



**Centrum voor Wiskunde en Informatica**  
Centre for Mathematics and Computer Science

---

V. Akman, P.J.W. ten Hagen

The power of physical representations

Computer Science/Department of Interactive Systems

Report CS-R8819

May

---

*Bibliotheek*  
Centrum voor Wiskunde en Informatica  
Amsterdam

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

69 K 14

Copyright © Stichting Mathematisch Centrum, Amsterdam

# The Power of Physical Representations †

Varol Akman

Paul J. W. ten Hagen

Department of Interactive Systems

Centre for Mathematics and Computer Science

Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

---

**Abstract:** Commonsense reasoning about the physical world as exemplified by “Iron sinks in water” or “If a ball is dropped it gains speed” will be indispensable in future programs. We argue that to make such predictions (viz. envisioning), programs should use abstract entities (such as the gravitational field), principles (such as the principle of superposition), and laws (such as the conservation of energy) of physics for representation and reasoning. This is along the lines of a recent study in physics instruction where expert problem-solving is related to the construction of physical representations that contain fictitious, imagined entities such as forces and momenta [cf. Jill H. Larkin, “The role of problem representation in physics,” pp. 75-97 in *Mental Models*, ed. D. Gentner and A.L. Stevens, Erlbaum, Hillsdale, N.J. (1983)]. We give several examples showing the power of physical representations.

**Categories and Subject Descriptors:** I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods

**Additional Key Words and Phrases:** Envisioning, Physics laws, Physical entities, Physical representations

---

## 1. Introduction

Leibniz's *An Introduction to a Secret Encyclopedia* includes the following marginal note [1]:

Principle of Physical Certainty: Everything which men have experienced always and in many ways will still happen: e.g. that iron sinks in water.

In our daily lives we use this principle routinely. Thus, we know that we can pull with a string but not push with it, that a flower pot dropped from our balcony falls to the ground and breaks, that when we place a container of water on fire, water may boil after a while and overflow the container.

The origin of such knowledge is a matter of constant debate. It is clear that we learn a great deal about the physical world as we grow up. However, even philosophers were always tricked by the mechanisms which achieve this [2]:

All our knowledge begins with sense, proceeds thence to understanding, and ends with reason, beyond which nothing higher can be discovered in the human mind for elaborating the matter of intuition and subjecting it to the highest unity of thought. At this stage of our inquiry it is my duty to give an explanation of this, the highest faculty of cognition, and I confess I find myself here in some difficulty.

In this paper, we shall argue that some difficulties regarding commonsense reasoning about the physical world can be overcome by using fictitious entities, laws, and principles of physics. This is along the lines of a recent study in physics instruction where expert problem-solving is attributed to the construction of physical representations that contain imagined entities [3].

---

† Presented at the *Second Eurographics Workshop on Intelligent CAD Systems: Implementational Issues*, Veldhoven, the Netherlands (11-15 April 1988).

After a short overview of the role of mental models, followed by the motivation for this research, the structure of this paper is as follows. In §2 we study the nature of physics and a simple abstraction mechanism based on essential attributes and influences. Envisioning and naive vs. physical representations are treated in §3. A brief account of the content of mechanics is given in §4. This is inspired by, yet basically incomparable to, Hayes' classification of the fluids domain in his seminal paper [4] and will be treated in greater detail in our future work. A set of problems showing the power of physical representations can be found in §5. Some interesting research efforts treating the related areas of naive physics and qualitative reasoning are examined in §6. Finally, §7 summarizes our points and suggests some directions.

### 1.1. Mental Models

Why do we advocate that a theory of commonsense reasoning about the physical world should be firmly based on physics? While we see the psychological literature on this subject as a valuable source of information, studies in learning show, maybe expectedly, that human subjects have fuzzy and even wrong ideas about the physics of everyday life. DiSessa [5] found out that a group of elementary school students learning to control a computer-simulated Newtonian object invariably had the wrong Aristotelian expectation that bodies must move in the direction they are last pushed. Another similar study by McCloskey [6] reports that assumptions of the naive theories of motion are quite consistent across college students. It turns out that the theories developed by different individuals are best described as different forms of the same basic theory. What is striking is that this basic theory is highly *inconsistent* with the fundamental principles of classical physics. McCloskey shows that this naive theory of motion is similar to a pre-Newtonian physical theory: the medieval impetus theory that the act of setting a body in motion "imprints" in the object a force, or impetus, that keeps the object in motion. Figure 1 illustrates a case which is very similar to the one examined by McCloskey. Asked about how a metal ball put into the end of the tube and shot out of the other end at high speed would behave, more than half of the subjects have drawn the incorrect path.

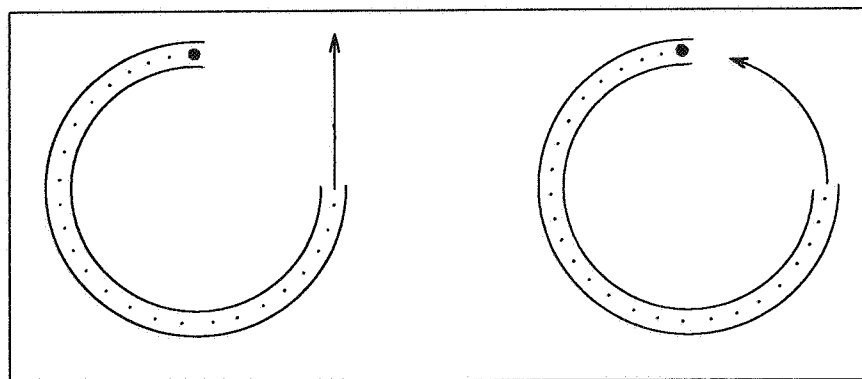


Figure 1. Correct and incorrect responses for the spiral tube problem

## 1.2. Motivation

Why is commonsense physical knowledge useful? In this paper we shall be concerned with mechanics. Therefore we may answer this question from a viewpoint focusing on mechanical design. We identify mainly two aspects:

- Mechanical design results in physical objects. After production a design object is left in the physical world. From that moment on it will interact with an environment where there are physical notions like force, motion, collision, etc. in existence. If we are interested in reliable products which continue functioning correctly under various disturbances, we must take heed of these notions and their effects on our design objects. Thus we may want to know, during the design stage, what happens to a nuclear reactor when a pressure regulator starts to malfunction. Depending on the outcome of such simulations we may embed more security checks and redundancies in our designs.
- All mechanical inventions are firmly based on a deep understanding of the physical world and its laws. If we want to design new devices automatically we need a design system which has an appreciation for physical phenomena. For example, even a simple can opener is a device unifying diverse physical notions such as friction, force, rotation, and cutting. When human designers invent a new device they use their physics knowledge in a fundamental way to reason about the functioning of the device under consideration.

The reader is also referred to [7-9] for detailed accounts of why commonsense physical knowledge may be crucial for realizing "intelligent" computer-aided design systems.

## 2. The Nature of Physics

Observing that all of our ideas in physics require a certain amount of commonsense in their application we see that they are not purely mathematical or abstract ideas. In fact, nearly every page of [10] has a caveat to this effect as the following excerpts show:

This system [a system of discourse about the real world] is quite unlike the case of mathematics, in which everything can be defined, and then we do not *know* what we are talking about. In fact, the glory of mathematics is that *we do not have to say what we are talking about*. The glory is that the laws, the arguments, and the logic are independent of what "it" is. If we have any other set of objects that obey the same system of axioms as Euclid's geometry, then if we make definitions and follow them out with correct logic, all the consequences will be correct, and it makes no difference what the subject was.

... [W]e cannot just call  $F = ma$  a definition, deduce everything purely mathematically, and make mechanics a mathematical theory, when mechanics is a description of nature. By establishing suitable postulates it is always possible to make a system of mathematics, just as Euclid did, but we cannot make a mathematics of the world, because sooner or later we have to find out whether the axioms are valid for the objects of nature. Thus we are immediately get involved with these complicated and "dirty" objects of nature, but with approximations ever increasing in accuracy.

Let us now study in detail why we cannot make mechanics a mathematical theory.

### 2.1. The Meaning of Physical Laws

Centuries of scientific activity gave rise to an enormous body of physical knowledge which can be found in textbooks. The aim is to provide an account of how the physical world behaves. Theoreticians realized that mathematics is an excellent tool for physics since all laws can be written in symbolic form with absolute clarity and economy. Yet physical formulas by

themselves do not provide enough insights [10]:

Although it is interesting and worthwhile to study the physical laws simply because they help us to understand and to use nature, one ought to stop every once in a while and think, "What do they really mean?" The meaning of any statement is a subject that has interested and troubled philosophers from time immemorial, and the meaning of physical laws is even more interesting, because it is generally believed that these laws represent some kind of real knowledge. The meaning of knowledge is a deep problem in philosophy, and it is always important to ask, "What does it mean?"

