THE LIMITS OF QUALITATIVE SIMULATION

Benjamin Kuipers

MIT Laboratory for Computer Science,
Cambridge, Massachusetts 02139.

Abstract:

Qualitative simulation is a key inference process in qualitative causal reasoning, In this paper, we present the QSIM algorithm, a new algorithm for qualitative simulation that generalizes the best features of rxisting algorithms, and allows direct comparisons among alternate approaches. QSIM is an efficient constraint-satisfaction algorithm that can follow either its standard semantics allowing the creation of new landmarks, or the {+, 0, -} semantics where 0 is the only landmark value, by changing a table of legal state-transitions. We argue that the QSIM semantics make more appropriate qualitative distinctions since the {+,0,-} semantics can collapse the distinction among increasing, stable, or decreasing oscillation. We also show that (a) qualitative simulation algorithms can be proved to produce every actual behavior of the mechanism being modeled, but (b) existing qualitative simulation algorithms, because of their local points of view, can predict spurious behaviors not produced by any mechanism satisfying the structural description. These observations suggest specific types of care that must be taken in designing applications of qualitative causal reasoning systems, and in constructing and validating a knowledge base of mechanism descriptions.

1 Introduction

An expert system is often a "shallow model" of its application domain, in the sense that conclusions are drawn directly from observable features of the presented situation. Many researchers believe that genuinely expert performance must also rest on knowledge of "deep models," in which an underlying mechanism, whose state variables may be not be directly observable, accounts for the observable facts [11].

One major line of research toward the representation of deep models is the study of qualitative causal models [1-19]. Research on qualitative causal models differs from more general work on deep models in focusing on qualitative descriptions of the deep mechanism, capable of representing incomplete knowledge of the structure and behavior of the mechanism. Symbolic manipulation of qualitative descriptions also corresponds well with explanations by human experts [16,17].

This research was supported in pari by the National Library of Medicine through NIH Grants LM 03003, LM 04125, and R01 LM 04374, and by the National Science Foundation through grants MCS-8303640 and DCR 8417034.

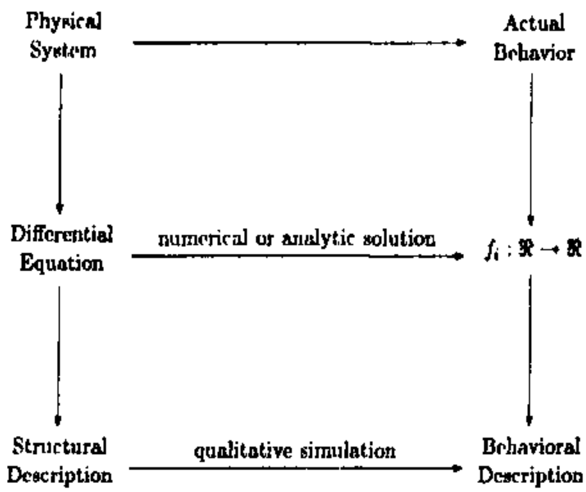

Figure 1: Qualitative simulation and differential equations are both abstractions of actual behavior.

A central inference within this approach is qualitative simulation: derivation of a description of the behavior of a mechanism from a qualitative description of its structure. Differential equations provide a useful comparison. A differential equation describes a physical system in terms of a set of state variables and constraints. The solution to the equation may be a function representing the behavior of the system over time. The qualitative structural description is a further abstraction of the same system, and qualitative simulation is intended to yield a corresponding abstraction of its behavior (figure 1).

All qualitative simulation systems predict multiple possible behaviors given certain structural descriptions and initial conditions. Researchers in this area (myself included) have hoped to prove that the predicted behaviors include all and only the possible behaviors of real mechanisms satisfying the given description. Half of this is correct: we can prove [15] that qualitative simulation cannot miss any actual behavior. However, because of the local nature of its decision criteria, qualitative simulation con predict behaviors that are not possible for any real mechanism satisfying the given description, and we construct a counterexample. We discuss the implications of these results for the construction of a qualitative causal reasoning system.

The QSIM algorithm has been implemented in Lisp on the Symbolics 3G00, and all examples in this paper have been run, as well as numerous others in elementary physics, nephrology, and cardiology.

1.1 Overview

This section provides an overview of qualitative simulation and the QSIM algorithm. Some of these concepts are defined more formally below. The complete formal treatment is presented in [15]. Qualitative simulation of a system starts with a description of the known structure of the system and its initial state, and produces a tree consisting of the possible future states of the system. The possible behaviors of the system are the paths from the root of this tree to its leaves.

The structural description consists of a set of symbols representing the physical parameters of the system (continuously differeiitiahle real-valued functions of time), and a set of constraints on how those parameters may be related to each other. The constraints are two- or three-place relations on physical parameters. Some specify familiar mathematical relationships: *DERIV(vel,acc)*. *ADD*(net, *out*, *in)*. *MULT (mass, acc, force), MINUS(fwd, rev)* Others assert qualitatively that there is a functional relationship between two physical parameters, but only specify that the relationship is monotonically increasing or decreasing: *M ' (price, power)* and *M (mph.mpg).* The constraints are designed to permit a large class of differential equations to be mapped straight-forwardly into structural descriptions.

