Axiomatizing Qualitative Process Theory

Ernest Davis*
Courant Institute
New York, New York

March 17, 1994

Abstract

We show that the type of reasoning performed by Forbus' [1985] Qualitative Process (QP) program can be justified in a first-order theory that models time and other measure spaces as real-valued quantities. We consider the QP analysis of a can of water with a safety valve being heated over a flame. We exhibit a first-order theory for the microworld involved in this example, and we prove the correctness of the first two transitions in the envisionment graph. We discuss the possibility of deriving the closure conditions in the theory via non-monotonic inference.

One way to increase confidence in a reasoning program is to show that the conclusions that it draws correspond to valid inferences within some easily intelligible logical theory. Such a correspondence has been shown for many of the best known physical reasoning programs. The calculations performed by QSIM [Kuipers, 86] correspond to theorems in real analysis, under a natural interpretation; [Duchier, 91] exhibits full first-order proofs of these. Likewise the reasoning in NEWTON [de Kleer, 77] and ENVISION [de Kleer and Brown, 85] can be shown to be valid for a simple physical theory, easily formalized in first-order logic, in which time and other physical parameters are viewed as real-valued quantities. (See, for example, [Rayner, 91], [Davis, 90, chap. 7].)

However, no adequate logical analysis has hitherto been given for the Qualitative Process (QP) program [Forbus, 85]. In QP, processes can come into and out of existence, and the topological structure of the physical system may change over time. Hence, the problem of finding an appropriate formulation of the necessary closed-world assumptions on processes and influences seemed daunting. In a previous analysis [Davis, 90] I was unable to find a reasonable characterization of these closure principles, and hence left them as intuitively plausible, but wholly unformalized, non-monotonic deductions.

In this paper, I show that this gap can be closed. It is possible to characterize inference in QP entirely in terms of a monotonic theory based on real analysis. There are two key points:

- For each parameter, the theory must give an exhaustive enumeration of the types of processes and parameters that can influence it. These axioms resemble to the kind of frame axioms advocated by Schubert [1991], which give necessary conditions for a fluent to change its value. They are also analogous to the circumscription over causes of change discussed in [Lifschitz, 87].
- Since QP uses only qualitative information as to the direction of change and influence, it is possible to combine influences using only existential criteria. A parameter may change in

*This research was supported by NSF grant #IRI-9001447.

> some direction if some influence is pushing it in that direction; it must change in a direction if there is some influence pushing it in that direction, and there is no influence pushing it in the opposite direction. These existential criterion means that it is not necessary to individuate different influences or to sum over different influences, which simplifies the theory.

Though QP theory centers on continuous parameters, these theories can be extended to include discrete change as well, as we shall show below.

At this point, let us briefly discuss what it is that we are axiomatizing. There is essentially no knowledge of physics built into QP. Rather, the QP representation gives a language in which (certain) physical theories can be expressed and associated physical situations can be described; and the QP algorithm uses the information to predict physical behavior over time. The user of QP must input both the specific scenario of the problem that interests him and also the physical theory to be used. Thus, the axioms that are common across all uses of QP are limited: they include the axioms of real analysis, some basic axioms of temporal reasoning, and a few general axioms constraining the possible behavior of a physical parameter, and relating it to the influences on it. The other physical knowledge needed is not associated with QP as it comes from the factory, so to speak; it is part of the user specification.

Thus, "axiomatizing QP theory" consists largely of showing that user specifications of physical domains can be translated into systems of axioms. One way of showing this would be to define formally how QP representations correspond to axioms; essentially, to specify a procedure for automatically translating QP representations into logical axioms. In fact, if we wished to prove formal properties of QP, such as soundness, we would be obliged to define such a translation.

But such a precise correspondence goes far beyond the purposes of this paper. Indeed for our purposes, it is almost immaterial whether such a translation is always possible or not,[1] and this paper does not discuss the specifics of the actual QP representation. Our object here is to show that physical theories like those expressible in QP can be expressed in simple physical axioms, and that the predictions like those made by QP can be justified as inferences. Whether there is a perfect correspondence between QP and axiomatic theories is relatively unimportant.

Therefore, the approach in this paper is to discuss the general form of a QP axiomatization, and then to give the specifics of a sample QP domain. The hope is that readers will be able to see how to generalize from this domain to other domains, without attempting to give an abstract description that would cover all domains.

Physical prediction programs vary in the degree to which they incorporate specific physical knowledge. ENVISION [de Kleer and Brown, 85] is like QP; it provides the user with a language (of components and connections) in which he can specify a physical theory. Thus, like QP, the axiomatic treatment of ENVISION consists of a demonstration that component specifications can be translated into physical axioms. Programs like NEWTON [de Kleer, 77], FROB [Forbus, 79] or CLOCK [Faltings, 87] do incorporate specific physical theories. Their axiomatic treatment consists of a specific set of physical axioms, together with a demonstration that user specifications of a particular scenario can be expressed as axioms. Since a scenario description can almost always be expressed in a collection of atomic ground formulas, supplemented with unique-names and closure axioms, this translation is much simpler than those of QP or ENVISION. QSIM [Kuipers, 86] is purely mathematics; it neither expresses nor incorporates any physical knowledge.

The remainder of this paper is organized as follows. Section 1 provides a high-level view of

[1] Translating from a representation intended for procedural use to a logical representation can be very tricky, even if the program is doing something "basically" deductive. For example, many such representations use negation as failure without worrying about it; such uses are often easy for procedures, but miserable to axiomatize. Some such gaps relate to clumsinesses in first-order logic; others relate to kludges in the program. My guess would be that there is at least a well-defined subset of QP for which a translation procedure could be defined.

the axiomatics. Section 2 deals with some fine points in defining certain properties of real-valued parameters. Sections 3 and 4 give a detailed analysis of a simple physical system combining continuous and discrete components: a boiling can of water with a safety valve. Section 3 presents a general language for QP theory and an axiomatization of the particular microworld used for this example. Section 4 specifies the particular scenario and shows that the predictions of QP theory can be justified in the logic. Section 5 discusses the application of non-monotonic logic to this theory. Section 6 discusses some features of the theory, and presents the conclusions.

1 Structure of the Theory

The ontology of QP follows familiar lines. The time line is taken to be isomorphic to the real line, with no branching. (Branching in envisionments corresponds to disjunctive uncertainty in prediction, rather than to actual branching in time.) The logic uses two kinds of temporal entities: *situations,* which are instants of time, and *time intervals,* which may be closed or open, bounded or unbounded.

Measure spaces other than time, such as temperature, mass, positional coordinate on some axis, and so on, are likewise taken to be isomorphic to the real line.

A *fluent* is a function from time to some range. A fluent with range {TRUE, FALSE} is called a *Boolean fluent* or *state*. A fluent from time to a measure space is called a *parameter*. If A is a state and S is a situation, then the predicate "holds(S, A)" means that A is TRUE in S. If F is a fluent other than a state, then the function "value_in(S, F)" gives the value of F in situation S. Alternatively, as a notational convenience, if term $\tau(\alpha_1 \ldots \alpha_k)$ denotes a fluent, we may add the situation as an additional argument, in the form $\tau(\alpha_1 \ldots \alpha_k, S)$. This will mean the same as either "holds($S, \tau(\alpha_1 \ldots \alpha_k)$)", if τ is a state, or as "value_in($S, \tau(\alpha_1 \ldots \alpha_k)$)", if τ is not a state. For example, we may say that Valve 1 is open in situation s0 either in the form "holds(s0, open(valve1))" or in the form "open(valve1,s0)".

A function or a predicate defined on a particular space may be extended in the natural way to take arguments that are fluents with range in that space. For example, if "square(X)" is a function mapping the reals to the reals, and F is a real-valued fluent, then "square(F)" is the fluent that, at any given instant gives the square of the value of F at that instant. If "$>$" is a predicate with two real valued arguments and $F1$ and $F2$ are real-valued fluents, then "$F1 > F2$" is the state that holds whenever the value of $F1$ is greater than the value of $F2$.

value_in(S,square(F)) = square(value_in(S, F)).
holds(S,$F1 > F2$) $\Leftrightarrow$ value_in($S, F1$) $>$ value_in($S, F2$).

Equality and inequality are exceptions to this. "$F1 = F2$" and "$F1 \neq F2$" are sentences, stating that $F1$ is the same fluent as $F2$, or $F1$ is a different fluent from $F2$, respectively. The state of the current value of $F1$ being equal to the current value of $F2$ is denoted "eql($F1, F2$)"; the state of the two values being different is denoted "neql($F1, F2$)".

A *process* is a particular category of state. For processes, we use the special predicate "active(S, P)" (process P is active in situation S); this is synonymous with "holds(S, P)" Besides processes, there are *events*, which occur over finite, non-point, intervals. We write "occur(I, E)" to mean that event E occurs over interval I. In this theory, we deal only with state, fluent, process, and event *types,* rather than tokens.[2]

Finally, there are physical objects. We use this term loosely to include practically any entity of physical interest that does not fall into the other categories. For example, in modelling water

[2] This differs from [Davis, 90] where a process was a state token.

flowing through a tank, one object could be a particular "piece" of water that comes in at one time and goes out at another; another object could be "the water in the tank", which has a mass that changes over time.

