

AD-A112 225

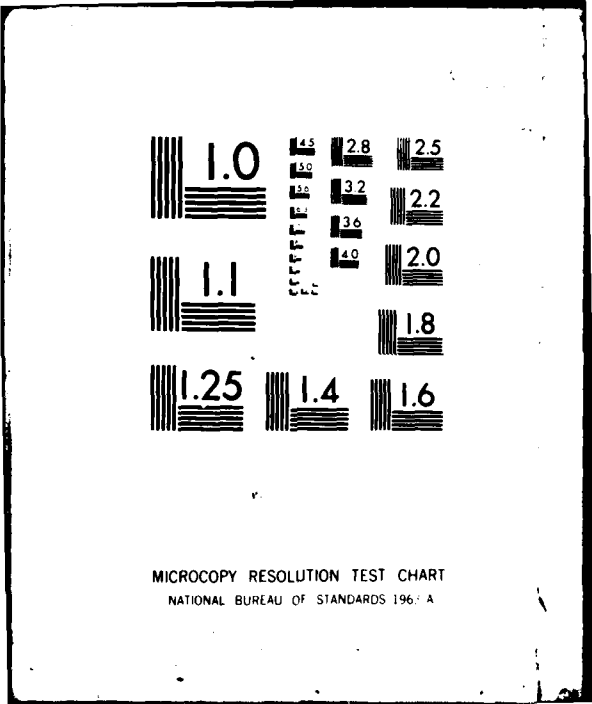
MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 9/4  
QUALITATIVE PROCESS THEORY. (U)  
FEB 82 K D FORBUS  
AI-M-664

N00014-80-C-0505  
NL

UNCLASSIFIED

1-1  
3/2/82


END  
DATE  
FILMED  
4 82  
DTIC



UNCLASSIFIED

12

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ADA 112225

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AIM 664	2. GOVT ACCESSION NO. AD-A112 225	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Qualitative Process Theory		5. TYPE OF REPORT & PERIOD COVERED memo
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kenneth D. Forbus		8. CONTRACT OR GRANT NUMBER(s) N0014-80-C-0505
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE February 1982
		13. NUMBER OF PAGES 54
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
15. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES None		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
Qualitative Physics Commonsense reasoning Causal reasoning Problem solving	Artificial Intelligence Knowledge representation Qualitative simulation Computer Aided Instruction	
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Things move, collide, flow, bend heat up, cool down, stretch, break, and boil. these and other things that happen to cause changes in objects over time are intuitively characterized as processes. To understand common sense physical reasoning and make machines that interact significantly with the physical world we must understand qualitative reasoning about processes, their effects, and their limits. Qualitative Process theory defines a simple notion of Reasoning about process also motivates a new qualitative representation for quantity, the		

DTIC  
MAR 22 1982  
H

DTIC FILE COPY

↓  
Quantity Space. This paper includes the basic definitions of Qualitative Process theory, describes several different kinds of reasoning that can be performed with them, and discusses its implications for causal reasoning. The use of the theory is illustrated by several examples, including figuring out that a boiler can blow up, that an oscillator with friction will eventually stop, and how to say that you can pull with a string, but not push with it. ↑

Accession For	
NTIS	<input checked="" type="checkbox"/>
ERIC	<input type="checkbox"/>
Unrestricted	<input type="checkbox"/>
Justification	
By	
Distribution	
Availability	
Date	

A

DTIC  
COPY  
INSPECTED  
2

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 664

February 1982

**Qualitative Process Theory**

Kenneth D. Forbus

**Abstract**

Things move, collide, flow, bend, heat up, cool down, stretch, break, and boil. These and other things that happen to cause changes in objects over time are intuitively characterized as processes. To understand common sense physical reasoning and make machines that interact significantly with the physical world we must understand qualitative reasoning about processes, their effects, and their limits. Qualitative Process theory defines a simple notion of physical process that appears quite useful as a language in which to write physical theories. Reasoning about processes also motivates a new qualitative representation for quantity, the Quantity Space. This paper includes the basic definitions of Qualitative Process theory, describes several different kinds of reasoning that can be performed with them, and discusses its implications for causal reasoning. The use of the theory is illustrated by several examples, including figuring out that a boiler can blow up, that an oscillator with friction will eventually stop, and how to say that you can pull with a string, but not push with it.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract number N00014-80-C-0505.

8 2 0 3 2 0 0 7 6

## CONTENTS

<b>1. Introduction</b> .....	4
1.1 Overview of the paper .....	5
<b>2. Quantities</b> .....	6
<b>3. Processes</b> .....	13
<b>4. Limit Analysis</b> .....	18
<b>5. Incremental and Differential Qualitative Analysis</b> .....	20
<b>6. Causality and Functional Dependence</b> .....	26
<b>7. Examples</b> .....	32
7.1 Boiler .....	32
7.2 Motion .....	36
7.3 Breaking string .....	38
7.4 An oscillator .....	40
7.5 Grain Elevator .....	45
7.6 Automobile Transmission .....	45
<b>8. Discussion</b> .....	47
8.1 Perspective .....	47
8.2 Common Sense Physical Reasoning .....	49
8.3 Reasoning about Engineered Systems .....	50
8.4 Economic modelling and Support Systems .....	51
8.5 Acknowledgements .....	52
<b>9. Bibliography</b> .....	53

## FIGURES

Fig. 1. Graphical Notation for a Quantity Space .....	7
Fig. 2. Combining $d_s$ Values .....	10
Fig. 3. Physical Process Definitions .....	15
Fig. 4. Linking derivatives with Inequalities .....	18
Fig. 5. Difference between IQ and DQ values .....	22
Fig. 6. Comparing alternate situations .....	24
Fig. 7. Constraint representation of relationships .....	27
Fig. 8. Model Fragments with possible Processes .....	28
Fig. 9. A tree of functional dependencies .....	30
Fig. 10. A simple boiler and its Quantity Spaces .....	32
Fig. 11. Sliding Block With Friction .....	41
Fig. 12. Process history for the oscillator .....	42

## 1. Introduction

Many kinds of changes occur in physical situations. Things move, collide, flow, bend, heat up, cool down, stretch, break, and boil. These and the other things that happen to cause changes in objects over time are intuitively characterized as processes. Much of formal physics consists of characterizations of processes by differential equations which describe how the parameters of objects change over time. But the notion of process is richer and more structured than this. We often reach conclusions about physical processes based on very little information. For example, we know that if we heat water in a sealed container the water can eventually boil, and if we continue to do so the container can explode. To understand common sense physical reasoning we must understand how to reason qualitatively about processes, their effects, and their limits. This paper describes a theory I have been developing, called Qualitative Process theory, for this purpose. I expect this theory, when fully developed, to provide a representational framework for programs that reason about complex physical systems as well as common sense reasoning. Programs that explain, repair and operate complex systems such as nuclear power plants and steam machinery will need to draw the kinds of conclusions discussed here.

Qualitative reasoning about quantities is a problem that has long plagued AI. Many schemes have been tried, including simple symbolic vocabularies (TALL, VERY TALL, etc.), real numbers, intervals, fuzzy logic, and so forth. None are very satisfying. The reason is that none of the above schemes makes distinctions that are relevant to physical reasoning. Reasoning about processes provides a strong constraint on the choice of representation for quantities. Processes usually start and stop when orderings between quantities change (such as unequal temperatures causing a heat flow). In Qualitative Process theory the value of quantities are represented by a partial ordering of other quantities determined by the domain physics. The representation appears both useful and natural.

The basic Qualitative Process theory described here is not intended to capture the full range of qualitative reasoning about the physical world. Instead it is concerned with describing the weakest kind of information that still allows useful conclusions to be drawn. There are two reasons why this weak level of description is interesting. First, conclusions from weak information are often required to drive the search for conclusions from more detailed information (an illustration is [deKleer, 1975]). More importantly, I believe that the basic theory can be used to write what corresponds to people's common sense physical knowledge. To capture more sophisticated kinds of physical reasoning (for example, how



an engineer makes estimates of circuit parameters or stresses on a bridge) extension theories containing more detailed representations of quantity, functions, and processes will be needed. By providing a shared basic theory, future studies of more sophisticated domains may yield a way to classify kinds of physical reasoning according to the extension theories they require.

### 1.1 Overview of the paper

Since the first published account of Qualitative Process theory [Forbus, 1981b], the ideas have been clarified and expanded considerably. At present an implementation is underway, but not complete.

The next two sections provide the basic definitions for the qualitative representation for quantities and the notion of a physical process. Quantities are discussed first because they are required for the process definitions. The three sections after that examine different kinds of reasoning that can be performed, including reasoning about the limits of processes ("What might happen if this valve is left open?"), consequences of alternate situations ("What would happen if the stove were hotter?"), and a discussion of issues involved in causal reasoning. Section 7 contains several extended examples, including modelling a boiler, motion, materials, and an oscillator. Finally the theory is placed into the perspective of similar work in Artificial Intelligence, and possible applications are discussed.

The exposition of the theory is intended to be informal but precise. Axioms are used only when they will lead to clarity. Although a full axiomatic description might be desirable, the technical details involved appear complex and the resulting gains appear small. When they appear axioms are usually written in a Lisp-like notation for clarity, with occasional lapses into standard predicate calculus notation. The underlying logic is assumed to be sorted.

## 2. Quantities

Processes affect objects in various ways. Most of these effects can be modelled by changing parameters of the object, properties whose value is drawn from a continuous range. The representation of a parameter for an object is called a quantity. Examples of parameters that can be represented by quantities include the pressure of a gas inside a container, one dimensional position, the temperature of some fluid, and the magnitude of the net force on an object.

A quantity consists of two parts, an amount and a derivative. These will be denoted  $A$  and  $D$  respectively. Amounts and derivatives are assumed to be some kind of number, in that they take on values, have distinguished parts sign and magnitude, and their values can be combined and compared in the same way numbers in mathematics are. The derivative of a quantity can in turn be the amount of another quantity (for example, the derivative of (one dimensional) position is the amount of (one dimensional) velocity). The notation for the parts of quantities is:

- $A_m$  - "magnitude of the amount"
- $A_s$  - "sign of the amount"
- $D_m$  - "magnitude of the derivative", or "rate"
- $D_s$  - "sign of the derivative"

The value of the amount of a quantity is defined in terms of its quantity space. A quantity space is a collection of quantities and numbers which form a partial order. Figure 1 illustrates the quantity space for the levels of fluid in two tanks A and B connected by a pipe. The elements which comprise the Quantity Space for a particular quantity will be drawn from the definitions of the kinds of processes and conditions that involve it. This means there will only be a finite number of elements in any reasonable Quantity Space, making it a good symbolic description.