Each physical parameter is a continuously differeiitiahle real-valued function of tnne with only finitely many critical points. Its value at any given point in time is specified qualitatively, in terms of its relationship with a totally ordered set of landmark values. Every critical value of the function is a landmark value. The landmark values may be described cither numerically (e.g. zero) or symbolically: their ordinal relationships are their essential properties. As the qualitative simulation proceeds, it can discover new critical points and thus add new landmark values to the sequence. The qualitative state of a parameter consists of its ordinal relations with the landmark values and its direction of change.

Time, within one possible behavior, is represented as a totally ordered set of symbolic distinguished time-points. The current time is either at or between distinguished time-points. All of the time-points are generated as a result of the qualitative simulation process.

At a distinguished time-point, if several physical parameters linked by a single constraint are equal to landmark values, they are said to have corresponding values which can be represented and used by the qualitative simulation. Corresponding values provide additional qualitative constraints on the behavior of structural relationships otherwise described only as M^+ or *M-.* The case of corresponding values (0,0) is sufficiently common to justify the special notation, M^+_0 and M_0~. When zero is the only landmark, useful corresponding values are seldom discovered.

A set of constraints on the physical parameters of the system is only valid in some operating region, denned by the legal ranges of values that some parameters may take on. The legal range of a parameter is a closed interval whose endpoints are landmark values of that parameter. These endpoints may be associated with transitions to other operating regions where a different set of constraints apply. The operating regions are designed as an interface to Forbus' [8,10] concept of *processes,* but that topic is beyond the scope of this paper.

The initial state of the system is defined by the operating region and a set of qualitative values for the physical parameters The qualitative simulation proceeds by determining all of the possible changes in qualitative value permitted to each parameter, then filtering the combinations by applying progressively broader constraints. If more than one qualitative change is possible, the current state has multiple successors, and the simulation produces a tree.

Two qualitative states in the same operating region are identical if all parameters are equal to the same landmark values, and all the directions of change are the same. If a state is identical to a direct predecessor, a cyclic behavior can be recognized.

2 Qualitative Behavior

In this* section, we present a more rigorous definition of qualitative description of continuous behavior, leading up to the QSIM algorithm and several theorems characterizing its strengths and limitations.

We reluctantly contribute to the proliferation of notations for qualitative description of continuous functions. The advantages of the notation used here are that it (1) naturally allows for an arbitrary and changing set of landmark values, (2) uses a single term for the qualitative description of a function's magnitude and derivative, and (3) emphasizes that the qualitative description of the derivative is of low and fixed resolution, while qualitative description of magnitude is of higher and possibly changing resolution.

Definition 1 . *Let $f : [a,b] \to \Re^*$ be a continuously differentiable function with only finitely many critical points in $[a,b]$. The* **landmark values** *of f are a finite set $l_1 < \cdots < l_k$ of values in the range of f including all its critical values. The landmark values always include $\{-\infty, 0, \infty\}$. The* **distinguished time-points** *of f are a finite set of $t_i \in [a,b]$ including all points where f takes on a landmark value.*

Definition 2 . *Let $l_1 < \cdots < l_k$ be the landmark values of $f : [a,b] \to \Re^*$. For any $t \in [a,b]$, $QS(f,t)$, the* **qualitative state of f at t**, *is a pair $\langle qval, qdir \rangle$, defined as follows:*

$$qval = \begin{cases} l_j & \text{if } f(t) = l_j, \text{ a landmark value} \\ (l_j, l_{j+1}) & \text{if } f(t) \in (l_j, l_{j+1}) \end{cases}$$

$$qdir = \begin{cases} inc & \text{if } f'(t) > 0 \\ std & \text{if } f'(t) = 0 \\ dec & \text{if } f'(t) < 0. \end{cases}$$

Definition 3 . *For adjacent distinguished time-points t_i and t_{i+1}, define $QS(f, t_i, t_{i+1})$, the* **qualitative state of f on (t_i, t_{i+1})**, *to be $QS(f,t)$ for any $t \in (t_i, t_{i+1})$.*

The Intermediate Value and Mean Value Theorems allow us to show that this definition is well-founded.

Definition 4 The qualitative behavior of f on [a, b] is the sequence of qualitative states of f:

$$QS(f,t_0), QS(f,t_0,t_1), QS(f,t_1) \dots QS(f,t_{n-1},t_n), QS(f,t_n)$$

an alternating sequence of qualitative states at distinguished time-points, and on intervals between distinguished time-points.

Definition 5 A system is a set $F = \{f_1 \dots f_m\}$ of functions f,: $[a,b] \rightarrow \Re^*$, each with its own set of landmarks and distinguished time-points. The distinguished time-points of a system F are the union of the distinguished time-points of the individual functions $f_i \in F$. The qualitative state of a system F of m functions is the m-tuple of individual qualitative states:

$$QS(F,t_i) = [QS(f_1,t_i), \dots QS(f_m,t_i)]$$

$$QS(F,t_i,t_{i+1}) = [QS(f_1,t_i,t_{i+1}), \dots QS(f_m,t_i,t_{i+1})]$$

The qualitative behavior of F is the sequence of qualitative states of F:

$$QS(F,t_0), QS(F,t_0,t_1), QS(F,t_1), \dots QS(F,t_n).$$

If t_t and/or $t_{t+}i$ are not distinguished time-points of a particular f_v then t_x and the interval (t_i, t_{i+1}) must be between two distinguished time-points of f_j, say t_k and t_{k+1}. Then $QS(f_j,t_i)$ and $QS(f_j,t_i,t_{i+1})$ are defined to be the same as the containing $QS(f_j,t_k,t_{k+1})$.