A axiomatic QP theory involving only continuous parameters contains axioms of the following forms:

1. **Process definitions.** Necessary conditions and sufficient conditions (they need not be the same) for a process of a given type to be active in a given situation.
2. **Direct influences.** For each parameter that is directly influenced, an exhaustive enumeration of the processes that influence it, with the directions of influence.
3. **Indirect influences.** For each parameter that is indirectly influenced, an exhaustive enumeration of the parameters that indirectly influence it, with the directions of influence.
4. **General axioms of influence.** Two axioms relating the behavior of a parameter to the influences on it:
   A. A parameter F can only change in direction G (up or down) if there is some influence on F pushing it in direction G.
   B. A parameter F must change in direction G if there are influences on F in direction G, and there are no influences on F in direction $-G$.
5. **Well-behavedness conditions.**
   A. Certain specified physical parameters are "well-behaved" functions of time.
   B. A well-behaved function is continuous.
   C. A well-behaved function does not asymptotically approach a value without attaining that value.
   D. At each instant, a well-behaved function is differentiable from the right and from the left. (See section 2.)
   E. States do not change infinitely often in finite time intervals. "States" here includes values of discrete fluents; order relations between parameters; and activity states of processes. (This condition will be discussed in detail in [Davis, in prep.])
6. **Unique names axioms.** Axioms specifying that objects, processes, and parameters with different names are unequal.
7. **Real analysis.** An axiomatic theory describing basic properties of the real numbers and of real-valued functions. In this paper, we will not spell out these axioms, which are well-known; rather, we will cite theorems from this theory *ad hoc* as needed.

If the theory contains discrete states that are changed by events, these are characterized by:

8. Necessary conditions and sufficient conditions for each discrete fluent to change its value.
9. Necessary conditions and sufficient conditions for each event to occur.

If the theory contains parameters that may change discontinuously, but are piecewise continuous, then [5A] above must be changed to

10. For each parameter, necessary conditions and sufficient conditions for the parameter to be discontinuous in a given situation.

An example of such a parameter is velocity in a theory of solid object dynamics with collisions. The use of axioms of this kind is discussed by Rayner [1991]. The example that we will discuss here does not include any discontinuous parameters.

There does not seem to be any physical need for parameters that are not piecewise continuous.

2 Finicky details about real-valued parameters

As often in using the real numbers as a basis for a physical theory, it is necessary to worry about fine details of small-scale topology to give precise and correct ontological definitions. Since this paper serves no purpose other than finicky precision, I need not apologize.

The problem is to define what it means for parameter to be "increasing," "decreasing," or "constant" at an instant of time. Most of the literature on qualitative reasoning uses the sign of the derivative of the parameter at the instant, which is perfectly fine as long as everything can be assumed to be everywhere differentiable. However, this assumption does not seem reasonable within all domains we would like to address in QP. Consider, for example, cutting a string supporting a weight at time $t = 0$. (Figure 1). The acceleration changes instantaneously (up to the precision of the model) from 0 to −g, so the velocity is not differentiable at $t = 0$.

How shall we characterize its behavior at $t = 0$? In fact, what we want to do depends on how we characterize the state of the string. If the string is whole for $t < 0$ and broken for $t \geq 0$, then we should say that the downward velocity is increasing at $t = 0$;, if the string is whole for $t \leq 0$ and broken for $t > 0$, then we should say that the velocity is constant at $t = 0$. How we want to characterize the string may in turn depend on considerations external to the QP analysis, such as the desired geometric theory.

One might be tempted, from this example, to refuse to deal with characterizing behavior at an instant, and demand that characterizations refer to open intervals. But that will hardly do. Very often, parameters are in a constant state only for an instant, such as a ball thrown in the air at the top of its path. Avoiding this would necessarily create a lot of clumsiness.

The solution we propose is as follows: Assume that every parameter is differentiable at every instant both from the right and from the left. We define the "true derivative" to be, disjunctively, either the derivative from the right or the derivative from the left. The disjunction allows the logic to "pick" whichever value will fit in better with the rest of the theory. "Increasing," "decreasing," and "constant" are then defined in terms of the sign of the "true" derivative. Thus, the velocity shown in figure 2 may be either increasing or constant at $t = 0$, whichever fits better with the rest of the world state. It cannot be decreasing, though.

Note that this means that there can be two parameters $F1$ and $F2$ that are always equal, but $F1$ is increasing at a time that $F2$ is decreasing. Thus, "increasing" and "decreasing" are properties of physical parameters, not of the associated functions of time.

3 The axiomatic theory

This section gives an axiomatic theory for the following example (Figure 3): A can of water is heated over a flame. The can has a safety valve with two states, open and closed. The valve opens when the pressure in the can exceeds a certain fixed pressure; it closes when the pressure drops below another (lower) fixed pressure. The processes we will model are the heat flow from the flame to the can, the heat flow from the can to its contents, the boiling of the water, and the flow of steam from the can through the safety valve to the outside air. We treat the flame as a heat reservoir, capable of supplying arbitrary heat-flow without being affected, and the outer air as a gas reservoir, capable of absorbing arbitrary gas-flow. We ignore the heat flow to the outside air. We make the idealization that water changes from liquid to gas only during a boiling process.

We use a sorted first-order logic with equality. The sorts of variables is indicated by the first letter of the variable name. We use the following sorts: situations (S), real-numbers (X, Y), signs (G) parameters (F), processes (P), states(A), objects(O). The signs are "pos", "neg" and "zero". For reasons of technical convenience (see axiom 5.5), we take pos and neg to be equal to 1 and -1 respectively, rather than being the intervals $(0, \infty)$ and $(-\infty, 0)$ as is more usual. For a microworld with events, it would also be necessary to include events and intervals.

The theory below contains only the physics needed for this particular example, and thus does not satisfy the "no function in structure" principle. Obvious extensions within the same general microworld, such as the processes of melting, freezing, condensing, or liquid flow, have not been included. However, it can be seen that these could be added with minor modifications to the analysis of this example.

3.1 Formal Language

The following non-logical primitives are used.

Arithmetic

$X1 < X2$ — Predicate. Order relation. Likewise the other order relations.
$X + Y, X - Y, X \cdot Y, X/Y$. — Functions. Plus, minus, times, divide.
within(Y, X, XD) — Predicate. Y is within XD of X. $X - XD < Y < X + XD$.
pos, neg, 0 — Constants. The three signs.

General properties of parameters and states

(Some of these are formally defined in axioms 5.1-5.11 below.)

holds(S, A) — Predicate. State A holds in situation S.
value_in(S, F) — Function. Value of parameter F in situation S.
continuous(F, S) —Predicate. F is continuous at time S.
one_side_deriv(F, S, X, G) — Predicate. F is differentiable from the side indicated by sign G at time S, and the derivative from that side is X.
direction(F) — Function. The fluent of the sign of the direction in which F is changing (pos if increasing, neg if decreasing, 0 if constant.)
no_asymptotic(F) — Predicate. As $t \leftarrow \infty$, F does not asymptotically approach a constant value with a fixed sign of derivative.
no_chatter(A, S) — Predicate. State A does not change infinitely often in the neighborhood of S.
good_param(F) — Predicate. F is a well-behaved parameter.

Influence

d_influence(P, F) — Fluent. The sign of the direct influence of process P on parameter F in each situation. 0 if no influence.
i_influence($F1, F$) — Fluent. The sign of the indirect influence of parameter $F1$ on parameter F in each situation. 0 if no influence.

influence(Q, F, S) — Fluent. The sign of the net influence of Q on parameter F in situation S. Q may be either a process or another parameter

directly_influenced(F) — F is the sort of parameter that is subject to direct rather than indirect influences.

Invariant Object and System Characteristics

boiling_point(O) — Function. Boiling temperature of object O.
heat_reservoir(O) — Predicate. O is a heat reservoir.
gas_reservoir(O) — Predicate. O is a gas reservoir.
valve_between($OV, O1, O2$) — Predicate. OV is a valve connecting $O1$ with $O2$.
thermally_connected($O1, O2$). — Predicate. $O1$ is thermally connected to $O2$.

Parameters

temperature(O) — Function. Fluent of temperature of object O.
heat(O) — Function. Fluent of heat of object O.
pressure(O) — Function. Fluent of the pressure of object O.
liquid_mass(O) — Function. Fluent of the mass of the liquid part of O.
gas_mass(O) — Function. Fluent of the mass of the gaseous part of O.

Object States

open(O) — Function. State of valve O being open.
conduit($OC, O1, O2$) — Function. State of OC serving as a conduit connecting $O1$ with $O2$.

Processes

heat_flow($O1, O2$) — Function. Process of a heat flow from $O1$ to $O2$.
boiling(O) — Function. Process of object O boiling.
gas_flow($O1, O2, OC$) — Function. Process of a flow of gas from $O1$ to $O2$ through conduit OC.

Envisionments

These are primitives that are useful in describing envisionments. They are not used either in the axioms describing the microworld or in the axioms describing the scenario. They are defined in axioms 9.1-9.4. Envisionments are described in terms of "modes", which are states. We use variables with initial letter M for modes.

throughout($S1, S2, A$) — Predicate. State A holds over the *open* interval ($S1, S2$).
dense($S1, S2, A$) — Predicate. State A holds over a dense subset of the interval ($S1, S2$).
borders(MA, MB, S) — Mode MA borders mode MB in situation S.
transition($M0, T, M1, M2 \ldots Mk$) — Mode $M0$ may transition to one of $M1 \ldots Mk$. If the Boolean argument T is "terminal," then $M0$ may be a terminal state; otherwise it cannot be.