Let us ask, "What is the meaning of the physical laws of Newton, which we write as  $F = ma$ ? What is the meaning of force, mass, and acceleration?" Well, we can intuitively sense the meaning of mass, and we can *define* acceleration if we know the meaning of position and time. We shall not discuss these meanings, but shall concentrate on the new concept of *force*. The answer is equally simple: "If a body is accelerating, then there is force on it." That is what Newton's laws say, so the most precise and beautiful definition of force imaginable might simply be to say that force is the mass of an object times the acceleration. Suppose we have a law which says that the conservation of momentum is valid if the sum of all external forces is zero; then the question arises, "What does it *mean*, that the sum of all the external forces is zero?" A pleasant way to define that statement would be: "When the total momentum is a constant, then the sum of the external forces is zero." There must be something wrong with that, because it is just not saying anything new. If we discovered a fundamental law, which asserts that the force is equal to the mass times the acceleration, and then *define* the force to be the mass times the acceleration, we have found out nothing. We could also define force to mean that a moving object with no force acting on it continues to move with constant velocity in a straight line. If we then observe an object *not* moving in a straight line with a constant velocity, we might say that there is a force on it. Now such things certainly cannot be the content of physics, because they are definitions going in a circle. The Newtonian statement above, however, seems to be a most precise definition of force, and one that appeals to the mathematician; nevertheless, it is completely useless, because no prediction whatsoever can be made from a definition. One might sit in an armchair all day long and define words at will, but to find out what happens when two balls push against each other, or when a weight is hung on a spring, is another matter altogether, because the way the bodies *behave* is something completely outside any choice of definitions.

For instance, the important thing about force is that it has a material origin. If a physical body is not present then a force is taken to be zero. If we discover a force then we also try to find something in the surroundings that is a *source* of the force. Another rule about force is that it obeys Newton's Third Law: the forces between interacting objects are equal in magnitude and opposite in direction (cf. §4.2). These concepts we have about force, in addition to a mathematical rule like  $F = ma$ , are the key elements in solving physics problems. It is only through combining the mathematical equations with concepts experts attempt to solve physics problems [3].

Consider, for example, the use of constraints to model physical laws. A constraint such as  $f(x_1, \dots, x_n) = 0$  can be used to determine any  $x_i$  if all the others are given a value. However, as de Kleer [11] points out:

There is much more information in an expression! If all we are interested in is solving a set of equations, looking at constraint expressions may be a valid perspective. However, [if] we are solving a physical problem in which a duality exists between the mathematical structure of the equations and the actual physical situation we have thrown away most of the information. To the sophisticated student this duality is very clear and the mathematical equation is far more than a constraint expression. For him, the expression encodes a great deal of qualitative knowledge and every mathematical manipulation

of the expressions reflects some feature of the physical situation.

For example, we shall see in §3.1 that whenever we calculate an imaginary number from a velocity equation, we decide that the body under consideration is not able to reach to a certain point. Similarly, we treat vector equations in a careful manner so that signs of the involved quantities make sense. If a force is trying to prevent the motion of a body, its direction is taken opposite to the direction of movement. The issue of what signs should be assigned to the quantities under consideration is a first sign of a correct attempt to solve a physics problem.

But there is a more important kind of knowledge encoded in a physics formula. Considering again  $F = ma$ ,  $F$  and  $m$  should be placed, during problem-solving, at the same conceptual “layer” since they can be determined independently from any other quantities [12]. However,  $a$  is placed at a higher layer since it is determined by the quantities in the previous layer. Quantities like  $F$  and  $m$  are allowed to be manipulated only externally (exogenous variables) whereas  $a$  is derived (endogenous). The only intended meaning of  $F = ma$  is that at any given moment, the net force on a body is equal to the product of its mass and acceleration.

The notion of causality is of key importance in physics although this is somewhat against Bertrand Russell’s famous advice [12]: “The law of causality, I believe, like much that passes to muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.” We tend to believe, as Shoham aptly stated in [12], that people decide what variables are to be exogenous, and then define causal rules correspondingly. The causal rules that are constructed depend on the theory we are dealing with. For example, we may choose, in case of  $F = ma$ , to write down two rules: one describing the influence of  $m$  on  $a$ , and the other describing the influence of  $F$  on  $a$ . Consequently, we omit a causal rule describing the effect of  $a$  on  $F$ ; that information is nevertheless present in the original formula.

## 2.2. Essential Attributes and Influences

Let us think of a simplifying way of looking at physics problems. We have usually various objects under consideration and we state several things about them. We can use predicates to talk about objects; this follows the advice of Hayes [4]. Thus,  $p(a)$  means that an object  $a$  has an attribute  $p$ . To denote that object  $a$  has no attribute  $p$  we shall write  $\bar{p}(a)$ . Predicates can be one-, two-, ... place depending on how many objects are required to form them. Thus the predicate “ $x$  is frictionless” is one-place, “ $x$  is longer than  $y$ ” is two-place, “ $z$  is faster than  $x$  but slower than  $y$ ” is three-place, and so on. We write, for the preceding examples, in predicate notation *frictionless*( $x$ ), *longer*( $x, y$ ), and *inbetween*( $z, x, y$ ). In general,  $p(a_1, \dots, a_n)$  means that an  $n$ -tuple of objects  $a_1, \dots, a_n$  has an attribute  $p$ ; in other words,  $a_1, \dots, a_n$  such that  $p$  holds.

By *abstraction*, we signify the mapping of a set of objects to a new set of objects which have some of their attributes ignored. For instance, we may take a block and map it to a point mass by simply ignoring its spatial dimension. Similarly, we may map a surface to a frictionless surface, a gas to an ideal gas, and a fluid to an incompressible fluid. By doing so we are creating predicates which are unverifiable, e.g. *frictionless*( $s$ ), *ideal*( $g$ ), *incompressible*( $f$ ) denote attributes which are not possible to observe. The objects denoted by abstract terms do not exist empirically because of the way they are constructed [13]; we *decide* to take certain objects and neglect some attributes of them.

Our view is that this is how physics experts start solving problems. That is, the mapping of

empirical objects to abstract objects is an important ingredient in solving a physics problem since the physical laws in general are stated over abstract entities and as state transformations. Using the attribute viewpoint above, it is easy to write down transformations. Assume that  $s_1$  is a state of an object at a certain moment and  $s_2$  is a state at a later moment. This transformation will be denoted by  $\rightarrow$ . To denote that object  $a$  exists we may write  $E(a)$ . Thus  $\bar{E}(a) \rightarrow E(a)$  denotes the creation of object  $a$  whereas  $E(a) \rightarrow \bar{E}(a)$  denotes the death of  $a$ . We may write similar statements about the attributes. Thus  $\bar{p}(a) \rightarrow p(a)$  stands for acquisition of an attribute while  $p(a) \rightarrow \bar{p}(a)$  stands for loss of an attribute by object  $a$ . It is messy to tell exactly when an object is created or annihilated. Considering a drop of water, if heat is given to it, the drop will acquire the attribute *hot* and lose the attribute *cold*. However, if enough heat is applied, the drop will turn to steam. There are two possibilities: either this is the same individual which was liquid before or this is a new individual.

By *essential influences* we shall mean those influences which play a central role in a situation. Consider the example of a pendulum clock [10]. If a pendulum clock is standing upright it works as expected, but if it is tilted nothing happens. Something else, something outside the machinery of the pendulum clock, is involved in the operation of the clock. This is the earth which is pulling on the pendulum. Once we encode this effect through the vector of gravitational force, we may take the earth out of the picture and consider our pendulum embedded in a field of gravitation — the essential influence for this problem. It is consequently an easy matter to deduce that e.g. the pendulum will tick with a different rate on the Moon.

Shape is a difficult attribute to deal with. Considering a rectangular block sliding along an inclined plane, one may use the abstraction that the block is a point mass. However if a ball is rolling down on an inclined plane, a point mass viewpoint may lead to an incorrect perspective since e.g. the angular momentum due to rotation is no more accountable. For the sliding block problem we take the mass of the block as a constant since we make the reasonable assumption that it is going at a low speed (compared to the speed of light) and its mass is independent of time. Similarly, for a block attached to a spring with stiffness constant  $k$ , we normally think that if the spring is stretched by  $x$  then we can find the force on the spring using  $F = -kx$ . This is based on the assumption that for small  $x$ ,  $k$  is constant. Note that the fact that the spring can break if the force is exceedingly large is not implicit in this formula.

Another way of looking at physical influences is to regard them as functions. Thus by  $(fx)$  let us denote that function  $f$  takes object  $x$  and gives  $fx$  as a new object. This is an intensional definition of a function as opposed to the usual, extensional way. Functional equivalence then becomes  $(fx) = (ghx)$ , i.e. apply  $g$  on the result of applying  $h$  to  $x$  to obtain a summary effect equivalent to  $(fx)$ . This notation helps one to formulate state transformations. Consider object  $x$  which is a block of ice. Now  $(fx)$ , where  $f$  is heat, renders a new object  $fx$  which is equal to water. Let us call this object  $\bar{x}$ . Applying  $f$  on  $\bar{x}$  gives a new object which is equal to steam. (Thus self-application of functions is sometimes meaningful.)