Table 1: The possible transitions

A continuously differentiable function $f : [a,b] \rightarrow \Re^*$ is restricted to the following set of possible transitions from one qualitative state to the next.

P-Transitions

Name	$QS(f,t_i)$	$\Rightarrow$	$QS(f,t_i,t_{i+1})$
P1	$\langle l_j, std\rangle$		$\langle l_j, std\rangle$
P2	$\langle l_j, std\rangle$		$\langle (l_j,l_{j+1}), inc\rangle$
P3	$\langle l_j, std\rangle$		$\langle (l_{j-1},l_j), dec\rangle$
P4	$\langle l_j, inc\rangle$		$\langle (l_j,l_{j+1}), inc\rangle$
P5	$\langle (l_j,l_{j+1}), inc\rangle$		$\langle (l_j,l_{j+1}), inc\rangle$
P6	$\langle l_j, dec\rangle$		$\langle (l_{j-1},l_j), dec\rangle$
P7	$\langle (l_j,l_{j+1}), dec\rangle$		$\langle (l_j,l_{j+1}), dec\rangle$

I-Transitions

Name	$QS(f,t_i,t_{i+1})$	$\Rightarrow$	$QS(f,t_{i+1})$
I1	$\langle l_j, std\rangle$		$\langle l_j, std\rangle$
I2	$\langle (l_j,l_{j+1}), inc\rangle$		$\langle l_{j+1}, std\rangle$
I3	$\langle (l_j,l_{j+1}), inc\rangle$		$\langle l_{j+1}, inc\rangle$
I4	$\langle (l_j,l_{j+1}), inc\rangle$		$\langle (l_j,l_{j+1}), inc\rangle$
I5	$\langle (l_j,l_{j+1}), dec\rangle$		$\langle l_j, std\rangle$
I6	$\langle (l_j,l_{j+1}), dec\rangle$		$\langle l_j, dec\rangle$
I7	$\langle (l_j,l_{j+1}), dec\rangle$		$\langle (l_j,l_{j+1}), dec\rangle$
I8	$\langle (l_j,l_{j+1}), inc\rangle$		$\langle l^*, std\rangle$
I9	$\langle (l_j,l_{j+1}), dec\rangle$		$\langle l^*, std\rangle$

In cases 18 and 19, / becomes std at l^*, a new landmark value such that $l_j < l^* < l_{j+1}$. In these cases, a previously unknown landmark value is discovered because other constraints force f'(t) to become zero.

These definitions give us a precise semantics for the qualitative description of continuous functions, and clarifies the concept of the "next state." Every state has a qualitative description QS(FJ), but that description changes only at discrete distinguished time-points, and remains constant on the open intervals between them. Thus the "next state" of a mechanism is more properly called the next distinct qualitative state description of the mechanism.

2.1 Qualitative State Transitions

The Intermediate Value Theorem and the Mean Value Theorem restrict the way a continuously differentiable function can change from one qualitative state to the next. There are two types of qualitative state transitions: P-transitions, moving from a time-point to a time-interval, and I-transitions, moving from an interval to a point. Table 1 specifies the set of possible transitions that can take place in the qualitative behavior of a single function. The table takes into account the possibility that not all landmark values are currently known, so that a new critical point might be discovered (18 and 19),

3 Qualitative Simulation

The qualitative simulation algorithm determines the possible qualitative behavior descriptions consistent with the initial state and the structural constraints. It is given the following structural description of a mechanism.

1. A set $\{f_1 \dots f_m\}$ of symbols representing the physical parameters in the system.
2. A set of constraints applied to the parameter symbols: $ADD(f,g,h)$, $MULT(f,g,h)$, $MINUS(f,g)$, $DERIV(f,g)$, $M^+(f,g)$, and $M^-(f,g)$. Each constraint may have associated corresponding values for its parameters.
3. Each parameter is associated with a totally ordered set of symbols representing landmark values. Each parameter may have upper and lower range limits, which are landmark values beyond which the current set of constraints no longer apply. A range limit may be associated with a new operating region which has its own constraints and range limits.
4. An initial time-point symbol, to, and qualitative values for each of the /, at to-

The result of the qualitative simulation is one or more qualitative behavior descriptions for the given parameters. Each qualitative behavior description consists of the following:

1. A sequence $\{io \dots t_n\}$ of symbols representing the distinguished time-points of the system's behavior.
2. For each parameter /,, a totally ordered set of landmark values, possibly extending the originally given set.
3. For each parameter, at each distinguished time-point or interval between adjacent time-points, a qualitative state description expressed in terms of the landmark values of that parameter.

3.1 The QSIM Algorithm.

The qualitative simulation algorithm. QSIM, repeatedly takes an active state and generates all possible successor states, filtering out states that violate a consistency criterion at one of several levels: individual parameter, individual constraint, pairwise constraint, or global state. Because the next state may not be determined uniquely, QSIM builds a tree of states representing the possible behaviors of the mechanism.

Place the initial state on the list ACTIVE, of states whose successors need to be determined. Repeat the following steps until ACTIVE becomes empty or a resource limit is exceeded.