3.2 Microworld Theory

We now enumerate the axioms for our microworld, organized according to the outline in section 1.

1. Process Definitions

We include here a number of atemporal axioms and state coherence axioms (axioms constraining the states that can hold in a single situation) constraining relations and states strongly associated with activation conditions.

1.1 $[\ \mathrm{thermally\_connected}(OS, OD) \wedge \mathrm{temperature}(OS, S) > \mathrm{temperature}(OD, S)\ ] \Rightarrow$
$\mathrm{active}(S, \mathrm{heat\_flow}(OS, OD))$.
(Sufficient condition for heat flow: If source OS is thermally connected to destination OD and OS is hotter than OD, then heat will flow from OS to OD.)

1.2 $\mathrm{active}(S,\ \mathrm{heat\_flow}(OS, OD)) \Rightarrow$
$[\ OS \neq OD \wedge \mathrm{thermally\_connected}(OS, OD) \wedge \mathrm{temperature}(OS, S) \geq \mathrm{temperature}(OD, S) \wedge$
$\neg \mathrm{active}(S, \mathrm{heat\_flow}(OD, OS))\ ]$
(Necessary conditions for heat flow: For heat to flow directly from OS to OD, they must be thermally connected; OS must be at least as hot as OD; and there must not be heat flow in the other direction.)

1.3 $\mathrm{thermally\_connected}(O1, O2) \Leftrightarrow \mathrm{thermally\_connected}(O2, O1)$.
(Thermal connections are symmetric.)

1.4 $\mathrm{active}(S, \mathrm{boiling}(OB)) \Leftrightarrow$
$[\ \mathrm{liquid\_mass}(OB, S) > 0 \wedge \mathrm{temperature}(OB, S) = \mathrm{boiling\_point}(OB) \wedge$
$\mathrm{direction}(\mathrm{heat}(OB), S) = \mathrm{pos}$.
(Necessary and sufficient conditions for boiling: An object OB will boil iff it is partially liquid and is at its boiling point and its heat is increasing.)

1.5 $\mathrm{liquid\_mass}(OB, S) > 0 \Rightarrow \mathrm{temperature}(OB, S) \leq \mathrm{boiling\_point}(OB)$.
(Constraint: An object can be partially liquid only if its temperature is below the boiling point.)

1.6 $\mathrm{active}(S, \mathrm{gas\_flow}(O1, O2, OC)) \Leftrightarrow$
$[\mathrm{conduit}(OC, O1, O2, S) \wedge \mathrm{gas\_mass}(O1, S) > 0 \wedge \mathrm{pressure}(O1, S) > \mathrm{pressure}(O2, S)]$
(Necessary and sufficient condition for gas-flow: Gas flows from $O1$ to $O2$ through OC if and only if OC is a conduit between $O1$ and $O2$, and $O1$ is partially gaseous, and the pressure in $O1$ is greater than that in $O2$.)

1.7 $\mathrm{conduit}(OC, O1, O2, S) \Leftrightarrow \mathrm{conduit}(OC, O2, O1, S)$.
(The conduit relation is symmetric in the two ends.)

1.8 $\mathrm{liquid\_mass}(O, S) \geq 0 \wedge \mathrm{gas\_mass}(O, S) \geq 0$.
(Masses are non-negative.)

1.9 $\mathrm{gas\_mass}(O) = 0 \Rightarrow \mathrm{pressure}(O) = 0$.
(If there is no gas, there is no pressure.)

2. Direct Influences

2.1 $[\mathrm{directly\_influenced}(F) \Rightarrow \mathrm{i\_influence}(F,S){=}0] \wedge$
$[\neg \mathrm{directly\_influenced}(F) \Rightarrow \mathrm{d\_influence}(F,S){=}0] \wedge$
$[\mathrm{directly\_influenced}(F) \Leftrightarrow \exists_O\ F{=}\mathrm{heat}(O) \vee F{=}\mathrm{liquid\_mass}(O) \wedge F{=}\mathrm{gas\_mass}(O)]$.
(Division of parameters into those that are directly influenced and those that are indirectly influenced, and an enumeration of the directly influenced.)

2.2 d_influence(P,heat(O),S) $=$ pos $\Leftrightarrow \neg$heat_reservoir(O) $\wedge \exists_{O1}$ P=heat_flow($O1, O$)
(Heat in objects that are not reservoirs is increased by incoming heat flow, and nothing else.)

2.3 d_influence(P,heat(O), S) $=$ neg $\Leftrightarrow \neg$heat_reservoir(O) $\wedge \exists_{O1}$ P=heat_flow($O, O1$)
(Heat in objects that are not reservoirs is decreased by outgoing heat flow, and nothing else.)

2.4 $\neg$d_influence(P,liquid_mass(O),S) $=$ pos.
(There are no processes, within this theory, that tend to increase liquid mass.)

2.5 d_influence(P,liquid_mass(O),S) $=$ neg $\Leftrightarrow$ P=boiling(O)
(Liquid mass is decreased by boiling, and nothing else.)

2.6 d_influence(P, gas_mass(O), S) $=$ pos $\Leftrightarrow$
$\neg$gas_reservoir(O) $\wedge$ [P=boiling(O) $\vee \exists_{O2,OC}$ P=gas_flow($O2, O, OC$)]
(Gas mass is increased by boiling and by incoming flow.)

2.7 d_influence(P, gas_mass(O), S) $=$ neg $\Leftrightarrow \neg$gas_reservoir(O) $\wedge \exists_{O2,OC}$ P=gas_flow($O, O2, OC$)
(Gas mass is decreased by outgoing flow.)

3. Indirect influences

3.1 i_influence(F,temperature(O),S) $=$ pos $\Leftrightarrow$ F=heat(O) $\wedge \neg$active(S,boiling(O))
(Heat is a positive influence on temperature, as long as the object is not boiling.)

3.2 $\neg$i_influence(F,temperature(O),S) $=$ neg.
(There are no negative indirect influences on temperature.)

3.3 i_influence(F,pressure(O),S) $=$ pos $\Leftrightarrow$
F=gas_mass(O) $\vee$ [gas_mass(O, S) $> 0.0 \wedge$ F=temperature(O)]
(Heat and gaseous mass are positive influences on pressure.)

3.4 $\neg$i_influence(F,pressure(O),S) $=$ neg.
(There are no negative indirect influences on pressure.)

4. General axioms of influence.

4.1 influence(Q, F, S)=G $\Leftrightarrow$
[directly_influenced(F) $\wedge$ active(S, Q) $\wedge$ d_influence(Q, F, S)=G] $\vee$
[$\neg$directly_influenced(F) $\wedge$ G=i_influence(Q, F) $\cdot$ direction(Q, S)]
(Definition: Q influences parameter F in direction G in situation S if Q is a process active in S that directly influences F in direction G, or if Q is a parameter whose change in S indirectly influences F in direction G.)

4.2. G=direction(F, S)$\neq 0 \Rightarrow \exists_Q$ influence(Q, F, S) $= G$.
(A parameter F can only change in direction $G \neq 0$ (pos or neg) if there is some influence on F pushing it in direction G.)

4.3 [$\exists_Q$ influence(Q, F, S)=G $\wedge \neg\exists_Q$ influence(Q, F, S)=$-G$] $\Rightarrow$ G=direction(F, S).
(A parameter F must change in direction G if there are influences on F in direction G, and there are no influences on F in direction $-G$.)

5. Well-behavedness conditions

5.1 good_param(F) $\Rightarrow$ continuous(F, S).
(Well-behaved parameters are continuous functions of time.)

5.2 continuous(F, S) $\Leftrightarrow$
$\forall_{XE>0}\ \exists_{XD>0}\ \forall_{S1}$ within($S1, S, XD$) $\Rightarrow$ within(value_in($S1, F$), value_in(S, F), XE).
(Standard delta-epsilon definition of continuity.)

5.3 within(X,Y,D) $\Leftrightarrow Y - D < X < Y + D$.

5.4 good_param(F) $\Rightarrow \exists_{X1,X2}$ one_side_deriv($F, S, X1$,pos) $\wedge$ one_side_deriv($F, S, X2$,neg).
(A well-behaved parameter is differentiable from the right and from the left.)

5.5 one_side_deriv(F, S, X, G) $\Leftrightarrow$
[$G \neq 0 \wedge$
$\forall_{E>0}\ \exists_{D>0}\ \forall_{S1}\ 0 < (S1 - S) \cdot G < D \Rightarrow$
within((value_in($S1, F$) $-$ value_in(S, F) / ($S1 - S$)), X, E)].
(Epsilon-delta definition of one-sided derivative.)

5.6 good_param(F) $\Rightarrow \exists_{G,X}$ one_side_deriv(F, S, X, G) $\wedge$ sign(X)=direction(F, S).
(Partially determined definition of direction: F is changing in direction G if G is the sign of either the derivative from the left or from the right. See section 2.)

5.7 good_param(F) $\Rightarrow$ no_asymptotic(F)
(A well behaved parameter does not asymptotically approach a value without attaining it.)

