Chapter

7

Qualitative Physics: Past, Present, and Future

Kenneth D. Forbus
Qualitative Reasoning Group
Department of Computer Science
University of Illinois at Urbana, Champaign

1 *Introduction*

Qualitative physics is concerned with representing and reasoning about the physical world. The goal of qualitative physics is to capture both the commonsense knowledge of the person on the street and the tacit knowledge underlying the quantitative knowledge used by engineers and scientists. The area is now a little over ten years old, which, at least measured in the span of AI, is a long time. So it makes sense to step back and try to systematize the work in the field and describe the current state of the art.

I'll start by describing what qualitative physics is, why one should be doing it, and where it came from. Then I'll sketch the current state of the art, at least the part that is now fairly stable. Then I'll describe what I think lies around the corner, including some pointers to recent work and some interactions between qualitative physics and other fields. Finally, I'll describe some open problems, each of which will probably require quite a few inspired Ph.D. theses to crack.

Qualitative physics is growing rapidly, and thus any survey is likely to become quickly dated. For example, several problems which were described as virgin territory when this material was presented at AAAI-86 have now been at least partially explored. Nevertheless, I think the general framework for understanding the area that was presented then remains sound, and so I have remained faithful to that organization.

2 *Why Qualitative Physics?*

Consider what we need to know about the physical world to make coffee. We know that to pour coffee from the pot into a cup requires having the cup under the spout of the kettle, and that if we pour too much in, there will be a mess on the floor. We know all this without knowing the myriad equations and numerical parameters required by traditional physics to model this situation.

Suppose we were going to build a household robot that, among other duties, made coffee. We might start by using traditional physics to model the situation. Immediately several problems arise. There are few formal axiomatic theories of physics. The formal aspects of physics, the equations, do not by themselves describe when they are applicable. What, for example, is the equation for the cup? There isn't one, per se, but rather various aspects of the cup potentially participate in several different equations describing "what happens" in the world. Many everyday physical phenomena, such as boiling, are not easily described by a single equation. And even when equations exist, people who know nothing about them can often reason fluently about the phenomena. So equations cannot be necessary for performing such reasoning.

But suppose for a moment that we had such a set of equations. Could we use them? Realistic equations rarely permit closed-form, analytic solutions. Even when they do, the high computational complexity of symbolic algebraic means it's not the sort of computation you want going on inside a robot engaged in real-time activity. An alternate route is numerical simulation. By plugging in numerical values, we could generate a very precise description of what will happen. But such simulations require immense computational resources. Worse yet, it assumes the existence of a complete set of accurate values for all input parameters. Typically we just don't have such accurate information, thus forcing us to search a space of parameters corresponding to the ranges the various input parameters may take. This increases the amount of computation even more, making numerical simulation infeasible.

Even if numerical simulation were technologically feasible, by say shirt-pocket supercomputers, or by allowing rough approximations, it still would be insufficient for our robot. First, we still need to interpret the output of the simulation. A list of numerical state parameters is not the most perspicuous representation of an event. Second, any run of a numerical simulator provides a specific set of predictions about what the system being simulated will do. This will suffice for some tasks, but not for all. Often we want to characterize the possibilities that might occur, with some guarantee of completeness. For instance, a fault-tree analysis of a power plant that captured only a small fraction of the failure modes of the system would be inappropriate. With numerical simulations it is often hard to tell when one has captured all of the possible behaviors.[1] In many situations one needs a rapid and rough estimate of what is possible, rather than a very precise prediction based on many unsupported assumptions. A robot pouring coffee should be cognizant of the possibility of overflow, and not spend its time calculating just how big the resulting puddle might be.

These problems are not specific to making coffee; they hold more generally whenever one tries to reason about the physical world. To summarize, these problems are:

1. *The modeling problem:* How does one map from real-world objects to the abstractions of one's physics?
2. *The resolution problem:* Carrying out numerical simulations requires more detail than is often available. Reasoning techniques that can exploit low resolution, partial information are required for commonsense reasoning.
3. *The narrowness problem:* Traditional simulation provides precise answers given a particular set of assumptions. Many reasoning problems require knowing alternative possibilities, rather than a single projection.

At first these problems may seem surprising. Physics, one of the crowning successes of the scientific method, has been carried on for hundreds of years. But consider: Physicists already have commonsense theories of the world. Their goal is to create models capable of more precise explanations. With few exceptions, the focus of formalization lies with building new models that have significantly better predictive and explanatory power than our implicit commonsense models. Qualitative physics arises from the need to share our intuitions about the physical world with our machines.

There are many potential applications of qualitative physics. As argued elsewhere [Gentner and Stevens, 1983; de Kleer and Brown, 1984; de Kleer, 1984], the tacit knowledge of engineers and scientists rests on this shared framework. If we are to build programs that capture this expertise, we must understand the foundation qualitative physics provides. We will return to this point after briefly summarizing the essence of qualitative physics.

1 It is said that if the angular increment in the simulation of the aerodynamic properties of the Boston John Hancock building had been halved, the fact that the building's windows would tend to pop out in high winds could have been predicted. Instead, it was discovered empirically.

2.1 *The Essence*

The key to qualitative physics is to find ways to represent continuous properties of the world by discrete systems of symbols. One can always quantize something continuous, but not all quantizations are equally useful. One way to state the idea is the *relevance principle: The distinctions made by a quantization must be relevant to the kind of reasoning performed* [Forbus, 1984b].

The idea is simple, but few quantizations satisfy it. Rounding to fewer significant digits, replacing numbers by arbitrary intervals, using simple symbolic groups like TALL, VERY TALL, and fuzzy logic do not satisfy it. Signs generally do, since different things tend to happen when signs change (balls fly up and then down, different kinds of things can happen if the level of coffee in a cup is rising versus falling). Inequalities do, since processes tend to start and stop when inequalities change (heat flows occur when there is a temperature difference, boiling occurs when the liquid's temperature reaches its boiling point).

Good quantizations allow more abstract descriptions of state, which in turn make possible more concise descriptions of behavior. If our state parameters are elements of $\Re$, there are potentially an infinite number of states. Replacing state parameters by floating-point numbers makes the number of potential states finite, but still numbering in the billions for many systems. In the quantizations of qualitative physics there may be as few as a dozen, or a hundred, or in some cases thousands. Each state in a qualitative physics typically corresponds to many states in a traditional description, each distinguished by having the same "meaningful behavior pattern" occurring in them.

Abstraction is a two-edged sword. While these abstract state descriptions succinctly capture possible behaviors, they tend not to prescribe exactly which behavior will occur. By themselves they typically cannot, for we have thrown away just that information required to settle such questions. Thus qualitative simulations tend to be ambiguous. Often such answers suffice, e.g., if a household robot cannot imagine any way for the house to burn down as a consequence of its plan to cook supper, then its plan is reasonably safe. However, if a house fire is a possibility, more knowledge must be invoked. The ability of qualitative physics to represent this ambiguity explicitly is beneficial, since it provides a signal to indicate when more detailed knowledge is required.

A central goal of qualitative physics is to achieve a degree of systematic coverage and uniformity far in excess of today's knowledge-based systems. In today's expert systems, knowledge is encoded about a particular domain for a particular purpose. Instead of continuing to build such systems, qualitative physics strives to create *wide-coverage, multi-purpose domain models.* By wide-coverage, we mean that there is some large but precisely characterizable set of systems that can be described by the domain model. It is assumed that every model for a specific system is built by instantiating appropriate elements of the domain vocabulary in appropriate ways. This will reduce the amount of hand-crafting required for new programs and will hopefully lead to "off the shelf" knowledge bases.

By multi-purpose, we mean that a domain model (or a model for a specific situation) can be used for more than one inferential task. Characterizing these *styles of reasoning* is another goal of qualitative physics. These styles of reasoning include qualitative simulation, interpreting measurements, planning, comparative analysis, and others. Developing domain-independent characterizations of these styles will hopefully lead to generic algorithms that can be used as modules in a variety of larger systems.

2.2 *Potential Applications*

To turn robots loose in unconstrained environments, we must teach them qualitative physics. Often we must enlist physical processes to carry out our plans. For example, if I want to make coffee in the morning, I need to use the stove to make boiling water. This requires filling the kettle, putting the pot on the stove, turning the stove on, and waiting for it to boil. One could imagine writing a little expert system to do this. It wouldn't take many IF-THEN rules to express this particular procedure. However, if you lived in my house you would prefer a robot to be reasoning from first principles. My stove is a little unusual: The surface that contains the burners retracts into the wall, under the oven. When the stove is retracted, the burners are directly under the electrical wiring for the oven. Having been designed in the 50's, it has no safety cutoff switch. Turning the burner on when the stove is retracted, or retracting the stove when the burner is still hot, is likely to burn the house down. It is doubtful that the designer of the IF-THEN rules could have taken my stove into account, so I would be very nervous about turning such a machine loose in my house. And houses are fairly stereotyped; consider such machines loose in a construction site. Clearly, such robots will need some form of qualitative physics

But qualitative physics has many other potential applications as well. The subject matter of many expert systems includes aspects concerned with the physical world, particularly in the sciences and engineering. Diagnosis and design are two obvious examples. As remarked above, qualitative physics identifies the "tacit knowledge" that engineers and scientists use to ground the formalisms they learn in school and on the job.

Consider for example the problem of building an intelligent tutoring system for propulsion systems. Figure 1 shows a simplified layout of a Navy propulsion system. Distilled water is fed into the boiler, heated by oil-fired burners, and turned to steam. The system operates at very high temperature and pressure (950° F, 1200 psi) to increase the amount of energy transferred per pound of steam. The steam is heated in the superheater, to impart even more energy. (By the time it leaves the superheater in a shipboard system, it is travelling faster than the speed of sound.) Here is a hard problem that instructors routinely ask about this situation: Suppose the feedwater temperature increases, as might occur when travelling in a warmer part of the ocean. What happens to the temperature at the superheater outlet?

This is a complicated situation, and most of us haven't had a lot of experience with it, so it hardly qualifies as commonsense physics. Yet qualitative reasoning suffices to answer it. In fact, qualitative reasoning is crucial: While a few numerical values have been provided, many critical ones have not, including how much the feedwater temperature rises! Here is the solution, according to instructors at the Navy Surface Warfare Officer's school in Newport, Rhode Island. The water coming into the boiler is now hotter. The boiling will occur at the same temperature, so this means that the amount of heat that must be added to get a piece of water to boil is reduced. This means the water will boil sooner, which means the rate of steam production increases. Assuming a constant load, this means the steam spends less time in the superheater. Since the amount of heat transferred to the steam in the superheater is a function of the time it spends in the superheater, and the starting temperature of the steam is the same, less heat is transferred. Thus the steam temperature at the superheater outlet falls when the feedwater temperature rises.

The ability to make these subtle, yet human-like, deductions makes qualitative physics an excellent candidate for a knowledge component in intelligent tutoring systems [Forbus and Stevens, 1981; Forbus, 1984a] and plant monitors. For example, Figure 2 shows an explanation generated by one of my programs a long time ago, as part of the STEAMER system. The valve shown is a spring-loaded reducing valve, and it converts 1200 psi steam to 12 psi steam at constant pressure, for a wide range of loads. The important thing to notice is that the terms of the explanation are those which are easily understood by human students and operators. No numerical values were used to generate these conclusions—just a very simple qualitative physics.[2]

Qualitative physics also has many potential applications in other aspects of engineering [Forbus, 1987b]. Consider a really smart mechanical design assistant that could generate a description of possible behaviors before detailed parameters were chosen. Suppose the desired behavior exists in the space of behaviors predicted by a qualitative simulation. Then the design effort proceeds by choosing parameters to force the desired behavior, and not the alternatives, to occur. If the desired behavior is not even possible, then it is clear that the design must be changed, even without more details. It does not take detailed

2 The physics used was the early de Kleer and Brown physics, which provided only perturbation analysis, not full dynamical reasoning. The limitations of this approach inspired my own qualitative process theory (and their confluences theory).

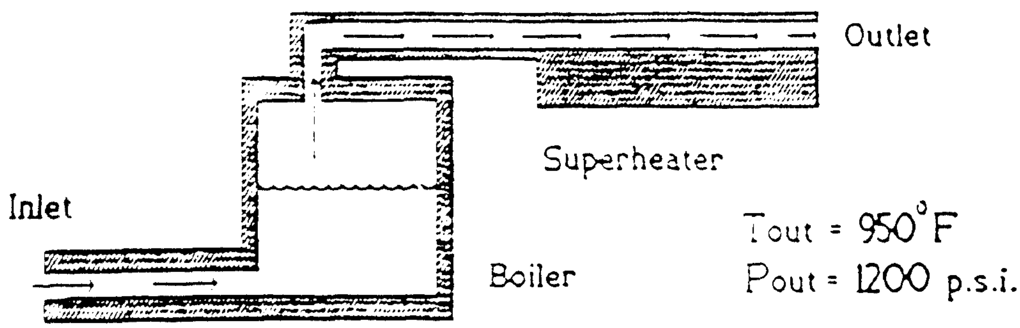


Figure 1. The SWOS Problem. Given that the temperature of the feedwater is increasing, what is the temperature at the superheater outlet? Instructors at the Navy Surface Warfare Officer's School say this is one of the hardest problems students are given, yet it can be answered with purely qualitative reasoning.

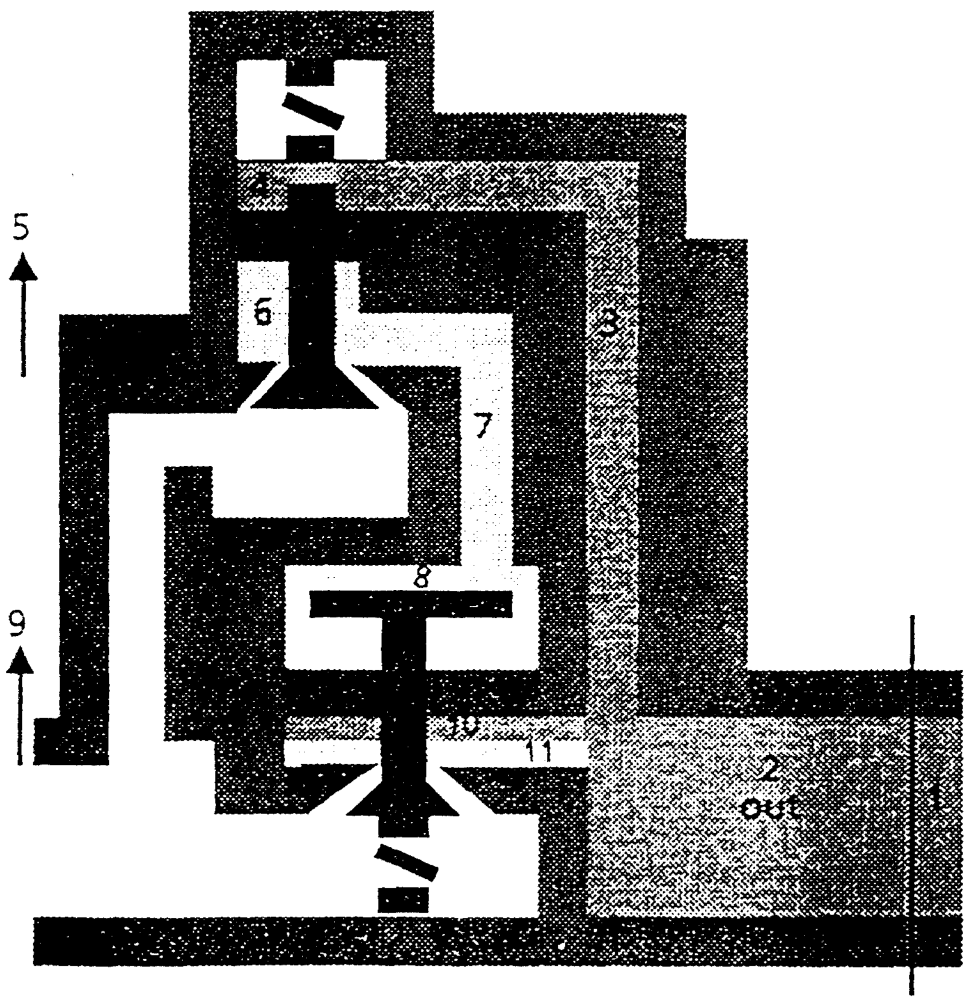


Figure 2. Qualitative physics can be used in intelligent tutoring systems

numerical simulation to ascertain, for example, that a pendulum is not a good oscillator to use in a wristwatch.