1. Select a qualitative state from ACTIVE.
2. For each parameter in the structural description, determine (from Table 1) the set of transitions possible from the current qualitative state.
3. For each constraint, aggregate the transitions associated with its arguments into 2-tuples and 3-tuples. The tuples can then be checked for consistency according to two criteria local to individual constraints.
   - The tuple of directions of change must be consistent with the constraint in the state resulting from the transition.
   - The result of the transition-tuple can be compared with corresponding values of the arguments to that constraint. Reference [15] demonstrates an efficient and verifiable method for this test, generalizing the Transition Ordering rules of Williams [18,19].
4. Perform pair-wise consistency (Waltz) filtering on the sets of tuples associated with the constraints in the system, applying the consistency criterion that adjacent constraints must agree on the transition assigned to the shared parameter. (The fact that we filter on transitions rather than states considerably simplifies the algorithm.)
5. Generate all possible global interpretations from the remaining tuples by depth-first traversal of the set of assignments of transition tuples to constraints. If there are none, mark the behavior as inconsistent. Create new qualitative states from each interpretation, and make them successors of the current state.
6. Apply global filtering rules to the new qualitative states, and place any remaining states on ACTIVE.

3.2 Global Filters

The completed qualitative state descriptions are mathematically plausible successors to the current state. However, several global filters arc applied (step 6 above) before a new state is added to ACTIVE. The mathematically valid filters applied at this stage are the following.

- (No Change.) Delete the new state if all transitions are in the set {11,14,/7}, because the new state description would be identical to its immediate predecessor, which therefore already captures its qualitative behavior. In other words, *something* must reach a limit point for an I-transition to take place.
- (Cycle.) If the new state is identical to one of its predecessors (all parameters have identical *landmark* values, and all directions of change are the same), then mark the behavior as cyclic, install a pointer to the identical predecessor, and do not add the new state to ACTIVE.
- (Divergence.) If any parameter takes on the value oo or - oo, the current time-point must be the endpoint of the domain, so the new state does not go onto ACTIVE.

The first filter does not reduce the number of behaviors described, but only eliminates a redundant description. The second detects when all the consequences of a particular state have already been determined, and need not be explored anew. The third determines when a state must be at the endpoint of the domain, and thus can have no successors.

The *pure* QSIM algorithm includes only these mathematically valid filters. For a particular application, additional heuristic filters may be added. Some possible heuristics include:

- (Quiescence.) If all parameters have derivative zero, conclude that the system is quiescent, the new time-point is the endpoint of the domain (possibly t = oo), and do not place the new state on ACTIVE.
- (No Divergence.) In physical systems, eliminate transitions in which any parameter goes to oo or -oo. A more accurate description of the system would include an operating region change corresponding to some component breaking.

3.3 Complexity

Suppose there are n parameters in the system, m constraints, and the longest behavior has length t. All steps except generating the global interpretations are linear in the number of constraints and/or parameters, In the worst cast, generating the global interpretations can be exponential, but in practice generating the successors of a given state appears to be approximately *o(mt)*. On the Symbolics 3G00, the Spring example below (3 parameters, 3 constraints) takes about 0.4 seconds, and the Starling mechanism (16 parameters, 14 constraints) [10,17] takes about 1.0 second. Thus, it is computationally feasible to run several simulations in the course of solving a single problem.

3.4 QSIM on the Ball

This section illustrates the QSIM algorithm deriving one step in the behavior of a ball, thrown upward under constant gravity. Its constraints are:

$$DERIV(Y,V)$$
$$DERIV(V,A)$$
$$A(t) = g < 0$$

Consider the second state, in which the ball is moving upward toward an (as yet undiscovered) peak.

$$QS(A,t_0,t_1) = \langle g, std\rangle$$
$$QS(V,t_0,t_1) = \langle (0,\infty), dec\rangle$$
$$QS(Y,t_0,t_1) = \langle (0,\infty), inc\rangle$$

First, determine the possible transitions for each parameter: $QS(A,t_0,t_1) \Rightarrow QS(A,t_1)$

$$I1 \quad \langle g, std\rangle \Rightarrow \langle g, std\rangle$$

$QS(V,t_0,t_1) \Rightarrow QS(V,t_1)$

$$I5 \quad \langle (0,\infty), dec\rangle \Rightarrow \langle 0, std\rangle$$
$$I6 \quad \langle (0,\infty), dec\rangle \Rightarrow \langle 0, dec\rangle$$
$$I7 \quad \langle (0,\infty), dec\rangle \Rightarrow \langle (0,\infty), dec\rangle$$
$$I9 \quad \langle (0,\infty), dec\rangle \Rightarrow \langle L^*, std\rangle$$

$QS(Y,t_0,t_1) \Rightarrow QS(Y,t_1)$

$$I4 \quad \langle (0,\infty), inc\rangle \Rightarrow \langle (0,\infty), inc\rangle$$
$$I8 \quad \langle (0,\infty), inc\rangle \Rightarrow \langle L^*, std\rangle$$

Next, each constraint forms a set of transition tuples. Filtering for consistency with the individual constraints eliminates those tuples marked with *c* below. Pairwise consistency filtering then eliminates those marked with *w*.