5.8 no_asymptotic(F) $\Leftrightarrow$
$\forall_{G \neq 0}$ [[$\forall_{S1} \exists_{S2>S1}$ G=direction($F, S2$)] $\wedge$ [$\exists_{S1} \forall_{S2>S1}\ -G \neq$ direction($F, S2$)]] $\Rightarrow$
$\forall_{S1,X} \exists_{S2>S1}$ sign(value_in($S2, F$) $-X$) = G.
(The "no asymptotic" property for a parameter F is as follows: If past a certain point, F never decreases, and there are points arbitrarily late where F is increasing, then F eventually exceeds any fixed value X. Likewise a decreasing function will eventually be less than any fixed value.)

5.9 no_chatter(A, S) $\Leftrightarrow$
$\forall_{G \neq 0}\ \exists_{S1}$ sign($S1 - S$)=$G \wedge$
$\forall_{S2}$ [sign($S2 - S$) = sign($S1 - S2$) $\Rightarrow$ [holds($S2, A$) $\Leftrightarrow$ holds($S1, A$)]].
(State A does not "chatter" around situation S if in some interval before S and in some interval after S it has a constant truth value.)

5.10 $\forall_{F1,F2,G}$ good_param($F1$) $\wedge$ good_param($F2$) $\Rightarrow$
$\exists_A$ [$\forall_S$ holds(S, A) $\Leftrightarrow$ G=sign(value_in($S, F1$) $-$ value_in($S, F2$))] $\wedge\ \forall_S$ no_chatter(A, S)
(The state defined by the order relations between two parameters does not chatter.)

5.11 $\forall_X\ \exists_F$ good_param(F) $\wedge\ \forall_S$ value_in(S, F)=X.
(Existence and good behavior of the constant parameters.)

5.12 good_param(temperature(O)) $\wedge$ good_param(heat(O)) $\wedge$ good_param(pressure(O)) $\wedge$
good_param(liquid_mass(O)) $\wedge$ good_param(gas_mass(O)).
(Physical parameters are well-behaved.)

5.13 no_chatter(conduit($OC, O1, O2$)) $\wedge$ no_chatter(open(O)) $\wedge$ no_chatter(heat_flow($O1, O2$)) $\wedge$
no_chatter(boiling(O)) $\wedge$ no_chatter(gas_flow($O1, O2, OC$)).
(Physical states are well-behaved.)

The "no chatter" axioms 5.9 and 5.10 are not generally needed for constructing envisionment graphs, but they are sometimes necessary for interpreting them. In particular, if an envisionment graph has a cycle, the "no chatter" rule may be needed to rule out histories in which the system traverses the cycle infinitely often in a finite interval, and then "appears" somewhere else in the graph.

6. Unique names

6.1 distinct(temperature($O1$), heat($O2$), pressure($O3$), liquid_mass($O4$), gas_mass($O5$)).

6.2 $O1 \neq O2 \Rightarrow$ temperature($O1$) $\neq$ temperature($O2$) $\wedge$ heat($O1$) $\neq$ heat($O2$) $\wedge$
pressure($O1$) $\neq$ pressure($O2$) $\wedge$ liquid_mass($O1$) $\neq$ liquid_mass($O2$) $\wedge$
gas_mass($O1$) $\neq$ gas_mass($O2$).
(Note that it is consistent with this axiom that the two parameters should sometimes be equal in value, or even that they should always be equal in value. All that the axiom says is that they are distinct entities.)

6.3 distinct(conduit(OA, OB, OC) open(OD), heat_flow(OE, OF), boiling(OG),
gas_flow(OH, OI, OJ)).

6.4 conduit($O1, O2, O3$) = conduit(OA, OB, OC) $\Rightarrow O1 = OA \wedge O2 = OB \wedge O3 = OC$.
open($O1, O2$)=open(OA, OB) $\Rightarrow O1 = OA \wedge O2 = OB$.
heat_flow($O1, O2$) = heat_flow(OA, OB) $\Rightarrow O1 = OA \wedge O2 = OB$.
boiling($O1$)=boiling(OA) $\Rightarrow O1 = OA$.
gas_flow($O1, O2, O3$) = gas_flow(OA, OB, OC) $\Rightarrow O1 = OA \wedge O2 = OB \wedge O3 = OC$.

These unique names axioms are not used in the proofs below. However, they could be important for other kinds of inference, such as the interpretation of a scenario description that specifies that the only active process is the boiling of water.

7. Real analysis

The usual axioms for real analysis. These are not enumerated here.

8. Discrete changes (Valves)

8.1 valve_connects($OV, O1, O2$) $\wedge$ pressure($O1, S$) $-$ pressure($O2, S$) $\geq$ open_pressure_diff(OV) $\Rightarrow$
open(OV, S).
(A valve OV must be open if the pressure difference exceeds the "open pressure.")

8.2 valve_connects($OV, O1, O2$) $\wedge$
pressure($O1, S$) $-$ pressure($O2, S$) $\leq$ close_pressure_diff(OV) $\Rightarrow$
$\neg$open(OV, S).
(A valve OV must be closed if the pressure difference is less than the "close pressure.")

8.3 [$S1 < S2 \wedge$ valve_connects($OV, O1, O2$) $\wedge$ $\neg$open($OV, S1$) $\wedge$ open($OV, S2$)] $\Rightarrow$
$\exists_S$ $S1 < S \leq S2 \wedge$ pressure($O1, S$) $-$ pressure($O2, S$) $\geq$ open_pressure_diff(OV).
(Frame axiom: The valve opens only if the pressure attains the open pressure.)

8.4 [$S1 < S2 \wedge$ valve_connects($OV, O1, O2$) $\wedge$ open($OV, S1$) $\wedge$ $\neg$open($OV, S2$)] $\Rightarrow$
$\exists_S$ $S1 < S \leq S2 \wedge$ pressure($O1, S$) $-$ pressure($O2, S$) $\leq$ close_pressure_diff(OV).
(Frame axiom: The valve closes only if the pressure difference falls under the close pressure.)

8.5 $0 <$ close_pressure_diff(OV) $<$ open_pressure_diff(OV).
(The close pressure is less than the open pressure.)

8.6 valve_connects($OV, O1, O2$) $\Rightarrow$ [conduit($OV, O1, O2, S$) $\Leftrightarrow$ open(OV, S)].
(A valve is a conduit for gas flow just if it is open.)

Definition of Envisionment Primitives.

9.1 throughout($S1, S2, A$) $\Leftrightarrow$
[$S1 < S2 \wedge \forall_S\ S1 < S < S2 \Rightarrow$ holds(S, A)].
(State A holds throughout the open interval ($S1, S2$).)

9.2 dense($S1, S2, A$) $\Leftrightarrow \forall_{SA,SB} S1 < SA < SB < S2 \Rightarrow \exists_{SZ}\ SA < SZ < SB \wedge$ holds(SZ, A).
(State A holds on a dense subset of ($S1, S2$).)

9.3 borders(MA, MB, S) $\Leftrightarrow$
[[holds(S, MA) $\wedge\ \exists_{S1>S}$ throughout($S, S1, MB$)] $\vee$
[[holds(S, MB) $\wedge\ \exists_{S1<S}$ throughout($S1, S, MA$)]] (In state S, the system goes from mode MA to mode MB.)

9.4 transition($M0, T, M1, M2 \ldots Mk$) $\Leftrightarrow$
[$\forall_S$ holds($S, M0$) $\Rightarrow$
[[T=terminal $\wedge\ \forall_{SA>S}$ holds($SA, M0$)] $\vee$
$\exists_{S1}$ [$S1 = S\ \vee$ throughout($S, S1, M0$)] $\wedge$
[borders($M0, M1, S1$) $\vee \ldots \vee$ borders($M0, Mk, S1$)]].
(If the system is in mode $M0$ then it may change to mode $M1$ or to mode $M2$ …or to mode Mk or, if T is "terminal" it may remain in $M0$ forever.)

4 Scenario Description and Envisionment

In this section, we first give a formal account of our sample scenario. Second, we define some of the modes of the systems; namely, those that can actually be attained from an initial state in which the water in the can is completely liquid and is below the boiling point of water. (Other modes do exist, such as modes in which the water is hotter than the flame and cooling down.) Figure 4 shows the envisionment graph for these twelve modes. Thirdly, we prove the first two outward transitions in the graph: mode 1 must be followed by mode 2; mode 2 must be followed by mode 3, mode 7, or mode 8.

4.1 Scenario Description

SC.1 in_scenario(O) $\Leftrightarrow$ O=oflame $\vee$ O=ocan $\vee$ O=owater $\vee$ O=ovalve $\vee$ O=outside_air.
(Enumeration of the objects in the scenario. Note: owater is the collective H_2O in the can, both liquid and steam. This decreases as steam is released through the valve.)

SC.2 in_scenario(O) $\Rightarrow$ [heat_reservoir(O) $\Leftrightarrow$ O=oflame]
(The flame is the only heat reservoir.)

SC.3 in_scenario(O) $\Rightarrow$ [gas_reservoir(O) $\Leftrightarrow$ O=outside_air]
(The outside air is the only gas reservoir.)

SC.4 thermally_connected(O,ocan) $\Leftrightarrow$ O=oflame $\vee$ O=owater.
(The flame and the water are the only things thermally connected to the can.)