Note that the orderings and other relations among elements in a quantity space need not be fixed over time, for the elements can be other quantities. A notation is needed to distinguish the different values a quantity has in different times and situations. In previous formalizations of common sense reasoning  $T$  is often used as an operator to tie the truth of a statement to a situation or possible world [Moore, 1979][McDermott, 1981]. An example from the blockworld is:

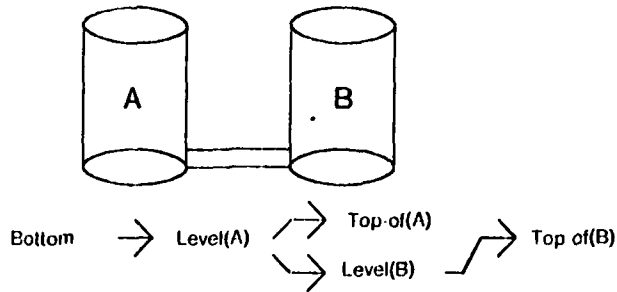
$(T (ON A B) (After (PUTON A B S0)))$

In addition to  $T$ , the operator  $M$  is introduced to denote "measuring" the value of a quantity or part of a quantity in a particular situation. The notation is:

$(M <quantity or part of quantity> <interval or instant>)$

Fig. 1. Graphical Notation for a Quantity Space

The arrow indicates that the quantity at the head is greater than the quantity at the tail. As drawn, Level(B) and Top-of(A) are unordered.



An example would be a statement about the effects of filling a container:

(greater-than (M A<sub>m</sub>(Level(c1)) (end (Filling c1))))

(M A<sub>m</sub>(Level(c1)) (start (Filling c1))))

Now the Quantity Space can be defined more formally. The Qspace of a quantity will consist of a set of elements (numbers, often the amounts of quantities) *N* and a set of orderings. The value of a quantity *Q* will be the ordering relations between *Q* and the other elements in the Qspace. The value is completely specified if the ordering between *Q* and every other element in *N* is known, and is incomplete otherwise. As with any other partial ordering, a quantity space can have a top and a bottom, such that

$\forall q \in \text{quantities}, t \in \text{times},$   
(and (not (less-than (M (A Q) t) (bottom (Qspace q))))

(not (greater-than (M (A Q) t) (top (Qspace q))))))

A quantity space can also have the distinguished element zero, such that

$\forall q \in$  quantities,

(implies zero  $\in$  (Qspace q)

$\forall t \in$  times,

(and (equiv (greater-than (M (A<sub>m</sub> q) t) zero)

(= (M (A<sub>s</sub> q) t) 1))

(equiv (less-than (M (A<sub>m</sub> q) t) zero)

(= (M (A<sub>s</sub> q) t) -1))

(equiv (equal-to (M (A<sub>m</sub> q) t) zero)

(= (M (A<sub>s</sub> q) t) 0))

(implies (greater-than zero (top (Qspace q)))

( $\forall t \in$  times (= (M (A<sub>s</sub> q) t) -1)))

(implies (less-than zero (bottom (Qspace q)))

( $\forall t \in$  times (= (M (A<sub>s</sub> q) t) 1)))

(implies (equal-to zero (bottom (Qspace q)))

( $\forall t \in$  times (or (= (M (A<sub>s</sub> q) t) 1)

(= (M (A<sub>s</sub> q) t) 0))))

(implies (equal-to zero (top (Qspace q)))

( $\forall t \in$  times (or (= (M (A<sub>s</sub> q) t) -1)

(= (M (A<sub>s</sub> q) t) 0))))))

Two points which are ordered and with no points in the ordering known to be between them will be called neighbor points. For the quantity space in figure 1, Level(A) has Bottom, Top-of(A), and Level(B) as neighbors, but not Top-of(B). Distinguishing neighboring points will be important in determining the ways that the processes acting in a situation can change.

It is important to provide "hooks" into the Quantity Space that relate it to domain concepts outside of QP theory. One example are objects with states that are defined in terms of parameter values and processes evolving, such as grain elevators being empty, full, or in between, or a four-cycle engine being in the expansion phase. Another example are operational criteria for a machine, such as keeping the fuel-air ratio in an engine with certain bounds to insure maximum fuel economy, or voltages in a circuit properly bounded so as to support the abstraction of logic signals. These connections are made by adding conditions on the quantity spaces of a situation. A condition consists of a test on quantities and

some assertion. Whenever the test holds, the assertion is true.<sup>1</sup> To say, for instance, that a grain elevator is full when it contains its capacity we would write:

```
(condition (equal-to Amount-of(grain,elevator)
                Capacity(elevator))
  (full elevator))
```

As would be expected, the change in a quantity is determined by its derivative. The contributions to the derivative of a specific quantity are represented by its set of influences. The derivative will be the sum of the numbers which are the members of this set. Often these numbers will be the amounts of other quantities, but they need not be. Determining the sign and possibly the magnitude of the derivative by examining the influence will be called resolving the influences, by analogy with resolving forces in classical mechanics. Finding and resolving the influences on the quantities of a situation is a key task in reasoning about quantities and processes. Consider for example the amount of fluid in a container where there are flows both inwards and outwards. The influences are the flow rates, and the change in amount at any time will be the sum of these rates. The sign of an influence will need to be specified as well, for a flow rate may be increasing the amount of something in one container while decreasing the amount of it in another. The cases where an influence is positive, negative, or unspecified will be written:

```
(I+ <quantity> <number>)
(I- <quantity> <number>)
(I <quantity> <number>)
```

Combining influences requires combining  $d_s$  values. Figure 2 illustrates how.

Before we can talk about integrability, we must define some simple ideas of time. An instant is a

---

1. It is easy to specify the inference in the reverse direction by defining the appropriate implication, so this definition loses no power. It is assumed that the implementation of the process theory is modular, so stating an implication as a condition allows the interpreter to know that this conclusion is one it must draw when relevant because it is needed elsewhere. An implication is used rather than logical equivalence to provide modularity; the conclusion can be mentioned in more than one condition statement rather than using a single statement with a disjunctive quantity condition if desired.

Fig. 2. Combining  $d_s$  Values

This table specifies how  $d_s$  values combine across addition and multiplication. deKleer's formulation used the symbol ? to denote the result for the cases which require information about amounts and rates.

	A	B	Result
D	0	0	0
s	0	1	1
	0	-1	-1
V	1	0	1
a	1	1	1
l	1	-1	(see below)
u	-1	0	-1
e	-1	1	(see below)
	-1	-1	-1

When equivocal,

if + then if  $d_m(A) > d_m(B)$  then  $d_s(A)$   
           if  $d_m(A) < d_m(B)$  then  $d_s(B)$   
           otherwise 0  
 if \* then if  $d_m(A) * \Lambda(B)$   
            $> d_m(B) * \Lambda(A)$  then  $d_s(A)$   
           if  $d_m(A) * \Lambda(B)$   
            $< d_m(B) * \Lambda(A)$  then  $d_s(B)$   
           otherwise 0

point in time.<sup>2</sup> The time of an instant is a mapping to an (implicit) global time scale. As with quantities, we assume for the sake of comparison and combination that times are drawn from the reals but that we don't know their values.<sup>3</sup> The usual intuitive relations and laws for instants will hold:

2. I do not agree with Allen's arguments [Allen, 1981] that including "points" in a temporal representation must lead to inconsistencies. Clearly certain events which are point-like in one view (such as a collision) really turn out to be a string of events happening over an interval, but that does not mean we should deny ourselves the convenience of the more abstract view. There also are some events which are fundamentally point-like, if we believe in continuity. Suppose a container which has been filling up suddenly starts emptying. If we assume its net flow is continuously changing, then there was some instant when it went to zero in changing from positive to negative.

3. This is different from [McDermott, 1981], which includes arbitrary known numbers, and thus can compute a numerical duration by subtracting the times for instants. One can easily imagine extending Qualitative Process theory by allowing information about numerical values, but that is not the goal of the base theory.

$\forall i_1, i_2 \in \text{instants.}$

(equiv before( $i_1, i_2$ ) (less-than (time  $i_1$ ) (time  $i_2$ )))

$\wedge$  (equiv after( $i_1, i_2$ ) (greater-than (time  $i_1$ ) (time  $i_2$ )))

$\wedge$  (equiv simultaneous( $i_1, i_2$ ) (= (time  $i_1$ ) (time  $i_2$ )))

An interval is a span of time, consisting of two distinguished instants start and end, with the set of instants in between denoted as its during. The duration of the interval I is defined by  
(= (duration I) (- (time (end I)) (time (start I))))

and of course

(implies (Interval I) (not (before (end I) (start I))))

or equivalently,

(implies (Interval I) (not (less-than (duration I) zero)))

The set of instants and intervals will be collectively known as times.

Now the relationship between amounts and derivatives can be defined. Basically, if the amount is increasing for a while there will be more of it, decreasing then less of it, and if it isn't changing it will remain the same.

$\forall q \in \text{quantities, } I \in \text{intervals,}$

(implies (and (constant-sign D I)

(not (equal-to (duration I) zero)))

(and (equiv (= (M  $D_S(q)$  (during I)) -1)

(< (M  $A_m(q)$  (end I)) (M  $A_m(q)$  (start I))))

(equiv (= (M  $D_S(q)$  (during I)) 0)

(= (M  $A_m(q)$  (end I)) (M  $A_m(q)$  (start I))))

(equiv (= (M  $A_m(q)$  (during I)) 1)

(> (M  $A_m(q)$  (end I)) (M  $A_m(q)$  (start I))))))

where

$\forall n \in \text{numbers, } I \in \text{intervals,}$

(equiv ( $\forall i_1, i_2 \in$  (during I) (= (M  $s(n)$   $i_1$ ) (M  $s(n)$   $i_2$ )))

(constant-sign n I))

This statement is very weak compared to our usual notion of integrability. It would be interesting to discover what other notions exist that are richer than this one but still weaker than the Calculus.

A key notion of this theory is that the physical processes in a situation induce functional dependencies between the parameters of a situation. In other words, by knowing the physics you can tell what, if anything, will happen to one parameter when you vary another. In keeping with the exploration of the weakest information we can, we define

( $\propto_Q Q_1 Q_0$ )

(read " $Q_1$  is qualitatively proportional to  $Q_0$ ") to mean "there exists a function induced by a process which

is increasing and monotonic in its dependence on  $Q_0$ , such that  $f$  determines  $Q_1$  and is dependent on at least  $Q_0$ ." In algebraic notation, we would write

$$Q_1 = f(\dots, Q_0, \dots)$$

If the function is decreasing monotonic, we will say

$$(\propto_{Q_0}^- Q_1 Q_0)$$

and if we don't wish to specify if it is increasing or decreasing,

$$(\propto_{Q_0}^* Q_1 Q_0)$$

This definition of  $\propto_{Q_0}$  is motivated by issues involved in causal reasoning, as will be made clear in Section 6. Aside from  $\propto_{Q_0}$  and its variants, the only other information that will be specified about the implicit function is a finite set of correspondences it induces between points in the two quantity spaces. An example of a correspondence is that the force exerted by an elastic band is zero when it is at rest. This would be written:

(correspondence (Internal-Force(band) zero)  
(Length(band) Restlength))



### 3. Processes

A physical situation is usually described in terms of a collection of objects, their properties, and the relationships between them. An important fact of the physical world is that things can change. The ways in which things change are intuitively characterized as processes. To build a theory of a particular kind of physical reasoning, a theory of the processes that occur in the domain at hand must be built. A process is something that acts through time to change the parameters of objects in a situation. Examples of simple processes include fluid and heat flow, boiling, motion, collisions, stretching and compressing.