3 *The Past*

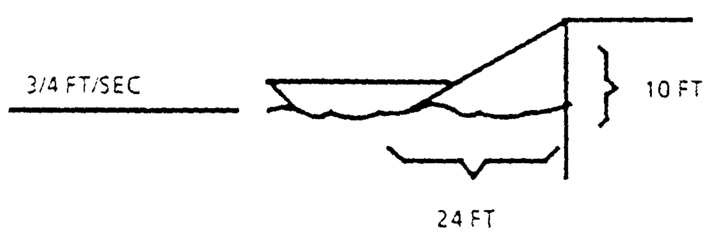
We will not attempt a complete historical survey or time line of qualitative physics. Instead, we will describe three early efforts, the "pre-history" of the area, that provide a background for making later work easier to understand.

Qualitative physics arose from attempts to build programs that could solve textbook physics and math problems. The earliest systems (STUDENT [Bobrow, 1968], CARPS [Charniak, 1968], MECHO [Bundy et al., 1979], ISSAC [Novak, 1976]) attempted to capture the full breadth of the problem, from parsing the initial problem description in natural language to generating diagrams. These programs could solve a variety of problems, but it was quickly discovered that the equations (explicit or implicit) were insufficient to solve most problems. Consider Figure 3 from the description of Charniak's CARPS program. To set up the equations properly required interpreting the phrase "approaching the dock," which here means the distance along the top of the water.

The easy answer, of course, is that more knowledge is needed. But what kind? de Kleer was the first person to characterize the relevant kind of knowledge. His work on the NEWTON program marked the beginning of qualitative physics. NEWTON was designed to solve problems concerning a single point mass sliding on a surface (see Figure 3).

A BARGE WHOSE DECK IS 10 FT BELOW THE LEVEL OF A DOCK IS BEING DRAWN IN BY MEANS OF A CABLE ATTACHED TO THE DECK AND PASSING THROUGH A RING ON THE DOCK. WHEN THE BARGE IS 24 FT FROM AND APPROACHING THE DOCK AT 3/4 FT/SEC HOW FAST IS THE CABLE BEING PULLED IN?

Make a sketch of this situation for yourself. Most all people will draw



Clearly when we say APPROACHING THE DOCK we mean at the level of the boat. Once again information of gravity would lead to this result.

Figure 3 Commonsense knowledge is needed to solve textbook problems. In extending STUDENT's techniques to handle calculus problems, Charniak found that more world knowledge was needed to properly interpret these problems.

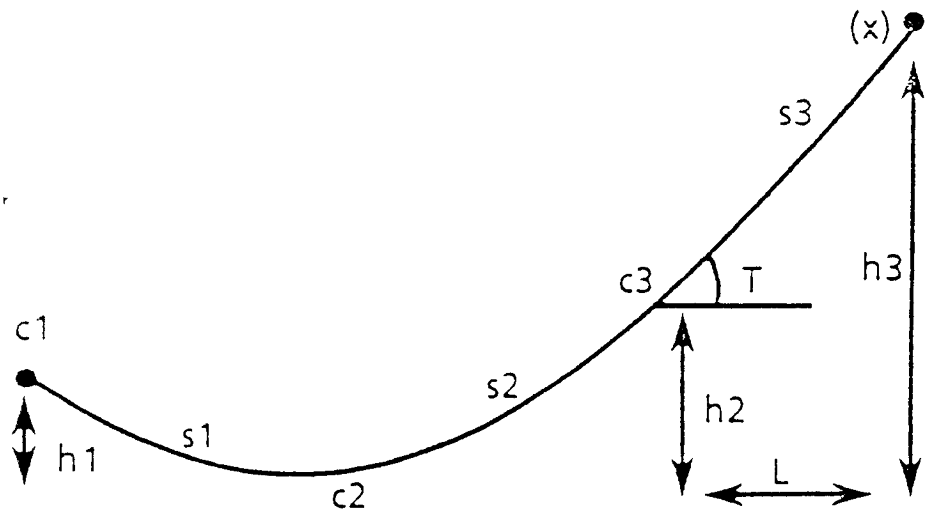


Figure 4 An example from NEWTON. de Kleer's NEWTON used a combination of qualitative and algebraic techniques to reason about a point mass moving on a surface.

When faced with a problem, NEWTON would begin by creating an *envisionment*, an explicit representation of all the different possible behaviors of the system. Figure 5 shows the envisionment for the problem in Figure 3. There are two things to note about this envisionment. First, in standard simulations there is a unique next state. In a qualitative simulation there can be more than one next state, due to the lack of resolution in the qualitative description. Second, the envisionment alone suffices to answer many questions about this domain. For example, if asked whether or not the mass could fly off segment S1 going to the right, NEWTON could answer "no," because no description matching that behavior can be found in the envisionment. To paraphrase de Kleer, an intelligent problem solver has to be able to answer stupid questions, and preferably with less work than it takes to answer subtle questions.

To answer more subtle questions, NEWTON performed algebraic manipulation. Consider the problem of determining conditions that will prevent the cart from flying off when it enters the right side of the track. There is a qualitative ambiguity in what happens after state S1, one branch corresponding to the cart flying off and the other branch to the cart sliding back. NEWTON used this qualitative ambiguity to index into a knowledge base of equations, which was then manipulated to derive an appropriate inequality.

The next event in the prehistory of qualitative physics was the Pat Hayes' Naive Physics Manifesto [Hayes, 1985]. This paper achieved wide informal circulation in 1978, and had a major impact. In particular, Hayes' notion of *histories* is central to qualitative physics. Figure 6 illustrates a fragment of the history for a liquid being poured from a container onto a table top. The basic idea of histories is that events should be represented as spatially bounded, but temporally extended, pieces of space-time. It is assumed that histories which do not intersect do not interact.

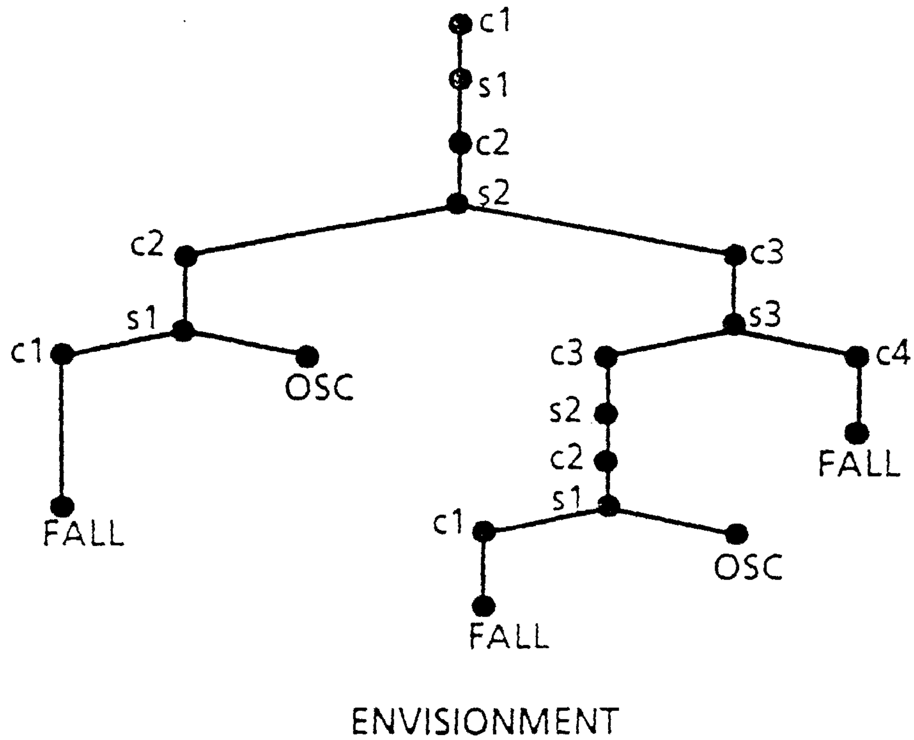


Figure 5. An Envisionment for a NEWTON problem.

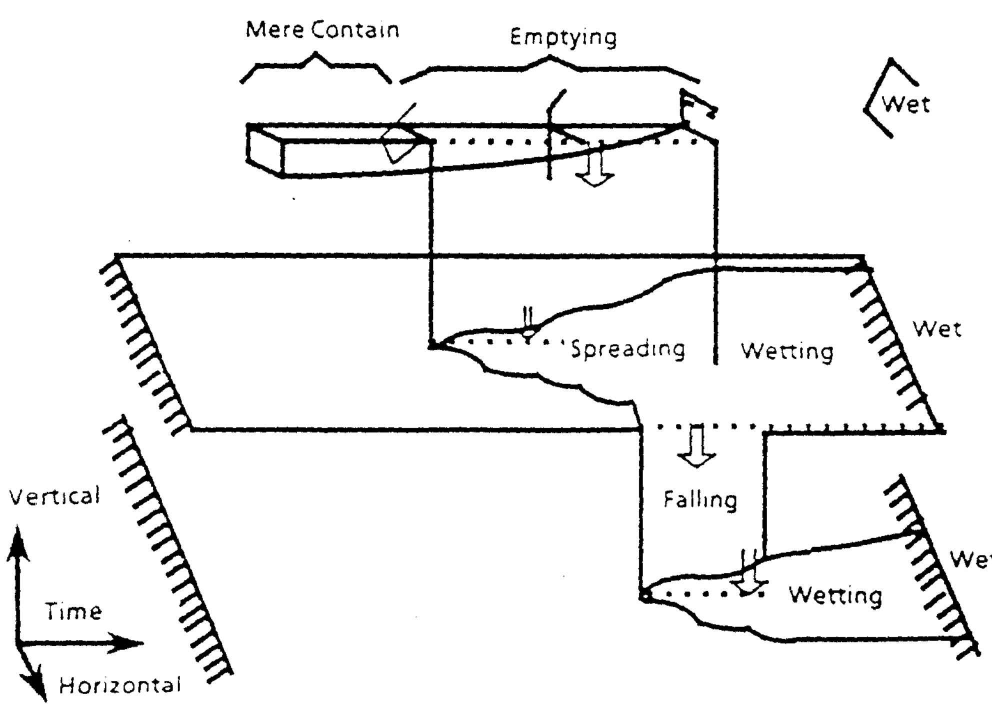


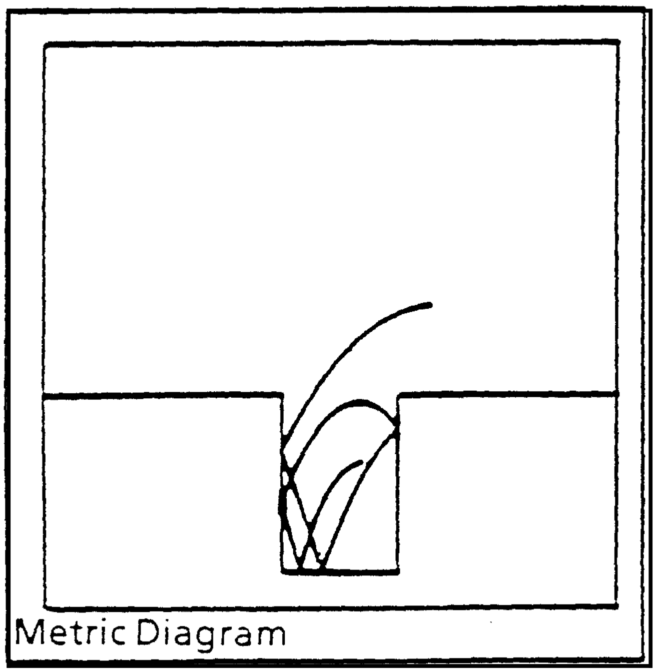
Figure 6 An example of Hayes' notion of histories.

Histories were designed to solve several problems with the situation calculus, especially the Frame Problem. Situation calculus provides no spatial boundaries for an event. In fact, the situation calculus describes what happens *between* events of some kind (such as the actions taken by an imaginary robot), not what happens *during* those events. This leads to several well-known problems, such as being forced to change situations whenever anything happens anywhere in the entire universe of discourse. There are two advantages to histories. Their being temporally extended means it is easier to talk about what is happening during some action (assuming appropriate temporal representations). Their being spatially bounded means that descriptions can be evolved locally, thus eliminating the requirement of global simulation (see [Hayes, 1979; Forbus, 1984b; Williams, 1986] for details).

While several aspects of Hayes' naive physics enterprise have been adopted enthusiastically in the qualitative physics enterprise, several have not. For instance, Hayes argued that implementation was an "unnecessary distraction." In qualitative physics, testing ideas via computer implementation is viewed as essential. As our models grow more complex, carrying out proofs by hand is burdensome. With abstruse mathematical constructs it is easy to maintain rigor, but with commonsense matters it is all too tempting to relax one's vigilance. Carefully written programs are superb bookkeepers, keeping one's theories honest. Furthermore, as discussed below, there are several *styles of reasoning* that use such knowledge. Identifying these problems and developing computational techniques to solve them is a worthwhile endeavor in its own right.

The third piece of prehistory is my FROB program [Forbus, 1980, 1981a] which reasoned about motion through free space. de Kleer's "roller-coaster" world was essentially one-dimensional, with the simulation halting whenever the cart left the surface. FROB worked with a true two-dimensional world, reasoning about balls bouncing around on surfaces (see Figure 7). The user could specify a scenario by drawing a diagram to specify the surfaces and introduce balls. The more information the user provides, the more FROB refines its descriptions. For example, FROB used a constraint language to determine, in conjunction with the diagram, the consequences of any numerical parameters provided. In addition to carrying out numerical analyses, FROB could answer questions like "where will this ball end up eventually?" and "can these two balls collide?" In all cases, FROB used minimal information to answer the question.