SC.5 thermally_connected(O,owater) $\Leftrightarrow$ O=ocan.
(The can is the only thing thermally connected to the water. We ignore any heat flows involving the valve or the outside air.)

SC.6 valve_connects(ovalve,owater,outside_air).
(The valve is a valve connecting the water in the can to the outside air.)

SC.7 conduit(OC,owater,OD,S) $\Rightarrow$ OC=ovalve $\wedge$ OD=outside_air.
(The valve is the only conduit from the water in the can to the outside air. The statement that the valve is a conduit when open is in axiom 8.6 above.)

SC.8 distinct(oflame, ocan, owater, ovalve, outside_air).
(Unique names.)

SC.9 boiling_point(ocan) $>$ temperature(oflame,$S1$) $=$ temperature(oflame,$S2$) $>$ boiling_point(owater).
(The temperature of the flame is constant, greater than the boiling point of water, and less than the boiling point of the can.)

SC.10 pressure(outside_air,$S1$) $=$ pressure(outside_air,$S2$)
(The pressure of the outside air is constant.)

SC.11 open_pressure=pressure(outside_air,S) $+$ open_pressure_diff(ovalve).
close_pressure=pressure(outside_air,S) $+$ close_pressure_diff(ovalve).
(Landmarks on the pressure of the steam in the can to open or close the valve.)

4.2 Mode Definitions

MD.1 holds(S,mode1) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
temperature(owater,S) $<$ boiling_point(owater) $\wedge$ liquid_mass(owater,S) $>$ 0.0 $\wedge$
gas_mass(owater,S)=0.0 $\wedge$ pressure(owater,S) $<$ open_pressure $\wedge$
$\neg$open(ovalve).
(The water is liquid and not boiling, the valve is closed.)

MD.2 holds(S,mode2) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
temperature(owater,S) $=$ boiling_point(owater) $\wedge$ liquid_mass(owater,S) $>$ 0.0 $\wedge$
pressure(owater,S) $<$ open_pressure $\wedge$ $\neg$open(ovalve).
(The water is boiling, the valve is closed. This is actually a superset of modes 5 and 6.)

MD.3 holds(S,mode3) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
temperature(owater,S) $=$ boiling_point(owater) $\wedge$ liquid_mass(owater,S) $>$ 0.0 $\wedge$
pressure(owater,S) $\geq$ open_pressure $\wedge$ open(ovalve).
(The water is boiling and the pressure is enough to open the valve.)

MD.4 holds(S,mode4) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
temperature(owater,S) $=$ boiling_point(owater) $\wedge$ liquid_mass(owater,S) $>$ 0.0 $\wedge$
open_pressure $>$ pressure(owater,S) $>$ close_pressure $\wedge$ open(ovalve).
(The water is boiling, the pressure is between the open and close pressures, and the the valve remains open.)

MD.5 holds(S,mode5) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
temperature(owater,S) $=$ boiling_point(owater) $\wedge$ liquid_mass(owater,S) > 0.0 $\wedge$
pressure(owater,S) $=$ close_pressure $\wedge$ $\neg$open(ovalve).
(The water is boiling and the pressure has fallen to the close pressure.)

MD.6 holds(S,mode6) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
temperature(owater,S) $=$ boiling_point(owater) $\wedge$ liquid_mass(owater,S) > 0.0 $\wedge$
open_pressure $>$ pressure(owater,S) $>$ close_pressure $\wedge$ $\neq$open(ovalve).
(The water is boiling, the pressure is between the open and close pressures, and the the valve remains closed.)

MD.7 holds(S,mode7) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
boiling_point(owater) $\leq$ temperature(owater,S) $<$ temperature(oflame,S) $\wedge$
liquid_mass(owater,S) $= 0.0$ $\wedge$ pressure(owater,S) $<$ open_pressure $\wedge$
$\neg$open(ovalve)
(The water has boiled away, and the valve is closed. This is a superset of modes 10 and 11.)

MD.8 holds(S,mode8) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
boiling_point(owater) $\leq$ temperature(owater,S) $<$ temperature(oflame,S) $\wedge$
liquid_mass(owater,S) $= 0.0$ $\wedge$ pressure(owater,S) $\geq$ open_pressure $\wedge$
open(ovalve).
(The water has boiled away, and the pressure is enough to open the valve.)

MD.9 holds(S,mode9) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
boiling_point(owater) $\leq$ temperature(owater,S) $<$ temperature(oflame,S) $\wedge$
liquid_mass(owater,S) $= 0.0$ $\wedge$ open_pressure $>$ pressure(owater,S) $>$ close_pressure $\wedge$
open(ovalve).
(The water has boiled away, the pressure is between the open and close pressures, and the the valve remains open.)

MD.10 holds(S,mode10) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
boiling_point(owater) $\leq$ temperature(owater,S) $<$ temperature(oflame,S) $\wedge$
liquid_mass(owater,S) > 0.0 $\wedge$ pressure(owater,S) $=$ close_pressure $\wedge$
$\neg$open(ovalve).
(The water has boiled away and the pressure has fallen to the close pressure.)

MD.11 holds(S,mode11) $\Leftrightarrow$
temperature(owater,S) $\leq$ temperature(ocan,S) $\leq$ temperature(oflame,S) $\wedge$
boiling_point(owater) $\leq$ temperature(owater,S) $<$ temperature(oflame,S) $\wedge$
liquid_mass(owater,S) $= 0.0$ $\wedge$
open_pressure $>$ pressure(owater,S) $>$ close_pressure $\wedge$ $\neg$open(ovalve).
(The water has boiled away, the pressure is between the open and close pressures, and the the valve remains closed.)

MD.12 holds(S,mode12) $\Leftrightarrow$
boiling_point(owater) $\leq$ temperature(owater,S) $\wedge$
temperature(owater,S) $=$ temperature(ocan,S) $=$ temperature(oflame,S) $\wedge$
liquid_mass(owater,S) $= 0.0$ $\wedge$ pressure(owater,S) $>$ close_pressure $\wedge$ open(ovalve).
(The water has attained the temperature of the flame. The valve is open.)

MD.13 holds(S,mode13) $\Leftrightarrow$
boiling_point(owater) $\leq$ temperature(owater,S) $\wedge$
temperature(owater,S) = temperature(ocan,S) =
temperature(oflame,S) $\wedge$
liquid_mass(owater,S) = 0.0 $\wedge$ pressure(owater,S) = close_pressure $\wedge$ $\neg$open(ovalve).
(The water has attained the temperature of the flame. The pressure has fallen to the close pressure.)

MD.14 holds(S,mode14) $\Leftrightarrow$
boiling_point(owater) $\leq$ temperature(owater,S) $\wedge$
temperature(owater,S) = temperature(ocan,S) = temperature(oflame,S) $\wedge$
liquid_mass(owater,S) = 0.0 $\wedge$
close_pressure < pressure(owater,S) < open_pressure $\wedge$ $\neg$open(ovalve).
(The water has attained the temperature of the flame. The valve is closed.)

4.3 Proof of the first two transitions

The presence of the two coupled heat flows, from the flame to the can, and from the can to the water, gives rise to complexities in the predictions and the proof. It is perfectly consistent with the above theory that the can should either attain the temperature of the flame, or that it should attain the temperature of the water. (These are achievable states even if axioms 1.1 and 1.2 are changed to read that no heat flow can occur unless there is a temperature differential.) In fact, the temperature of the can can do anything it wants to as long as it stays between the temperature of the flame and the temperature of the water. If the can gets as hot as the flame, then the heat-flow from the flame to the can may cease. It can only cease for an instant, though, because the heat flow from the can to the water will bring down the temperature of the can immediately. Similarly, if the can gets as cool as the water, then the heat flow from the can to the water will cease, and the temperature of the water will stop rising; but, again, this can only happen for an instant. (This problem was called "stutter" in [Forbus, 85].) Therefore, some of our results are stated, not in the form "Such and such a condition must hold throughout an interval," but in the form, "The condition must hold over a dense subset of the interval." It is possible to prove mathematically that these conditions must, in fact, hold almost everywhere on the interval. Indeed, if we impose the "no-chatter" condition (axiom 5.9), it follows that they must hold at all but finitely many points in the intervals. However, since neither of these stronger conclusions give us any additional leverage, we have not included them in the proof below.

Lemmas of a purely mathematical content are merely stated and not proven below.

Lemma 1:
temperature(owater,S) < temperature(ocan,S) $\Rightarrow$ active(S,heat_flow(ocan,owater)).
(If the water is cooler than the can, there must be a heat flow from the can to the water.)

Proof: 1.1, SC.5. □

Lemma 2:
temperature(ocan,S) < temperature(oflame,S) $\Rightarrow$ active(S,heat_flow(oflame,ocan)).
(If the can is cooler than the flame, there must be heat flow from the flame to the can)

Proof: 1.1, SC.4. □

Lemma 3:
good_param($F1$) $\wedge$ good_param($F2$) $\wedge$ throughout($S1, S2$,eql($F1, F2$)) $\Rightarrow$
dense($S1, S2$,eql(direction($F1$),direction($F2$)))
(Mathematical. If functions $F1$ and $F2$ are equal throughout the interval $(S1, S2)$ then their direc-

tions have to be equal on a dense subset.)