Qualitative Process theory is based on the assumption that processes are the sole cause of changes in the parameters of objects.<sup>1</sup> A physics for a domain consists of a description of the class of objects included in it and a vocabulary of processes that occur in the domain. A situation would then be described as a collection of objects and their properties, the relations between them, and the processes that are occurring. If it is assumed that the process vocabulary for a domain is complete, then all the ways a quantity can change are known. This makes reasoning by exclusion possible. Without this form of "closed world" assumption (see [Moore, 1975] or [Reiter, 1980]) it is hard to see how a reasoning system could debug or extend its physical knowledge. The further consequences of this assumption will be explored after simple processes are defined.

A process is specified by five things:

- o The individuals it applies to,
- o A set of preconditions, which are statements about the individuals and their relationships other than quantity conditions,
- o A set of quantity conditions, which are either assertions of inequalities between quantities belonging to the individuals (and perhaps some domain-dependent constants) or assertions that particular processes are active.

---

1. When viewed as a possible psychological model, I call this assumption the strong physical causality conjecture. The weak version is that processes are the major cause of changes in objects, but certain changes are considered "magic". A priori, the weaker version appears a better psychological model.

o A set of relations the process imposes between the parameters of the individuals, along with new entities that might be created.

o A set of influences (see previous section) imposed by the process on the parameters of the individuals.

A process will act on any individuals to which it can apply, exactly whenever both the preconditions and the quantity conditions are true. Preconditions are those factors that are external to Qualitative Process theory, such as someone opening or closing a valve to establish a fluid path. The quantity conditions are those limits and conditions that can be deduced within Qualitative Process theory, such as requiring the temperature of two bodies to be different for heat flow to occur, or boiling to occur as a prerequisite to generating steam.

The set of relations associated with a process are the constraints it imposes between the parameters of the objects it is acting on. Relations usually concern amounts and rates, but can include the introduction of new entities. Examples are flow rates and the amount of steam produced by a boiling. The influences specify how these new entities affect the quantities of the objects involved in the process - for example, the flow rate will correspond to the increase in the amount of "stuff" at the destination of a flow and to the decrease in the amount of "stuff" at the source. To find out how a quantity is actually changing requires summing all of the influences on it, since several processes may be acting at once. Figure 3 illustrates process specifications for heat flow and boiling. Qualitative Process theory concerns the form of physical theories, not their specific content. For example, the heat flow process illustrated adheres to energy conservation, and does not specify that "stuff" is transferred between the source and destination. The language provided by the theory also allows a heat flow process that violates energy conservation and transfers "caloric fluid" between the source and destination to be written.<sup>1</sup> The assumptions made about the content of physical theories are weak indeed.

The quantities and constants that are compared to a particular quantity by quantity conditions in the process vocabulary are included in its quantity space (the other source of quantity space elements

---

1. Another example is the "kludge" in the boiling process - adding a negative influence to the heat of the water equal to the flow rate prevents deducing that the temperature of the water rises during boiling. A better solution would be to explicitly use energy transfer, but I haven't worked out the details.

## Fig. 3. Physical Process Definitions

Heat-Flow(s,d)  
 individuals (s object with heat)  
               (d object with heat)  
 Precondition:  $\exists(p) \text{Heat-Path}(p,s,d)$   
 QuantityCondition:  $A_m(T(s)) > A_m(T(d))$   
 Relations: Let fr be a number, (greater-than fr zero)  
               ( $\propto_Q$  fr (-  $A_m(T(s))$   $A_m(T(d))$ ))  
 Influences: (I- Heat(s) fr)  
               (I+ Heat(d) fr)

Boiling (w)  
 individuals (w contained fluid)  
 Precondition:  $\exists(f) \text{AirSpace}(f)$   
                $\wedge \text{Shared-Face}(f,w)$   
 QuantityCondition:  $\exists(s) \text{Heat-Flow}(s,w)$   
                        $A_m(T(w)) = A_m(T(\text{boil}(\text{made-of}(w))))$   
 Relations:  $\exists(g \text{ individual}) \text{Gas}(g) \wedge \text{made-of}(g) = \text{made-of}(w)$   
               Let gr be a number.  
               ( $\propto_Q$  gr  $D_m(\text{Heat}(w))$ )  
                $T(g) = T(w)$   
 Influences: (I- Heat(w) fr(Heat-flow(s,w)))  
               (I- A(w) gr)  
               (I+ A(g) gr)

A useful law for dealing with heat is

$$\forall(s) (\propto_Q T(s) \text{Heat}(s))$$

To model a heat source,

$$D_S \text{Heat}(s) = 0 \text{ and so } D_S(T(s)) = 0$$

are from the condition mechanism). Because they correspond to discontinuous changes in the processes that are occurring, they are called limit points. Limit points serve as boundary conditions. For example, the temperature quantity space for a fluid would include the limit points:

$$T(\text{ice}) \rightarrow T(\text{boiling})$$

where temperatures at  $T(\text{ice})$  and below correspond to the solid state, temperatures of  $T(\text{boiling})$  and above correspond to the gaseous state, and any temperature in between to being a liquid.

Just as the ontology for physical situations at a particular instant is extended by the addition of processes, the representation of change must also be extended. The history representation introduced by Hayes in the axioms of liquids [Hayes, 1979] is assumed as a starting point. To summarize, a history of an object consists of episodes, each of which corresponds to "something happening" to an object. Each episode consists of a where and when, comprising the spatial-temporal bounds of that occurrence. Interactions can occur only when episodes intersect. The kind of histories introduced in the liquids paper will be called parameter histories, because they describe the change in a particular parameter of the object. A process history is the description of the processes happening to a particular collection of

objects, where the *when* of an episode in the process history is the interval over which that instance of the process is acting. The history for an object is just the union of all its parameter histories and the process histories it participates in. The use of histories is illustrated in the examples section.

The language of processes is made considerably more useful by applying ideas about languages gleaned from computer science. A good language should include some notion of primitives, means of combining these primitives, and means of abstraction which allow new primitives to be defined. The things we intuitively characterize as processes are of course primitives, but the descriptions of objects may also be viewed as primitives. The "process of being a gas", for instance, implies that the qualitative relation

$$P(G) * V(G) \cdot \Lambda(G) * r(G)$$

holds. Presumably a richer process vocabulary would contain the "mechanisms" that induce this relationship, but there is no reason to always include such detail. Consider for example a resistor in a circuit that never exceeds its electrical capacity. The detailed mechanics of conduction hinder rather than help when calculating the current that will result from a voltage across it. This way of including relationships between parameters that hold by "the nature of" an object allows some of the processes causing changes to remain implicit.

The means of combining processes is by sharing parameters and by sequentiality, and the means of abstraction is giving these combinations names. An example of a shared parameter combination is the compression stroke of a four stroke engine. An example of a sequential combination is a cycle in an oscillator. Treating these combinations as new processes then allows properties of the system they describe to be reasoned about, such as deducing the conditions under which a pumped oscillator will remain stable.<sup>1</sup> Note that sequential combinations require including a behavioral description - the process history for the individuals - in the relations of the compound process. Encapsulating behavior in the definition of a process also allows certain primitive processes to be described, such as collisions (see below).<sup>2</sup>

---

1. One use of compound processes would be representing the device models in deKleer and Brown's theory of machines [deKleer & Brown, 1982]. The preconditions and quantity conditions of the compound process would correspond to their assumptions about the validity of the device model.

2. This should also allow the representation of diSessa's "phenomenological primitives" [diSessa, 1982]

It is apparent that the notion of process should be extended to form a hierarchy via classification as well as by composition. For example, there are many kinds of motion - flying, sliding, swinging, and rolling. Sliding and rolling are examples of motion along a surface, and along with swinging form motions involving constant contact with another object. Having explicit abstract descriptions of processes should be useful because they are often easier to rule out than more detailed descriptions. If, for instance, there is no path between two places through which an object can be moved, it cannot get there by sliding, flying, rolling, or any other kind of motion that might exist. The notation for hierarchial processes is still under development.

## 4. Limit Analysis

The definitions of quantities and processes above provide enough formal structure to deduce, given a physics and a very general description of a situation, what processes are occurring and the changes they will cause. The preconditions and quantity conditions can be used to determine what processes are operating within the situation. This information can in turn be used to deduce changes in the properties of the situation (such as a temperature rising or an amount dropping) and the limits of the processes involved.

To infer the limits of a process, first find the influences on all affected quantities and determine the resulting  $D_s$  value. Then find the neighboring points within the quantity space. If there is no neighbor in a direction, then motion in that direction cannot affect the process. The ordering between each neighbor and the current amount of the quantity can be combined with the  $D_s$  values of each to determine if the relationship will change (see Figure 4). If the neighbor is a limit point, some process may end there and others begin. Thus the set of possible changes in orderings involving limit points becomes the ways the current set of active processes might change. This assumes that rates are non-infinitesimal, so

Fig. 4. Linking derivatives with Inequalities

This table summarizes how the ordering relationship between two quantities may change according to the sign of their derivatives over some interval.

	A	>	B	Next Relation
D	0	0	0	>
s	0	1	1	=
	0	-1	-1	>
V	1	0	0	>
a	1	1	1	If $D_m(A) > D_m(B)$ , .> ; < implies =, = implies >
l	1	-1	-1	>
u	-1	0	0	=
e	-1	1	1	=
	-1	-1	-1	If $D_m(A) > D_m(B)$ , .> ; < implies >, = implies >

	A	=	B	Next Relation
D	0	0	0	=
s	0	1	1	<
	0	-1	-1	>
V	1	0	0	>
a	1	1	1	If $D_m(A) > D_m(B)$ , .> ; < implies <, = implies =
l	1	-1	-1	>
u	-1	0	0	<
e	-1	1	1	<
	-1	-1	-1	If $D_m(A) > D_m(B)$ , .> ; < implies >, = implies =

that if a quantity is moving towards some point in its space it will actually reach that value in some finite time.<sup>1</sup>

More than one change is typically possible, as the examples in Section 7 will illustrate. There are three reasons for this. First, if the ordering within a quantity space is not total more than one neighbor can exist. Second, a process can influence more than one quantity. Finally, more than one process can be occurring at once. For some kinds of tasks just knowing the possible changes is enough (such as envisioning, in [deKleer, 1975]). If required, knowledge outside the scope of Qualitative Process theory can be used to disambiguate the possibilities. Depending on the domain and the style of reasoning to be performed there are several choices: simulation [Forbus, 1981], algebraic manipulation [deKleer, 1975], teleology [deKleer 1979], or possibly by default assumptions or observations (discussed in [Forbus, 1982]).