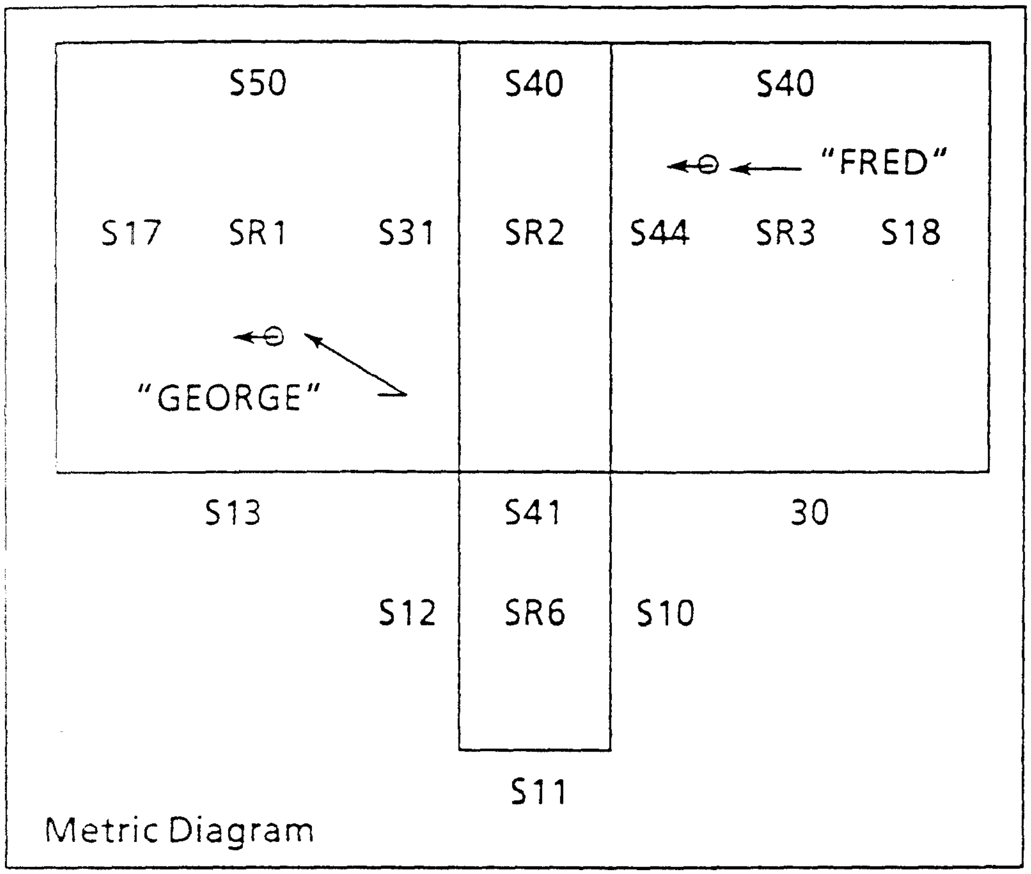
FROB's spatial reasoning worked by calculating a qualitative vocabulary of *places* from the surfaces in the diagram. Combined with symbolic descriptions of activity (such as `FLY` and `COLLIDE`) and velocity (e.g., `(LEFT UP)`), these places provided the framework for qualitative spatial analysis. Consider the problem of determining whether or not the two balls in Figure 8 will collide. To collide, two balls must be in the same place at the same time. If all we know is that both balls are going to the left, then they might collide, since the union of the places they might be overlap. But if we also assume that FRED never gets to S31, then a collision is ruled out, since the two balls can never be in the same place.




```
->> 'Motion-Summary-for b1)

FOR G0364
THE BALL WILL EVENTUALLY STOP
IT IS TRAPPED INSIDE (WELL0)
AND WILL STOP FLYING AT ONE OF (SEGMENT 11)
NIL
```

Figure 7 FROB reasoned about motion through space.




```
->>(collide? fred george)
(POSSIBLE AT SEGMENT 50 SEGEMNT 17 SEGMENT 13 SREGION)
->>(cannot-be-at fred segment 31)
(SEGMENT 31)
UPDATING ASSUMPTIONS FOR (>> INITIAL-STATE FRED)
CHECKING PATH OF MOTION AGAINST ASSUMPTIONS
->(collide? fred george)
NO
->>(what-is (>>state initial-state fred)
(>>STATE INITIAL-STATE FRED) = (FLY (SREGION3) (LEFT))
NIL
->>(what-is (>>state initial-state george))
(>>STATE INITIAL-STATE GEORGE) = (FLY (SREGION) (LEFT))
NIL
```

Figure 8 Collision problem.

FROB advanced the state of the art in several ways. First, it demonstrated that Hayes' notion of histories was indeed useful. There was perhaps more numerical information in FROB's histories than in Hayes' original conception, but they are histories nonetheless. Second, FROB was based on a theory of spatial reasoning that divided the problem into two parts, using a diagrammatic representation to provide quick answers to a class of geometric questions, and a qualitative description of places computed from the diagram. Third, it demonstrated that qualitative ambiguities could be resolved by numerical calculation, just as NEWTON demonstrated that symbolic algebra could resolve them. And finally, the notion of envisionments was generalized from the trees used in NEWTON to full graphs. This allows many properties of the behavior, such as final states and oscillations, to be characterized by properties of the envisionment graph (e.g., end states and cycles) rather than by explicit nodes as in NEWTON.

At this point we draw our pre-historic retrospective to a close. NEWTON and FROB were organized around using a combination of qualitative and quantitative techniques to solve particular classes of problems. It became clear around this time that simply understanding the nature of qualitative representation was a full-time effort, and that a domain-independent, general qualitative physics could exist. Research effort turned to finding such a physics—or, more correctly, understanding the space of such systems of physics—and we now turn to this exploration.

4 The State of the Art

Work in qualitative physics may be roughly divided into three areas: *qualitative dynamics*, *qualitative kinematics*, and *styles of reasoning*. In traditional physics,

> Dynamics deals with the causes of motion, as opposed to kinematics, which deals with its geometric description, and to statics, which deals with the conditions for the lack of motion [Considine, 1983].

Dynamics is used generically to describe the study of forces on systems (e.g., fluid dynamics), and typically includes statics. Hence qualitative dynamics is concerned with what causes systems to change over time, ignoring geometry except as a source of boundary conditions.

Qualitative kinematics is concerned with the spatial reasoning required by commonsense physics. Not all commonsense spatial reasoning is qualitative kinematics—counterexamples include navigation, spatial planning, and controlling arm motions. Carrying the distinction between dynamics and kinematics into qualitative physics is not an arbitrary choice, as we will argue in Section 4.2.

Styles of reasoning, of course, concern how to exploit the knowledge of qualitative physics. There is no direct analog in traditional physics, except insofar as physicists and educators have attempted to formalize their problem-solving methods in order to teach them more readily. But studying styles of reasoning is crucial for qualitative physics, since representation without reasoning is an idle exercise.

4.1 *Qualitative Dynamics*

Qualitative dynamics studies how physical systems change. It addresses the problem of how to represent differential equations qualitatively, and how to organize such knowledge in a usable form. We begin by surveying qualitative representations for numbers and time-varying differential equations. Ontological issues are discussed next, since providing a formalism for organizing knowledge is a central job of qualitative physics. Finally we take a brief look at two other issues, the role of continuity and how such equations are given causal interpretations, since these topics are often misunderstood.

But before we start: A variety of notations have been used in qualitative physics. While terminology differences can be bewildering to the uninitiated, and standardization has been suggested ([Bobrow, 1984], p. 5), it is doubtful that the situation will improve soon. In fact, two facts suggest that standardization is not an urgent issue. First, there is already significant overlap. Second, the lack of a single standardized notation has not seemed to retard progress in traditional mathematics, in which there are still over six different notations for derivatives, despite its being hundreds of years older than qualitative physics. We will sometimes point out variations, but will not attempt a complete concordance.

4.1.1 *Numbers* Three representations for number have proven useful so far in qualitative physics: *signs*, *inequalities*, and *orders of magnitude*. We describe each in turn.

Signs Reducing numbers to signs is the simplest qualitative representation for number [de Kleer, 1979b, 1984b; Williams, 1984]. For example, we might say that the level of water in a container is -1, 0, or 1, depending on whether or not the level is lower, the same as, or higher than a desired height. If the comparison is chosen carefully, we can satisfy our desiderata of capturing relevant distinctions while not introducing irrelevant ones.

Signs of derivatives form a natural indicator of change [Forbus, 1981b; de Kleer, 1984b; Williams, 1984]. We will use the notation of qualitative process (QP) theory and denote the sign of the derivative of a quantity Q by Ds[Q]. If the sign of the derivative is −1, then the quantity is decreasing, if 0 then it is constant, and if 1 then it is increasing. Since change is intuitively important, and the direction of change determines what boundary conditions might change, signs carry critical information about derivatives.

The earliest use of signs in qualitative physics was de Kleer's QUAL program [de Kleer, 1979a], where signs were interpreted as the difference between an original equilibrium value and the new equilibrium value reached as the result of a perturbation (the *incremental qualitative value* (IQ) interpretation). The semantics of this representation were slightly problematic: For example, it was not clear what the IQ value should be if the system went through several behavioral states before settling into an equilibrium value.

The major advantage of the sign representation is simplicity. We are taught the method of substitution very early in mathematics, and sign values provide a concrete object that may be "plugged in" to qualitative equations of whatever form. However, signs alone are often not enough. Consider the problem of figuring out what might happen if we have three tanks F, G, and H with pipes hooked up between them. Given some initial level of water in each, we turn on all the valves in the pipes between them. To determine how the water would flow requires comparing the pressures in the tanks that are linked together.

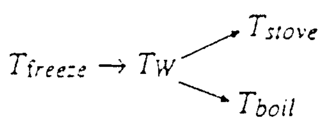
A sign value encodes a comparison of a magnitude with a single reference value. Suppose tank G is connected by pipes to both F and H. Clearly no sign representation of pressure will suffice for the pressure in G, since we must compare the pressure with two reference values, the pressures in F and G. The fact that these reference values are themselves changing is yet another complication. It seems counterintuitive to say that the value of pressure in G is changing simply because the pressure in F is changing.

One representational "trick" sometimes suggested to work around these problems, albeit unnaturally, is to rewrite a quantity as a constellation of signed quantities. For example, a given quantity Q might be represented by new quantities $Q_1 \ldots Q_n$, one for each comparison Q is involved in. This does violence to the notion of quantity. Furthermore, it makes the number of pseudo-quantities needed to describe a quantity vary with the situation, rather than with the type of object. The next section describes a more natural representation for such circumstances.

Inequalities Comparing the value of a quantity with several other parameters is a common occurrence in physics. For example, to determine the phase of a piece of stuff, one determines the relationship of its temperature to the boiling temperature and freezing temperature of that substance for the appropriate conditions (such as pressure). Worse yet, the parameters that it makes sense to compare a value with can change as conditions change. For example, if we discover a leak in tank G in the previous example, we should also consider the relationship between the pressure at the leak and the surroundings.

These considerations suggest collecting a set of inequalities to describe a quantity. This set of inequalities is called its *quantity space* [Forbus, 1981b]. Inequalities makes sense for several reasons. First, they provide a means to partition numerical values, and thus express boundary conditions for behavior. For example, when two objects in thermal contact are at different temperatures, there will be a heat flow from the object with higher temperature to the object with lower temperature. Second, a quantity can participate in any number of inequalities, thus providing the variable resolution we desire. Third, if numbers are combined by addition, inequality information often suffices to determine the sign of the outcome. If, for instance, there is flow into a tank and flow out, the relative magnitudes of the flows determine whether the level of the tank is rising or falling.

Here is a simple quantity space that describes the temperature of water *W* in a pot on the stove.



A simple quantity space. The significant relationships involving the temperature of a piece of water (T_W) can be expressed as inequalities. Here, the temperature is above freezing (T_{freeze}) and less than the temperature of the stove and its boiling temperature.

The arrows represent inequalities, with the quantity at the head of the arrow being greater than the quantity at the tail of the arrow. Thus *W* is warmer than freezing, and cooler than both its boiling temperature and the temperature of the stove. Importantly, quantity spaces need not be complete—notice that in this diagram we do not know the relationship between the temperature of the stove and the boiling point of *W*. The ability to represent this ambiguity allows us to accumulate partial information, and detect when more information is required.

What should a number be compared to? One source of quantity space elements are parameters representing domain-specific boundary conditions. An example of such *limit points* are the boiling temperature of a substance or the fracture stress of a material [Forbus, 1981b]. Some comparisons are required due to the specifics of a situation, such as a comparison between the rate of flow into and out of a container. We will adopt the terminology of [Kuipers, 1986] and refer to the elements of a quantity space generically as *landmark values* for the quantity, whether or not they are limit points.

Landmarks versus limit points Two distinct semantics have been used for landmark values in the literature. The distinction has often been misunderstood, via a type/token confusion, and we undertake to clarify it here. We call a description *temporally generic* if it refers to a class of temporal behaviors, rather than just a single behavior. A description of a single behavior we will call *temporally specific*. The script of a play is a temporally generic description, while a videotape of its performance is temporally specific. Limit points are temporally generic, as are comparisons between rates, since there are classes of situations where liquids boil and flows occur. The value of the boiling temperature at 3 PM is temporally specific—we are referring to a single situation, and hence a single specific value.

Most systems of qualitative physics use only temporally generic landmarks. But temporally specific landmarks can be critical for many reasoning tasks: For example, it may be crucial for a doctor to compare a patient's cholesterol level today with the specific cholesterol level last week, not just with some generic "safe" value. Kuipers' QSIM generates such temporally specific landmarks. These landmarks do not correspond to "discovering" new limit points, as originally claimed. Rather, they are the equivalent of a qualitative "strip chart" that describes a specific behavior of a system. QSIM thus provides an automatic naming facility to support reasoning about temporally specific values.

Although temporally specific landmarks are essential for some inferences, they introduce a new level of computational complexity. Consider for example a decaying oscillation, such as a ball bouncing up and down, each time rising only some fraction of the height it reached before. Each height is a new landmark value. Thus an infinite behavior can sometimes lead to an infinite number of landmark values (see Section 4.3.2).

The quantity space is now a standard feature of qualitative physics [Kuipers, 1984, 1986; Simmons, 1983; Weld, 1986]. It addresses the resolution problem by providing the ability to incrementally accumulate information about a number, thus simplifying the modeling task. However, manipulating sets of statements describing a value is more complicated than treating values as atomic objects, as the sign representation allows. Quantity space implementations require efficient application of the laws of transitivity, typically obtained by separate inferential mechanisms [Forbus, 1984c; Simmons, 1983; Forbus, 1988].

Several useful variations of the quantity space have been developed. For instance, Kuipers requires quantity spaces to be totally ordered [Kuipers, 1984], which simplifies the representation into a collection of intervals. Simmons [1986] augments inequalities with numerical intervals, thus providing a simple way to integrate empirical bounds.

Orders of magnitude Sometimes saying that N_1 is greater than N_2 is not enough: One may need to say that N_1 is so large compared to N_2 that N_2 may be ignored. For instance, the effect of evaporation on the level of a lake may be ignored if the dam holding it has burst. In everyday life, engineers rely on the ability to distinguish a value that is significantly out of range from a normal variation. One way to represent such information is to extend the range of comparative relationships to include *orders of magnitude*. Three such representations, FOG [Raiman, 1986], O[M] [Mavrovouniotis and Stephanopolous, 1987], and Davis' infinitesimal theory [Davis, 1987] have been developed in qualitative physics. We begin with FOG and O[M] since they share intended use, and then describe Davis' system.

FOG introduces three new relationships, in addition to the traditional order relations. They are:

$A << B$: A is negligible compared to B.
$A \equiv B$: A is very close to B.
$A \sim B$: A is the same order of magnitude as B.

Raiman has developed a consistent formalization that captures the intuitive meaning of these statements, using infinitesimals as a model. The effect of these relationships is to stratify values into equivalence classes, thus providing the means to say that values are very different. For example, in the DEDALE diagnosis system [Dauge et al., 1987], this vocabulary is used to describe the typical relationships between values in component models.

The O[M] is based on assigning labels to ranges of ratios. For example, the relationship

$A \sim< B$ (read A is slightly smaller than B)

is true exactly when

$$\frac{|A|}{|B|} < (1 + e)$$

where e is a domain-specific parameter. This mapping simplifies the laws of the system and potentially allows a variety of quantitative information to be easily incorporated. O[M] also uses physical units to reduce inferential complexity: only parameters of the same units may be compared.