### 3. Envisioning

In envisioning we deliberately ignore the values of the problem variables — we let them take any value. This permits us discover all the possibilities for a given problem. Only one of these possible answers is observable for a given set of initial conditions. The idea of envisionment will now be made clearer by studying a pioneering program.



### 3.1. NEWTON

The first attempt for predicting the interaction of moving objects in a simple, idealized world has been made by Johan de Kleer and a program called NEWTON was implemented. NEWTON works in a 2-dimensional world called the "roller-coaster" world. We start with a classical problem used by de Kleer [11]. In Fig. 2, a small sliding block (idealized as a point) of mass  $m$  starts at point  $c_1$  along a frictionless surface consisting of three parts:  $s_1$  and  $s_2$  which are concave, and  $s_3$  which is straight. We want to know if the block will reach point  $c_4$ .

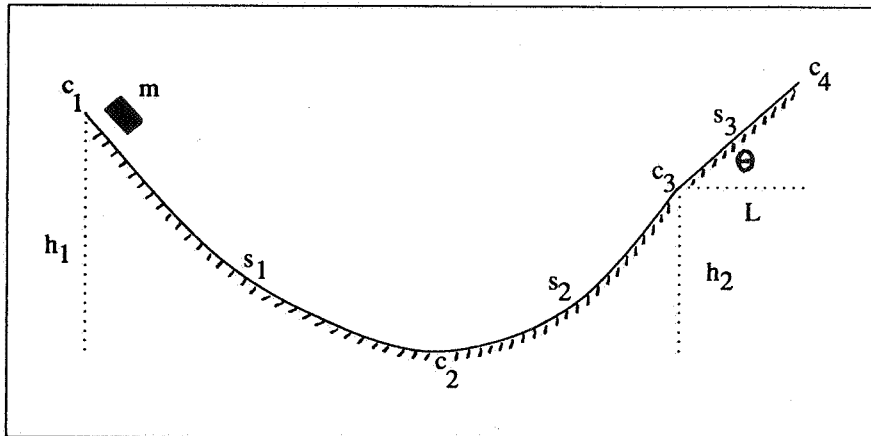


Figure 2. The sliding block problem

In NEWTON, a production rule system looks at the local geometric (topological) features to predict what might happen next. The left hand side of a production describes the features of the environment (e.g. the shape of the paths, the velocity of the block) and the right hand side lists the consequences (e.g. sliding, falling). A closed world assumption<sup>1</sup> is in effect: the actions on the left hand side never produce a change in the features (e.g. a block sliding on a segment does not cause a change in the shape of the segment).

We may summarize the reasoning suggested by de Kleer as follows. Without falling off or changing direction  $m$  will start to slide down the surface. After reaching the bottom  $c_2$ , it will start going up. It still will not fall off, but it may start sliding back. If  $m$  ever reaches the straight section (i.e. the segment  $s_3$ ) it still will not fall off there, but it may change the direction of its movement. To determine exactly whether  $m$  reaches  $c_4$  we must study its velocity. The velocity at the bottom can be computed by using the conservation of energy:  $v_{c_2} = \sqrt{2gh_1}$ . Similarly, using  $v_{c_2}$  and conservation of energy we can set up an equation which can be solved for the

<sup>1</sup> Davis gives, in a related work, the following justification for the closed world assumption [14]: "The "frame" or "persistence" problem of determining what remains true over time requires no special treatment in our logic. We can avoid this problem, not by virtue of any special cleverness on our part, but by virtue of the structure of the domain. The predicates in the domain are divided into two classes. The first class includes predicates which depend on position and velocity of objects. These are not assumed to remain constant over any interval unless they can be proven to be so. The second class includes structural predicates, depending only on the shape and other material properties of the objects. These are defined to be constant over the problem, and so are defined atemporally. Similar considerations would seem to apply to any closed world, complete physical theory; it is not clear why the frame problem should ever create trouble in such a domain."

velocity  $v_{c_3}$  at the start of the straight section:  $\frac{1}{2}mv_{c_3}^2 = \frac{1}{2}mv_{c_2}^2 - mgh_2$ . If  $v_{c_3}$  is imaginary, we know that the straight segment is never reached. At the straight section we would use kinematics to find out whether the block ever reaches  $c_4$ . The acceleration of the block along  $s_3$  must be  $a = g \sin\theta$ . The length of the straight segment is  $\frac{L}{\cos\theta}$ , so using the kinematic equation relating acceleration, distance, and velocities we get:  $v_{c_4}^2 = v_{c_3}^2 - 2Lg \tan\theta$ . Again, if  $v_{c_4}$  is imaginary,  $c_4$  is not reachable.

This line of reasoning can be conveniently depicted in the *envisionment* tree of Fig. 3. The branches of the tree correspond to the different situations that may arise depending on the solutions of the equations mentioned above. Basically, we have a set of worlds  $W$  and a set of times  $T$ . There is a linear ordering  $<$  (meaning ‘prior to’) over  $T$ . We can then think of a 2-dimensional space where we have differing worlds as we move along one axis and differing times as we move along on the other axis. Any particular point in this space can be thought of as a pair of coordinates  $(w, t)$  for some  $w \in W$  and  $t \in T$ . This index denotes a point whose location is determined by which world it is on on one hand, and by what time it is on the other hand. The ‘possible worlds’ in Fig. 3 are then as follows, in a left-to-right order:

- $w_1$ : Slide until at most  $c_3$ , slide backwards, oscillate around  $c_2$ .
- $w_2$ : Slide until at most  $c_3$ , slide backwards, fall off from  $c_1$ .
- $w_3$ : Slide until after  $c_3$  but before  $c_4$ , slide backwards, oscillate around  $c_2$ .
- $w_4$ : Slide until after  $c_3$  but before  $c_4$ , slide backwards, fall off from  $c_1$ .
- $w_5$ : Slide until  $c_4$ , fall off from  $c_4$ .

For example, denoting the moments that  $m$  passes through the points  $c_i$  by  $t_i$ , it is seen that  $m$  will be at  $c_2$  at  $t_2$ , at some point of  $s_2$  for any  $t$  such that  $t_2 < t < t_3$ , and then for all  $t > t_3$  on  $s_1$  or  $s_2$ , if it follows the prediction denoted by  $w_1$ .

At each fork of the envisionment tree there are several possibilities. A block would take one of the branches always and would go down the tree, ending up in a leaf. The forks correspond to the points where we have to *disambiguate* the ambiguities whereas leaves correspond to particular behaviors. Disambiguation is achieved by solving some equations. Looking at the formula giving  $v_{c_2}$  we see that this velocity can never be imaginary. Thus the block would always reach  $c_2$ ; it would never stop there since it would, at that point, have a positive velocity. Let us look at the next equation giving the value of  $v_{c_3}$ . It can be rewritten, after manipulation, as  $v_{c_3}^2 = 2g(h_1 - h_2)$ . This equation has a simple message. If  $c_3$  is located higher than  $c_1$  then there is no way that  $m$  would reach  $c_3$ . Until now we did not really need any numerical knowledge other than  $h_1 > h_2$  to decide that  $m$  will arrive at  $c_3$ . However, after  $c_3$  we need to know the values of  $L$  and  $\theta$  to discover that  $c_4$  is reachable. Assume that  $m$  has made it to  $c_4$  barely, i.e.  $v_{c_4} \geq 0$ . Then the inequality  $h_1 - h_2 \geq L \tan\theta$  should be satisfied.

### 3.2. Naive vs. Physical Representations

NEWTON has been a corner-stone in the search for knowledge representation and reasoning methodology for physical domains. A later work on mechanics problem-solving in the style of NEWTON is MECHO [15]. This program tries to improve the envisioning process by making it more goal-directed. The main improvement is that MECHO does not generate the envisionment tree but only the parts needed<sup>2</sup> to answer a question. Thus, for a suitable set of values of the