---

1. Note that relaxing this assumption would result in only one additional state in the possibilities returned by the limit analysis - that the current set of active processes never changes.

### 5. Incremental and Differential Qualitative Analysis

An important technique for understanding physical systems is Incremental Qualitative (IQ) analysis, introduced in [deKleer, 1979]. Qualitative Process theory provides an interesting way to understand Incremental Qualitative analysis, and suggests a variant reasoning technique that appears quite useful.

In Incremental Qualitative analysis, the value of a quantity is represented by the sign of its change from some earlier situation - either "increasing", "decreasing", or "same". Note that an IQ value of a parameter is not the same as the sign of the parameter's derivative, but a comparison of the amount at two different times. The insight underlying this choice is that an important part of understanding machines (or other physical situations) is to understand how they respond to a perturbation.

The IQ model of a system is built from device models for its parts, each of which defines the relationships it imposes between its parameters. A resistor, for example, would be modeled as:

Voltage Change  $\Leftrightarrow$  Current Change

The analysis proceeds by assigning an increase or decrease to the input of the machine and propagating the effects of this change through a model of the structure of the device. Causality is imposed to the sequencing of events in the simulation, although the qualitative nature of the description usually requires extra assumptions or information to reduce the result down to a single sequence representing the behavior of the device. deKleer elegantly demonstrated that causal arguments of this form, along with the assumption that every device in a circuit serves some purpose, suffices to recognize a large class of electronic circuits. The results of IQ analysis can also be used to generate English explanations with interleaved animation to provide intelligent Computer Aided Instruction [Forbus, 1980].

The lack of an explicit representation of processes severely limits the applicability of IQ analysis. IQ analysis implicitly assumes that the system is in some equilibrium, and that the changes in the system don't cause the "region of behavior" of the device models to change. This corresponds to assuming a fixed process description for the situation. The only encoding of such state information is in state variables of the objects in the situation being modelled. For instance, a simple diode model is:

```
if diode is ON, Change in Voltage = 0
if diode is OFF, Change in Current = 0
```

The first case corresponds to a current flow, the second to a situation where the reversed voltage means



there is no flow. deKleer determined values for the state variables in the system by looking for globally consistent interpretations under the teleological assumption mentioned previously. Another way to determine the active processes that appears much simpler (and is available even when the system is not an engineered device) is to make assumptions about the relative magnitudes of the quantities in the system.<sup>1</sup>

Deducing that a state variable will change as a result of events in the qualitative simulation - true time domain analysis - is impossible with IQ analysis alone. As an experiment, deKleer tried to capture such changes by adding rules to the device models which stipulated cases where change was possible [deKleer, 1979]. The graph of possibilities was then pruned by using consistency relations between states. The computation proved unwieldy. From the present perspective we can see why - a process corresponds to a consistent set of states, and there can be several state variables which depend on a process. Processes introduce a new "locality" which makes sense physically and reduces the combinatorics of the problem considerably. In addition, the richer notion of time and quantity in Qualitative Process theory may make true qualitative time domain analysis possible.

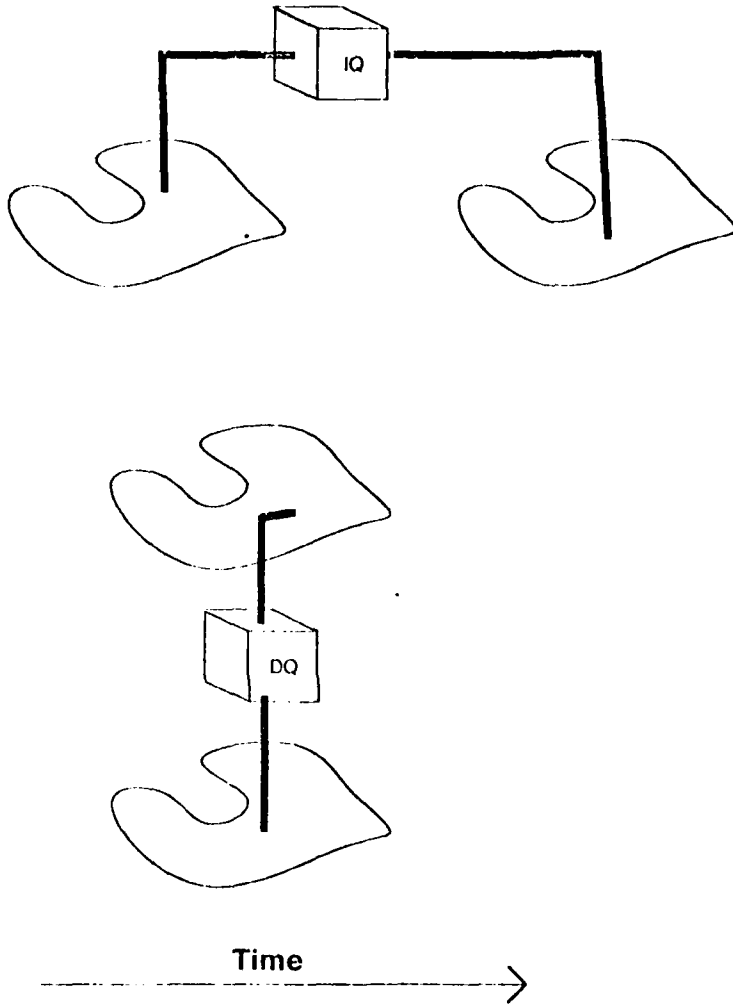
The idea of a comparison in IQ analysis suggests a complementary qualitative reasoning technique. IQ analysis concerns the relationship between two situations, one of which is a consequence of things happening in the other. Another case of interest concerns situations which are just slightly different from one another. For instance, we often have an idea of the different consequences that would result if something were changing a bit faster - if we put the heat up on the stove the water in the kettle would boil sooner, and if our arm were quicker the serve would have been returned. The language in which such conclusions are expressed is in part the same as that used in IQ analysis - amounts are either the same, increased, decreased, or indeterminate as compared with the old situation. The difference is in "where the measurements are taken", as suggested by Figure 5. Answering these kinds of questions will be called differential qualitative analysis.

Let us consider a situation A. If we get a new situation B by changing some ordering in A or by changing a single process in A, we will call B an alternative to A. There are two kinds of changes which may occur as a result of perturbing A. First, the process history for the situation itself may change, apart

---

1. Jerry Roylance has suggested that an expert circuit designer has a good enough idea of the numerical values in a circuit to do this [personal communication]

Fig. 5. Difference between IQ and DQ values



from any changes made to define B in the first place. An example would be punching a hole in the bottom of a kettle, which could let all the water drain out before a boiling occurs. Even changes in orderings can lead to historical consequences - if we reduce the intensity of a flame while still agreeing that it will be turned off in five minutes, boiling may again be prevented. For simple DQ analysis we are

interested in the case where the process history remains the same.<sup>1</sup>

Let  $DQ(q, A, B)$  for some quantity  $q$  be the sign of the difference between two situations  $A$  and  $B$  that are alternatives. Then the inequality order between them defines DQ values, as follows:

(greater-than (M q A) (M q B))  $DQ(q, A, B) = 1$

(less-than (M q A) (M q B))  $DQ(q, A, B) = -1$

(equal-to (M q A) (M q B))  $DQ(q, A, B) = 0$

The inequality orderings for instants must of course be extended to apply over intervals. For equality this is simple:

$\forall q_1, q_2 \in \text{quantities}, i \in \text{intervals}$

(equiv (equal-to (M  $q_1$  i) (M  $q_2$  i)))

( $\forall i_1 \in (\text{during } i) (\text{equal-to (M } q_1 \text{ } i_1) (M \text{ } q_2 \text{ } i_1))))$ )

For the other cases the choice is less clear. The strongest version of *greater-than* is having it hold over every instant in the interval:

$\forall q_1, q_2 \in \text{quantities}, i \in \text{intervals}$

(equiv (greater-than (M  $q_1$  i) (M  $q_2$  i)))

( $\forall i_1 \in (\text{during } i) (\text{greater-than (M } q_1 \text{ } i_1) (M \text{ } q_2 \text{ } i_1))))$ )

but for extending our notion of integrability, the following will also suffice:

$\forall q_1, q_2 \in \text{quantities}, i \in \text{intervals}$

(equiv (greater-than (M  $q_1$  i) (M  $q_2$  i)))

(and ( $\exists i_1 \in (\text{during } i) (\text{greater-than (M } q_1 \text{ } i_1) (M \text{ } q_2 \text{ } i_1))))$ )

( $\forall i_1 \in (\text{during } i) (\text{not (less-than (M } q_1 \text{ } i_1) (M \text{ } q_2 \text{ } i_1))))$ )

A version of *less-than* for intervals may be similarly defined.

Let us use DQ analysis to express the relationship between rate, duration, and "distance" for a quantity that is changing during an interval. Intuitively we know that if the rate increases or decreases, the duration of time will decrease or increase, or the "distance" the value moves will increase or decrease for the same duration. Implicit in this simple intuition is the restriction that the rate is constant during the interval, i.e., that the function defining the change of the quantity is linear and time invariant. This often is not the case, so we must require that either the beginning or the end of the two episodes being compared are the same. If we apply DQ analysis only to alternative situations this restriction will be

---

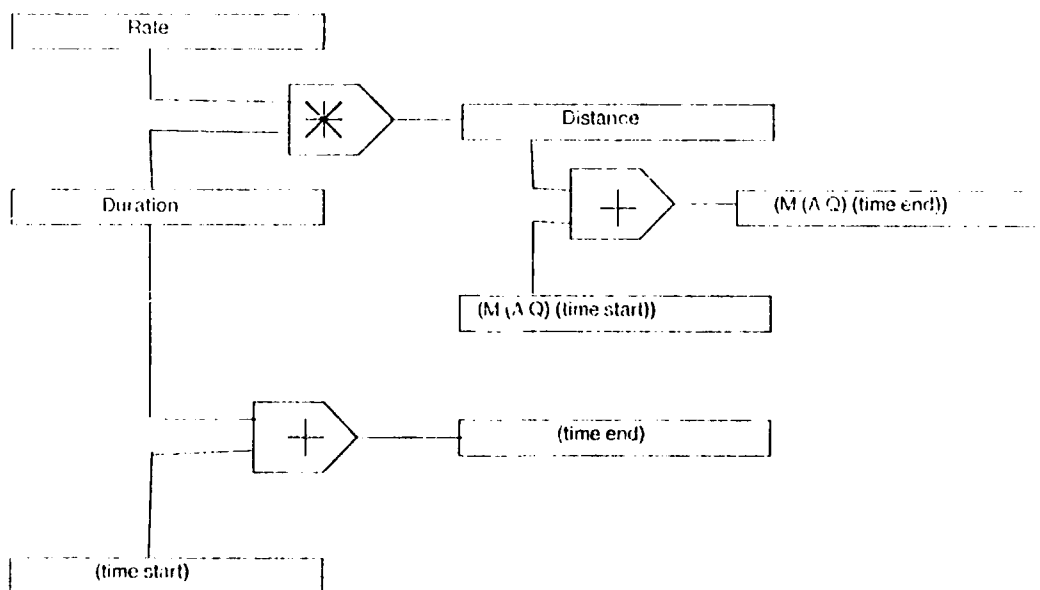
1. Analyzing changes in the process structure requires a better vocabulary for histories than I have at present.

satisfied. The desired relationship is simple - the difference in "distance" is just the product of the differences in rates and durations. The DQ values combine as do  $d_s$  values, and this sort of relationship lends itself easily to deduction using constraint network techniques, as illustrated in Figure 6.