Lemma 4:
[good_param(F) $\wedge$ throughout($S1, S2$,eql(direction(F),0))] $\Leftrightarrow \exists_X$ throughout($S1, S2$,eql(F, X)).
(Mathematical: A parameter is constant over an open interval just if its direction is always 0.)

Lemma 5:
throughout($S1, S2$,eql(temperature(oflame),temperature(ocan))) $\wedge$
throughout($S1, S2$,temperature(ocan) $>$ temperature(owater)) $\Rightarrow$
throughout($S1, S2$,heat_flow(oflame,ocan)).
(If the can and the flame are the same temperature and hotter than the water throughout an open interval, then there must be heat flow from the flame to the can throughout the interval. Note: This does not apply to a closed interval.)

Proof: By Lemma 1, there is a heat-flow from the can to the water. By 2.2, SC.1, SC.2, this is a negative influence on heat(ocan)

By SC.9. temperature(oflame) is constant, so, by assumption, temperature(ocan) is likewise constant. By Lemma 4, the direction of temperature(ocan) is 0. By SC.10 and 1.4, the can is not boiling. By 3.1, 3.2, heat(ocan) is an influence and the only influence, on temperature(ocan). By 4.1, 4.3, the direction of heat(ocan) is 0. Since we know that there is a negative influence on heat(ocan), by 4.3, there must be a positive influence on heat(ocan). By 2.2, this must be a heat flow into ocan. By 1.2 and SC.4, the only possible heat flow into ocan is from oflame.□

Lemma 6:
direction(temperature(O),S)=pos $\Rightarrow \exists_{O1}$ active(S,heat_flow($O1, O$)).
(The temperature of O can increase only if there is a heat flow into it.)

Proof: From 3.1, 3.2, 4.1, 4.2, the temperature of O can increase only if the heat of O increases. From 2.2, 4.1, 4.2, heat(O) can increase only if there is a heat flow into O.□

Lemma 7:
direction(temperature(O),S)=neg $\Rightarrow \exists_{O1}$ active(S,heat_flow($O, O1$)).
(The temperature of O can decrease only if there is a heat flow out of it.)

Proof: From 3.1, 3.2, 4.1, 4.2, the temperature of O can decrease only if the heat of O decreases. From 2.2, 4.1, 4.2, heat(O) can decrease only if there is a heat flow out of O.

Lemma 8:
throughout($S1, S2$,temperature(oflame) $>$ temperature(ocan)) $\wedge$
throughout($S1, S2$,eql(temperature(ocan), temperature(owater))) $\Rightarrow$
$\exists_S$ $S1 < S < S2$ $\wedge$ active(S,heat_flow(ocan,owater)).
(If the can and the water are the same temperature and cooler than the flame throughout an open interval, then there is heat flow from the can to the water at some time during that interval.)

Proof: By Lemma 2, there is a heat flow from oflame to ocan. By 1.2 and SC.4, the only possible heat flow out of ocan is to owater.

We prove by contradiction that at some time between $S1$ and $S2$ there must be a heat flow from ocan to owater. Suppose not. Then, from the above remark, there is no heat flow out of ocan. By 2.3, there is no negative influence on the heat of ocan, and by 2.2 there is a positive influence. By 4.3, the heat of ocan is rising throughout the interval ($S1, S2$). By SC.9 and 1.4, the can is not boiling, so by 3.1 and 3.2, the heat of the can is the unique influence on temperature. Therefore, by 4.3, the temperature of the can rises throughout ($S1, S2$). By Lemma 3, since the assumptions specify that the temperature of ocan and owater are equal throughout ($S1, S2$), it follows that the temperature of owater is rising at a dense subset of ($S1, S2$). By lemma 6, there must be a heat flow into owater from somewhere. By 1.2 and SC.5, the only possible source for a heat flow into owater is ocan; but by assumption there is no such heat flow. This completes the contradiction.□

Lemma 9:
[throughout($S1, S2$,temperature(oflame) $\geq$ temperature(ocan)) $\wedge$
throughout($S1, S2$,temperature(ocan) $\geq$ temperature(owater)) $\wedge$
throughout($S1, S2$,temperature(oflame) $>$ temperature(owater))] $\Rightarrow$
$\exists_S$ $S1 < S < S2$ $\wedge$ active(S,heat_flow(oflame,ocan)) $\wedge$ active(S,heat_flow(ocan,owater)).
(If the temperature of the can is (not strictly) between the temperature of the flame and the temperature of the water throughout an open time interval, then at some time in between there must be both heat flow from the flame to the can, and heat flow from the can to the water.)

Proof: There must be some subinterval SA, SB of $S1, S2$ throughout which one of the following holds:

- temperature(oflame) $>$ temperature(ocan) $>$ temperature(owater).
  By lemmas 1 and 2, the two heat flows are active.
- temperature(oflame) $=$ temperature(ocan) $>$ temperature(owater).
  By lemmas 1 and 5, the two heat flows are active.
- temperature(oflame) $>$ temperature(ocan) $>$ temperature(owater).
  By lemmas 2 and 6, the two heat flows are active.□

Lemma 10:
[$\forall_{SA,SB}$ throughout($SA, SB, A1$) $\Rightarrow \exists_S$ $SA < S < SB$ $\wedge$ holds($S, A2$)] $\Rightarrow$
[$\forall_{SA,SB}$ throughout($SA, SB, A1$) $\Rightarrow$ dense($SA, SB, A2$)].
(Mathematical. If every interval satisfying A throughout contains a point satisfying B, then every interval satisfying A contains a dense collection of points satisfying B.)

MODE1.1:
throughout($S1, S2$,mode1) $\Rightarrow$
dense($S1, S2$,heat_flow(oflame,ocan)) $\wedge$ dense($S1, S2$,heat_flow(ocan,owater)).
(If mode 1 holds throughout an interval, then there is heat flow from the flame to the can and from the can to the water over a dense subset.)

Proof: Immediate from MD.1, Lemma 9, Lemma 10.□

Lemma 11:
[active(S,heat_flow(ocan,owater)) $\wedge$ $\neg$active(S,boiling(owater))] $\Rightarrow$
direction(temperature(owater,S)) = pos.
(If there is a heat-flow from the can to the water, and the water is not boiling, then the temperature of the water is rising.)

Proof: By 1.2 and SC.5, there cannot be any heat flow out of the water. By 2.2, 2.3, 4.1, and 4.3, heat(owater) must be increasing. By 3.1, 3.2, this is the only influence on the temperature of the water. By 4.1 and 4.3 temperature(owater) must be rising.□

Lemma 12:
good_param(F) $\wedge$ $G \neq 0$ $\wedge$ dense($S1, S2$,eql(direction(F),G)) $\Rightarrow$
throughout($S1, S2$,neql(direction(F),$-G$)).
(Mathematical: If direction(F) has non-zero value G over a dense subset of ($S1, S2$), then it cannot be $-G$ anywhere on ($S1, S2$).)

MODE1.2:
throughout($S1, S2$,mode1) $\Rightarrow$
dense($S1, S2$,eql(direction(temperature(owater),pos))) $\wedge$

throughout($S1, S2$,neql(direction(temperature(owater),neg)))
(In mode 1, the temperature of the water is increasing over a dense set, and it is never decreasing.)

Proof: From MODE1.1, MD.1, Lemma 11, and Lemma 12.

Lemma 13:
good_param($F1$) $\wedge$ good_param($F2$) $\wedge$
$S1 < S2 \wedge$ value_in($S1, F1$) $\leq$ value_in($S1, F2$) $\wedge$ value_in($S2, F1$) $>$ value_in($S2, F2$) $\Rightarrow$
$\exists_{SA}$ $S1 < SA < S2 \wedge$ value_in($SA, F1$) $>$ value_in($SA, F2$) $\wedge$
[direction($F1, SA$)=pos $\vee$ direction($F2, SA$)=neg]
(Mathematical: If $F1 \leq F2$ at time $S1$ but $F1$ is greater than $F2$ later, then there is a time later when $F1$ is greater than $F2$ and either $F1$ is increasing or $F2$ is decreasing.)

Lemma 14:
temperature(oflame,S) $\geq$ temperature(ocan,S) $\geq$ temperature(owater,S) $\Rightarrow$
$\forall_{S1>S}$ temperature(oflame,$S1$) $\geq$ temperature(ocan,$S1$) $\geq$ temperature(owater,$S1$).
(If the temperature of the can is (not strictly) between the temperature of the flame and the temperature of the water, then these inequalities will hold at all future times.)

Proof: By contradiction. Suppose that this does not hold for some particular S and $S1$. Then in $S1$ either (a) the water is hotter than the can; or (b) the water is not hotter than the can, but the can is hotter than the flame. We consider these two possibilities in turn.

A) By lemma 13, there is some time SA between S and $S1$ during which the water is hotter than the can and the temperature of the water is increasing. But (lemma 6) the temperature of the water can increase only if there is heat flow into the water, which is impossible by 1.2 and SC.5.

A) By lemma 13, there is some time SA between S and $S1$ during which the can is hotter than the flame and the temperature of the can is increasing. But (lemma 6) the temperature of the can can only increase if there is heat flow into the can, which means (1.2 and SC.4) that the water must be hotter than the can, contrary to assumption.

This completes the contradiction.□.