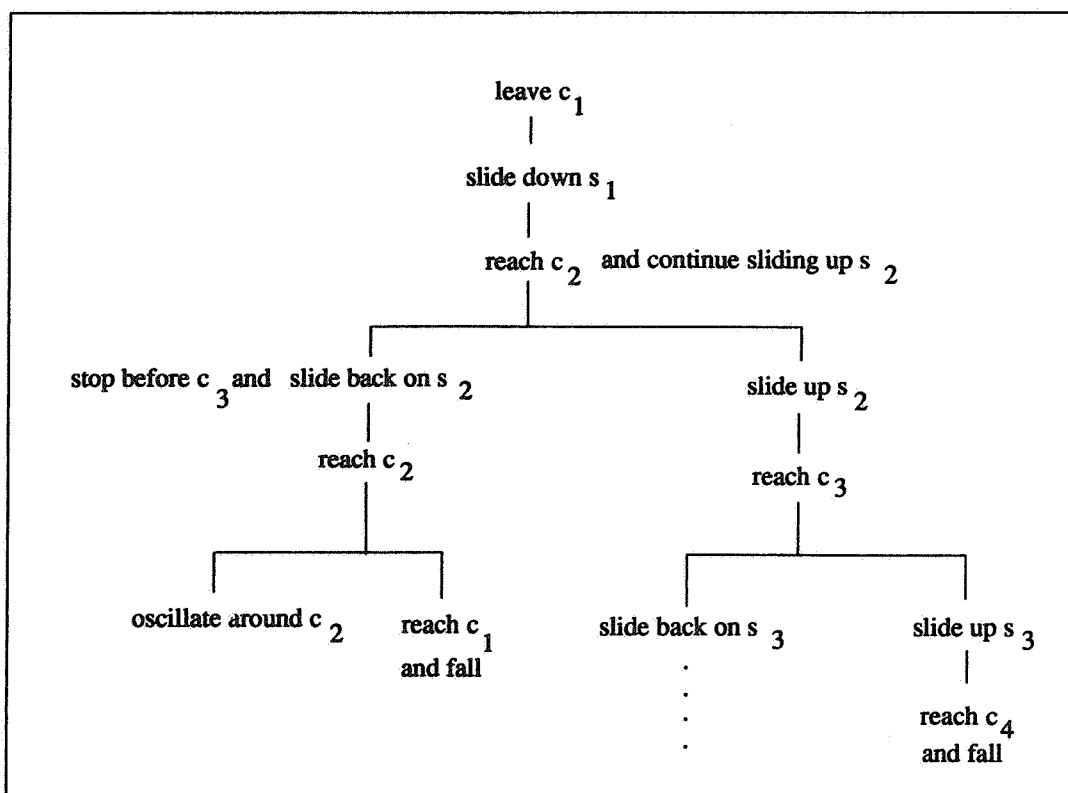


Figure 3. The envisionment tree for the sliding block problem

problem parameters above MECHO would only generate the real path that would be taken by the block, say the left-most.

Larkin [3] points out that NEWTON has a naive way of representing physical knowledge. In particular, it has an internal representation which contains direct representations of the visible entities mentioned in the problem description. As a result, it performs simulation — NEWTON's inferencing via an envisionment tree follows the direction of time flow. What is more serious is the lack of deeper physics principles in NEWTON. For instance, Larkin notes that a physicist encountering the above problem about the sliding block might reason as follows [3]. (It should probably be remarked that the following protocol is hypothetical and does not originate from an interview with an expert physicist.) The energy at  $c_1$  consists of kinetic energy, zero because  $m$  is at rest, and potential energy determined by  $h_1$ . At  $c_2$  the potential energy is zero, because the block is at the bottom and the kinetic energy is unknown because the speed is unknown. At  $c_3$  the potential energy is determined by  $h_2$  and the kinetic energy is still unknown. At some point  $c$  (which may be above or below  $c_4$ ), the block stops<sup>3</sup> and the kinetic energy is again zero. Writing down the equation  $0 + mgh_1 = 0 + mg(h_2 + X \sin\theta)$ , where  $X$  is the distance  $m$  travels along  $s_3$  after point  $c_3$ , immediately leads to a solution. (It is noted that in the last equation, inside the parentheses, Larkin writes  $\frac{X}{\sin\theta}$  which is wrong.) Briefly, if  $X > \frac{L}{\cos\theta}$  then  $m$  will reach  $c_4$  and fall off.

<sup>2</sup> It is arguable that an envisioner should generate only the relevant parts of an envisionment tree. Goal-directed envisioning is fine as long as questions like "What happens next?" are avoided.

<sup>3</sup> It is incorrect (contrary to what [3] suggests in the context of its hypothetical expert protocol) to think that the

Compared to NEWTON's solution, there are good insights in this solution. The fact that a body has speed is identified with its having kinetic energy. That a body is at a given height is identified with its having potential energy. In writing down the preceding equation, the absence of friction (and hence no loss of energy due to heat dissipation) is used to state a simpler law of conservation of energy. The fact that potential energy and kinetic energy are convertible to each other is used implicitly. The expert is also aware that  $m$ , which is initially at rest, is brought to motion by the gravitational field of the earth which is constant and equal to  $g$  in distances not too far away from the surface of earth. As another pointer to the use of deep knowledge, consider the decision of the expert to select  $c_2$  as a "standard" point for potential energy calculation. If he had used any other point, say  $c_3$ , instead of  $c_2$ , it is obvious that the potential energy is changed only by the addition of a constant. Since the energy conservation law cares only about *changes*, it does not matter if a constant is added to the potential energy.

NEWTON's use of quantitative knowledge (represented as frames) to disambiguate ambiguities remains as a very important contribution. Frames in NEWTON are not procedures but describe dependencies among variables. They are similar to Minsky's frames [16] in that they are used to chunk physical formulas of the same nature. Since there are many different equations applicable to a problem, grouping equations in frames help isolate the relevant ones with less effort. Dependencies among variables are then searched for a solution (the goal variable). For example, a kinematics frame<sup>4</sup> knows about the usual equations of motion [17]:

**frame kinematics of *object*, *surface*,  $t_1$ ,  $t_2$**   
**variables:** ( $v_1$ : velocity of object at time  $t_1$ ,  $v_2$ : velocity of object at time  $t_2$ ,  
 $d$ : distance of *surface*,  $t$ : time between  $t_1$  and  $t_2$ ,  $a$ : acceleration of *object*)  
**equations:** ( $v_2 = v_1 + at$ ,  $v_2^2 = v_1^2 + 2ad$ ,  $d = v_1 t + \frac{1}{2}at^2$ )

Frames use two kinds of variables: the names of the objects and their essential attributes (such as velocity or acceleration). Each equation of a frame referencing the goal variable is a possible way to determine that variable. On the other hand, unknown variables in equations referencing the goal variable must be given a value (possibly using other frames or asking the user) since every unknown in the equation must be determined before achieving the goal.

#### 4. The Content of Mechanics

This section gives, rather tersely, the building blocks of classical mechanics. It is neither complete nor uniform as it stands. However, it should give an idea about what kind of knowledge should be formalized for a deep coverage of mechanics. It is noted that we are not *yet* concerned with a concrete knowledge representation scheme. This is thought to be in agreement with [4] where the building of mini-theories (clusters) are given priority over favorite notations: "Initially, the formalizations need to be little more than carefully worded English sentences. One can make considerable progress on ontological issues, for example, without actually formalizing anything, just by being *very* careful what you say."

---

body stops. Since there is no friction the body will certainly oscillate indefinitely. Thus, an expert would only consider the extremal points of the oscillation where the kinetic energy is indeed zero.

<sup>4</sup> De Kleer published a revised, shorter version of [11] later in [17]. An interesting decision has been made in the second version to call the RALCMs (Restricted Access Local Consequent Methods) of the first version FRAMEs.

#### 4.1. Fictitious Entities

*Description of Motion:* Speed as a derivative ( $v = \lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t}$ ), distance as an integral ( $s = \lim_{\Delta t \rightarrow 0} \sum_i v(t_i) \Delta t$ ), and acceleration as the derivative of speed are the basic notions. We distinguish velocity from speed, which is the magnitude of velocity. (We shall sometimes denote vectors with bold letters.)

*Pseudo Vectors:* Ordinary vectors are e.g. the coordinate, force, momentum, velocity, etc. Vectors which are obtained as a cross product are artificial, e.g. torque  $\tau$ , angular velocity  $\omega$ , and angular momentum  $L$ . The corresponding equations are  $\tau = r \times F$ ,  $v = \omega \times r$ , and  $L = r \times p$ , respectively, where  $p$  is the momentum and  $r$  is the radius.

*Work:* If a body is moving along a curved path, then the change in kinetic energy as it moves from one point (1) to another (2) is equal to  $\int_1^2 \mathbf{F} \cdot d\mathbf{s}$ .

*Torque:* Torque bears the same relationship to rotation as force does to linear movement, i.e. a torque is the thing which makes something rotate or twist.

*Energy:* In mechanics problems, energy takes basically the following forms: gravitational energy, kinetic energy, heat energy, and elastic energy.

*Power:* Power equals work done per second. In other words, the rate of change of kinetic energy of a body is equal to the power used up by the forces acting on it.

*Field:* We need two kinds of laws for a field. The first law gives the response to a field, i.e. the equations of motion. The second law gives how strong the field is, i.e. the field equation. In other words, "One part says that something *produces* a field. The other part says that something is *acted on* by the field. By allowing us to look at the two parts independently, this separation of the analysis simplifies the calculation of a problem in many situations" [10].

The idea of a field is closely associated with potential energy. We note that the gravitational force on a body is written as mass times a vector which is dependent only upon the position of the body:  $\mathbf{F} = m \mathbf{C}$ . This lets us analyze gravitation by imagining that there is a vector  $\mathbf{C}$  at every position which effects any mass placed there. Since the potential energy can be written as  $m \int (\text{field}) \cdot (d\mathbf{s})$ , we find that the potential energy of a body in space can be written as mass times a function  $\Psi$ , the potential. To get the force from the potential energy we use:  $F_x = -\frac{\partial U}{\partial x}$  (and similarly for other directions). To get the field from the potential we do the same thing:  $C_x = -\frac{\partial \Psi}{\partial x}$  (and similarly for other directions). More succinctly,  $\mathbf{F} = -\nabla U$  and  $\mathbf{C} = -\nabla \Psi$ .