Differential Qualitative analysis should prove useful in characterizing other kinds of similar situations. Part of the job of describing the states of a complex system can be performed by describing DQ values for quantities in situations where different states hold (see Transmission example below). DQ values are also useful in analyzing behavior of a system during the construction of compound processes

Fig. 6. Comparing alternate situations

This network can be interpreted by the usual conventions of constraint networks to yield the kinds of deductions possible about the relationships between the rate of change during the interval, the duration of the interval, and the change of value for a quantity during the interval. The particular argument shown below characterized as "If it had moved faster it would have gotten there sooner".



Either (Time end) or (Time start) must be constant.  
 Assume constant distance,  
 Rate 1 => Duration -1 => (time end) -1 or (time start) 1  
 Rate -1 => Duration 1 => (time end) 1 or (time start) -1  
 Similarly for constant duration.

Kenneth D. Forbus

-25-

QP Theory

(see oscillator example below).

## 6. Causality and Functional Dependence

Causality is an important concept in understanding physical systems. One component of our notion of causality is found in IQ analysis, in particular knowing that "changing A will cause B to change". This kind of causality will be called incremental causality. It can be usefully defined within Qualitative Process theory, and doing so explains a problem found by several workers in implementing systems to perform causal arguments.

I claim the statement "A change in A causes a change in B" is equivalent to "The processes in the situation induce a functional dependence of B on A". Making B functionally dependent on A insures that a change in A will result in a change in B. Requiring the function to be a consequence of processes occurring in the situation introduces a mechanism for the change. To see the importance of having an underlying physical mechanism for the change, consider an abstract rectangle. By definition, its area is the product of the length of its sides. If we then imagine a longer rectangle, we know its area is larger. There is no sense of causality in the change between the two rectangles. Now imagine the first rectangle to be made of an elastic material, and the second rectangle obtained by stretching the first. In this case it makes sense to say "the increase in length causes the area to increase". The relational description is the same in both cases, only the assumption of an underlying process is different.

This example illustrates a problem that has arisen in implementing systems to construct causal arguments.<sup>1</sup> The physical system under study is modelled by the relationships between its parameters, and Incremental Qualitative analysis is used to construct descriptions of how the system changes. Each change in the model is interpreted as a change in the system, with the order of computation being identified with the order of events in the system. Figure 7 contains fragments from two of the models.<sup>2</sup> Sometimes an assumption is needed to make further deductions. Suppose while using the "stuff" model we have reached the fragment shown in the top of the figure, concluding that the heat is increasing. We

---

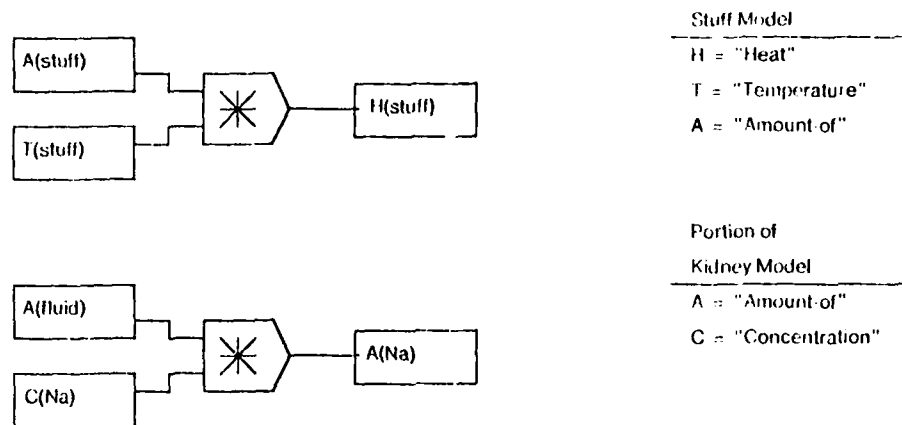
1. The problem was observed in implementing the model of a student's understanding of a heat exchanger described in [Williams, et. al., 1982], in my own work on understanding Automatic Boiler Control systems, and in the kidney model described in an unpublished paper by Irwin Asbell.

2. The systems were implemented in CONLAN [Forbus, 1980], a constraint language. The notation is similar to that of logic diagrams, except that the terminals are given explicit names and the devices are multi-functional.

Fig. 7. Constraint representation of relationships

(a) is drawn from the model for a piece of "stuff" used in an effort to represent a student's understanding of heat exchangers.

(b) is drawn from an IQ model of a kidney to be used in explaining the syndrome of inappropriate secretion of anti-diuretic hormone (SIADH).

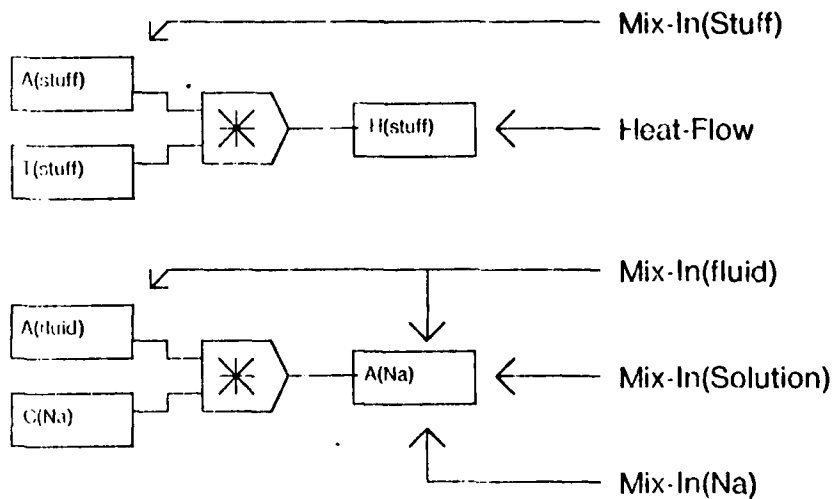


are at an impasse, for either the change in amount or the change in temperature must be known to deduce the change in the other. A further assumption is both reasonable and prudent, but which one? We could assume that the temperature is constant, in which case we deduce an increase in the amount of "stuff". While a reasonable deduction, it is not the case that "the increased heat causes the amount of stuff to increase". The only causal interpretation is backwards - that "the increased heat must have come about because the the amount of water increased, because the temperature is constant". A similar problem occurs in modelling the effects of sodium retention in the kidney model, a fragment of which is shown in the bottom of the figure. Increasing sodium will cause the amount of water to increase, if the rest of the kidney is working as it should - but the mechanism involved is a complicated feedback system which depends first on detecting the increased concentration, not the definition of concentration itself! The only proper causal conclusion that can be drawn from this fragment is that the concentration is increasing. How can we determine in general which assumption to make?

The answer is to explicitly consider the processes that cause the changes in the system. Figure 8

**Fig. 8. Model Fragments with possible Processes**

Here are the models from the previous figure with the quantities annotated with the (likely) processes that might affect them. Note that certain quantities (temperature, concentration) cannot be directly changed. These are dependent quantities, and should not be the subject of assumptions in building causal arguments.



illustrates the fragments shown previously with annotations corresponding to a hypothesized set of processes that could affect them. The assumption that a quantity is constant is the assumption that either no process is influencing it or that the influences on it cancel. In the "stuff" model, for instance, it is reasonable to assume (unless we know otherwise) that no mixing is occurring. This leads to the conclusion that the temperature is increasing. Assumptions to further causal arguments should be about processes, not the values of quantities! In particular, quantities which cannot be directly affected by a process should not be the subject of assumptions in causal reasoning, for they are dependent on other quantities. The enriched ontology provided by processes also allows explanations of anomalous conditions should a causal argument be found in error. If for instance there is a heat flow into some stuff and the temperature is still not rising then either the stuff is changing state or enough colder stuff is being added to swamp the influence of the heat flow. Knowing just that the constant amount assumption was violated begs the question of by what. A mixing or a boiling is something that can be observed, or at least has other



measurable effects - if a program were dealing with a physical system it could try to find out which of the alternatives is occurring.

One goal of research into common sense physical reasoning is to develop a theory of observation - a set of techniques for planning experiments on a system and using the results to figure out how it works. One prerequisite of such a theory is a theory of what the result looks like. Qualitative Process theory appears to be a good candidate for a target language, since it dictates the form of the theories more than their content. The notion of incremental causality is also important, for we often experiment with a system by changing something and seeing what happens. Ascribing causality to a change is assigning credit for the observation to some theory of the situation. Some notation for local causal connections is required to express these simple observations. This requirement was a major motivation in the definition of  $\alpha_Q$  (see Section 2), which asserts that a process induces a functional dependence between two quantities. If whenever parameter A in a system is poked parameter B changes, the result can be expressed as  $(\alpha_Q B A)$ .

More powerful statements about a system being understood will require extensions of  $\alpha_Q$ . To see what is involved, consider the analogous situation of learning how a typewriter works.<sup>1</sup> If the space bar is pushed, the carriage will move to the left. This is the kind of statement that can be made with  $\alpha_Q$ . But lots of other things can happen to move the carriage, namely all of the letter keys and a few more. Thus it would be useful to be able to state all of the influences (at least, within the current grasp of the situation) on some particular parameter. Suppose also that we just wanted to move the paper up without changing anything else. The return bar would move the paper up, but before doing so would return the carriage to the right. Being able to say there are no (known) intervening parameters is then also a useful ability.

To see how these notions can be expressed, consider the collection of  $\alpha_Q$  relations that hold at some instant in time. For any quantity, the  $\alpha_Q$  statements relevant to it can be thought of as a tree with the dependent quantity at the root and the "independent" quantities at the leaves.<sup>2</sup> A plus or minus

---

1. This is not proposed as a serious example because the quantity definitions and  $\alpha_Q$  would apply only in a very abstract sense.