$DERIV(Y,V)$		$DERIV(V,A)$	
(I4, I5)	c	(I5, I1)	c
(I4, I6)	c	(I6, I1)	
(I4, I7)		(I7, I1)	
(I4, I9)	w	(I9, I1)	c
(I8, I5)	w		
(I8, I6)			
(I8, I7)	c		
(I8, I9)	c		

The remaining tuples can be formed into only two global interpretations:

Y	V	A
I4	I7	I1
I8	I6	I1

The first of these interpretations yields a qualitative state description identical to the preceding state, so the No Change filter applies. The only remaining possibility then specifies the unique successor state:

$$QS(A,t_1) = \langle g, std\rangle$$
$$QS(V,t_1) = \langle 0, dec\rangle$$
$$QS(Y,t_1) = \langle Y_{new}, std\rangle.$$

The new landmark $0 < Y_{new} < \infty$ has been discovered.

4 Should Simulation Create Landmarks?

The most important semantic difference between QSIM and other approaches to qualitative simulation is that QSIM ran create new landmark values during the simulation, while the other algorithms require all landmarks to be specified when the structure is defined. The inability to create new landmark values makes it impossible to express certain important qualitative distinctions, such as that between increasing, decreasing, and stable oscillation. The fixed landmark assumption is particularly deeply embedded in de Kleer's approach, which depends on arithmetic operators defined over the fixed set of *qualitative values,,* {+,0,-}, produced by the single landmark 0. A change in landmarks would change the set of qualitative values, and thus require the operators to be redefined. Such a redefinition is not always possible.

Within QSIM, it is possible to experiment with the {+,0,-} semantics for qualitative simulation simply by replacing Table 1 with an alternate table of legal transitions.

The Bouncing-Ball system is described by the following constraints:

$$DERIV(Y,V)$$
$$DERIV(V,G)$$
$$G = G^* < 0$$
$$Y \geq 0$$

with an instantaneous bounce simulated by an operating region transition at which, if $Y = 0$ with negative velocity, the sign of the velocity is inverted. If the qualitative description does not specify the relation between the magnitudes of the velocity before and after the bounce, there are three possible behaviors: the second bounce could be higher, lower, or the same as the first.

Figures 2(a), 2(b). and 2(c) show the way the standard QSIM semantics expresses the three possibilities. In these *qualitative plots,* a point is plotted at, or halfway between, two landmark values on the vertical axis and two distinguished time-points on the horizontal axis. At the peak of the first bounce, *Y* has a critical point, so a landmark value, Y1, is created. After the bounce, both *Y* and *V* move toward limiting values. The second peak is higher, lower, or equal to the first according to whether *Y* reaches *Y1* before, after, or at the same time as *V* reaches 0. QSIM gives distinct representations to the three possible behaviors. If further information is available to exclude one or two alternatives, the significance of the remaining ones is clear.

Figure 2(d) shows the same mechanism simulated using {+,0,-} semantics. There is only a single behavior since the peak value is represented by the qualitative value +, which corresponds to the entire interval (0,oo). The description captures the repeated up-and-down motion of the bouncing ball, but fails to make the important qualitative distinction between a higher, lower, or equal bounce. Thus three qualitatively distinct behaviors are collapsed into a single description.

Similarly, De Klcer and Bobroar [4] present an example of a spring with frictional damping, wliose actual behavior is a decreasing oscillation. Because the maximum amplitude of the oscillation is represented by the qualitative interval +, the behavioral description derived is cyclic. The cyclic description accurately captures the repetitive scries of increase and decrease in the different parameters. However, without being able to assign a symbolic *name* to the critical values, it does not express the distinction between increasing, decreasing and steady amplitude, and so cannot even ask which qualitative behavior is correct.

A related problem is that de Kleer and Bobrow [4] recognize the cycle by matching the *qualitative descriptions* of states. However, if some parameters are in the intervals between landmarks, then apparently matching states may not be identical, leading decreasing oscillation to be taken as a cycle.

The heart of both problems is the inability to create new landmarks, or equivalently, to give *names* to newly discovered critical values. Without representing the initial value (and subsequent critical values) of a parameter in a way that permits ordinal comparison, it is not possible to ask whether the next repetition of a cycle leaves that parameter increased, decreased, or stable. Thus, we argue that the {+,0, -} semantics, and in fact any semantics with a fixed set of landmarks, can collapse importantly distinct behaviors. The QSIM semantics, by providing for discovery and naming of new landmarks, allows more appropriate qualitative distinctions to be made.

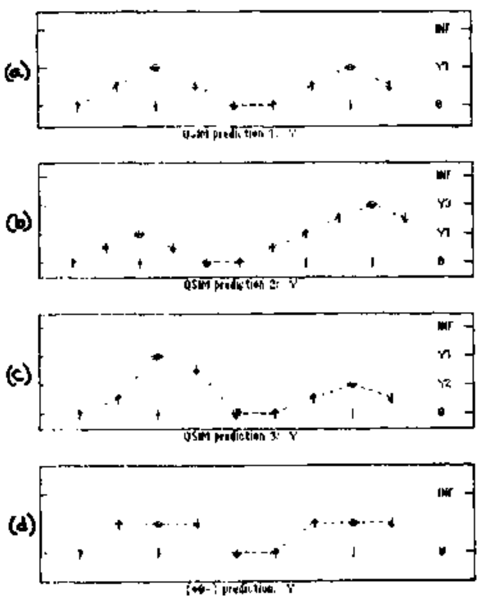

Figure 2: Simulation of the bouncing ball under the standard QSIM semantics distinguishes three possible behaviors which are collapsed into a single description under the {+,0, -} semantics. This figure includes only the qualitative plots for *Y*.