The definition of orders-of-magnitude relations in O[M] in terms of ranges simplifies the mapping from numerical values, a problem for which FOG provides little guidance. However it also allows a large but finite number of negligible values to add up to something that is significant, which violates the intuitions underlying such reasoning. This cannot happen in FOG. The relative advantages of the two systems remain to be explored.

Davis [1987] describes another formalism for orders-of-magnitude which, like FOG, is based on infinitesimals. He reconstructs a qualitative calculus to include infinitesimal values for both numbers and as durations of intervals. Thus he can talk about changes taking infinite (or very short) time.

4.1.2 *Equations* Equations are the hallmark of physics. Just as qualitative physics restricts the accuracy to which numerical values are known, the notions of equations developed in qualitative physics are also typically weaker. These weaker constraints can better capture partial knowledge and simplify inference, thus addressing the resolution problem.

Arithmetic operations Every system of qualitative physics includes at least addition and subtraction. Multiplication is often introduced as well. While the operations are familiar, the effects of weakening the values they are performed on has profound consequences. First, ambiguities can arise, even with complete initial information. If one only knows that A is greater than zero and B is less than zero, for instance, then the sign of $A + B$ cannot be determined. In this case knowing the relative magnitudes of A and B can provide the answer, but in general, algebraic inequalities are required. But since most qualitative values do not form a field, algebraic manipulations must be performed with care.

In [de Kleer and Brown, 1984], equations involving sign values are called *confluences.* Confluences are solved by propagation of constraints, using generate and test when unresolvable simultaneities occur. Under certain conditions, Dormoy has shown that sets of confluences can be solved by a variant of Gaussian elimination [Dormoy and Raimen, 1987]. Confluences have also been used with the FOG formalism, where the comparison is made between the actual value of a parameter and its nominal value [Dauge et al., 1987].

Monotonic functions One of the weakest statements that can be made about the relationship between two quantities is that when one increases, the other tends to increase. This level of knowledge is captured by *monotonic functions*, which are used as a primitive in several systems of qualitative physics and mathematics. Monotonic functions provide a means of approximating complicated or unknown functions with minimal commitment.

If $y = f(x)$ then $f(x)$ is *increasing monotonic* if whenever x increases, y increases. $f(x)$ is *decreasing monotonic* if whenever x increases, y decreases. Often there is no reason to name the function involved, so various notations for anonymous functions have been developed. For example, Kuipers [1984, 1986] uses $M^+(x, y)$ to denote an increasing monotonic connection between x and y, and $M^-(x, y)$ to denote a decreasing function.

QP theory allows the partial specification of monotonic functions through *qualitative proportionalities.* Formally, $y \propto_{Q+} x$ indicates $y = f(\ldots, x, \ldots)$, where f is some function which is increasing monotonic in its dependence on x. Similarly, $y \; \alpha_{Q+} \; x$ indicates that the function involved is decreasing monotonic in x. To determine the complete specification of functional dependence in any particular situation requires a closed-world assumption.[3]

The advantage of qualitative proportionalities is composability; the knowledge of a function can be decomposed and distributed appropriately through a representation, to be assembled as needed by the reasoning system. For example, parameters may be selectively ignored (such as the effect of pipe resistance on the rate of liquid flow, if the fluid is moving very slowly) by "turning off" the description that contributes them to the function. Qualitative proportionalities can also be used to express intermediate hypotheses in a learning system. For example, ABACUS [Falkenhainer, 1985] searches for them as the first step in finding equations to describe numerical data. The disadvantage is that ambiguities arising from them cannot be settled by just inequality information. Consider for instance

$$C \; \alpha_{Q-} \; A \;\wedge\; C \; \alpha_{Q-} \; B \;\wedge\; \mathrm{Ds}[A] = \mathrm{Ds}[B] = 1$$

No additional sign or inequality information suffices to determine Ds[C], unlike subtraction or multiplication.

We have found it useful to allow two other kinds of information to be specified about monotonic functions. First, *correspondences* are introduced to propagate inequality information. Intuitively, a correspondence fixes a point on the curve relating two (or more) parameters. For instance, when a spring is at its rest length it exerts no force. Suppose the force is α_{Q-} its length (i.e., stretching it produces a force that tends to make it return to its rest length). These two facts together allow us to deduce that if we push a spring to be shorter than its rest length, we will cause it to exert a positive force (i.e., push against us). A detailed discussion of correspondences can be found in [Forbus, 1984b; Kuipers, 1986]. Second, functions can be named, so that inequality information can be propagated across distinct individuals [Forbus, 1984b]. For example, the function that determines the pressure of a contained liquid in terms of its level is the same for all containers, and hence information about differences in level can be mapped into differences in pressure.

Of course, many functions required in modeling the physical world are not monotonic. Such functions can be represented by decomposing them into monotonic segments. Providing a framework for explicitly describing the assumptions underlying this decomposition is one of the roles played by ontology in qualitative physics.

3 A language for framing more complete hypotheses about functional dependence is described in [Forbus, 1984b], Section 5.3.

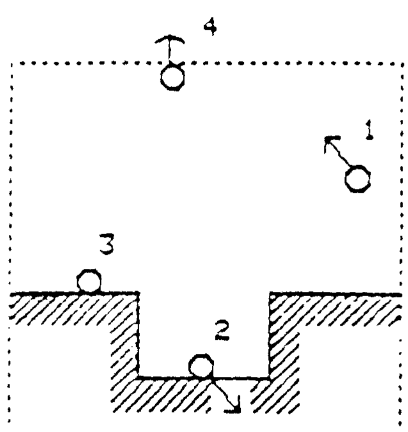
4.1.3 *Ontology* Ontological choices are central to qualitative physics. Along with space and time, ontology provides the organizational structure for everything else. Continuous properties are properties of something, and equations hold as a result of that. Usually developing the appropriate ontology is the most difficult part of formalizing a domain.

If we are to build a complete qualitative physics, one that covers the breadth and depth of our commonsense knowledge of the physical world, we must discover and utilize common abstractions. Generating an ad hoc model for each scenario is impractical and unreliable. Two such ontological abstractions, *devices* and *processes*, have been widely used in qualitative physics. We describe them here, after briefly reviewing a simple precursor.

4.1.4 *Qualitative State Vectors* The *qualitative state vector* ontology was the earliest used in qualitative physics. It was the ontology used in both NEWTON [de Kleer, 1975, 1979a], and FROB [Forbus, 1980, 1981a]. The idea is to decompose system behavior into segments, each described by a list of symbols. This symbolic state vector contains two types of elements:

1. A quantization of the traditional state variables.
2. A symbolic description of the type of activity.

In traditional physics, we might state informally what kind of system we are reasoning about (say, a ball bouncing on a surface), describe the initial values for the state parameters, and state what equations will be used to describe the different things a ball can do (i.e., fly through space and collide with surfaces). In the corresponding qualitative description, we would quantize position into symbolic places, velocities into symbolic directions, and add a symbol for the type of behavior. For example, we might say a ball is in `REGION0`, going `(LEFT UP)`, and `FLY`ing (see Figure 9).




```
1. (FLY REGION3 (LEFT UP))
2. (COLLIDE S11 (RIGHT DOWN))
3. (STOP S13 NIL)
4. (CONTINUE S49 (UP))
```

Figure 9 An example of qualitative state vectors.

The need for the first class of constituent is obvious, since some representation of state variables is needed to capture the behavior. The second type explicitly describes that which is left implicit in the traditional representations. Roughly, the symbolic description of activity should change whenever the quantitative equations traditionally used to describe the behavior will change. Since we do not have equations, we must provide instead a set of *qualitative simulation rules*. These rules take a state and produce the set of states which can occur next. As mentioned previously, more than one state may be possible due to the coarse grain of the representation. The particular content of the rules is highly domain-specific, but typically a small set of rules suffices for each class of behavior. (Hayes' conception of reasoning with histories by "gluing them together" fits within this framework as well.)

The qualitative state vector representation has three useful properties. First, it is quite natural. The notion of state is central in any account of physics, traditional or qualitative. Second, it is very compact. Each state can be succinctly described by a short list of symbols, and hence envisioning is very cheap. Third, it provides an easy means to combine dynamic and kinematic representations, something which is more difficult with the other ontologies.

The difficulty with this ontology is that it lacks composability. To describe a complex system directly is often too difficult. Instead, one decomposes it into smaller parts, models each of those parts and the relationships between them, and then combines these models into a model of the whole system. The advantages of such modular approaches are well known; the pieces can often be reused to describe yet more systems. But we have placed little constraint on the actual contents of states and simulation laws, and so we have no methodology for combining them.

For example, suppose we wish to combine the states in NEWTON and FROB. Each simulation stops when it reaches conditions that make the other appropriate, so one might imagine using the union of their simulation laws to more fully describe the behavior of a point mass. But not all combinations are so simple. If we glue the point mass onto a stick that is attached to a pivot (thus creating a pendulum), both sets of laws are simply wrong. Each new condition we add requires reorganizing our vectors and simulation laws in some ad hoc fashion.

Hayes' axioms for liquids do not escape this problem, either. First, Hayes himself points out there are many cases where his theory cannot make predictions (such as pouring water into a leaky cup). Second, adding new phenomena, such as solutions, would require wholesale reorganization of the theory. No theory is completely composable, of course. What we seek is an organizing principle, a methodology that simplifies combination as much as possible. Patterns of history combinations (or, equivalently, tables of qualitative simulation laws) are not constrained enough.

In traditional physics, composability is arranged by sharing parameters. The equations for distinct parts are combined by identity of names in some cases, and by new equations describing the relationship between the parts in others. Qualitative versions of such theories thus require both a qualitative representation of equations, and an organizing structure to place them in. This generative power is exactly what is required to provide composability. The other two ontologies exploit this idea.

4.1.5 The Device Ontology System dynamics [Shearer et al., 1971] is an engineering methodology which provides a common set of abstractions that encompass a variety of domains, including many electrical, thermal, mechanical, and acoustical systems. This modeling paradigm has been widely used in qualitative physics as well, the principle advocates being de Kleer and Brown [de Kleer, 1979b; de Kleer and Brown, 1984; de Kleer, 1984a] and Williams [1984]. These theories replace the quantitative equations of system dynamics with qualitative equations, and have developed new inference techniques for using these descriptions.

The basic idea is to view a system as constructed from a collection of *devices*, such as transistors and resistors. The behavior of a device is specified by internal laws, often decomposed into distinct states or operating regions. Each device has some number of *ports*, and all interaction between devices occurs through these ports. To model a particular system, one builds a network of devices. The device network is then analyzed by using the combined equations from the devices and interconnections, either by constraint propagation or symbolic relaxation.

Consider, for example, the bipolar transistor common emitter amplifier in Figure 10. The catalog of domain devices will include descriptions of transistors and resistors, and descriptions of what parameters are shared when terminals are connected together. A typical conclusion (but not the only kind) that can be reached with this description is how the circuit might respond to a change in input. This reasoning is accomplished by "perturbing" a declared input parameter, and using the laws associated with devices and interconnections to propagate effects through the system. For instance, suppose the input voltage increased. This will cause the base-emitter current to increase, which (due to the way transistors work) will cause the collector-emitter current to increase. This in turn will cause the collector voltage to drop, which will in turn cause the output voltage to go down.

This example has been deliberately simplified; detailed descriptions can easily be found in the literature (see [de Kleer and Brown, 1984; Williams, 1984]). However, it illustrates two important properties of this ontology. First, once a model is created, most inferential work occurs by local propagation within the model. Such antecedent reasoning is easy to control and can be made to work very efficiently. Second, we have assumed that *flow of informa-*

tion in the model of the system directly mirrors *flow of causality* in the world. The ramifications of this assumption are discussed in Section 4.1.7.

One additional complexity that bears mention is that devices can have *states*, corresponding to different modes of a device. For example, a valve may be OPEN, CLOSED, or PARTIALLY-OPEN. Each device state is characterized by a different set of laws (see Figure 11). The state of a device is invariably predicated on the (qualitative) value of a numerical parameter.

The device ontology has three advantages. First, the fixed network topology provides a substrate for efficient computations. All references within laws are strictly local, and hence resolving them is straightforward. This simplifies implementation. Second, composability is maintained by having all information transferred through local connections. Given a correct catalog of device models and interconnections, one could in principle model an arbitrarily complex system by connecting together the corresponding device models.

The third advantage is that system dynamics is a widely used traditional engineering methodology. Consequently, there are generally accepted standards for structural descriptions (i.e., schematics) and standard quantitative models for many domains which can be used as a starting point for creating qualitative models. The translation of such quantitative to qualitative models is not trivial, since new device states may have to be introduced (see [de Kleer and Brown, 1984] for details). However, most of the ontology can be inherited from system dynamics intact, thus simplifying the modeler's task and providing greater confidence in the result.

However, there are two serious disadvantages to this ontology. First, the device ontology provides no guidance for the construction of the network model itself. This is not a problem in some domains, such as electronics, where the mapping from objects and relationships in the world is straightforward. In manufacturing electronic components, great care is taken to ensure that the physical objects perform much like their idealizations, within reasonable limits. But for most domains this aspect of the modeling process is problematic.

Consider, for example, the block shown in Figure 12(a). If the block is sitting on a table and we push it, then we probably want to model it as an idealized mass. But if we push it while it is resting against a wall, then we will probably want to model it as an idealized spring (albeit very stiff). If we immerse the block in water and push on it, then we will probably model it as an idealized damper. Thus we see that the same physical object can be modeled by three distinct abstract devices, depending on the conditions in the system.

The advice given in system dynamics texts is to figure out how the object behaves, and then select the right device model. This advice is fine for human engineers, since their goal is to produce quantitative analyses and they presumably already have some idea of the system's qualitative behavior. But the goal of qualitative physics is to produce precisely those qualitative descriptions of behavior, and hence we are left in the position of needing the answer before

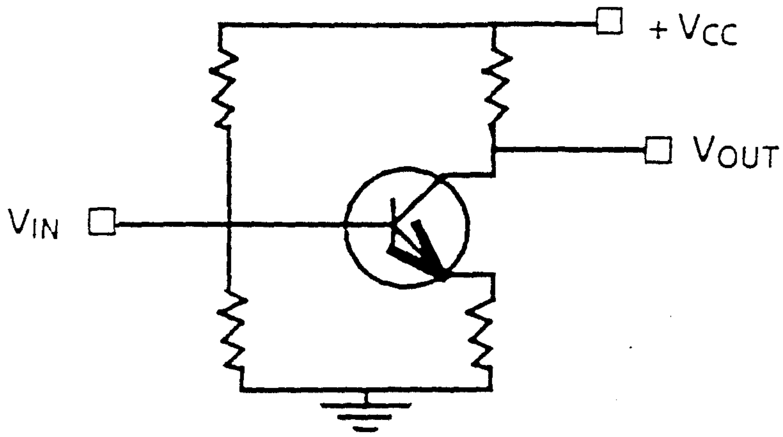


Figure 10 An example of the device ontology.

State	Condition		
OPEN:	$[A = A_{max}]$	$[P] = 0$	$\partial P = 0$
PARTIALLY-OPEN:	$[0 < A < A_{max}]$	$[P] = [Q]$	
CLOSED:	$[A = 0]$	$[Q] = 0$	$\partial Q = 0$

Figure 11 A device model for a valve. This simple model of a valve is drawn from *Confluences*. *A* refers to the area of the valve, relative to some maximum area A_{max}. *P* refers to the pressure across the valve, while *Q* refers to the flow rate of gas through the valve.