2. Actually a directed graph with cycles can be formed, as for instance in a control system

denotes the sense of the connection (whether or not it will reverse the sign of the change in the input). ( $\alpha_Q \cdot Q_1 Q_0$ ), then, only specifies that  $Q_1$  is on some branch "above"  $Q_0$ .

Figure 9 illustrates such a dependency tree. Suppose we are trying to cause  $Q_0$  to change. If we don't want to change  $Q_2$ , then  $Q_3$  or  $Q_1$  are our only choices. We need a way to express that (at least within our knowledge of the situation) there are no intervening parameters. To say this, we use

( $\alpha_Q$ -direct  $Q_0 Q_1$ )

which can be modified by +, -, or \* as before.  $\alpha_Q$ -direct adds a single link to the tree of dependencies. Another problem is to find all the ways to bring a change about, or to prove that changing one thing won't cause a change in some other quantity of interest. We do this by stating that a particular collection of quantities together "closes off" the tree - there will be exactly one quantity for each branch. Our notation will be

( $\alpha_Q$ -all <quantity> <plus-set> <minus-set>)

which means that there is a function induced by a process which determines the quantity, and which relies on the quantities in the two sets solely. If a quantity is not mentioned in a  $\alpha_Q$ -all statement, then either it is irrelevant to the quantity of interest, it depends on some quantity in the  $\alpha_Q$ -all statement (above the slice of the tree which it makes), or some quantity in the  $\alpha_Q$ -all statement depends on it. By ruling out the other two possibilities, independence can be established.

As a rule  $\alpha_Q$  statements will not hold for all time. In the typewriter analogy, imagine the carriage at the end of its travel - hitting the space bar will no longer result in movement. More to the point, consider  $Q_0$  given by:

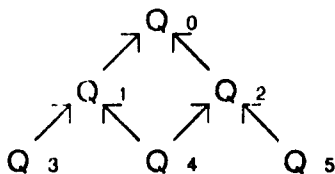


Fig. 9. A tree of functional dependencies

$$Q_0 = (a - b \cdot Q_2) \cdot Q_1$$

if  $a > b \cdot Q_2$ , ( $\alpha_Q Q_0 Q_1$ )

$a = b \cdot Q_2$ , (not ( $\alpha_Q \cdot Q_0 Q_1$ ))

$a < b \cdot Q_2$ , ( $\alpha_Q \cdot Q_0 Q_1$ )

In the case of equality,  $Q_0$  and  $Q_1$  are not related at all, and in the other two cases the sign of the function connecting them is different. Thus the collection of  $\alpha_Q$  statements which are true for a system can vary as a function of the values of the quantities as well as changes in the process structure of the situation. The collection of  $\alpha_Q$  statements that holds for some class of situation will define a mode of the system being described. Multi-mode systems include four stroke engines and automobile transmissions.

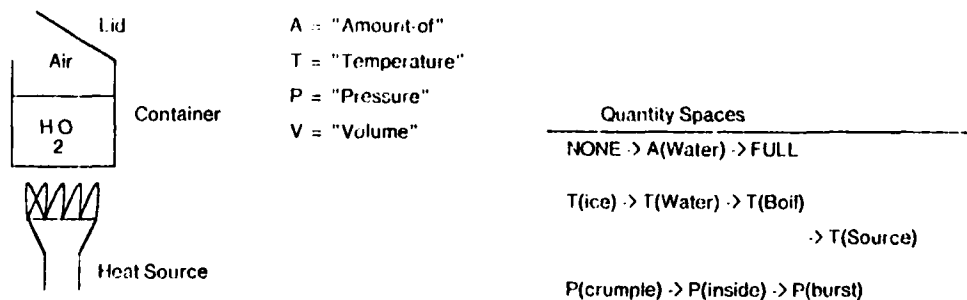
## 7. Examples

At this point a great deal of formal machinery has been introduced. It is time to illustrate how QP theory can be used in physical reasoning. The examples will be fairly informal for two reasons. First, several of the domains involve issues of spatial reasoning that are still under study.<sup>1</sup> The second reason is that the theory is not yet implemented (although a program is in progress). Still, they should provide some kind of indication as to the theory's utility.

### 7.1 Boiler

Let us consider the possible consequences of the situation shown in Figure 10. The situation consists of a container partially filled with water that can be heated by a flame; the container has a lid which can be sealed and is surrounded by air. The initial amounts are assumed to be those of standard temperature and pressure, all  $D_s$  values are initially 0. At some point in time the heat source is turned on. We will stipulate that if boiling occurs, the lid will be closed and sealed. Some of the physics required for

Fig. 10. A simple boiler and its Quantity Spaces



1. At present work is focusing on the Mechanism World, which includes the BlocksWorld but also more complex shapes and some non-rigid materials. The aim of the project is to understand devices such as mechanical watches and automobile transmissions.

this problem is contained in Figure 1. The rest of the details, and especially formalizing the geometry involved, will be ignored in this example.

When the heat source is activated, there will be a heat path between the source and the container. Assuming standard temperature and pressure in the environment, and assuming both that our process vocabulary is complete and that there are no unstated processes occurring (closed world assumption), if  $T(\text{source}) > T(\text{water})$  there will be a heat flow from the source to the water. Being a temperature source, the influence of the loss on the temperature is ignored and  $D_s(T(\text{source})) = 0$ . The only influence on  $T(\text{container})$  is that of the heat flow, so  $D_s(T(\text{container})) = 1$ . This in turn will cause a heat flow to the air surrounding the cup, the air inside the container, and the water. Most of these temperatures will be ignored. The temperature quantity space looks like:

$$\begin{aligned} T(\text{ice}) &\rightarrow T(\text{water}) \rightarrow T(\text{source}) \\ &\rightarrow T(\text{boil}) \end{aligned}$$

If  $T(\text{source}) > T(\text{boil})$  and the process is unimpeded (i.e., the preconditions for the heat flow remain true), the next process that will occur is a boiling.

Before considering the boiling, we can examine what happens to the air inside the container.

The relationship between the parameters of air due to its gaseous state can be expressed as:

$$P(\text{air}) * V(\text{air}) = A(\text{air}) * T(\text{air})$$

While the water is heating,

$$D_s(V(\text{air})) = 0 \text{ and}$$

$$D_s(T(\text{air})) = 1$$

$$\Rightarrow D_s(P(\text{air})) = D_s(A(\text{air})) + 1$$

Changes in pressure and amount of something usually result from a flow. If there is a flow then it must be either inward or outward. First assume no flow occurs. Then because the only way for Amount-of of the air to change is a flow,

$$D_s(A(\text{air})) = 0, \text{ so } D_s(P(\text{air})) = 1$$

But initially  $P(\text{air}) = P(\text{outside})$  so the conditions for a flow are established, contradicting the assumption. Can the flow be inward? If so,  $D_s(A(\text{air})) = 1$ . This requires  $D_s(P(\text{air})) = 1$ , which enables an outward flow, again a contradiction. Finally, if there is an outward flow then  $D_s(A(\text{air})) = -1$ . From the combination table for  $D_s$  values we know  $D_s(P(\text{air}))$  could be -1, 0, or 1.  $D_s(P(\text{air})) = -1$  can be ruled out because in that case an inward flow would also be occurring, violating our assumption of

outward flow. By exclusion we accept it, realizing that some ambiguity still exists.<sup>1</sup>

Suppose the preconditions for the heat flow continue to be met and boiling occurs. The amount quantity space looks like:

NONE  $\rightarrow$  A(water)  $\rightarrow$  (M A(Water) Initial)  
 $\rightarrow$  A(steam)  $\rightarrow$

The influence of the boiling on A(water) moves it towards NONE. So one of the ways the process might end is that all of the water is converted to steam. However, we must deduce the effects of the change in A(steam) to be sure we have all of the possibilities.

Because the steam is still in contact with the water their temperatures will be the same, and under normal conditions the boiling point of water is constant. However, we assumed that the container would be sealed when the boiling began. The only influence on A(steam) is from boiling because the geometry of the situation makes gas flow impossible. So  $D_S(A(steam))=1$ . If we think about what is happening in some particular instant of time we can first assume  $D_S(T(steam))=0$ , and since  $T(steam)=T(water)$ ,  $D_S(T(steam))=0$ . Steam is a gas, so its parameters are related by:

$$P(steam) * V(steam) = A(steam) * T(steam)$$

and by substitution,

$$D_S(P(steam)) + D_S(V(steam)) = 1.$$

Since the container holds only the water and steam (ignoring the air), geometry tells us

$$V(Inside(container)) = V(steam) + V(water)$$

and because the container is rigid,  $D_S(V(Inside(container)))=0$ . Also, from physics we know

$$(\propto_Q V(water) A(water))$$

and from the process description  $D_S(A(water))=-1$ . Therefore  $D_S(V(water))=-1$  and  $D_S(V(steam))=1$ .

Examining the combination table for  $D_S$  values reveals that for  $D_S(V(Inside(container)))=0$  to hold, it must be that

$$D_m(V(steam)) = D_m(V(water))$$

---

1. The details of how the pressure changes with time depend on more geometry than we have here. For example, if the top is very small the pressure might build up for a while, but if it is very large then the pressure might be essentially constant. The important point is that each model for outward flow is consistent.

What does this imply about the pressure?

If  $D_s(P(\text{steam}))=0$ , then from physics we know that for some amount of water boiled off,

$$V(\text{steam}) \gg V(\text{water})$$

which means,

$$D_m(V(\text{steam})) \gg D_m(V(\text{water}))$$

which cannot be. The pressure must supply some influence on the volume, in order to make the rates equal. Suppose  $D_s(P(\text{steam}))=-1$ . Then for a particular amount of steam at a particular temperature the gas law tells us the influence of pressure on volume:

$$D_s(P(\text{steam})) + D_s(V(\text{steam})) = 0$$

This means the influence of the pressure change would result in  $D_s(V(\text{steam}))=1$ , which does not help. On the other hand,  $D_s(P(\text{steam}))=1$  means  $D_s(V(\text{steam}))=-1$ , which provides a negative influence on  $D_s(V(\text{steam}))$  that can cancel the difference in rate. So  $D_s(P(\text{steam}))=1$ .

Because the steam touches the water and the container,

$$P(\text{steam})=P(\text{water})=P(\text{Inside}(\text{container}))$$

This means that  $D_s(P(\text{water}))=1$ , and because physics tells us

$$(\propto_Q T(\text{boil}) P(\text{water}))$$

we conclude  $D_s(T(\text{boil}))=1$ . This means more heat can flow from the source and the boiling can continue at a higher temperature and pressure. Since the same conditions hold for the new temperature and pressure, the increase will be continuous.