(a) QSIM prediction 1: the second bounce is the same height as the first.

(b) QSIM prediction 2: the second bounce is higher than the first.

(c) QSIM prediction 3: the second bounce is lower than the first.

(d) {+,0, -} prediction: the three behaviors are collapsed into a single description.

5 Does Qualitative Simulation Find the Real Behaviors?

Ideally, qualitative simulation will find *all and only* the actual behaviors of a mechanism being simulated. We take as our "gold standard[1]" the solutions to the ordinary differential equation describing the mechanism. A qualitative structure description is less restrictive than a differential equation, so we can expect multiple behaviors produced by the simulation, varying according to factors not captured in the structure description. (For example, if I throw a ball upward with a velocity described only as "positive", it might hit the ceiling, or it might not.) First, we want to know that all behaviors that are actually possible arc found by the qualitative simulation. Second, we want to know that every possibility predicted by the qualitative simulation can actually happen.

In QSIM, care has been taken so that the structural description of a mechanism is an abstraction *of* its differential equation. The algorithm generates the space of all possible next states of the system given its current state (Table 1), and each filtering step removes only states which are internally inconsistent. Thus, we prove in detail in [15],

Theorem 1. Each actual behavior of the system is necessarily among those produced by the simulation.

One of the attractive applications of qualitative simulation is to predict possible future states, particularly to warn of surprising or disastrous events. Although we can trust qualitative simulation to produce every real behavior, the converse is not true: some predictions can be spurious. It is possible for the QSIM algorithm, and local qualitative simulation algorithms in general, to produce behaviors which are not actual behaviors for *any* physical system satisfying the structural description.

Consider a mass on a spring, oscillating on a frictionless Burface. The qualitative structural description of this system is

$$\begin{array}{l} DERIV(X,V) \\ DERIV(V,A) \\ M_0^-(A,X), \end{array} \qquad (1)$$

which might also be written in the form of a second-order differential equation:

$$\frac{d^2X}{dt^2} = -M_0^+(X). \qquad (2)$$

With initial state $X(t_0) = 0$, $V(t_0) = V_{init} > 0$, $A(t_0) = 0$, this system is periodic for any function $A = -M_0^+(X)$, because if we define total energy as

$$TE(x,v) = \int_0^x M_0^+(y)dy + \frac{1}{2}v^2,$$

then equation (2) implies that $\frac{d}{dt}TE = 0$.

Starting with *X,* V, and *A* equal to 0, V_{init} and 0, respectively, QSIM proceeds straight-forwardly through most of the cycle, predicting a unique successor to each state. Finally, all three parameters approach their initial values. *X* and *A* must reach zero together, but there is not enough information to determine whether *V* reaches its limit earlier, later, or at the same time.

Figure 3: QSIM predicts three behaviors for the Spring: one valid and two spurious. The three columns represent the three behaviors. Look particularly at the qualitative plots for V (the middle row).

1. V and X reach their limits at the same time (the only real possibility).
2. X reaches 0 before V reaches V_{init}: decreasing oscillation (impossible).
3. V reaches V_{init} before X reaches 0: increasing oscillation (impossible).

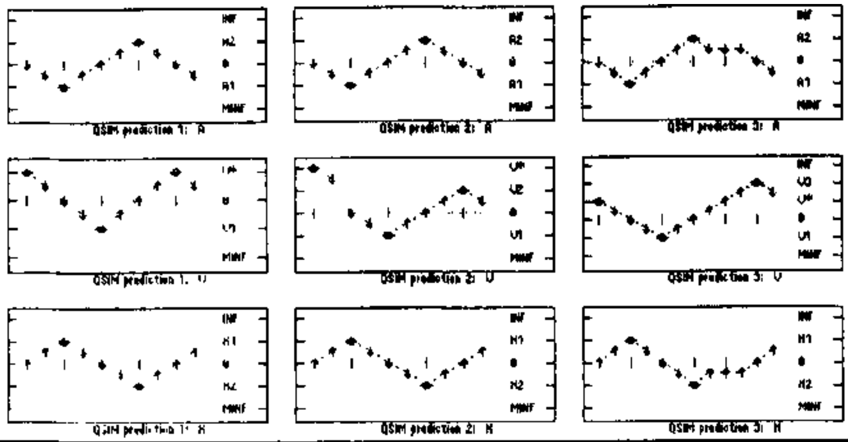

Figure 4: $\{+,0,-\}$ semantics also predicts three behaviors for the Spring: one valid and two spurious. However, we must simulate the initial landmark value of V by defining $W(t) = V(t) - V_{init}$. Compare the qualitative plots for W (the third row) across the three behaviors to distinguish whether V passes its initial value.

1. W and X reach zero at the same time (the only real possibility).
2. X reaches 0 before W does: decreasing oscillation (impossible).
3. W reaches 0 before X does: increasing oscillation (impossible).