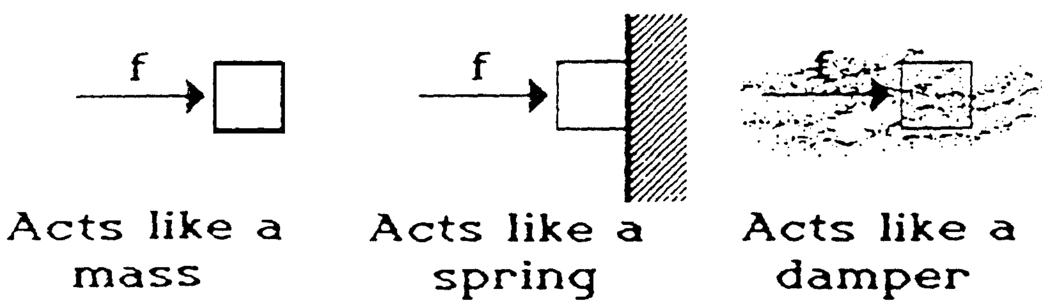


Figure 12 System dynamics doesn't capture modeling assumptions

we can compute it. Consequently, the standard device ontology fails to completely address the modeling problem, since it does not formalize the critical task of model creation.

The second disadvantage is that, in many cases, the device ontology is unnatural. Consider the situations in Figure 13. We can consider the water in the pot on the stove (Figure 13(a)) to be an object. If the water boils, this object will decrease in size until it vanishes. It is hard to think of this system as a collection of devices, since the reasoning requires "clipping" a device out of the network when the water vanishes. Such changes in the network topology lie outside the device formalism. Similarly, the bouncing ball in Figure 13(b) illustrates that what an object interacts with can change drastically. It is difficult to see any elegant representation for this system in the device ontology.

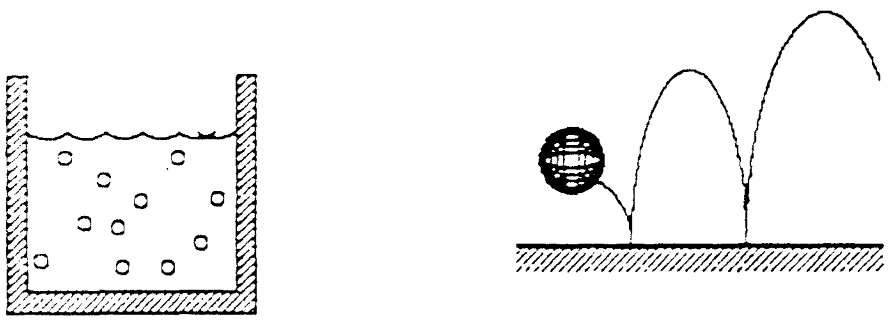


Figure 13 System dynamics cannot model many interesting systems.

4.1.6 *Processes* Informally, people often describe changes in the physical world in terms of *processes*. Examples include motion, liquid flow, heat flow, boiling, bending, compressing, and expanding. This notion has been formalized in qualitative physics as an ontological commitment. Consider a cup under a faucet. If the faucet is turned on, there will be a *process* of liquid flow occurring from the faucet, through the fluid path formed by the space above the cup, to the cup itself. This liquid flow is not a property of either the cup, the faucet, the water, or the space above the cup. It is a new type of entity, with properties of its own, such as the rate of water flow.

In this ontology, processes like liquid flow provide the notion of mechanism for physical situations. All changes, ultimately, are assumed to be caused directly or indirectly by physical processes. A model of a domain includes a description of the kinds of objects there are, the kinds of relationships that hold between them, and the kinds of processes that can occur. To describe a specific situation, models for each of the parts and relationships are asserted. Importantly, the modeler does not directly specify what processes are possible in each situation. Instead, the process specifications in the domain model state the conditions under which they can occur, and the inference system uses these specifications to automatically generate descriptions of the possible processes.

This notion of process was introduced by *qualitative process* (QP) theory [Forbus, 1981b, 1984b], and has been used in various forms by several researchers in qualitative physics, including Simmons [1983], Weld [1986], Mohammed and Simmons [1986], and Schmolze [1986]. Some of these theories describe the effects of processes continuously over time (such as QP theory), while others describe processes discretely by the net effect they have over an interval of time [Simmons, 1983; Weld, 1986]. (The earliest attempts to formalize physical processes in AI preceded qualitative physics. Hendrix [1973] described processes as STRIPS-like operators augmented with equations for use in planning. Brown, Burton, and Zdybel [1973] represented processes as finite-state automata, for instructional purposes. Neither representation used qualitative information, in the current technical sense of the term.)

Figure 14 illustrates a simple model of liquid flow expressed in QP theory. The *individuals* specification provides a form of quantification. An *instance* of a process is said to exist for every combination of objects in a scenario that matches the individual specifications. The *preconditions* and *quantity conditions* together determine when the process is active. Roughly, quantity conditions can be inequalities and whether or not other processes are active, and preconditions are external conditions. `Aligned`, for example, means that all valves in the path are open. A QP model can predict that pressures will change, but not that a sailor may walk by and close a valve.

The *relations* field describes what holds when the process is active. This field can declare local quantities and constraints, as well as information relevant to external representations (such as appearances). Here, the local quantity `flow-rate` is introduced and is declared to be equal to the difference in pressures. Together with preconditions, the relations field provides a means of interfacing QP theory to other representations.

The direct effects of a process are specified by the *influences* field. Every process must have at least one direct influence, and only processes can have direct influences. Direct influences, noted by *I+* and *I–*, specify the derivative of their first argument. Here, the amount of liquid in the source will tend to decrease, and the amount of liquid in the source will tend to increase. Like qualitative proportionalities, direct influences must be composed to compute the total derivative by making closed-world assumptions. But unlike qualitative proportionalities, where no commitment is made to the method of combination, direct influences are additive. So if we knew that in fact some other process were influencing the amount at the destination (an instance of liquid flow corresponding to a leak, say), then by knowing the relative flow rates we could predict how the amount of water in the destination will actually change. (This solves the problem with Hayes' leaky cup, mentioned earlier.)

The process ontology has several advantages. First, the notion of process is intuitively appealing for many domains. Objects can come into existence and vanish, for example, something that is not allowed in the device ontology. Second, processes provide a simple notion of causality by imposing a distinction between independent variables (those which are directly affected by processes) and dependent variables (those which are affected as a consequence of the independent variables changing). The next section examines this issue in detail.

The third advantage of the process ontology is that it allows explicit representation of modeling conditions and assumptions, via the individuals and preconditions fields. This means the program can take on more of the modeling burden. Instead of demanding a complete initial description, a program using the process ontology can "fill in" the user-supplied description of a particular situation with the kinds of processes that can occur. Potentially, this flexibility provides considerable power. For example, the *class-wide assumptions* described informally in [de Kleer and Brown, 1984] can be formally expressed by combinations of individuals and preconditions specifications in QP theory.

Of course, nothing comes for free—the process ontology also has some disadvantages. First, in some domains (like electronics) the distinction between dependent and independent parameters changes according to the kind of analysis being performed. Process descriptions are very hard to write for such cases. Second, the process ontology requires more inference, and the manipulation of quantified descriptions, to set up the model. This complicates the design of programs using the process ontology, and often results in longer run times. And third, the process ontology has not been formally explored as much as the device ontology. There is no process-oriented equivalent engineering formalism to system dynamics, no off-the-shelf models to adapt.

```
Process    Liquid-Flow(?src ?sub   ?dst ?path)
           Individuals:            ?src a container
                                   ?dst a container
                                   ?sub a substance
                                   ?path a fluid-path,
                                           Connects(?path,?src,?dst)
           Preconditions:          Aligned(?path)
           Quantity Conditions:    A[Pressure(C-S(?sub,liquid,?src))]
                                           >A[Pressure(?dst)]
           Relations:              Quantity(flow-rate)
                                   flow-rate = Pressure(C-S(?sub,lqiuid,?src))
                                           - {ressire)?dst)
           Influences:             I+(Amount-of-in(?sub,liquid,?dst),A[flow-rate])
                                   I-(Amount-of-in(?sub,liquid,?src),A[flow-rate])
```

Figure 14 A description of liquid flow.

4.1.7 *Other Issues* A common misconception is that the different theories described in the literature are merely notational variants for "the" qualitative physics, or that eventually only one theory will be proven to be "right." Such a view ignores the rich variety of the phenomena we are trying to model (from the patchy, incomplete theories constructed on the fly by the person on the street to the integrated, broad theories formulated explicitly by world-class engineers and scientists) and the range of potential applications we are addressing (from student modeling in intelligent tutoring systems to monitoring process plants to scientific discovery).

As the earlier sections indicate, there are a variety of choices for representations of quantity, equation, and ontology. Different combinations of these choices correspond to different systems of qualitative physics. I claim the best way to view research in qualitative physics is to think of it as describing this space of possible theories and their properties. By understanding the alternatives and trade-offs, we can select the best combination of choices for particular purposes.

The next two issues apply this viewpoint to two controversial issues in the current state of the art: continuity and causality.

Continuity Continuity is a formal way of enforcing the intuition that things change smoothly. A simple consequence of continuity, respected by all systems of qualitative physics, is that, in changing, a quantity must pass through all intermediate values. That is, if $A < B$ at time t_1 then it cannot be the case that at some later time t_2 that $A > B$ holds, unless there was some time t_3 between t_1 and t_2 such that $A = B$.

This law has consequences for computing state transitions, since changing inequality relations (or just comparisons with zero, in the case of sign representations) herald state transitions. If $X > Y$ and $D[X] < D[Y]$, for instance, then the relation between X and Y could change to =. Similarly, if $X = Y$ and the same relationship held between their derivatives, then the relationship would change to <.

The details of computing state transitions are the same for all the existing theories, with one exception—how long these transitions will take. The second kind of transition, changes *from* equality, everyone agrees will occur in an instant. The first kind of transition, in every theory right now but QP, always takes an interval of time. In QP theory it takes an interval of time if the difference is finite, but only an instant if the difference is infinitesimal.

Invoking infinitesimals is an unusual step. The motivation is to capture the commonsense intuition that "if you kick something only for a moment, you can kick it back quickly," a kind of symmetry in duration. If you influence a quantity away from equality for only an instant, one should be able to push it back in an instant. In my first implementation of QP theory, GIZMO, this model caused cycles of behavior whose states only lasted for an instant (called *stut-*

ter). These cycles could then be merged into single states, expressing a changing equilibrium [Forbus, 1984b]. Unfortunately, in at least some of the examples studied the instant-instant transitions were violating continuity on derivatives, and a more accurate implementation (QPE) fails to show stutter. At this point it is not clear whether or not stutter will always be ruled out by such constraints,[4] and whether or not it will appear in "natural" models.

The more general question is, are infinitesimal models useful? Or should we simply adopt classical continuity universally? There are two arguments for continuing to pursue alternatives to classical continuity. The first is that infinitesimal models are proving their worth in other areas of qualitative physics (see Section 4.1.1 and [Weld, 1987]). The second is that classical continuity alone is inadequate to model the full range of phenomena in qualitative physics. Impulses, for instance, are part of every engineer's vocabulary. Yet they violate classical continuity, by allowing instantaneous transitions *to* equality. Other similar phenomena have been explored recently by Nishida and Doshita [1987]. Continuity, while significantly tamed through the efforts of a few hundred years of mathematics and physics, still has some unexplored territory.

Causality By any account, causality remains unruly, even after a long history of investigation. A recent public exchange between de Kleer and Brown and Iwasaki and Simon in the AI Journal unfortunately may have shed more heat than light on the matter. At the risk of unleashing yet more rhetoric, I will attempt to clarify the issues here.

The necessary framework to understand these issues appears in [Forbus and Gentner, 1986b], where Dedre Gentner and I analyze the various notions of causal reasoning about quantities used in qualitative physics. The goal of that analysis is to isolate some distinctions that may be useful in understanding human reasoning. Roughly, these distinctions are: the temporal aspects relating cause and effect (the *measurement scenario*), whether or not the ontology contains an explicit class of mechanisms or not, and whether or not the primitives for describing equations include presuppositions about the direction of effect (*directed* versus *non-directed primitives*). The second two factors will be the most relevant for this discussion.

We assume that some notion of *mechanism* underlies all causal reasoning (see [Forbus and Gentner, 1986a]). However, accounts differ in their construal of what mechanisms are. In *explicit-mechanism* theories, the notion of mechanism is tied to particular ontological classes. For example, in QP theory, processes are the mechanism: they are the source of all changes. In *implicit-mechanism* theories, such as de Kleer and Brown's confluence theory, the notion of mechanism arises from the interactions of the system's parts. They

4 Cycles of length 2 are forbidden, but longer sequences look plausible.

assume that flow of information in the model of the system directly mirrors "flow of causality" in the world. To see the differences, consider a liquid flow between two containers. In QP theory all changes would be caused by an instance of the `liquid-flow` process. In a confluence model the changes would arise from the interaction of the constitutive equations.

The difference between directed and non-directed primitives can be illustrated again by comparing QP theory and Confluence theory. The *influences* used in QP theory (and others) to represent equations are directed primitives. Influences include qualitative proportionalities and direct influences (`I+` and `I-`) needed to specify derivative relationships. We might represent the relationship between level and pressure in a contained liquid `WC` as:

$$\text{pressure(WC)}\ \alpha_{Q+}\ \text{level(WC)}$$

indicating that a change in level could cause a change in pressure, but not the reverse. In Confluences (and others), the primitives are non-directed since they do not carry a presupposition of causality. Thus we might say

$$\text{pressure(WC)} = \text{level(WC)}$$

but would be equally willing to say a change in pressure causes a change in level as the reverse. Notice that, at least in this case, there is a clear, intuitive direction.

Any causal analysis must determine which way the primitives in its representation are to be used. In theories with explicit mechanisms, what is an independent parameter is determined by what the mechanism directly affects. In QP theory, for instance, the *causal directedness hypothesis* [Forbus, 1984b] expresses causality:

> Changes in physical situations which are perceived as causal are due to our interpretation of them as corresponding either to direct changes caused by processes or propagation of those direct effects through functional dependencies.

A process directly affects something by supplying its derivative. (Since it can supply a derivative of 0, the same notion suffices to impose causality on static situations.)

By contrast, in theories with implicit mechanisms, some other means of specifying independent parameters must be found. For example, the confluence model critically relies on an input perturbation for causal analysis. The choice of input parameter provides significant constraint on the direction of propagation (which is interpreted as the direction of causation) in the system. This constraint is not quite sufficient, since it is necessary to annotate some parameters as independent, to prevent inappropriate causal deductions ([de Kleer and Brown, 1984], page 73).

Now we are in a position to understand the *causal ordering* proposal of Iwasaki and Simon [1986]: They propose to use directed primitives, similar to qualitative proportionalities, but without associating a sign of effect (i.e., α_Q, but not α_{Q+} or α_{Q-}). The exogenous variables of the system are used as the independent variables. Given these independent parameters, the technique of causal ordering will produce a graph of dependencies by manipulating the quantitative equations describing the system. To get the direction of change imposed by each connection, they propose to use the method of *comparative statics*, which uses quantitative information to produce a sensitivity analysis. The end result will be much the same as the graph of influences that holds for the corresponding situation in a QP model. The possibility of incorrect causal arguments seems to be avoided by detecting when the system of equations is underdetermined: It is exactly in such cases that an assumption must be made, and an external knowledge source (such as the user) can determine which assumption will lead to correct arguments.