Lemma 15:
temperature(owater,S) $<$ boiling_point(owater) $\Rightarrow$
direction(liquid_mass(owater),S) = direction(gas_mass(owater),S) = 0.
(If the water is cooler than boiling temperature, then neither liquid mass nor gas mass are changing.)

Proof: From 1.4, the water is not boiling in S. From 2.4, 2.5, 4.1, 4.2, the liquid mass of the water is not changing. From 2.6, 2.7, 4.1, 4.2, the gas mass of the water is not changing either.□

MODE1.3:
holds(S,mode1) $\Rightarrow$
direction(liquid_mass(owater),S) = direction(gas_mass(owater),S) = 0.
(In mode 1, neither liquid mass nor gas mass is changing.)

Proof: Immediate from Lemma 15.□

Lemma 16:
[gas_mass(owater,$S1$) = 0.0 $\wedge$
[$\forall_S$ $S1 \leq S < S2 \Rightarrow$ temperature(owater,S) $<$ boiling_point(owater)]] $\Rightarrow$
pressure(owater,$S2$) = pressure(owater,$S1$).
(If no part of the water is gaseous at $S1$, and the temperature of the water remains below boiling until $S2$, then there is no change in pressure.)

Proof: If the pressure changes between $S1$ and $S2$, then by lemma 4, it must have a non-zero direction at some time in between. From 3.3, 3.4, 4.1, 4.2, the pressure can change only if the gas mass is changing or if the gas mass is greater than 0.0 and the temperature is changing. From Lemma 15, the gas mass is never changing. From lemma 4, this implies that the gas mass remains equal to 0.0 throughout $(S1, S2)$. Thus the result follows.

MODE1.4:
holds(S,mode1) $\Rightarrow$ direction(pressure(owater),S) $=$ 0.
(In mode 1 there is no change in pressure.)

Proof: Immediate from lemma 16.□

MODE1.5
holds(S,mode1) $\Rightarrow \exists_{S>S1}$ $\neg$holds(S,mode1)
(Mode 1 cannot be a final state.)

Proof: Suppose that mode 1 were a final state. Then, by MODE1.2, the temperature of owater would be forever rising. By 5.8, it would eventually exceed boiling_point(owater), but then the system would no longer be in mode1, which is inconsistent.□

Lemma 17:
good_param(F) $\wedge$ $X1 <$ value_in(S, F) $< X2 \Rightarrow \exists_{S1>S}$ throughout($S, S1$,$X1 < F < X2$).
(Mathematical: If F has value X strictly between $X1$ and $X2$ in situation S, then it continues to lie between $X1$ and $X2$ for some interval after S.)

Lemma 18:
good_param($F1$) $\wedge$ good_param($F2$) $\wedge$ throughout($S1, S2$, $F1 \leq F2$) $\Rightarrow$
value_in($S1, F1$) $\leq$ value_in($S1, F2$) $\wedge$ value_in($S2, F1$) $\leq$ value_in($S2, F2$).
(Mathematical: If a non-strict inequality holds over an open interval, it holds at both end points.)

Corollary 19:
good_param($F1$) $\wedge$ good_param($F2$) $\wedge$ throughout($S1, S2$, eql($F1, F2$)) $\Rightarrow$
value_in($S1, F1$) $=$ value_in($S1, F2$) $\wedge$ value_in($S2, F1$) $=$ value_in($S2, F2$).
(If two parameters are equal over an interval, they are equal at the endpoints. Corollary of lemma 18.)

MODE1.6: transition(mode1,nonterminal,mode2).
Mode 1 must transition to mode2.

Proof: By MODE1.5, mode 1 cannot be a final state. Let S be a state in which mode 1 holds, and let $S1$ be the greatest lower bound of all situations greater than S in which mode 1 does not hold. Then mode 1 holds in S and over the open interval $(S, S1)$; but either mode 1 does not hold in $S1$ or there is no interval $(S1, S2)$ such that mode 1 holds throughout $(S1, S2)$.

Let us begin by considering what is the situation in $S1$. If $S1 = S$, then, of course, mode 1 holds in $S1$. Suppose $S1 > S$, so that mode 1 holds throughout the interval $(S, S1)$. By lemma 14, in $S1$ the temperature of the water must still be less than or equal to the temperature of the can, and the temperature of the can must be less than or equal to the temperature of the flame. By lemma 16, the temperature of the water in $S1$ is less than or equal to the boiling point of water. By MODE1.3, MODE1.4, and lemma 4, the liquid mass, the gas mass, and the pressure are all constant over the interval $(S, S1)$, so by lemma 16 they are unchanged in $S1$. Therefore the constraints liquid_mass(owater,$S1$) $>$ 0.0, gas_mass(owater,$S1$) $=$ 0.0, pressure(owater,S) $<$ open_pressure must all hold. By 8.3, since the pressure stays less than open_pressure throughout $(S, S1]$, the valve cannot open.

Putting these together, we conclude that in $S1$, either the temperature of the water has reached the boiling point and the system is in mode 2, or it has not and the system is in mode 1. In the first case, there is a transition from mode 1 to mode 2. We will show that the second case is impossible by considering what happens in short intervals following $S1$. By lemma 17, if the temperature of the water is below boiling in $S1$ then there is an interval $(S1, S2)$ during which it remains below boiling. By lemmas 15 and 16, the directions of change of the liquid mass, the gas mass, and the pressure are zero throughout $(S1, S2)$. Hence, by lemma 4, these parameters remain constant. By same argument as above, the valve remains closed throughout $(S1, S2)$. Hence, the system remains in mode 1 throughout $(S1, S2)$, contrary to assumption.□.

MODE2.1:
throughout($S1, S2$,mode2) $\Rightarrow$
dense($S1, S2$,heat_flow(oflame,ocan)) $\wedge$ dense($S1, S2$,heat_flow(ocan,owater)).

Proof: Immediate from MD.2, Lemma 9, Lemma 10.□

MODE2.2:
throughout($S1, S2$,mode2) $\Rightarrow$
dense($S1, S2$,boiling(owater)) $\wedge$ dense($S1, S2$, eql(direction(heat(owater)),pos)) $\wedge$
dense($S1, S2$, eql(direction(liquid_mass(owater)),neg)) $\wedge$
dense($S1, S2$, eql(direction(gas_mass(owater)),pos)) $\wedge$
dense($S1, S2$, eql(direction(pressure(owater)),pos)).
(If mode 2 holds during an open interval, then the water is boiling, the liquid mass of the water is decreasing, and the temperature, gas mass, and pressure of the water are increasing over a dense subset of the interval.)

Proof: From MODE2.1, there is a heat flow from the can to the water at a dense set of instants. Let S be any such instant. By 1.2 and SC.5, there cannot be any heat flow out of the water in S. Hence, by 2.2, 2.3, 4.1, 4.3, the heat of the water is increasing. By MODE2.1, MD.2, 1.4., the water must be boiling in S. By 2.4, 2.5, 4.1, 4.3, the liquid mass of the water is decreasing in S. By SC.7, 8.6, the valve is the only possible conduit for gas flow, and only when it is open. Since, by MD.2, it is not open in mode 2, there is no possible conduit for gas flow, so by 1.6 there is no gas flow. Hence, by 2.6, 2.7, the only influence on gas mass is positive, so by 4.1, 4.3, gas mass must be increasing in S. By definition of mode 2, the temperature is constant, and so (lemma 4) it is neither increasing nor decreasing during $(S1, S2)$. There is thus one positive influence on the pressure, and no negative influences (3.3,3.4) so the pressure must be increasing (4.1, 4.3).□

MODE2.3: transition(mode2, nonterminal, mode3, mode8, mode9)

Proof: Mode 2 cannot be a terminal mode. Since liquid mass steadily decreases, by 5.8 it must eventually attain zero, at which point (if not before) the system is no longer in mode 2.

Let S be a situation in which mode 2 holds, and let $S1$ be the greatest lower bound of situations after S in which mode 2 does not hold. Thus, if $S1 > S$, then mode 2 holds throughout the interval $(S, S1)$. By lemma 9, the water is cooler than the can which is cooler than the flame in $S1$. By lemma 19, the temperature of the water is equal to the boiling point of water in $S1$. By lemma 17, the liquid mass of the water is greater than or equal to zero in $S1$, and the pressure is less than or equal to the opening pressure, since both of these non-strict inequalities hold over $(S, S1)$. By 8.1, if the pressure in $S1$ is equal to the opening pressure, then the valve must be open in $S1$; by 8.3, if the pressure in $S1$ is less than the opening pressure, then the valve must be closed in $S1$. (Note that by definition of mode 2, the valve is closed in S, and the pressure is less than the opening pressure throughout $(S, S1)$.

Combining these conditions, it follows that the system in $S1$ is either in mode 2, mode 3, mode 8, or mode 9. It remains to eliminate the first of these possibilities, by showing that if mode 2 holds in $S1$, then it holds over some interval $(S1, S2)$, contrary to hypothesis. From lemma 18, we know that if the strict inequalities liquid_mass(owater) > 0 and pressure(owater) $<$ open_pressure hold in $S1$ then they must hold for some interval after $S1$. Therefore, by 8.3, the valve remains closed throughout $(S1, S2)$. The temperature of the water cannot fall below the boiling point, since there is no negative influence on it, and it cannot rise above the boiling point by 1.5 The inequalities on the temperature of the water, the can, and the flame continue to hold by lemma 9. Thus, all the conditions of mode 2 are satisfied, and mode 2 continues through $(S1, S2)$, contrary to hypothesis.□

5 Non-monotonicity

There are (at least) two possible roles that non-monotonic inference could play in QP theory:

1. It may be possible to infer parts of a theory like that above, particularly closure conditions, by applying non-monotonic inference to a simpler theory.
2. It may be desirable to modify a theory like that above by changing some of the axioms from deductive rules to default rules, thus allowing inferences to be drawn provisionally and withdrawn if they lead to contradictions with other information.