How might all of this end? Unless there is some outside factor, either:

1.  $A(\text{water})=\text{NONE}$ , boiling stops and steam heats up to  $T(\text{source})$ .
2.  $T(\text{water})=T(\text{source})$ , boiling stops, thermal equilibrium achieved.
3.  $P(\text{Inside}(\text{container}))=P(\text{burst})$ , container explodes!

To actually determine which of these occurs requires more information, but at least we have a warning of potential disaster.

## 7.2 Motion

Motion is perhaps the most commonly occurring process. Motion is also very complex because it is intimately connected with our concepts of space and shape. Qualitative Process theory provides a way of expressing dynamics, not kinematics. To illustrate, we can consider the case of a single object moving in one dimension. The most abstract description of such motion can be written as:

Motion(B,dir)

Preconditions: free-direction(B, dir)

QuantityConditions: (greater-than  $A_m[Vel(B)]$  zero)

Influences: (I+ Pos(B) Vel(B))

An additional assumption of course is that the only influence on Pos(B) is some kind of motion. This model is Newtonian, but other theories of motion can of course be expressed as well. Aristotelian dynamics only requires changing the quantity conditions:

Aristotelian: (greater-than  $A_m(F_{net}(object))$  zero)

The dynamics of an Impetus theory, which appears to be the common sense model for motion used by many people (see [McClosky, 1982]), can also be modelled. An impetus quantity must be introduced (impetus is the "force" an object carries along with it that maintains its motion) and the quantity conditions for motion become having non-zero impetus. Non-zero impetus also forms the quantity condition for a "decay" process which reduces the impetus with time.

In Newtonian dynamics the process of acceleration provides the sole influence on velocity.

Acceleration is brought about by a non-zero net force in some direction in which the object is free to move. These facts can be written as:

Acceleration(B,dir)

Preconditions: free-direction(B,dir)

QuantityConditions: (greater-than  $A_m[F_{net}(B)]$  zero)

Relations: Let Acc be a number

( $\propto_Q$  Acc  $F_{net}(B)$ )

(correspondence (Acc zero)

( $F_{net}(B)$  zero))

Influences: (I+ Vel(B) Acc)

Moving friction can be modelled as a process that occurs during motion that involves a surface contact, and produces a force on the moving object in the opposite direction of the motion. Static friction can be modelled as a process that occurs when no motion is occurring but there is a force component on the object along the surface.



Collisions are complicated. The simplest version just involves a reversal of velocity:

```
Collide(B,C,dir)
Precondition: direction-towards(B,C,dir)
QuantityCondition: Motion(B,dir)
Relations: (= (M Vel(B) start) (- (M Vel(B) end)))
           (= (M Vel(B) during) zero)
           (= duration zero)
           (T direction-towards(C,B,dir) end)
```

where *direction-towards(C,B,dir)* asserts that the object is moving in direction *dir* from C to B, *start*, *end*, *during* and *duration* define the temporal aspects of an episode in a process history that corresponds to this process occurring. Even our more complicated models of collisions appear to use such behavioral descriptions, such as a compound process consisting of contacting the surface, compression, expansion, and finally breaking contact. The preconditions for the type of collision must also include some reference to the particular theory of materials it assumes for the object.

It is possible that the preconditions for motion could be mapped into QP theory by representing the position of an object by an element in a Place vocabulary. The Quantity Space for position would be given by the ordering imposed along a direction, with the ambiguity resulting from the qualitative description of position and direction being reflected in the lack of order between the corresponding constants in the Quantity Space. The advantage is that the possibilities imposed by the geometric constraints of the problem would be expressed in the results of limit analysis. Experiment will be required to ascertain the value of this technique.

The process vocabulary for motion presented above is quite abstract. The particular kind of motion - flying, sliding, rolling, or swinging - is not mentioned. These motions would be specializations of the motion process considered above, defined by additional preconditions and relations (sliding and rolling require surface contact and could involve friction, for instance). Previous qualitative descriptions of motion centered around the idea of a qualitative state ([deKleer, 1975], [Forbus, 1981]). A qualitative state would consist of a statement that a particular type of motion was occurring in some "place" and in some quantized direction. The knowledge of motion was encoded in simulation rules that mapped a qualitative state to a set of possible next states. If the preconditions are mapped into the limit analysis as proposed above, then the simulation rules can be viewed as a compilation of the cases produced by the limit analysis for a given process vocabulary for motion. Making motion into an explicit process should

allow greater flexibility in reasoning about motion than qualitative simulation rules provide. For instance, we can conclude that if an object is kicked in a direction then it will move unless something is in the way, without knowing enough to specify the particular kind of motion that occurs.

### 7.3 Breaking string

Consider a string tied to the top of a block. We know a number of things might happen if we try to move the string. We know that we can lift the block by pulling up on the string, unless the block is so heavy that the string breaks. We know that a string, if taut, can transmit a pull, but under no circumstances can you push with it. We can use the notions of quantity and process provided by Qualitative Process theory to state these facts.

Let us consider what happens when we pull on something. If it doesn't move, then its internal structure is "taking up" the force (this can happen even if it does move - try hitting an egg with a baseball bat - but here we ignore this case). Three things can happen - (1) it can do nothing (rigid behavior), (2) it can stretch (elastic behavior) or (3) it can break. For a push, (2) becomes compression and (3) becomes crushed. We can express the changes between these kinds of behavior by creating a quantity space for forces on an object (in a full qualitative theory of materials there would be different quantity spaces for different directions on the objects).

Each force quantity space will include zero. For a breakable rigid object, there will also be  $\text{max}f_c$  and  $\text{max}f_b$  corresponding to the force needed to crush or break the object, respectively. By convention, forces into an object (pushes) will be negative and applied forces directed outwards (pulls) will be positive. The Quantity Space for forces on a breakable rigid object then looks like:

$$\text{max}f_c \rightarrow \underline{\text{zero}} \rightarrow \text{max}f_b$$

If the object can stretch or compress, the quantity space must also contain  $f_{\text{stretch}}$  and  $f_{\text{compress}}$ , which denote the boundary between rigid behavior and elastic behavior. If the force is less than the appropriate value the object will be rigid and if above, it will stretch or compress. The Quantity Space for forces on a partially elastic object looks like:

$$\text{max}f_c \rightarrow f_{\text{compress}} \rightarrow \underline{\text{zero}} \rightarrow f_{\text{stretch}} \rightarrow \text{max}f_b$$

There are of course correspondences with the quantity space for length:

(correspondence (Force(0)  $\text{max}f_b$ ) (Length(0)  $\text{maxlength}$ ))

(correspondence (Force(0)  $\text{max}f_c$ ) (Length(0)  $\text{minlength}$ ))

The following is a partial specification of the processes of stretching and compressing:

Stretching(0)

individuals: elastic object 0

Precondition: (Pull-on 0)

QuantityCondition: (not (less-than  $A_m(\text{Force}(0))$   $f_{\text{stretch}}$ ))

Relations: ( $\alpha_Q$   $A_m(\text{Length}(0))$   $A_m(\text{Force}(0))$ )

Compressing(0)

individuals: elastic object 0

Precondition: (Push-on 0)

QuantityCondition: (not (less-than  $A_m(\text{Force}(0))$   $f_{\text{compress}}$ ))

Relations: ( $\alpha_Q$   $A_m(\text{Length}(0))$   $A_m(\text{Force}(0))$ )

Writing process descriptions for stretching, compressing, breaking, and crushing is complicated. Stretching and compressing involve a notion of internal force if the object is elastic, and breaking and crushing involve deformation of shape and the transformation of one object into several. As with kinematics, these issues are beyond the scope of QP theory.

A classic conundrum for AI is to be able to express in some form usable by a program that "you can pull with a string, but not push with it". This fact can be succinctly described using Qualitative Process theory. First, consider what pushes and pulls are. Both concepts imply one object making contact with another to apply force. If the force applied is into the object it is being applied to it is a push, and if out of the object (in the vector sense) it is a pull. Obviously push can occur with any kind of contact, but pulls cannot occur with an abutting.

Understanding how pushes and pulls are transmitted is fundamental to understanding mechanisms. For a first pass model, consider the notion of *push-transmitters* and *pull-transmitters*. We will say an object is a push transmitter if when it is pushed, it will in turn push an object that is in contact with it, in the direction between the two contact points. Pull transmitters can be similarly defined. This particular set of definitions is obviously inadequate for mechanisms (consider for example a rocker arm or an object which is tied to a string by another object. In the first case a push will be transmitted in a different direction, and in the second case it will be transformed into a pull), and is only for illustration. Note also that push-transmitters and pull-transmitters need not be reflexive relations. An exceptional case are rigid objects:

$\forall o \in \text{objects}$

(rigid o)  $\Rightarrow$  ( $\forall c_1, c_2 \in \text{contact-points}(o)$ )

```

(push-transmitter c1 c2)
^ (push-transmitter c2 c1)
^ (pull-transmitter c1 c2)
^ (pull-transmitter c2 c1)

```

Strings, however, are more complicated. A string can never be a push-transmitter:

$\forall s \in \text{strings}$

```

 $\forall t \in \text{times}$  (T (and (not (push-transmitter (end1 s) (end2 s)))
(not (push-transmitter (end2 s) (end1 s)))) t)

```

But if it is taut it can be a pull transmitter:

$\forall s \in \text{strings}$

```

 $\forall t \in \text{times}$  (implies (T (taut s) t)
(T (and (pull-transmitter (end1 s) (end2 s))
(pull-transmitter (end2 s) (end1 s))) t))

```

Now the problem becomes how to define taut. We can do this by writing conditions on the quantity spaces of lengths and distance from the ends of the string:

```

(condition (less-than  $A_m(\text{ends-distance}(s))$   $A_m(\text{Length}(s))$ )
(not (taut s)))
(condition (not (less-than  $A_m(\text{ends-distance}(s))$   $A_m(\text{Length}(s))$ ))
(taut s))

```

This model assumes that only the ends of the string contact other objects - it would fail for a rope hanging over a pulley, for instance. But it does illustrate how the concept of quantity spaces makes the problem much easier.

#### 7.4 An oscillator

Consider the block B connected to the spring S in figure 11. Suppose that the block is pulled back so that the spring is extended. Assume also that the contact between the block and the floor is frictionless. What happens?