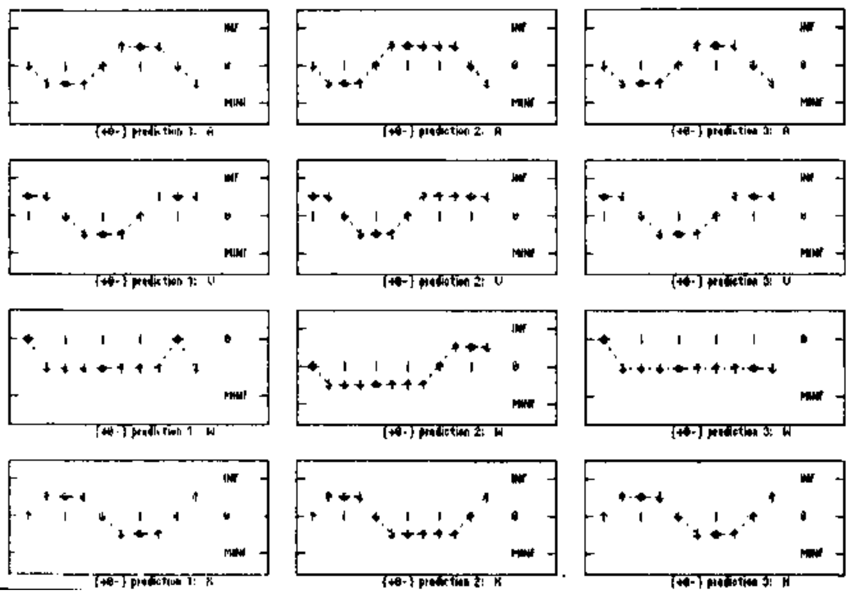

Figure 3 shows how QS1M branches to yield three behaviors corresponding to increasing, decreasing, and stable oscillation. Figure 4 shows how, using a translated variable W to simulate the non-zero landmark value V_{init}, the same indeterminacy exists under the {+,0, -} semantics.

As we have seen, only the stable periodic behavior is an actual behavior possible for this structural description. Thus,

Theorem 2. There are behaviors predicted by qualitative simulation which do not correspond to the behavior of any system satisfying the qualitative structure description.

The fundamental problem is that simulation, qualitative or quantitative, is inherently local: the successors to the current state arc computed given only the information in the current state. Given a descriptive framework consisting of the functions and constraints describing the mechanism, and the states to be linked, there is simply not enough information available to eliminate all spurious behaviors.

An incorrect solution to this problem is to employ a qualitative description so coarse that the alternate behaviors are simply collapsed into a single one. Figure 5 shows the single behavior predicted for the Spring under the {+,0, -} semantics without an initial landmark value. Precisely like the bouncing ball discussed in the previous section, the three alternative behaviors are collapsed into a single description, so the problem is not solved, simply inexpressible.

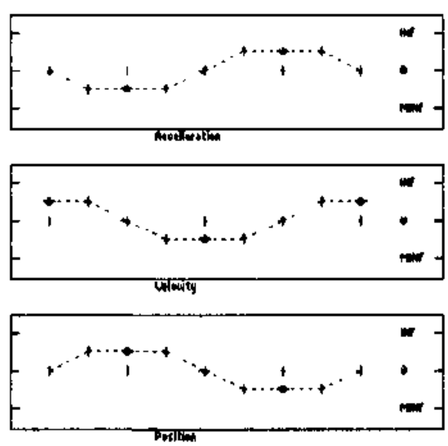

Figure 5: The stable, increasing, and decreasing oscillatory behaviors of the Spring are collapsed into one by the {+,0, -} semantics.

Changing the problem to be solved [de Kleer, personal communication] can sometimes avoid this difficulty in familiar cases. By changing the Spring description to take into account the conservation of total energy, an expanded view allows QSIM to determine that there is a single, periodic behavior (Figure 6). A physicist can look at equation (2) and recognize or derive the fact that it represents an energy conserving system, and therefore that the behavior must be periodic. Part of this knowledge is the ability to recognize the physical system described by a structure, and to know that there is a better structural description for it. However, this means that external information is needed to set up the correct problem and reach a useful conclusion in such a case.

These observations yield some important warnings about the proper use of qualitative descriptions of mechanisms, and the result of their simulation.

- The two theorems above have a corollary that highlights their implications for knowledge engineering.
  Corollary. If a structural description is consistent, and if QSIM predicts a single behavior, then that behavior represents the actual behavior of the mechanism.
- The structural description must be shown to be consistent, perhaps by demonstrating that it is an abstraction of an accurate quantitative description, to guarantee that the qualitative simulation will include a genuine behavior.
- If qualitative simulation yields several possible behaviors, further analysis is required before concluding that each represents a possible future.

In developing a knowledge base of kidney mechanisms, our experience suggests that the most useful knowledge base consists of a collection of first-order views, focusing on a small portion of the overall mechanism. Each simulation yields a single behavior, which we therefore know to be correct (modulo the assumptions behind the first-order model).

Thus, much as backward-chaining search provides the mathematical underpinnings for search in a rule-based expert system, so qualitative simidation may provide the mathematical underpinnings for theory evaluation in a causal reasoning system. If the underpinnings are solid and well-understood, then the semantics of the knowledge base is clear, a necessary condition for the creation of powerful reasoning systems.