The distinction between these two roles mirrors a division throughout the NML literature between non-monotonic theories as *abbreviations for monotonic theories* and non-monotonic theories as *theories of defeasible inference.* (I don't present this as a technical distinction, but as a difference of objective and outlook.) The former approach treats non-monotonic inference as an expansion of a partial theory into a more complete monotonic theory that is done once and for all at the beginning of inference. Examples include the application of the closed-world inference to a static database; the usual view of circumscription; and the inferal of frame laws from causal laws, as in [Lifschitz, 87], and [Lin and Shoham, 91]. The latter approach treats non-monotonic inference as occurring in the midst of the deductive process, or as part of a time-varying system. Examples include the application of the closed-world assumption to a dynamic database; the usual view of Reiter's [1980] default logic; most procedural implementations of non-monotonic inference, including negation as failure and non-monotonic truth maintenance systems; solutions to the frame problem where non-monotonic frame inferences are constructed for the particular scenario as in [Shoham, 88]; and the chronological minimization of discontinuity in [Sandewall, 89]. On the whole the former type of inference is easier to reason about than the latter.

Regarding the first role: Clearly many of the closure conditions in the theory and in the scenario description can be omitted and derived via non-monotonic inference of a standard kind. Specifically:

- In cases (the majority) where it is possible to state conditions that are both necessary and sufficient for the activity of a process, it will suffice just to state them as sufficient conditions. That they are also necessary can then be derived by circumscribing "active". For example, in axioms 1.4 and 1.6 above, one could just state the axiom with the left-pointing arrow, and derive the right-pointing arrow by circumscribing "active".
- In the enumeration of influences, it would be possible just to state axioms of the form "Process P or parameter $F1$ has influence G on parameter $F2$," and then derive that these are the only influences by circumscribing "d_influence" and "i_influence". For example, in the above theory, one would replace axioms 2.2 and 2.3 by the axioms

  $\neg$heat_reservoir(O) $\Rightarrow$ d_influence(heat_flow($O1, O$),heat(O),pos).
  $\neg$heat_reservoir(O) $\Rightarrow$ d_influence(heat_flow($O, O1$),heat(O),neg).

- Every ground instance of the unique-names axioms 6.1-6.4 can be derived via the usual unique-names assumptions on ground terms.

- It is probably possible to derive the frame axioms 8.3 and 8.4 by choosing a suitable causal language and applying circumscription to the causal axioms 8.1 and 8.2, along the lines of [Lifschitz, 87] and [Lin and Shoham, 91].
- The exhaustive enumeration of the heat and gas reservoirs in the scenario (SC.1, SC.2, SC.3) can be achieved by stating that the flame is a heat reservoir and that the outside air is a gas reservoir, and then circumscribing over those two predicates. Likewise, the exhaustive enumeration of thermal connections (SC,4,SC.5) can be achieved by stating that the flame is connected to the can, and the can to the water, and then circumscribing over that predicate.
- The unique-names axiom on the objects in the scene (SC.8) is an instance of the unique-names assumption on constant symbols.

Non-monotonic inference thus allows us to start with a theory that is clearly substantially simpler. It is also more *additive,* in the following sense: If we wish to add a new process to the theory, all that is required is to add axioms describing its activation conditions and its influences and to "re-run" the circumscription. None of the existing axioms have to be changed. By contrast, adding a new process to the monotonic theory will, in general, require rewriting the closure conditions. Consider, for example, adding "freezing(O)" as a new process and "solid_mass(O)" as a new parameter. In the monotonic theory, axiom 2.5, which states that the only negative influence on liquid_mass(O) is boiling(O), is no longer true. The axiom must be weakened to read that the only negative influences are boiling(O) and freezing(O). By contrast, none of the axioms of the non-monotonic theory become false. The weakening of the closure conditions happens automatically as a result of strengthening the positive part of the theory (in this case, adding the fact that freezing is negative influence on liquid mass.) Likewise, expanding the scenario by adding new thermally connected objects would require rewriting SC.4 and SC.5 in the monotonic theory, but only requires adding the new objects in the non-monotonic theory.

As regards the second type of inference: The modification of the theory so that the closure assumptions are merely defeasible inferences seems attractive in many instances. For example, the assumption that all the relevant influences on a parameter have been enumerated could be made a defeasible inference, that could be withdrawn if the observed behavior of a parameter violated the predicted behavior. If it is observed that the water is not heating up, contrary to prediction, then we must posit that the closure assumption was mistaken and some additional process is active. In terms of circumscription, this would require circumscribing "active" over a theory that included these very observations. However, this kind of inference tends to be prone to anomalies like the Yale Shooting Problem, and a careful analysis would be required to determine whether the theory leads to all and only the reasonable conclusions.

6 Remarks on the theory

Some particular features of the above theory and inference process are worth noting.

The theory largely achieves the objective of *locality*. It would be possible to posit two separate scenario running simultaneously side by side, and to reason about them separately. It would even be possible to posit that these same objects were involved in an entire separate collection of parameters and process (e.g. electrical processes). The validity of the proof above would not be affected as long as these additional processes and parameters do not influence our original processes and parameters. (Influence in the opposite direction would be OK.) Nowhere did either the theory or the scenario description assert that these are the only objects in the world, or that these are the only processes or process types.

As remarked above, the monotonic theory does not have the property of *additivity*, either in expanding the list of processes known to affect a given parameter, or in expanding the list of objects in a scenario. This additivity can be largely achieved, however, if the monotonic theory is derived from an underlying non-monotonic theory.

The natural form of reasoning in this theory has a somewhat different flavor from the reasoning in QP, even in doing the same task of prediction. QP always starts with a complete qualitative description of some mode, and calculates the next mode. In using the logic, the natural way to proceed is to develop lemmas that start with partial characterizations of an interval of time, and derive other partial characterizations. QP, so to speak, works vertically from one time period to the next; logical inference works most comfortably horizontally, building up constraints among intervals of time.

For this reason, certain inferences that require special mechanisms in QP do not require any special treatment in the logic. For example, it is a fact that the cycle "mode 3 $\rightarrow$ mode 4 $\rightarrow$ mode 5 $\rightarrow$ mode 6 $\rightarrow$ mode 3" cannot persist indefinitely, since the liquid mass drops throughout and must eventually attain zero. This fact cannot even be expressed in a simple envisionment graph. However, in the logic, it takes the form of the lemma, "If, throughout an interval interval, the liquid mass is alway positive and always dropping, then the interval must be finite," which is a simple consequence of axiom 5.8.

In a recent extended e-mail discussion, a number of people expressed doubts as to whether QP could be justified within a well-defined logical theory. This paper, I believe, has answered that question in the affirmative. Whether this adds to our understanding of QP is another question, of course.

7 References

E. Davis (1990) *Representations of Commonsense Knowledge,* Morgan Kaufmann, San Mateo, CA.

E. Davis (In preparation). "Infinite Loops in Finite Time."

J. de Kleer (1977) "Multiple Representations of Knowledge in a Mechanics Problem Solver," *Proc. IJCAI-77,* pp. 299-304.

J. de Kleer and J.S. Brown (1985) "A Qualitative Physics Based on Confluences," in D. Bobrow (ed.) *Qualitative Reasoning about Physical Systems,* M.I.T. Press, Cambridge, MA.

D. Duchier (1991) "Logicalc: An Environment for Interactive Proof Development," Yale Computer Science Dept., Research Report #862.

B. Faltings (1987) "Qualitative Kinematics in Mechanisms," *Proc. IJCAI-87*, pp. 436-442.

K. Forbus (1985) "Qualitative Process Theory," in D. Bobrow (ed.) *Qualitative Reasoning about Physical Systems,* M.I.T. Press, Cambridge, MA.

B. Kuipers, (1986) "Qualitative Simulation," *Artificial Intelligence,* vol. 29, pp. 289-338.

F. Lin and Y. Shoham (1991) "Provably Correct Theories of Action," *Proc. AAAI-91*, pp. 349-354.

M. Rayner (1991) "On the applicability of nonmonotonic logic to formal reasoning in continuous time," *Artificial Intelligence,* vol. 49, pp. 345-360.

E. Sandewall, (1989) "Combining logic and differential equations for describing real-world systems," in R. Brachman, H. Levesque, and R. Reiter (eds.) *Proc. First International Conference on Principles of Knowledge Representation and Reasoning,* Morgan Kaufmann, San Mateo, CA, pp. 412-420.

L.K. Schubert, (1990) “Monotonic solution of the frame problem in the Situation Calculus: An efficient method for worlds with fully specified actions,” in H. Kyburg, R. Loui and G. Carlson (eds.), *Knowledge Representation and Defeasible Reasoning,* Kluwer, pp. 23-67, 1990.

Y. Shoham (1988) *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence,* MIT Press, Cambridge, MA