Whether or not causal ordering is useful in analyzing a particular example depends on the availability of two things: a set of quantitative equations and knowledge about which variables are exogenous. For many circumstances equations are available, but for many simple circumstances (such as boiling) they aren't. Often the available equations are too complicated to use: A high-accuracy differential equation model of a coal-fired power plant, for instance, can be dozens of pages long. Basing the notion of causal independence on exogenous parameters limits causal ordering to creating models of specific systems in specific modes of behavior. The limitation to specific systems comes from the fact that what is exogenous often changes when a system becomes part of a larger system. Thus we cannot carry our analysis of, say, a heat exchanger, intact to the analysis of a larger system including it. The limitation to specific modes of behavior comes from the fact that the equations describing a system or object can change drastically (phase changes in fluids and turbulent versus non-turbulent flow are two examples).

While causal ordering satisfies several intuitions about commonsense reasoning, it also violates two others. First, since it requires quantitative equations, it cannot explain how commonsense physics comes about—after all, people reason causally about quantities long before they can do symbolic algebra. Second, it also does not assign causality in feedback systems ("a chicken and egg problem," [Iwasaki and Simon, 1986]), although such descriptions are common in informal descriptions of how systems work.[5]

5 There is no obvious reason why it couldn't; in classical simulation paradigms such "loops" in the equations are broken by delay elements (i.e., integration operators), and similar techniques can be used in qualitative equations (e.g., the QP theory notion of direct influence).

I believe that, while the techniques Iwasaki and Simon describe seem to have only limited usefulness as simulation tools, they could be quite valuable in the context of knowledge acquisition. Consider the problem of acquiring knowledge from textbooks. Two kinds of knowledge must be encoded. The formal aspects, the equations, must be transformed into qualitative laws. The informal aspects, the contents of the text, must be transformed into the organizational structure (typically ontological) that tells when these laws are appropriate and useful. Causal ordering and comparative statics may be useful techniques in translating the explicit, formal knowledge of a domain. By combining these techniques with a system that can induce representations for the implicit knowledge, we might be able to develop tools to semiautomatically acquire qualitative models by interacting with human experts.

4.2 *Qualitative Kinematics*

There has been significant progress in qualitative dynamics. Several representations for ontology, number, and equations have been explored, a number of successful programs developed to test these ideas, and there are high expectations of future progress. Unfortunately, the same cannot be said for qualitative kinematics. This section explores why, and describes some progress made since the original survey talk upon which this essay is based.

To begin with, we must refine what we mean by qualitative kinematics. We exclude problems like navigation, manipulator-level planning, and layout design simply because they overlap to a greater degree with robotics and engineering problem solving than with qualitative physics per se. By qualitative kinematics I mean the spatial reasoning aspects of qualitative physics. Examples include reasoning about motion, the geometry of fluid flow, the shape of charge distributions, and so forth. Most efforts have focused on the simplest of these, reasoning about motion. And recently, significant progress has been made on reasoning about *mechanisms*, in the classical sense—gears, transmissions, mechanical clocks, and the like.

I mentioned before that the dividing line between "prehistory" and the present in qualitative physics lay in the decision to explore purely qualitative representations. This tactic was reasonably successful in qualitative dynamics. I claim this hasn't happened in spatial reasoning because it cannot be done. We conjecture that there is no purely qualitative kinematics (the *poverty conjecture* [Forbus et al., 1987]).

This idea takes some explaining. Consider FROB. It did some fairly sophisticated spatial reasoning, including understanding collisions and the notion of being trapped in gravity wells. But to arrive at this understanding took a metric diagram, which contained a significant amount of quantitative information.

Thus FROB itself is not purely qualitative.[6] But in fact purely qualitative representations suffice for a surprising number of inferences about dynamics. Sadly, it just doesn't seem to be the case for qualitative kinematics.

The poverty conjecture is based on three arguments. First, no one to date has developed a purely qualitative kinematics. For example, I've spent years trying to develop one, and I've talked to a number of other people who have as well, with little success.

Naturally, this is a weak argument. Negation by failure is rarely safe scientifically, and part of my motivation for making this conjecture is the hope that someone will succeed in proving me wrong! But the second argument makes me skeptical. Much of the power of qualitative dynamics comes about from partial orders. Time, as Allen [1984] showed, can be nicely modeled in terms of temporal relations where transitivity provides significant constraint. Inequalities, while individually weak descriptions, combine via transitivity to yield often powerful conclusions. But these are both one-dimensional problems. There is a result in dimension theory which states that partial orders don't work for higher dimensions. Try it yourself: Create a vocabulary of spatial relationships between 2D figures like Allen's relationships for time, such as EQUAL, INSIDE, ABUT, OVERLAP, and so forth. You'll find the only entries in a transitivity table for such relationships that provide significant constraint are those which impose a partial order (in this case, EQUAL and INSIDE). With the others (e.g., ABUT, OVERLAP), just about anything is possible.

While stronger, this second argument still does not clinch the matter. After all, there might be some other powerful idea, some new formalism that will provide the "right" quantization for shape and space independent of an initial quantitative description.[7] But the third argument is that we have no reason to think that such a formalism necessarily exists, because people appear to perform poorly at spatial reasoning without the "moral equivalent" of a diagram. There is a large literature on the psychology of visual imagery, and while it must be interpreted with care, it seems to indicate that some kind of quantitative information plays an important role in human spatial reasoning. In addition to imagery, people resort to sketches, models, looking at the object itself, and so forth—in short, we harness our perceptual apparatus in service of spatial reasoning.

This apparent reliance on perceptual apparatus motivated FROB's metric diagram, and we believe that this model can be extended productively into a general model for qualitative kinematics (the *MD/PV model* [Forbus et al.,

6 If quantitative dynamics worked that way, there would be no qualitative simulators per se. Instead, we would always have to provide numerical simulation routines and lots of numerical parameters to get any predictions. (Or use symbolic algebra—as mentioned earlier, not every symbolic description is qualitative, and this is a good example.)

7 As shown previously, useful qualitative descriptions for space can be *computed* from quantitative ones—but the goal in this argument is to avoid using a metric diagram altogether.

1987]). By this account, spatial reasoning requires at least two representations. The first is a metric diagram, which includes quantitative information and can answer geometric questions by some form of calculation or measurement. The metric diagram attempts to describe the functionality of the visual system in human spatial reasoning. One operation that can be done with a metric diagram is computing a *place vocabulary*, which quantizes space by some relevance criteria. Figure 15 shows how this model was instantiated in FROB.

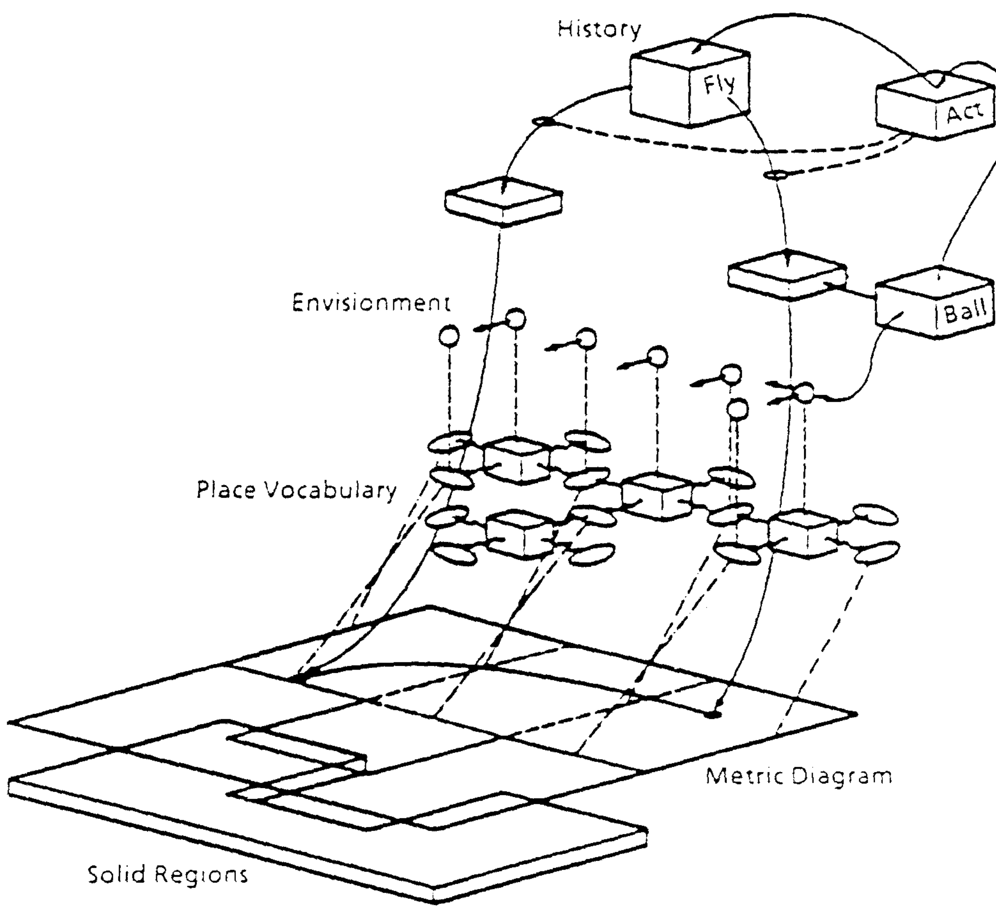


Figure 15 FROB illustrated the MD/PV model of spatial reasoning. This picture illustrates what is "under the hood" in FROB. The metric diagram provides a means of communicating with the user, a means of answering quantitative spatial queries, and a substrate for computing a qualitative description of space. The first step in computing this place vocabulary is to ascertain the *solid regions*, where free space isn't. Next, it breaks up the free space into regions, in a way that simplifies the description of possible motions. These regions plus symbolic descriptions of their connectivity form FROB's place vocabulary.

It seems that all spatial reasoning projects to date fit the MD/PV model fairly well. For example, the (earlier) natural language understanding program by Waltz and Boggess [1979] used a metric diagram in constructing models of sentences like "A fly is on the table." Geoff Hinton [1979] developed an elegant theory of imagery that used a mixture of propositional and numerical representations to explain phenomena that simpler theories based on array representations cannot explain. In reasoning about geological processes, Simmons [1983] compared quantitative calculations with a diagram to check the correctness of qualitatively plausible histories. Stanfill [1983] used symbolic descriptions with numerical parameters to reason about simple pistons and bearings. Davis [1987] argues that purely qualitative representations are "too weak" to support reasoning about motion involving solid objects.

4.2.1 *Reasoning About Mechanisms* There has been renewed interest in spatial reasoning recently, particularly in understanding mechanisms. Gelsey [1987] uses a constructive solid geometry CAD description as his metric diagram, and computes motion envelopes to recognize kinematic pairs. The place vocabulary in his system consists of regions that involve interactions between parts. Joskowicz [1987] has proposed to analyze single interactions in a mechanism by recognition, describing kinematic pairs by patterns in configuration space. (Configuration space was first used in robotics for motion planning problems, see [Lozano-Perez, 1983]).

In our own CLOCK project, Faltings [1986, 1987a, 1987b] has developed a general theory of place vocabularies for mechanisms. Faltings observes that the important distinctions for quantizing shape must come from *pairs* of objects, rather than objects in isolation, since it is their interaction that determines whether or not a pair of objects will move together or bind. In mechanisms, each part has only one degree of freedom, so a configuration space for a pair of objects is two-dimensional. The place vocabulary for an entire mechanism (such as a clock) is the combination of the place vocabularies for the pairs of parts. Faltings also observes that symbolic algebra can be used to parameterize place vocabularies, thus increasing the potential for their use in mechanical design. Faltings's theory has been tested by an implementation on a wide range of examples, including gears, ratchets, escapements, and the complete set of kinematic pairs for a mechanical clock [Faltings, 1987b].

Of course, Faltings's theory only solves half of the problem: It describes what contact relationships are possible, and what might be reached if movement occurs in a particular direction. To integrate this information with a qualitative dynamics requires imposing reference frames in order to describe forces and motions. Nielsen, in his part of the CLOCK project, has developed a theory of qualitative vectors and reference frames. Such vectors are used for representing contact directions, forces, velocities, and other parameters. He has used these techniques in a qualitative theory of rigid-body statics [Nielsen, 1987].

which can determine what directions an object is free to move in as well as what movement will occur. This theory has been implemented and has successfully answered questions about the stability of Blocks World structures, in addition to gears and escapements.

4.3 *Styles of Reasoning*

The purpose of representation is reasoning. This section describes some of the styles of reasoning that have been explored in qualitative physics to date. Because there has been confusion about the relationship between envisioning and other forms of qualitative simulation, this issue is discussed in detail. I will ignore diagnosis, since an adequate treatment is well beyond the scope of this survey.

4.3.1 ***Qualitative Simulation*** The result of a standard numerical simulation is a list of state vectors, each vector representing the system being simulated at some particular Δt. Qualitative simulations differ from numerical simulations in two respects. First, time is individuated by the occurrence of interesting events, rather than some regular, fixed increment. Second, the reduced precision of qualitative representations often requires branching to represent alternate possible futures.

It is important to note that some qualitative simulators do not produce specific histories at all! This is a subtle point that is often misunderstood. A history describes a specific behavior of an object. While a history is (at least potentially) infinite, it typically consists of only a finite number of distinguishable episodes. Referring back to Section 4.1.1, we say that two episodes are distinguishable exactly when they differ in some limit point (i.e., temporally generic landmark). The implication is that each episode can be described as an occurrence of one of a finite set of abstract *qualitative states*. This assumes there are a finite number of properties, and a finite number of values for each property, and hence only a finite number of combinations of these properties. Similarly, for any finite collection of objects we can define qualitative states that describe consistent collections of every possible distinguishable episode for each object.

Qualitative states can be defined without recourse to histories. In fact, the notion of qualitative state was developed earlier than histories, as Section 3 indicates. The graph formed by the collection of all qualitative states of a system and the transitions between them is called an *envisionment*. The notion of envisionment is due to de Kleer [1975]. The process of constructing an envisionment, *envisioning*, was the first method of qualitative simulation. Roughly, each history corresponds to some path through the envisionment, but the converse is not true, as we will see shortly.

A further distinction between envisioners is whether they start from a given initial state or from all possible states. The former are said to produce *attainable* envisionments, the latter *total* envisionments. Total envisionments are usually larger than attainable envisionments, but are more useful for certain tasks. A number of envisioners of each type have been built for different theories. NEWTON [de Kleer, 1975] and FROB [Forbus, 1980] both produced attainable envisionments for different kinds of motion problems. QUAL [de Kleer, 1979b] produced attainable envisionments for electronics, while ENVISION produced total envisionments for system-dynamics-like models (see Section 4.1.5) For qualitative process (QP) theory, GIZMO [Forbus, 1984c] produced attainable envisionments, while QPE [Forbus, 1988] produces total envisionments.

Several programs produce histories directly. FROB, for instance, used a constraint-based numerical simulation to generate histories. In several important applications, histories are specified as part of the description of a problem, as in integrated circuit fabrication [Mohammed and Simmons, 1986] or hypothesized on the basis of other knowledge [Simmons, 1983]. Kuipers's QSIM system, of course, generates histories directly.

4.3.2 ***Envisioning Versus History Generation*** The relationship between envisionments and histories is more subtle than first suspected, and is still being explored. Some aspects are clear; for instance, I've defined a *logic of occurrence* [Forbus, 1987a] that specifies how a history may be related to an envisionment so that general behavioral constraints (such as assuming classes of behavior must or may not occur) can be enforced. Sometimes there have been simple terminological confusions, such as de Kleer and Brown [1984] calling their qualitative states "episodes," Kuipers [1986] calling his account of history generation a "deeper semantics" for envisioning, or Collins and Forbus [1987] calling their MC envisioning a history. Other aspects, however, are genuinely problematic and have become fertile areas of research.