*Pseudo Force:* A well-known pseudo force is what is often called *centrifugal* force. If we are in a rotating coordinate system, we experience a force throwing us outward. Using the pseudo force we can explain several interesting problems (Fig. 4). In this figure, adapted from [10], a container of water is pushed along a table, with acceleration  $a$ . The gravitational force acts downward but there is in addition a pseudo force acting horizontally. The latter is in a direction opposite to  $a$ . As a result, the surface of the water will be inclined at an angle with the table, as shown in the left part of the figure. If now we stop applying a push, the container will slow down because of friction, and the pseudo force will reverse its direction, causing the water stand higher in the forward side of the container.

Among the forces that are developed in a rotating system, centrifugal force is not the only one. There is another force called *Coriolis* force. It has the property that when we move a body

in a rotating system, the body seems to be pushed sidewise. If we want to move something radially in a rotating system, we must also push it sidewise with force  $F_c = 2m\omega v_r$ . Here  $\omega$  is the angular velocity and  $v_r$  is the speed the body is making along the radius.

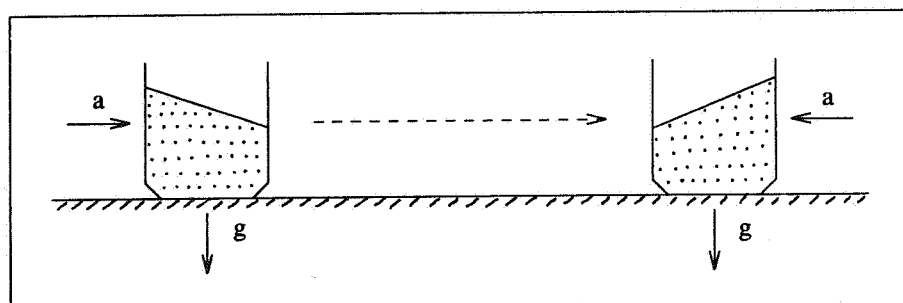


Figure 4. Pseudo force acting on a container of water

**Conservative Force:** If the integral of the force times the distance in traveling from a point to another is the same (regardless of the shape of the curve connecting them) then the force is *conservative* (e.g. gravity).

**Center of Mass:** Given a rigid body there is a certain point such that the net external force produces an acceleration of this point as if the whole mass is concentrated there. The point does not have to be in the “material” of the body but can lie outside.

## 4.2. Laws of Mechanics

**Newton's Laws of Dynamics:**

**First Law (The Principle of Inertia):** If an object is left alone (not disturbed), it continues to move with a constant velocity in a straight line if it was originally moving, or it continues to stand still if it was standing still.

**Second Law:** The motion of an object is changed by forces in this way: the time-rate-of-change of momentum is proportional to the force.

**Third Law (The Principle of “Action equals reaction”):** Suppose we have two small bodies and suppose that the first one exerts a force on the second one, pushing it with a certain force. Then simultaneously the second body will push on the first with an equal force, in the opposite direction. These forces effectively act in the same line.

**Conservation of Linear Momentum:** If there is a force between two bodies and we calculate the sum of the two momenta, both before and after the force acts, the results should be equal.

**Conservation of Angular Momentum:** If no external torques act upon a system of particles, the angular momentum remains constant. In other words, the external torque on a system is the rate of change of the total angular momentum:  $\tau_{ext} = \frac{dL_{tot}}{dt}$ .

**Conservation of Energy:** There is an abstract quantity that does not change in all the natural phenomena which the world undergoes: energy.

**(Galilean) Relativity Principle:** The laws of physics look the same whether we are standing still or moving with a uniform speed in a straight line.

**Work Done by a Force:** If the force is in one direction and the object on which the force is

applying is displaced in a certain direction, then only the component of force in the direction of the displacement does any work: i.e. physical work is expressed as  $\int \mathbf{F} \cdot d\mathbf{s}$ .

*Work Done by Gravity:* The work done in going around a path in a gravitational field is zero. This implies that we cannot make perpetual motion in a gravitational field.

*Hooke's Law (The Law of Elasticity):* The force in a body which tries to restore the body to its original condition when it is distorted is proportional to the distortion. This holds true if the distortion is small. If it gets too large the body will be torn apart or crushed.

### 4.3. Principles

Although they are classified under different headings in this paper, it should be noted that the boundary between physics laws and physics principles is not well-defined. We suggest that by principles we should refer to the problem-solving elements, e.g. the principle of virtual work below. In any case, this section is very unsatisfying in its present form and should include more material.

*Superposition:* The total field due to all the sources is the sum of the fields due to each source. Suppose that we have a force  $F_1$  and have solved for the forced motion. Then we find out that there is another force  $F_2$  and solve for the other forced motion. Using the superposition of solutions we can now predict what would happen if we had  $F_1$  and  $F_2$  acting together. The solution is  $x_1 + x_2$  if  $x_i$ 's are the individual solutions for forces  $F_i$ , respectively. In general, a complicated force can be divided into a set of separate forces each of which is simple (in the sense that we can solve for the forced motion they cause).

*Equivalence of Simple Harmonic Motion and Uniform Circular Motion:* Uniform motion of a body in a circle is closely related to oscillatory up-and-down motion. Although the distance  $y$  means nothing in the oscillator case, it can still be artificially given in order to model oscillation in terms of circular motion.

*Virtual Work:* We imagine a structure moves a little — even though it is not really moving or even movable. We use this small imagined move to apply the law of energy conservation. This principle is especially useful in problems of the sort depicted in Fig. 5 where we are asked to find the value of weight  $W$  such that the system is in equilibrium. Noting that a small move of  $W$  toward the bottom should be counteracted by weights  $W_1$  and  $W_2$ , we find  $hW = \frac{d_1}{d} hW_1 + \frac{d_2}{d} hW_2$ , or  $W = \frac{W_1 d_1 + W_2 d_2}{d}$ .

*Analogs:* We have already seen examples of analogs when we talked about force vs. torque, linear momentum vs. angular momentum, etc. in §4.1. In general, analogy refers to relating two domains which are at first sight dissimilar and using the tools of one domain to solve the problems of the other. The analogy between mechanics and electricity is well-known, as Table 1 partially illustrates [10]. As a result, not only in problem-solving but also in design, analogy has found its deserved place. The following excerpt is especially illuminating in this respect [10]:

Suppose we have designed an automobile, and want to know how much it is going to shake when it goes over a certain kind of bumpy road. We build an electrical circuit with inductances to represent the inertia of the wheels, spring constants as capacitances to represent the springs of the wheels, and resistors to represent the shock absorbers, and so on for parts of the automobile. Then we need a bumpy road. All right, we apply a *voltage* from a generator, which represents such and such a kind of bump, and then look at how the left wheel jiggles by measuring the charge on some capacitor. Having measured it (it is easy to do), we find that it is bumping too much. Do we need more shock absorber, or less



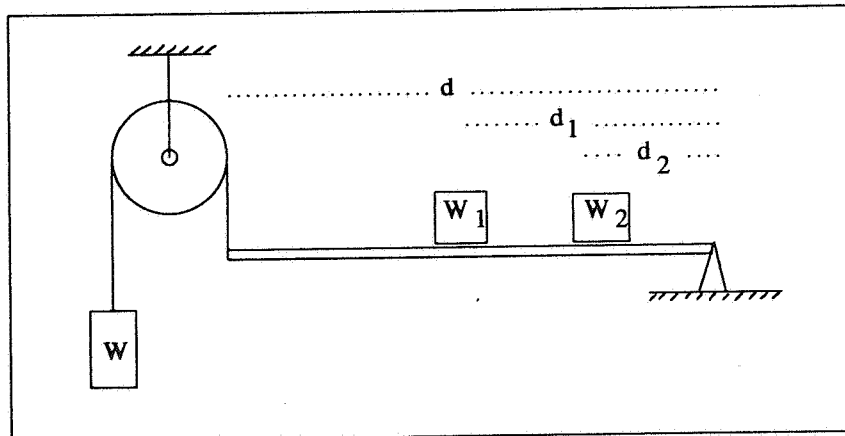


Figure 5. A system of blocks in equilibrium

shock absorber? With a complicated thing like an automobile, do we actually change the shock absorber, and solve it all over again? No!, we simply turn a dial; dial number ten is shock absorber number three, so we put in more shock absorber. The bumps are worse — all right, we try less. The bumps are still worse; we change the stiffness of the spring (dial 17), and we adjust all these things *electrically*, with merely a turn of a knob.