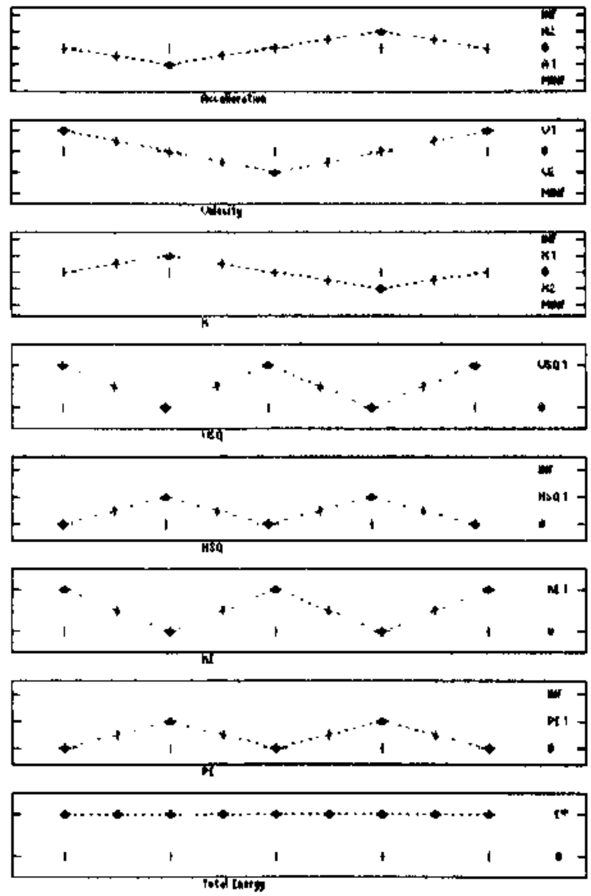

Figure 6: Adding conservation of energy

$$(TE = KE + PE = M_0^+(V^2) + M_0^+(X^2))$$

yields a single correct behavior, but requires domain-specific problem-formulation expertise.

Acknowledgements

The environment of the MIT LCS Clinical Decision Making Group provides valuable support and stimulation for this research. Christopher Eliot is responsible for the implementation of the QSIM algorithm with a very comfortable interface, and for a number of valuable technical contributions. I am also grateful for helpful discussions and observations on qualitative simulation in general and QSIM in particular from Daniel Bobrow, Johan de Kleer, Ken Forbus, Walter Hamscher, Ramesh Patil, Elisha Sacks, Mike Wellman, and Brian Williams.

Bibliography

[1] J. dc Kleer. 1977. Multiple representations of knowledge in a mechanics problem-solver. Proceedings of the Fifth International Joint Conference on Artificial Intelligence.

[2] J. de Kleer. 1978. Causal and teleological reasoning in circuit recognition. MIT Artificial Intelligence Laboratory TR-529.

[3] J. de Kleer. 1979. The origin and resolution of ambiguities in causal arguments. Proceedings of the Sixth International Joint Conference on Artificial Intelligence. Tokyo, Japan.

[4] J. de Kleer and D. G. Bobrow. 1984. Qualitative reasoning with higher-order derivatives. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-8J).* Los Altos, CA: William Kaufman.

[5] J. de Kleer and J. S. Brown. 1983. The origin, form, and logic of qualitative physical laws. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence.* Los Altos, CA: William Kaufman.

[6] J. de Kleer and J. S. Brown. 1984. A qualitative physics based on confluences. *Artificial Intelligence* 24: 7 - 83.

[7] K. D. Forbus. 1981. Qualitative reasoning about physical processes. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence.* Los Altos, CA: William Kaufman.

[8] K. D. Forbus. 1983. Measurement interpretation in qualitative process theory. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence.* Los Altos, CA: William Kaufman.

[9] K. D. Forbus. 1984. Qualitative process theory. *Artificial Intelligence* 24: 85 - 168.

[10] K. D. Forbus. 1984. Qualitative process theory. Ph.D. thesis, MIT Department of Electrical Engineering and Computer Science.

[11] D. Centner and A. Stevens, (Eds.). 1983. *Mental Models.* Hillsdale, NJ: Erlbaum.

[12] P. J. Hayes. 1979. The naive physics manifesto. In D. Michie (Ed.), *Expert Systems in the Micro-electronic Age.* Edinburgh: Edinburgh University Press, 1979.

[13] B. J. Kuipers. 1982. Getting the envisionment right. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-82).* Pittsburgh, Pennsylvania, August 1982.

[14] B. J. Kuipers. 1984. Commonsense reasoning about causality: deriving behavior from structure. *Artificial Intelligence* 24: 169 - 204.

[15] B. J. Kuipers. 1985. Qualitative Simulation of Mechanisms. Cambridge, MA: MIT Laboratory for Computer Science TM-274.

[16] B. J. Kuipers and ,1. P. Kassirer. 1983. How to discover a knowledge representation for causal reasoning by studying an expert physician. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence.* Palo Alto: William Kaufman.

[17] B. J. Kuipers and J. P. Kassirer. 1984. Causal reasoning in medicine: analysis of a protocol. *Cognitive Science* 8: 363-385.

[18] B. Williams. 1984a. The use of continuity in a qualitative physics. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-84)-* Los Altos, CA: William Kaufman.

[19] B. Williams. 1984b. Qualitative analysis of MOS circuits. *Artificial Intelligence* 24: 281 - 346.