In a correct envisionment, every possible history can be expressed as a path. Various properties of the graph correspond to important behavioral distinctions. For example, states with no transitions from them represent final states for the system, and cycles correspond to oscillations.

Originally, de Kleer [de Kleer and Brown, 1984; de Kleer, 1984a] claimed that, just as every history corresponds to a path through the envisionment, so every path through the envisionment must correspond to a physically realizable history. Kuipers [1986] shows this is incorrect. The counterexample he uses is shown in Figure 16 (this envisionment was generated with QPE [Forbus, 1988]). The parameter Z is a function of position, and should be compared with Z', but is otherwise unconstrained. By declaring the comparison between

Z and Z' as interesting, we will cause a state transition to occur whenever the relationship between them changes. There are other transitions that will occur due to the way motion and acceleration are modeled (see [Forbus, 1984c] for details).

To generate a history from an envisionment, begin by selecting a start state. That state forms what occurs at the first episode in the history, the duration of the episode being the duration of the corresponding qualitative state (i.e., either an interval or instant). If there are no transitions from the chosen state, then that episode is the end of the history. If there are, select one of the transitions as representing what actually occurs. Then continue as before, starting from the state resulting from the transition.

Carrying out this procedure on the envisionment of Figure 16 reveals a variety of possible histories. For example, the sequence of states S_1, S_4, S_7, S_{10}, S_{13}, S_{16}, S_{19}, S_{22} corresponds to a legal history, as does S_3, S_6, S_9, S_{12}, S_{15}, S_{18}, S_{21}, S_{24}. Other legal histories correspond to variations of these where Z changes in its relationship to Z' within the range of variation for X. For example, the sequence S_3, S_6, S_8, S_{10}, S_{13}, S_{16}, S_{20} ,S_{24} corresponds to the case where Z equals Z' when X equals zero.

All of the histories mentioned so far are legitimate. But consider again the transitions from, say, S_6. Each time around the cycle, one of these transitions must be chosen. In the algorithm specified, which corresponds to the original de Kleer claim, each such choice is independent. Thus we are free to choose another transition the next time we reach S_6, which will give us an illegitimate history. The problem can arise even on a single cycle: the sequence S_3, S_6, S_8, S_{10}, S_{13}, S_{16}, S_{17}, S_{18} ,S_{21}, S_{24} is inconsistent because the S_6, S_8 ,S_{10} subsequence assumes $Z = Z'$ when X = ZERO, while the S_{16}, S_{17}, S_{18}, S_{21} is based on the assumption that Z reaches Z'D before X reaches ZERO. The choices are not in fact independent, and treating them as such can lead to incorrect predictions.

In this simple case, the solution seems clear: Each choice of transition implies additional information about the functional relationship between X and Z. For example, assuming that the transition from S_6 to S_8 occurs "fixes" a point on the (implicit) graph defining their relationship: in particular, $Z = Z'$ when X = ZERO. (Assuming that one of the other transitions occurs requires introducing a new constant related either to X or to Z, but the principle is the same.) These constraints must then be respected in successive choices. For example, choosing the transition from S_{12} to S_{11} forces the later transition of S_{16} to S_{17}. However, it is not straightforward to generalize this technique to all situations.

To summarize: With no information, we can get incorrect predictions. If we had a fully specified correct quantitative model, there would be no ambiguity and hence we would always get correct histories. The open research question right now is, just how much information, and in what form, suffices to generate histories correctly from envisionments?

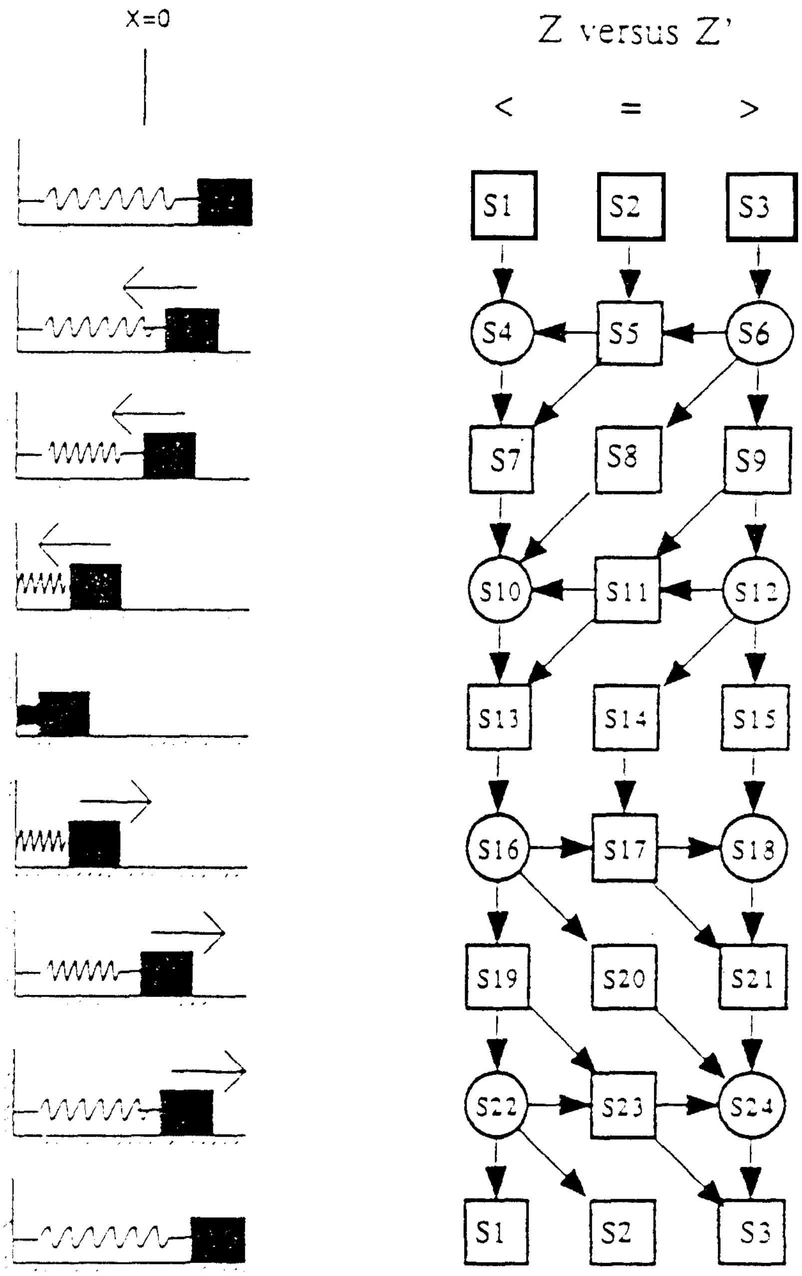


Figure 16 Generating histories from envisionments can be difficult. An envisionment for a modified spring-block oscillator is shown below. The modification consists of an extra parameter *Z*, which is a function of *X* and is compared with an arbitrary constant *Z*'. Each row is labelled with a picture indicating the general position and velocity of the block in the states of that row. Each column indicates the relationship *Z* has with *Z*' in those states. Arrows denote locally consistent transitions between states. Circles indicate states that last over an interval, while squares indicate states lasting only for an instant.

This problem arises even without envisionments; direct history generation must also take into account constraints imposed by earlier choices. In QSIM, for example, new named values can be introduced at every step of the computation, corresponding to the value a quantity takes on in a particular episode of the history (more on this below). Since the algorithm can introduce a new value between any two adjacent previous values, the number of possible episodes can (and does) grow exponentially without bound. This means that QSIM also produces incorrect histories. Several pruning techniques to weed out incorrect histories have been investigated, including problem-specific constraints [Lee et al., 1987], algebraic manipulation [Kuipers and Chiu, 1987], and quantitative knowledge [Chiu, 1987], but so far these results have been mixed. (For instance, Struss [1987] points out several limitations of qualitative mathematics, such as sensitivity to the form of equations, which indicate that algebraic manipulation of qualitative equations is often unsafe.)

Both envisionment and direct history generation have their role to play in the arsenal of qualitative physics. The notion of envisionment is a superb theoretical tool, providing a simple way to think about classes of behaviors. Envisioning is a good methodological tool for qualitative model development, since it exercises domain theories in obscure cases that the model builder might otherwise ignore. But envisioning is unlikely to be the desired solution for quick on-line computation: After all, it corresponds to explicitly generating the entire problem space for some class of problems! In such cases history generation, perhaps combined with heuristics, seems to make sense. The space/time trade-offs in qualitative simulation have only begun to be explored. One can imagine compiling envisionments "offline," for example, or the envisionment of a system at a high level of abstraction being used to guide direct history generation at a lower level.

4.3.3 *Recognition* Engineers are good at explaining how things work. Often, this occurs by recognition "Oh, it's a proportional-action controller"—they redescribe the system in terms drawn from a *functional vocabulary*. This functional vocabulary appears to help organize their knowledge for several purposes. In diagnosis, symptoms might be computed by comparing current behavior against the standard behavior stored with the functional description. In design, a functional vocabulary provides an intermediate goal that constrains the search space. The designer might decide what combination of functional blocks would achieve her purpose, and then figure out how to implement this functionality with the available components. Capturing this ability to map from *structure* to *function* was an early focus of qualitative physics.

The most successful work in this area is still that of de Kleer [1979b, 1984a], who originally pointed out the problem as well. His theory is that to perform recognition, engineers first figure out how the system behaves, and then use that description of behavior to "retrieve" into a functional vocabulary.

A transistor circuit that behaves in a particular manner, for instance, might be recognizable to an engineer as a "common-emitter amplifier." One elegant aspect of de Kleer's work was how he constrained the result of qualitative simulation. The simulation proceeded by determining how the system would respond to "poking" its input. He noted that any sensible engineer wouldn't include parts that didn't help the circuit perform its function. Thus, any interpretation of the circuit's behavior that did not include every component could be ruled out on *teleological* grounds. In almost all of the electronic circuits he examined, this principle sufficed to rule out all but one interpretation.

While this work was one of the early successes of qualitative physics, little has been done by way of follow-up. What is needed is the formalization of rich functional vocabularies, and this problem has received little attention. Recent work by Chandreskaran [Sembugamoorthy and Chandrasekaran, 1984] and Doyle [1986] can be viewed in this light.

4.3.4 *Measurement Interpretation* Ideally, we would like our programs to gather their own data about the world. A program that works in a power plant, for instance, should have the ability to "read the gauges" to find out what is happening inside the plant. This is the problem of *measurement interpretation.* My ATMI theory [Forbus, 1986a, 1987c] describes how to interpret measurements taken over a span of time in terms of qualitative states. This theory is very general, requiring only domain-specific procedures for performing an initial signal/symbol translation and that an envisionment (potentially) exists. An implementation has been demonstrated that works on multiple ontologies (i.e., both QP models and FROB models). However, at this writing it has only been tested on simulated data without gaps, and does not specify control strategies for handling noisy data.

Yet a different kind of measurement interpretation was studied by Simmons in the GORDIUS program [Simmons, 1983]. The specific problem he addressed was evaluating whether or not a hypothesized sequence of geological events could account for the strata at a particular place. Knowing how the sequence came about is important economically, since some sequences will result in oil as a byproduct and others won't. A map built up out of well measurements represents the final state of this behavior. The program accumulated constraints on the size and shapes of maps that could result from the proposed history, and checked the actual map to see if it was consistent with these constraints.

5 *The Frontier*

The previous sections examined where qualitative physics came from, and where it is now. I have tried to paint a coherent picture of the state of the art, indicating the alternatives that have been explored and where substantial progress has been made. But no survey is complete without looking at the boundaries: areas which right now are relatively unexplored, and are thus fertile ground for new investigations.

5.1 *The Near Future*

I'll begin by describing some areas that are likely to see rapid progress. It would surprise me to not see significant advances in these areas in the next three years or so.

5.1.1 *Improved Domain Models* A central activity of qualitative physics is developing a variety of models for physical phenomena and engineered systems. However, building good domain models is very difficult, and even with good tools takes much longer than one would expect. Nevertheless, the next few years should see significant advances in the kinds of physical phenomena that we can represent. For example, initial forays into reasoning about granularity and composition [Bunt 1985; Schmolze, 1986; Raulefs, 1987] may provide tools for reasoning about nonrigid objects. I suspect that progress in modeling powders and clays will require developing more sophisticated geometric representations to describe deformations, sheer, stress planes, and the other constructs of materials science. In modeling fluids, we still do not have a good theory of mixtures that describes exactly how different stuffs affect each other inside a container. An especially fertile ground is chemistry, which is interesting both industrially and intellectually, since it requires integrating discrete structures and geometry with reasoning about continuous systems.

5.1.2 *Implementations* I expect that implementations will steadily improve in performance and storage economy—we haven't been building qualitative simulators for very long, after all, and are still discovering the right techniques. This trend, combined with the rising tide of improvements in computer technology, suggests that the range of problems we can tackle will continue to expand.

As we understand styles of reasoning better, the kinds of programs used in qualitative physics will become more diverse as well. Problems like design, for instance, require a detailed accounting of how different properties of the components and their interconnections relate to properties of the behavior produced. Keeping track of these justifications, especially in the presence of feedback, is a difficult problem. Williams's [1986] temporal constraint propagator TCP is the first system that does this correctly. Widespread application of these techniques should improve the sophistication possible in qualitative analyses.

One of the advantages of envisioning is that it postpones worrying about control issues. Alas, such issues cannot be put off forever. Solving problems by explicitly generating the entire search space simply is not a viable long-range alternative. Notice that history generation, per se, is not the answer— these approaches are already plagued with control problems, since they can lead to infinite descriptions of behavior. (In fact, a resource limit is often imposed for control purposes.) An attractive alternative is to generate generic qualitative states by heuristic search, applying the standard AI techniques to minimize effort. This subset of the envisionment can then be used as a framework for constrained generation of temporally specific landmarks, if needed.

Of course, this is just one alternative. Another idea is to decompose a complex system (such as a power plant) into a collection of semi-independent pieces, produce envisionments for each of the pieces, and glue them together as needed to provide a description of the whole plant. A few theoretical ideas have been proposed for such decompositions (e.g., the notion of *p-component* in [Forbus, 1984b]), but the bulk of the work remains to be done.

Another control issue that must be faced concerns domain models which are potentially infinite. Consider this simple model: An object consists of a set of parts, each of which is itself an object. This simple recursive structure will kill every existing qualitative simulator in which it can be stated (it cannot even be stated in most), and hence such models have been avoided. However, such descriptions are sufficiently useful that techniques for controlling their instantiation should be explored.

5.1.3 *Ontological Shifts* It is unlikely that we have exhausted the space of ontological choices. Furthermore, not much is known about the relationship between various ontologies. For example, aside from a few rules of thumb, we cannot precisely characterize when to use a device-centered ontology instead of a process-centered ontology.

In examining human reasoning, it seems ontological shifts occur in the course of solving a single problem. Recall the SWOS problem from Section 2.2. Most people implicitly use two distinct ways of looking at fluids to solve this problem. To establish directions of flow and the fact of boiling required looking at "the stuffs" in different parts of the system—the water in the boiler is turning into steam, the lower pressure in the load means there will be a flow of steam from the boiler through the superheater, and so on. To figure out how the temperature actually changed, however, required thinking of a little piece of stuff travelling through the system.