Table 1. Analogs		
General Characteristic	Mechanical Property	Electrical Property
independent variable	time ( $t$ )	time ( $t$ )
dependent variable	position ( $x$ )	charge ( $q$ )
inertia	mass ( $m$ )	inductance ( $L$ )
force	force ( $F = ma$ )	voltage ( $V$ )
velocity	velocity ( $v$ )	current ( $I$ )
resistance	drag coefficient ( $c = \gamma m$ )	resistance ( $R = \gamma L$ )
stiffness	stiffness ( $k$ )	(capacitance) <sup>-1</sup> ( $1/C$ )
period	$t_0 = 2\pi\sqrt{m/k}$	$t_0 = 2\pi\sqrt{LC}$

## 5. Some Examples

In this section we give some example problems and their solutions. Our aim is to demonstrate the use of physical representations. We regard the following problems as difficult problems for envisioners and expect that they will constitute part of a test set for future programs.

### 5.1. Rocket Problems

These are taken from [10]. How fast do we have to send a rocket away from the earth so that it leaves the earth? The problem can be stated as a functional requirement: the rocket must leave the earth. We are now asked to find an attribute (the speed) of the rocket. Notice that many other attributes of the rocket has been left out and a pair of objects, the earth and the rocket, have been identified. The essential influence is the earth's gravitational field.



Kinetic energy plus potential energy of the rocket must be a constant. Let us exaggerate and imagine the rocket in two extreme positions. When it is far away from the earth it will have zero potential energy. Besides its kinetic energy will also be zero since we may assume that it barely left the earth. On the other hand, initially it has the total energy  $\frac{1}{2}mv^2 - G\frac{mM}{R}$ , where  $m$  is the rocket's mass,  $M$  is the earth's mass, and  $R$  is the earth's radius. The conservation of energy gives  $v = \sqrt{2gR}$ , where  $g = \frac{GM}{R^2}$ .

Changing the problem a little, at what speed a satellite should travel to keep going around the earth? It turns out that the conservation of energy is not the right way to approach this problem. A force approach written as an equality between centrifugal and gravitational forces,  $\frac{mv^2}{R} = \frac{GMm}{R^2}$ , gives  $v = \sqrt{gR}$ . The reason we have thought that a force expression is more convenient here is due to the nature of the problem; there is a "rotating" object and this guides our search among the mathematical expressions applicable to the problem. This should be compared with de Kleer's search strategy [11].

Now let us take a look at the following problem. A rocket of large mass  $M$  ejects a small piece of mass  $m$  with a velocity  $V$  relative to the rocket. Assuming it were standing still, the rocket now gains a velocity  $v$ . Using the law of conservation of momentum, this velocity is seen to be  $v = \frac{m}{M}V$ . Thus, rocket propulsion is essentially the same as the recoil of a gun. It does not need air to push against [10].

Let us suppose that the two objects are exactly the same, and then we have a little explosion between them. After the explosion one will be moving, say toward the right, with a velocity  $v$ . Then it appears reasonable that the other body is moving toward the left with a velocity  $v$ , because if the objects are alike there is no reason for right or left to be preferred and so the bodies would do something that is symmetrical.

## 5.2. The Ball with Strings

Consider Fig. 6 which is taken from den Hartog [18]. A heavy ball of weight  $W$  is suspended by a thin thread and has an identical thread hanging down from it. When we start pulling down on the lower string, which of the two strings will break first?

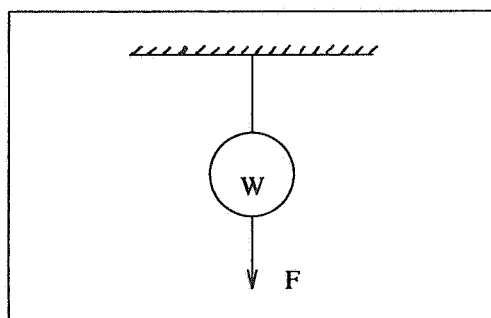


Figure 6. The ball with strings

We assume that we apply force  $F$ . Then the bottom string experiences  $F$  whereas the upper

string experiences  $W + F$ . Thus the upper string will break first. According to den Hartog, this happens only if we pull down slowly. If we instead give a sudden, sharp pull to the lower string, it will break and the ball remain suspended. This has to do with the fact that the threads are elastic and have a certain elongation associated with the force sustained by them. By giving a quick pull to the lower thread, the force in the lower thread can be made quite large and this force will accelerate the ball downward. But this takes some time and before any appreciable downward displacement is observed the string is broken.

Den Hartog relates a similar experiment. Consider a ball, with a single string attached to it, lying on a table. By a slow pull, one can drag the ball on the table with a uniform speed. In this case the applied force is equal to the friction force between the ball and the table. A quick pull, on the other hand, will break the thread in an instant. The ball is subjected, for a short time, to a large force which subjects it to acceleration. However the ball will hardly move since the time interval is small. Instead its velocity will be destroyed by the retarding action of the friction force.

### 5.3. The Open Water Jar

A major weakness of the current envisioners is their inability to switch between macro- and micro-worlds. In other words, the "individuals" that an envisioner knows about are either in the world as we see it or underlying the world (atomic processes). (In fact, Schmolze's thesis is the only attempt, albeit a rather restricted one, to deal with atomic processes [19].) To our best knowledge, there is no current envisioner which can predict that e.g. water would slowly evaporate if it is left in an open jar in room temperature. This is due to the fact that there is no change that is immediately observable at a macro level — the realm which the envisioner talks about.

Physicists think that our knowledge of atoms, that all things are made of atoms, particles that move around in perpetual motion, attracting or repelling each other depending on the distance between them, gives us very useful information [10]. The atomic hypothesis describes processes very well. For example, let us observe the above jar with some water in it. What will happen as time passes? The water molecules are constantly moving around. From time to time, one on the surface is hit strong enough so that it flies out. Thus, molecule by molecule the water evaporates.

If we cover the jar with a lid, we observe no change because the number of molecules leaving the surface are equal to the number of those coming back. Thus, in the long run nothing happens. Note on the other hand that we should find a large number of molecules amongst the air molecules. If we take the lid away and push dry air instead of the moist air, again water will evaporate. The number of molecules leaving the surface is still the same but not as many are coming back.

### 5.4. The Heated Rubber Band

This is taken from [10] and is another good example as to the convenience of thinking in terms of atomic processes. If we apply a gas flame to a rubber band holding a weight, the band contracts abruptly, as shown in Fig. 7. How can we explain this?

A molecular explanation would be as follows. Rubber band consists of a tangle of long chain of molecules, not unlike a molecular spaghetti, with cross links. When such a tangle is pulled out, some of the chains line up along the direction of pull. At the same time chains are

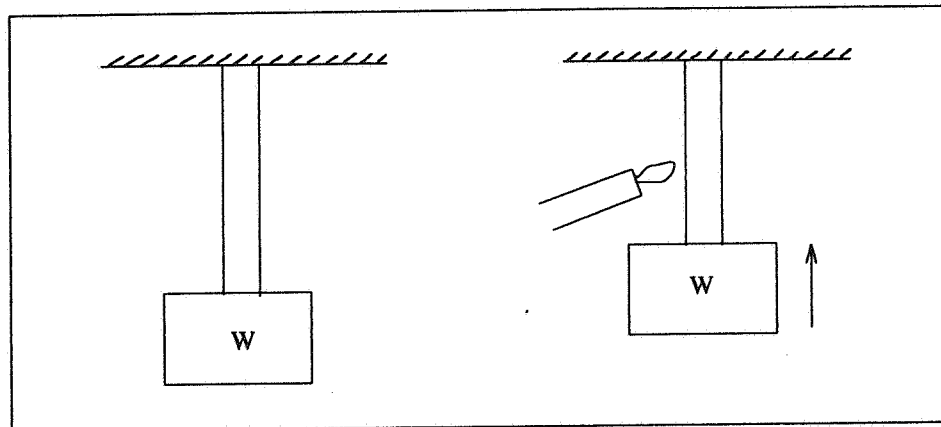


Figure 7. The heated rubber band

hitting each other continually, due to thermal motion. Thus, if a chain is stretched, it will not by itself remain stretched; it would be hit from the sides by other chains which would tend to unstretch it again. When we heat the rubber band, we increase in effect the amount of bombardment on the sides of the chains.

Using thermodynamics, the above explanation can be made more quantitative. When heat  $\Delta Q$  is delivered, the internal energy is changed by  $\Delta U$  and some work is done. The work done by the rubber band is  $-F \Delta L$  where  $F$  is the force on the band and  $L$  is the length of the band. Note that  $F$  is a function of temperature  $T$  and  $L$ . Now we have  $\Delta U = \Delta Q + F \Delta L$ . Additionally, we have  $\Delta Q = -T \frac{\partial F}{\partial T} \Delta L$ , which tells us that if we keep the length fixed and heat the band, we can calculate how much the force will increase in terms of the heat needed to keep the temperature constant when the band is stretched a bit.

