethink previous answers or answer new questions, until he or she earns enough data from the problem to find what it requires. Repeated runs are generally different, as run \( N \) can benefit from data asserted into the database in run \( N - 1 \). We generated Figure 1 by running the software three times, with the answer finally appearing near the bottom, in the line describing how \( x \text{(porsche, went_farther)} \) is found.

Running the Software
The complete Prolog code for our work is available at \text{http://ocean.physics.calpoly.edu/prolog}. You can download the code, \text{phys.pl}, and the supporting input/output routines, \text{io.pl}, and save them in a common folder. With Swi Prolog installed, you can load and compile the code by typing [\text{phys}], in response to Prolog’s \text{-} prompt. You can then run the software by typing \text{solve(porsche)} or \text{solve(stone)} following the next \text{-} prompt, to start the program working on what’s computable for that object. Finally, you can rerun the software as needed—via additional \text{solve(porsche)} or \text{solve(stone)}, entries—at each subsequent \text{-} prompt.

Prolog’s input/output interface is extremely fragile. All “yes/no” responses must be a single, lowercase “y” or “n” followed by a period (that is, \text{y. or n.}) and the return key. All short-response inputs must be in the form of a valid Prolog atom construct, followed by a period—that is, all lowercase letters, with no spaces or symbols other than the underbar (\_), such as \text{at_the_top_of_the_window.} (with a period at the end).

Although we could have written similar software using a procedural language, Prolog’s natural ability to handle symbolic information in a strict logical setting makes it a compelling platform for this work. We’re particularly impressed with the pattern-matching ability of Prolog’s inference engine, yet find it ever-difficult to write code that maximally exploits this after a lifetime of procedural programming. We’re also intrigued by Prolog’s \text{call} predicate, which allows data to be executed as code; this enables compact representation of the \text{physics_laws} by allowing each member of a \text{physics_law} list (that is, each data element) to be executed as a Prolog query (code).
Our future work includes adapting this code into a Web-ready form, so students can access it via a Web browser. In addition, we plan to adapt the code to handle kinematics problems involving more than one object, such as “two trains passing each other” and so on. We also see potential for this software to help grade-school students with the dreaded “story problems”—that is, the word problems involving age, money, simple motion, and so on that often stand as a barrier between those who “get math” and those who don’t.

References


Thomas J. Bensky is an associate professor of physics at California Polytechnic State University in San Luis Obispo, California. His research interests include optics of the ocean and the use of computer graphics and artificial intelligence in undergraduate education. Bensky has a PhD in physics from the University of Virginia. Contact him at tbensky@calpoly.edu.

Catherine A. Taff is a mission systems engineer at The Boeing Company. Her research interests are in computer modeling and simulation. Taff has an MS in physics from San Francisco State University. Contact her at ctaff@sfsu.edu.

Selected articles and columns from IEEE Computer Society publications are also available for free at http://ComputingNow.computer.org.
TIMELY, ENVIRONMENTALLY FRIENDLY DELIVERY

DIGITAL EDITIONS

Subscribe to the interactive digital versions of Computer and IEEE Security & Privacy, and access the latest news and information whenever and wherever you want it.

Computer
The IEEE Computer Society's flagship publication, Computer magazine publishes peer-reviewed technical articles on all aspects of computer science, computer engineering, technology, and applications.

Industry professionals, researchers, and managers rely on Computer to keep current on research developments, trends, best practices, and changes in the profession.

Upcoming theme issues include:
- Extreme-scale computing,
- Multi- and many-core,
- Biometric identification,
- Nano-architecture.

To see what you're missing, check out selected Computer articles for free in Computing Now, and then subscribe to the digital edition to get full access right away.

$29.95 for 12 issues!

IEEE Security & Privacy
IEEE Security & Privacy brings together the practical and the leading edge advances in security, privacy, and dependability.

IEEE Security & Privacy covers and influences policy in the enterprise and the government—from basic training and attack trends to the US's cyberattack policy and telephone wiretapping. S&P brings guidance from some of the leading thinkers in the field. Bruce Schneier, Steve Bellovin, Gary McGraw, and Mike Howard have you in mind when writing their columns!

Upcoming theme issues include:
- The insider threat,
- Mobile device security, and
- The security and privacy of cloud computing.


$19.95 for 6 issues!

The latest content at your fingertips within minutes.
Email notification. Receive an alert as soon as each digital edition is available. Links will take you directly to the enhanced PDF edition OR the web browser-based edition.
Quick access. Download the full issue in less than two minutes with a broadband connection.
Convenience. Read your digital edition anytime -- from your home PC, at work, or on your laptop while traveling.
Digital archives. Subscribers can access the virtual archive of digital issues dating back to Jan./Feb. 2007.

To subscribe, go to: computer.org/digitaleditions