Early on, Hayes [1985] identified these ideas as the *contained liquid* ontology and *piece of stuff* ontology, respectively. Most qualitative physics work has used the contained liquid ontology. Recently John Collins and I developed a specialization of the piece of stuff ontology, the *molecular collection* ontology, to capture the kind of reasoning engineers do about thermodynamic cycles. The idea is to define a little piece of stuff, MC, which is large enough to have macroscopic properties yet small enough never to split up when traversing a fluid system.

How is an MC envisionment generated? Since qualitative representations are not detailed enough to provide local gradients, what MC does is computed from an envisionment generated using a contained stuff ontology. We suspect this is exactly the kind of ontological shift occurring in examples like the SWOS problem.[8]

Even considering fluids, many ontological questions remain open. For example, what other specializations of Hayes' piece of stuff ontology are useful? Spatially extended pieces of stuff appear essential to modeling mixing and weather patterns—how are they to be individuated and combined? I am sure that as we attempt to build more sophisticated domain models, we will uncover many new ontological issues, many of them revolving around spatial reasoning.

5.1.4 *Hypothesizer* One particularly interesting potential application is a kind of monitoring task, using a module I call a hypothesizer.[9] The goal is to merge measurement interpretation with explanation in order to improve plant operations and fault management.

Suppose you have someone controlling a large, complicated system, such as a production line in a chemical plant, and some condition arises that must be dealt with. Operators in such circumstances will often seize upon the first theory they generate about what is going on, and stick with it even in the face of contradictory data. Imagine a program that could critique an operator's theory. Such a program, if done properly, could have two benefits. First, it would force the operator to be explicit about his theory of what is wrong. Second, the program could compare the consequences of the theory with measurements, point out discrepancies, and suggest further experiments and modifications. Besides being used for diagnosis, it would not surprise me if this kind of module became one of the first applications of qualitative physics. Providing human-understandable explanations is the forte of qualitative physics, after all.

5.1.5 *Planning* Realistic planning requires knowing what the physical world will do, with and without the planner's actions. How can we best use qualitative physics in planning?

One way is to transform the domain model into something the planner can use. Hogge's *domain compiler* [Hogge 1987a, 1987b] takes as input a QP domain model, and produces rules suitable for a temporal planner. (The planner derives from [Allen and Koomen, 1983], adding inference rules and other extensions—see [Hogge, 1987c] for details.) Given a description of liquid

8 Techniques for comparative analysis in [Weld, 1987] provide another piece of the puzzle. It is not known at this writing if together these techniques are sufficient to solve the SWOS problem.
9 Mike Williams of IntelliCorp calls it a "Doubting Thomas" system.

flow, for instance, the domain compiler produces an inference rule describing what it takes to cause a liquid flow to happen. When these rules are added to other inference rules and a specification of the actions an agent may take, the planner can create plans which involve processes as intermediaries, such as filling a kettle by moving it under a faucet and turning it on.

While elegant, this approach requires more research to live up to its promise. The large descriptions produced by the domain compiler, and the complex inferences required (especially transitivity), tend to choke the temporal planner. Compiling can also produce oversimplified models. For instance, the rules implicitly assume that any influence they impose on a quantity will actually succeed in changing that quantity. Thus a planner using these rules might assume that it can prevent an ocean liner from sinking by bailing with a teaspoon. Such limitations do not appear impossible to overcome, and no doubt there are other valuable approaches to be explored as well.

There is also a second kind of planning problem that I think ultimately is going to be extremely important, yet has received little attention to date—the problem of *procedure generation.* When you design a new engineering system, you don't just design the object, you have to develop procedures for operating it, for maintaining it, for diagnosing problems with it. If we are trying to get our computers to help us design complex systems, we need to find ways to have them generate such procedures automatically. If the design system knew the kinds of actions the system operators can take and their limitations, its output could include not just the blueprint, but the operations manual, the maintenance manual, and the diagnosis manual (or expert systems that provided the same service). Furthermore, safe operation could be posted as an explicit constraint on the design of the plant.

5.1.6 *Connections with Traditional Physics* Understanding the kind of reasoning scientists and engineers do was the original motivation for qualitative physics. To fully capture what they are doing, we must extend qualitative physics in the direction of traditional physics. This section describes two exciting recent efforts in this area.

In traditional physics, a set of equations can be solved analytically or by simulation to derive the behavior of a system. Similarly, qualitative equations are typically derived from an ontology in order to generate behavior via qualitative simulation (either envisioning or history generation, see above). Sacks [1985] has developed an analytic technique that generates qualitative descriptions from traditional equations. His initial QMR system could solve a variety of systems, including models of a dampened oscillator and heat dissipation. One limitation of this approach is that most interesting equations do not have analytic solutions. Sacks's [1987] solution is to decompose more complex systems into piecewise linear approximations, use QMR on each piece, and reconstruct the global solution from the local solutions.

Yip [1987] has a complementary approach to a similar problem. *Phase portraits* are a geometric technique traditionally used in mathematics to describe complex dynamics. Yip has created a vocabulary of qualitative descriptions of phase space that formalizes the intuitions mathematicians bring to bear in understanding such portraits. Given a numerical simulation of a non-linear system, he uses this vocabulary to interpret the particular behavior, and make predictions about what the other parts of phase space must be like. Ultimately, these predictions will form the basis of additional numerical experiments.

Williams [1988] has developed an elegant formalism that combines qualitative and quantitative algebra. Potentially, this theory could greatly extend the range of qualitative reasoning.

5.1.7 *Learning* Creating a complete qualitative physics is a herculean task; it will become much easier if our machines can help. Several workers are tackling different aspects of this problem. Langley, Simon, Bradshaw, and Zytkow [1987] have studied various aspects of scientific discovery of physical laws. So far, their work has focused on equational and discrete symbolic (as opposed to qualitative) models. Kokar [1987] describes a methodology for determining limit points using dimensional analysis. Falkenhainer's ABACUS [Falkenhainer, 1985] program uses qualitative proportionalities as an intermediate representation in inducing equations from numerical data. Mozetic [1987] describes how hierarchy can be exploited in automatically acquiring qualitative models, demonstrating his techniques with a model of the heart. Rajamoney and DeJong [1987] have tackled the problem of debugging qualitative theories, providing a theoretical classification of bug types, including strategies for detecting and fixing them.

At Illinois we are taking two different approaches to understanding learning in physical domains. The first is psychological: Dedre Gentner and I are combining QP theory and her Structure-Mapping theory of analogy [Gentner, 1983, 1987, 1988] in an attempt to account for experiential learning in physical domains [Forbus and Gentner, 1986a]. We suspect the kinds of representation and reasoning explored by qualitative physics to date actually appear rather late in human learning, with two other stages postulated for both computational reasons and to explain certain psychological findings. Right now we are exploring these ideas through both cognitive simulation (using SME, a cognitive simulation of Gentner's analogy theory [Falkenhainer et al., 1986, 1988]) and psychological experiments.

The other approach, the Automated Physicist project, is being carried out in collaboration with Jerry DeJong. The idea is to build a series of machine learning systems that learn by experimentation and observation and by solving textbook problems. The dream behind the AP project is to build a sort of "Sherlock Holmes" of physics—it it begins by sitting back in its armchair and trying to explain reported behavior in the physical world. If it can explain a report no learning takes place. But if it cannot, then it tries to fix its model. Our ultimate goal is to have a program which designs and builds its own experimental apparatus, analyzes real data, and so forth.

The first such programs are due to Falkenhainer and Rajamoney. Falkenhainer's PHINEAS program has demonstrated how QP models can be learned with his theory of *verification-based analogical learning* [Falkenhainer, 1987]. Given a new behavior, PHINEAS attempts to use its current domain model to explain the behavior. If it cannot, PHINEAS accesses a database of previously observed behaviors with associated explanations. An important aspect of PHINEAS is that it performs analogical matching on the *behaviors* first, to guide the transfer of a QP model from an understood domain to explain the new one. The new model is tested to see if it can explain the observations. Often, the model has to be "fixed up" in various ways. Rajamoney's ADEPT system provides exactly the right functionality, since it has the ability to generate potential improvements and the conceptual specifications of experiments required to decide between them. The two programs have been successfully linked and tested on several examples [Falkenhainer and Rajamoney, 1988].

5.2 *Open Problems*

I would like to finish with a set of open problems. While we will make significant progress on these problems in the near term, they are sufficiently deep and tough not to yield to short assaults. I suspect each of them will take a few generations of Ph.D. theses to solve.

5.2.1 ***Spatial Quantities*** There are no doubt other representations lying between the poverty of signs and the richness of $\Re$ that remain to be discovered. And no doubt there will be advances in qualitative representations for time-varying differential equations as well. But the real frontier is now partial differential equations, especially quantities that vary by space instead of time. Formalizing these *spatial quantities* will allow us to describe a vastly wider range of phenomena than at present. These phenomena include the flow over an airplane wing, the distribution of electric fields due to a distribution of charges, and the stresses on different parts of a bridge.

I suspect the problem decomposes into two parts. The first is the formalization of partial derivatives in general. While this part may have many technical obstacles, it seems likely that the current theories can be gracefully extended in this direction. The second problem appears to me to be much harder: the problem of choosing the appropriate axes and frames of reference to simplify computations and produce perspicuous results.

5.2.2 ***What Kinds of Numbers Are There?*** Imagine what we know about the space of representations for number. Let sign values be at the top and elements of $\Re$ be at the bottom, so that increased height corresponds to increased degree of abstraction. Inequalities are high in this structure, almost up to sign values. Floating point numbers and other simple truncations of $\Re$ lie toward the bottom. You may choose for yourself where to put the order of magnitude formalisms that have been developed recently. The question is, what else is in there? How many different representations for number remain to be developed, and what do they look like?

It would not surprise me if several more useful representations of number were developed. Some, like fuzzy numbers [D'Ambrosio, 1987], will be imported from other branches of AI and mathematics. A better understanding of the tradeoffs and systems that integrate several types of numerical reasoning (like [Simmons, 1986]) are necessary.

5.2.3 *What Kinds of Functions Are There?* A related question is, what sort of functions are there? Traditional physics relies heavily on the *analytic functions*, i.e., combinations of +, −, *, polynomials, trigonometric functions, and so on. These lie at the most precise end of an abstraction continuum. At the other end are qualitative proportionalities, where a closed world assumption is required to even determine what parameters affect a given quantity. How many representations for functions remain to be developed?

I suspect the answer is very few, much fewer than for numbers. Functions and algebras have been well explored by mathematicians for a long time, and while we may harvest a few new things from their efforts, I doubt there will be much because the class of analytic functions is so large. But it is an empirical question.

5.2.4 *Large-Scale Organization of Qualitative Models* Almost all of the models we have built to date are quite simple (on the order of 300 or so axiom-equivalents) compared to the scope of human commonsense or expert knowledge of the physical world. Building such a massive knowledge base will be impossible on an ad hoc basis. Ontology provides one source of organizing principles, but there are no doubt others.

Hierarchy plays an important role in organizing many other AI knowledge bases, and it is likely to do so in qualitative physics as well. Making qualitative simulations work with multiple levels of detail is an important problem (see [Weld, 1986; Kuipers, 1987] for some initial forays).

At least two other organizational ideas appear necessary as well. First, we need to formalize the idea of *structural abstractions*, the conceptual objects used in our representations, as distinct from their real-world counterparts. This separation is needed in order to provide an input language for systems that is reasonably independent of the theoretical commitments of a particular model. It is seductive to consider a transistor as identical to our model of it, and as long as we limit our analysis to a particular frequency range this conflation does little harm. But more sophisticated reasoning about circuits, and any consideration of almost any other engineering domain (e.g., fluid systems, thermal systems, motion) requires more work to map from a relatively neutral description of the physical system to the kind of model used for a particular level of analysis.

The second organizational tool is a language of simplifying assumptions. Rather than build distinct models for different purposes, we should instead use explicit assumptions to turn off and on different parts of a model. For instance, in reasoning about thermodynamic cycles one often invokes a "steady-state assumption—the amount of fluid in each part of the system remains constant, despite flows. Human engineers constantly use assumptions like this to drastically reduce the number of possible states, making analysis of complex systems more feasible. Our models will have to be designed in a way that allows our programs to do the same. We have recently developed some conventions for representing such assumptions in QP theory, and tested them on a large multigrain, multiple perspective model of a Navy propulsion plant [Falkenhainer and Forbus, 1988]. These conventions are a solid first step, but much research remains.

As qualitative physics becomes ready for widespread application, we will face the same kinds of validation issues now confronting other kinds of expert systems. Most engineering disciplines have validation procedures in place, and standards on the quality of model that must be used for a particular level of safety desired. We will have to fit qualitative models into such schemes, somehow.

5.2.5 *Integration with Vision and Robotics* Vision and robotics are, in principle, closely tied to qualitative physics. Qualitative physics can tell a robot where something might go if it is dropped, and what it has to do in order to boil water. As mentioned in the introduction, some form of qualitative physics will be needed by robots that work in unconstrained environments (although in general the useful representations may be more like *protohistories* and the *causal corpus* [Forbus and Gentner, 1986a] than like the current state of the art). But qualitative physics also needs vision and robotics. The poverty conjecture suggests that advances in spatial reasoning and vision will help drive qualitative kinematics. For instance, Ullman's theory of visual routines [Ullman, 1985] can be viewed as a theory of human metric diagrams. Knowing what the visual system computes can suggest what primitives are likely to be useful, and conversely, knowing the computational requirements of qualitative kinematics may in turn suggest what spatial descriptions people might be computing. Eric Saud [1987] has in fact proposed an "information rich spatial representation," using the various representations postulated for human vision to support spatial reasoning.

5.2.6 ***A Complete Qualitative Physics*** Today qualitative dynamics and kinematics are typically pursued in isolation. Integrating them is crucial to building a complete qualitative physics. A full understanding of an internal combustion engine, for instance, cannot be gleaned without understanding how physical processes and geometry interact. Efforts like the CLOCK project are a step, but just a first step, in this direction.

And, finally, of course, there is the ultimate goal. The holy grail of qualitative physics is a complete set of models, spanning the space of all the physical domains people know, able to characterize human models from the person on the street up to the best experts, capable of supporting efficient application programs, and so forth. Like traditional physics, we will probably never get there. But we will certainly learn interesting things on the way.

Acknowledgments

I would like to thank Johan de Kleer, Dedre Gentner, Paul Nielsen, John Collins, Brian Falkenhainer, and Ernie Davis for useful comments and discussions. Support for this work has come from the Office of Naval Research (Contract No. N00014-85-K-0225, Contract No. N00014-85-K-0559), and the National Aeronautics and Space Administration (Contract No. NASA-NAG-9137).

References

Allen, J., 1984. Towards a general model of action and time. *Artificial Intelligence* 23(2).

Allen, J. and Koomen, J., 1983. Planning using a temporal world model. In *Proceedings of IJCAI-83*, Karlsruhe, West Germany. San Mateo: Morgan Kaufmann Publishers.

Bobrow, D., 1968. Natural language input for a computer problem-solving system. *Semantic Information Processing*, M. Minsky, ed. Cambridge, Mass.: MIT Press.

Bobrow, D., ed., 1984. *Qualitative Reasoning About Physical Systems*. Cambridge, Mass.: MIT Press.

Brown, J., Burton, R. and Zdybel, F., 1973. A model-driven question-answering system for mixed-initiative computer-assisted instruction. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(2).

Bundy, A., Byrd, L. Luger, G., Mellish, C., Milne, R. and Palmer, M., 1979. MECHO: A program to solve mechanics problems. Working Paper 50, Department of Artificial Intelligence, Edinburgh University.