## 6. Other Related Research

A discussion of naive physics and qualitative physics is given in the dissertation of Schmolze [19]. Here we offer a shorter discussion for completeness and refer the reader to the above reference for an analysis of these areas. (N.B. We omit a discussion of Forbus' *Qualitative Process Theory*, notwithstanding its importance [20].) The articles by de Kleer and Brown [21], Kuipers [22, 23], Forbus [20], Forbus and Gentner [24], and Hayes [4] form the basis of naive and qualitative physics. Incidentally, it is not easy to delineate the areas covered by these terms; we propose that naive physics should be understood as the construction of knowledge representation methods while qualitative physics should cover reasoning techniques.

Patrick Hayes proposed, by the term *naive physics*, the construction of a formal, sizable portion of commonsense knowledge about the physical world [4]. This should include knowledge about objects, shape, space, movement, substances, time, etc. As for the knowledge representation language to be used Hayes is not specific; a collection of assertions in logic, for example, may be sufficient (cf. §2.2). He is not, at this preliminary stage, interested in the efficiency of reasoning with this body of knowledge. What he is really after is "the extent to which it [a naive physics theory] provides a vocabulary of tokens which allows a wide range of intuitive concepts to be expressed, to which it then supports conclusions mirroring those which

we find correct or reasonable" [4].

*Qualitative physics* provides an account of behavior in the physical world. Unlike conventional physics qualitative physics predicts and explains behavior in qualitative terms. Although the behavior of a physical system can be given by the precise values of its variables (temperatures, velocities, forces, etc.) at every moment, such a description fails to provide insights into how the system works. Important concepts causing change in physical systems are concepts like momentum, force, feedback, etc. which can be understood intuitively [21]. They are in conventional physics embedded in a framework of continuous differential equations. In qualitative physics each measurable property such as the speed of a ball is represented in two parts: a quantity plus its rate of change. The representation is qualitative since the quantity values are selected from a discrete quantity space. For example, the quantity space for water may have only two values: the freezing and the boiling temperatures. Rates of change are in general limited to three cases: increasing, decreasing, and constant.

### 6.1. ENVISION

De Kleer and Brown [21] introduced qualitative differential equations (confluences) and implemented them in a program called ENVISION. To obtain confluences, we let continuous variables take discrete values from a quantity space such as  $\{0, +, -\}$ . Let  $[x]$  denote the qualitative value of an expression  $x$  with respect to the quantity space. Then the proposition " $x$  is increasing" is written as  $\frac{dx}{dt} = +$  and we can define arithmetic to deal with  $[x] + [y]$  or  $[x][y]$ , although there will be cases where ambiguities will have to be resolved using numeric values (e.g. when the operation is addition and  $[x] = +$  and  $[y] = -$ ). Let  $\partial x$  denote  $[\frac{dx}{dt}]$ . Using confluences, we can provide explanations as follows. Consider the pressure regulator in Fig. 8 which is adapted from [21]. A confluence such as  $\partial P + \partial A - \partial Q = 0$  where  $P$  is the pressure across the valve,  $A$  is the area available for flow, and  $Q$  is the flow throughout the valve describes this device in qualitative terms. It is seen that an increase in pressure at  $a$  causes an increase in pressure at point  $b$ . This generates more flow through  $b$ . As a result pressure at  $c$  increases and this is felt at  $d$ . Then the diaphragm  $e$  will press downward, causing the valve to close somewhat. As a result, constant pressure will be maintained at  $c$  although the pressure at  $a$  is fluctuating. Since a single confluence may not characterize the behavior of a component over its *entire* range of operation, the range is divided into subregions, each characterized by a different component state where different confluences apply. When the valve is closed the correct confluence should read  $\partial Q = 0$ ; we simply do not say anything about  $P$ . Similarly, when the valve is open the confluence becomes  $\partial P = 0$ .

ENVISION has a component library where the components relevant to the domain of reasoning are stored. Another module holds the "topology" of the device under consideration. The input to ENVISION are the input signals and boundary conditions. The output is a behavioral prediction along with a so-called causal explanation. De Kleer and Brown's principle of *no-function-in-structure* requires that the laws of the parts of the device may not presume the functioning of the entire device. Various class-wide assumptions are made to avoid tricky situations, e.g. in fluid flow there are always enough particles in a pipe so that macroscopic laws hold; the mean free path of the particles is small compared to the distances over which the pressure appreciably changes; dimensions of an electrical circuit are small compared to the wavelength associated with the highest frequency; etc.

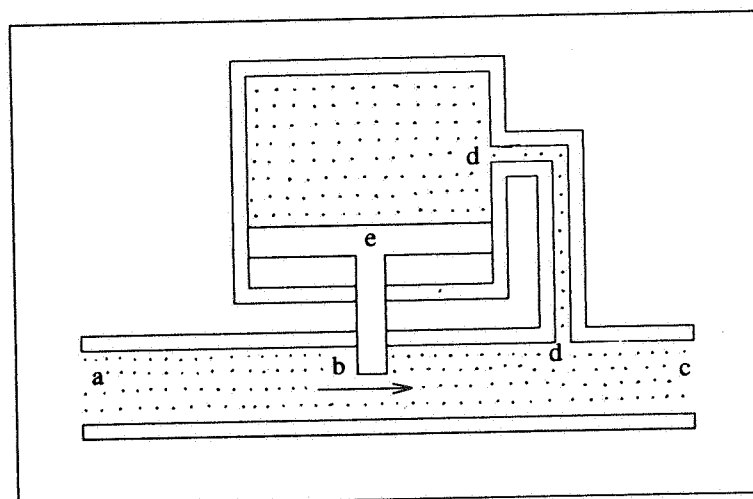


Figure 8. A pressure regulator

## 6.2. QSIM and CA

*Qualitative simulation* can also be defined as the derivation of a description of the behavior of a mechanism from a qualitative description of its structure. Kuipers thinks that causality can be taken as identical to value propagation with constraints [23]. Hence, Kuipers has a constraint-centered ontology [22]. (On the other hand, de Kleer and Brown [21], and Forbus [20] have device- and process-centered ontologies, respectively.) Kuipers' qualitative simulation framework is a symbolic system which solves a set of constraints obtained from differential equations. His QSIM algorithm is guaranteed to produce a qualitative behavior corresponding to any solution to the original equation. He also shows that in some cases a qualitative description of structure is consistent with intractably many behavioral predictions. A couple of techniques representing different trade-offs between generality and power have been proposed for "taming" this intractable branching [25].

As an example of how QSIM can be used consider the problem of throwing an object vertically into air from some height. Using a problem description language this is first described to QSIM. The description includes, in addition to the initial values of the problem variables, the physical constraints (law-like knowledge) acting on the ball:  $\frac{ds}{dt} = v$ ,  $\frac{dv}{dt} = a$ , and  $a = g < 0$  where  $s$ ,  $v$ , and  $a$  stand for displacement, velocity, and acceleration, respectively. These are entered to the system by the user. (In the future, it may be possible to extract the relevant portion of a "physics knowledge base" to use it in a problem.<sup>5</sup>) From this initial information, QSIM reasons in an intuitive way and finds that the ball will rise to a particular height, stop, and then fall to the earth again. The *landmark* point where the velocity becomes zero is discovered by

<sup>5</sup> An important issue is then the composition of the mechanism under consideration: a system's behavior should be deducible from its structure (e.g. components and their connections), as mentioned in the case of ENVISION. Analogs of this principle are used in various domains. In linguistics, it is assumed that the semantic value of any expression is a function of the semantic values of its syntactic constituents. Thus, the semantic rules will compute the semantic values of increasingly longer parts of a statements. Admittedly, this is sometimes a dangerously mechanistic view.

QSIM. Landmark values are important in identifying the different regions of behavior for a mechanism.

In our view, the main problem with QSIM lies in its essence. Obviously QSIM has no knowledge of physics whatsoever. In the above example, QSIM cannot possibly know that the object may leave the earth to orbit around it or even go to the infinity if it is given enough speed initially (cf. §5.1). This is because QSIM assumes that for all cases  $a$  is constant — it does not know that the value  $g$  for  $a$  is applicable only near the surface of the earth and that  $a$  decreases inversely with the square of distance. Thus, by submitting a problem to QSIM it is assumed that all relevant knowledge is given; any physics law left unspecified cannot be used. It is only fair to note that this is not a total condemnation of QSIM. The simple yet mathematically elegant framework of QSIM has much to offer. We think that one can “teach” QSIM the necessary domain-specific knowledge (in our case classical mechanics but in another case e.g. economical laws) and thus employ it in a more useful manner. It remains to be seen whether QSIM’s problem description style is sufficient enough to codify laws such as energy conservation. (The answer is probably no.)

An interesting extension of qualitative simulation is Weld’s “comparative analysis” [26] which was implemented in a program called CA. CA deals with the problem of predicting how a system will react to perturbations in its parameters and why. For example, comparative analysis explains why a ball would go up higher if it is thrown with a greater speed. Weld also studies *exaggeration*. Consider a question like “What happens to the period of oscillation of a block attached to a spring on a frictionless table as the mass of the block is increased?” Exaggeration suggests that if the mass were infinite, then the block would hardly move and thus the period would be infinite. Therefore, had the mass increased a little the period would increase as well. (For another example use of exaggeration, cf. §5.1.)

## 7. Summary

Contrary to their public image, it is admitted that expert systems are not true experts in their fields [27]. This contributed to the emergence of dichotomies such as “deep vs. shallow knowledge” in AI. While there are problems with using these adjectives, it is understandable that an expert system is not a “deep” model of its domain of expertise. An expert system’s *if-then* rules cannot capture but the superficial characteristics of a domain. On the other hand, a real expert has profound thoughts in his domain of expertise. Several criteria have been suggested for true expertise; we shall only give a representative list:

- Experts can explain their way of reasoning in a logical way. They do this in a way markedly different from the current expert systems. When asked about how a certain result was obtained, an expert gives information related both to the real world and its abstract models — not just a list of the rules used.
- Experts use multiple models of a domain to classify and solve problems in an efficient manner. They solve difficult problems using difficult methods. However, for easy problems they either reduce the complicated methods to simpler versions or use already available simple methods. Furthermore, experts can predict. Before they delve into a problem they have some general idea about how the solution should look like.
- Experts can discover the inconsistencies in ill-defined problems. They can judge and eliminate irrelevant or contradictory information. Presented with incomplete problem statements, they make reasonable (default) assumptions. As a result, they can work in a

“nonmonotonic” mode, occasionally revising their beliefs during the problem-solving process.

In this paper we argued that an expert theory of envisioning should mainly be based on physics knowledge. This theory satisfies the above criteria. The fictitious entities of physics such as energy, work, force, etc. make up the things that the envisioner reasons about. Abstract principles such as the principle of superposition and laws such as Newton’s laws are used as the essential tools for reasoning. Simplifying techniques such as essential attributes and influences are advocated as preliminary aids to problem-solving. Our work is in the precise spirit of Larkin’s study [3] in expert-novice difference in problem-solving performance and is based on her distinction between naive vs. physical representations:

Many problems remain to be solved. Central among them is the problem of choosing a concrete knowledge representation method. We think that frames are appropriate for this purpose [16]. Another problem is to define a core subset of physics which can be used effectively for envisioning in *different* areas. Until now, we have mainly studied the domain of mechanics, following the AI tradition. Yet another project is to study the limits of reasoning *without* recourse to law-like knowledge. To invert a remark of Larkin [3], Why are people good at predicting the outcome of physical interactions in the world around them, while they are so bad at physics, even the branch of physics (mechanics) that deals with interaction of everyday objects? Are there other, fundamentally different ways of looking at the physical world? For example, John McCarthy writes that a commonsense knowledge “database would contain what a robot would need to know about the effects of moving objects around, what a person can be expected to know about his family, and the facts about buying and selling” in addition to other information [28]. Here the problem lies in integrating these diverse (and obviously, not all physical) domains of knowledge using a base language — a problem McCarthy calls *generality in AI*. He then adds: “This does not depend on whether the knowledge is expressed in a logical language or in some other formalism.” We agree but cannot help to point out that this brings us to the dangerous waters of sense experience, learning, causality, etc. [29]:

In the notice that our senses take of the constant vicissitude of things, we cannot but observe that several particular, both qualities and substances, begin to exist, and that they receive their existence from the due application and operation of some other being. From this observation we get our *ideas* of *cause* and *effect*. That which produces any simple or complex idea we denote by the general name, *cause*, and that which is produced, *effect*. Thus, finding that, in that substance which we call wax, fluidity, which is a simple idea that was not in it before, is constantly produced by the application of a certain degree of heat, we call the simple idea of heat, in relation to fluidity in wax, the cause of it, and fluidity the effect.

#### ACKNOWLEDGMENTS

This paper owes a great deal to [10] as it is evident from the amount of material cited. The interpretations given here, however, are our own and should not be taken as representing the views of the authors of [10]. The first author wishes to thank Tetsuo Tomiyama (University of Tokyo) for his encouragement and invaluable advice.



**References**

1. Gottfried Wilhelm Leibniz, *Philosophical Writings*, Everyman's Library, London (1984).
2. Immanuel Kant, *Critique of Pure Reason*, Everyman's Library, London (1984).
3. Jill H. Larkin, "The role of problem representation in physics," pp. 75-97 in *Mental Models*, ed. D. Gentner and A.L. Stevens, Erlbaum, Hillsdale, N.J. (1983).
4. Patrick J. Hayes, "The second naive physics manifesto," pp. 467-485 in *Readings in Knowledge Representation*, ed. R.J. Brachman and H.J. Levesque, Morgan Kaufmann, Los Altos, Calif. (1985).
5. Andrea A. diSessa, "Unlearning Aristotelian physics: A study of knowledge-based learning," *Cognitive Science* 6, pp. 37-75 (1982).
6. Michael McCloskey, "Naive theories of motion," pp. 299-323 in *Mental Models*, ed. D. Gentner and A. Stevens, Erlbaum, Hillsdale, N.J. (1983).
7. Bart Veth, "An integrated data description language for coding design knowledge," pp. 295-313 in *Intelligent CAD Systems I: Theoretical and Methodological Aspects*, ed. P. ten Hagen and T. Tomiyama, Springer-Verlag, Heidelberg (1987).
8. Varol Akman, Paul ten Hagen, and Tetsuo Tomiyama, "Design as a formal, knowledge engineered activity," Technical Report No. CS-R8744, Center for Mathematics and Computer Science, Amsterdam, the Netherlands (1987).
9. Varol Akman, Paul ten Hagen, Jan Rogier, and Paul Veerkamp, "Knowledge engineering in design," *Knowledge-Based Systems* 1(2), pp. 67-77 (1988).
10. Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman Lectures on Physics, Volume I (Mainly Mechanics, Radiation, and Heat)*, Second Printing, Addison-Wesley, Reading, Mass. (1966).
11. Johan de Kleer, "Qualitative and quantitative knowledge in classical mechanics," Technical Report No. AI-TR-352, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. (1975).
12. Yoav Shoham, "Reasoning about change: Time and causation from the standpoint of artificial intelligence," Doctoral Dissertation, Technical Report No. YALEU/CSD/RR#507, Department of Computer Science, Yale University, New Haven, Conn. (1986).
13. Aleksandr A. Zinoviev, *Logical Physics*, trans. from the Russian by O.A. Germogenova, ed. R.S. Cohen, Reidel, Dordrecht, Holland (1983).
14. Ernest Davis, "A logical framework for solid object physics," Technical Report No. 245, Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York (1986).
15. Alan Bundy, "Will it reach the top? Prediction in the mechanics world," *Artificial Intelligence* 10, pp. 129-146 (1978).
16. Marvin Minsky, "A framework for representing knowledge," pp. 211-277 in *The Psychology of Computer Vision*, ed. P.H. Winston, McGraw-Hill, New York (1975).
17. Johan de Kleer, "Multiple representations of knowledge in a mechanics problem-solver," pp. 299-304 in *Proceedings of the Fifth International Joint Conference on AI (IJCAI-77)*, ed. R. Reddy, Massachusetts Institute of Technology, Cambridge, Mass. (1977).



18. J.P. den Hartog, *Mechanics*, Dover, New York (1961).
19. James G. Schmolze, "Physics for robots," Doctoral Dissertation, Technical Report No. 6222, BBN Laboratories Inc., Cambridge, Mass. (1987).
20. Kenneth D. Forbus, "Qualitative process theory," *Artificial Intelligence* 24, pp. 85-168 (1984).
21. Johan de Kleer and John Seely Brown, "A qualitative physics based on confluences," *Artificial Intelligence* 24, pp. 7-83 (1984).
22. Benjamin Kuipers, "Commonsense reasoning about causality: Deriving behavior from structure," *Artificial Intelligence* 24, pp. 169-203 (1984).
23. Benjamin Kuipers, "Qualitative simulation," *Artificial Intelligence* 29, pp. 289-338 (1986).
24. Kenneth D. Forbus and Dedre Gentner, "Learning physical domains: Toward a theoretical framework," Technical Report No. UIUCDCS-R-86-1247 (also UILU-ENG-86-1774), Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Ill. (1986).
25. Benjamin Kuipers and Charles Chiu, "Taming intractable branching in qualitative simulation," pp. 1079-1085 in *Proceedings of the Tenth International Joint Conference on AI (IJCAI-87)*, ed. J. McDermott, Milan, Italy (1987).
26. Daniel S. Weld, "Comparative analysis," Technical Report No. AIM-951, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. (1987).
27. Daniel Bobrow, Sanjay Mittal, and Mark Stefik, "Expert systems: Perils and promise," *Communications of the ACM* 29(9), pp. 880-894 (1986).
28. John McCarthy, "Generality in artificial intelligence," *Communications of the ACM* 30(12), pp. 1030-1035 (1987).
29. John Locke, *An Essay Concerning Human Understanding*, Abridged Edition, Everyman's Library, London (1985).