First, the spring "process" includes:

Relations:

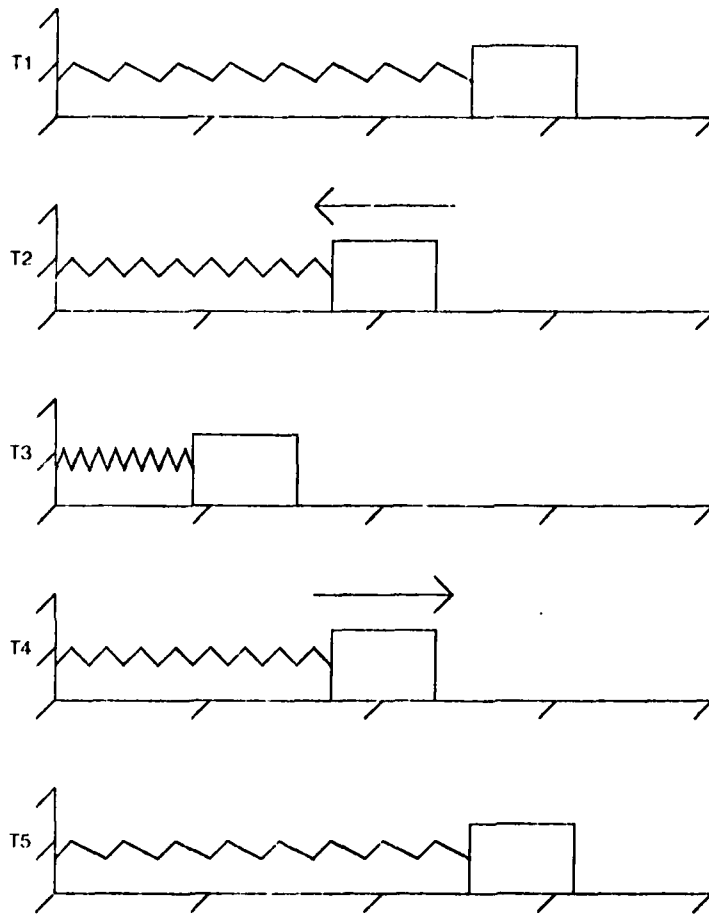
```

 $D_s(L_{rest}(s)) = 0$ 
Let  $Disp(s) = (- A(L(s)) L_{rest}(s))$ 
( $\propto_0$   $F_s(s)$   $Disp(s)$ )
(correspondence ( $F_s(s)$  zero) ( $Disp(s)$  zero))

```

where  $F_s$  is the internal force due to the composition of the spring. Since  $Disp(s)$  is greater than zero.

Fig. 11. Sliding Block With Friction



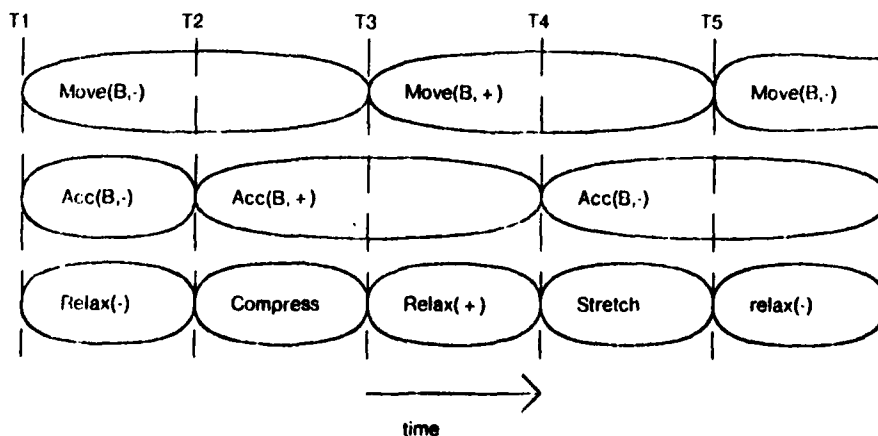
the spring will exert a force. Because the block is rigidly connected to the spring, the net force on the block will be negative and since the block is free to move in the direction of the force, an acceleration will occur. The acceleration will in turn cause the velocity to move from zero, which will in turn cause  $D_s(Pos(B)) = -1$ . By rigid contact,  $D_s(L(S)) = -1$  and by the  $\alpha_Q$  relation with displacement,  $D_s(F_{net}(B)) = 1$ . The processes occurring are motion(B, -), relaxing(S, -), acceleration(B, -). The next process limit occurs when  $L(s) = L_{rest}(s)$ , ending the relaxing. The correspondence tells us the force on the block becomes

zero, so the acceleration will end as well. However, the motion does not. Setting aside the details, the next set of processes are Motion(B, -), compress(S), and acceleration(B, +). The only limit point in the quantity spaces that are changing is the zero velocity point (assuming the spring is unbreakable), so the motion will continue until the velocity is zero. The conclusion that the next set of processes are Motion(B, +), relaxing(S, +), acceleration(B, +) and then Motion(B, +), stretching(S, +), acceleration(B, -) follows in the same way. At the end event of the last set of processes, the orderings on the quantity spaces and the processes evoked are the same as the initial instant. Thus we can conclude that an oscillation is occurring. Note that the processes need to be the same, because the preconditions might have changed. Figure 12 illustrates the process history for the oscillator.

Some of the assumptions made in producing the process history can now be relaxed to discover the effects of a more detailed physical model. First, suppose the spring is breakable and/or crushable. Then there are limit points around  $t_{rest}(s)$  corresponding to breaking and crushing. For crushing, it seems an assumption is in order. If we can deduce that the block will go no further out than it was originally, then we can claim breaking will not occur since it didn't break in the first place. In other words, we need to show that  $\text{DO}(\text{Pos}(B), t_1 t_5) = 1$  is not the case, where the situations being compared are denoted by the instants  $t_1$  and  $t_5$  (see the previous figure).

This deduction requires an energy argument. If we ignore the mass of the spring the energy for the combination at any particular instant in time is given by

Fig. 12. Process history for the oscillator



$$E(\text{System}) = E_k(B) + E_p(S)$$

where  $E_k(B)$  is the kinetic energy of the block and  $E_p(S)$  is the potential energy of the spring. By the definitions of kinetic and potential energy, we have:

$$(\propto_Q E_k(B) A_m(\text{Vel}(B)))$$

$$(\text{correspondence } (E_k(B) \text{ zero}) (A_m(\text{Vel}(B)) \text{ zero}))$$

and

$$(\propto_Q E_p(S) A_m(\text{Disp}(S)))$$

$$(\text{correspondence } (E_p(S) \text{ zero}) (\text{Disp}(S) \text{ zero}))$$

In  $t_1$  the block is still but the spring is stretched, that is:

$$(\text{= } (M A(\text{Vel}(B)) t_1) \text{ zero})$$

$$(> (M \text{Disp}(S) t_1) \text{ zero})$$

which means

$$(> (M E(\text{System}) t_1) \text{ zero})$$

If there is no external source of energy, conservation tells us

$$\forall t \in \text{times } (\text{not } (> (M E(\text{System}) t) (M E(\text{System}) t_1)))$$

This rules out  $DQ(\text{Pos}(B), t_1, t_5) = 1$ , because the energy of the system would then be higher.

The notion of a system in Qualitative Process theory is captured by the idea of a compound process. The previous arguments provide a set of assumptions which can serve as preconditions and quantity conditions for the compound process - the material composition of the spring being such that the spring will not crush is a precondition, and the lack of external energy sources (other processes acting on the system) is a quantity condition for the new process. The abstraction allows the explicit representation of properties over an interval of the cycle, such as energy lost and maximum displacement.

In particular, suppose we conclude:

$$(\propto_Q \text{MaxDisp}(\text{Obj}) E(\text{System}))$$

$$(\text{correspondence } (\text{MaxDisp}(\text{Obj}) \text{ zero}) (E(\text{system}) \text{ zero}))$$

This relation makes it possible to deduce that if friction were introduced (i.e.,  $D_s(E(\text{system})) = -1$ ) the oscillation process will eventually stop, and that if the system is pumped without friction, or pumped with increasing amounts of energy (i.e.,  $D_s(E(\text{system})) = 1$ ), that the materials involved in the oscillator may break in some way.<sup>1</sup> Suppose the oscillator is now pumped with some fixed amount of energy per cycle, as it would be in a mechanism such as a clock. Is such a system stable? If there is no friction, then we have seen already that it is not, for the sole influence on energy will be the pumping and the energy will increase until something breaks. Suppose there is friction. The only things we will assume about the

---

1. The Tacoma bridge phenomena, something every engineer should know about.

friction process is that

$$\begin{aligned} & (I - E(\text{System}) - E(\text{loss})) \\ & (\propto_0 E(\text{loss}) - E(\text{System})) \\ & (\text{correspondence } (E(\text{loss}) \text{ zero}) (E(\text{System}) \text{ zero})) \end{aligned}$$

where  $E(\text{loss})$  is the net energy lost due to friction over a cycle of the oscillator process. The loss being qualitatively proportional to the energy is based on the fact that the energy lost by friction is proportional to the distance travelled, which in turn is proportional to the maximum displacement, which itself is qualitatively proportional to the energy of the system, as stated above.

The lower bound for the energy of the system is zero, and an upper bound for energy is implicit in the possibility of the parts breaking. The result, via the  $\propto_0$  statement above, is a set of limits on the quantity space for  $E(\text{loss})$ . If we assume  $E(\text{pump})$ , the energy which is added to the system over a cycle, is within this boundary then there will be a value for  $E(\text{System})$ , call it  $E(\text{stable})$ , such that:

$\forall t \in \text{C intervals}$

$$\begin{aligned} & (\text{implies } (= (M E(\text{System}) t) (M E(\text{stable}) t)) \\ & (= (M E(\text{loss}) t) (M E(\text{pump}) t))) \end{aligned}$$

If the energy of the system is at this point, the influences of friction and pumping will cancel and the system will stay at this energy. Suppose

$$(> (M E(\text{System}) t) (M E(\text{stable}) t))$$

over some cycle. Then because the loss is qualitatively proportional to the energy, the energy loss will be greater than the energy gained by pumping, i.e.,  $D_s(E(\text{System})) = -1$ , and the energy will drop until it reaches  $E(\text{stable})$ . Similarly, if  $E(\text{System})$  is less than  $E(\text{stable})$  the influence of friction on the energy will be less than that of the pumping, thus  $D_s(E(\text{System})) = 1$ . This will continue until the energy of the system is again equal to  $E(\text{stable})$ . Therefore for any particular pumping energy there will be a stable oscillation point.<sup>1</sup>

---

1. This is a qualitative version of the proof of the existence and stability of limit cycles in the solution of non-linear differential equations. Uniqueness is implied by the monotonicity of the function implicit in



### 7.5 Grain Elevator

In reasoning about the commodities market, it is often necessary to represent the physical limitations of the parts of the economic situation. When the grain elevators available for storage are full, for example, any excess grain must be sold off, which can cause a drop in price. These kinds of deductions are an important part of what a human expert knows about the economic world [Stanfield, 1980].

In terms of the Qualitative Process theory, the problem is to express that a filling of a grain elevator (or some more abstract storage facility) may end by the elevator becoming full. This can be done by including in the quantity conditions for a filling the following:

(less-than  $A_m$ (Amount-of(stuff, container)) capacity(container))

At first glance making (not (full container)) part of the preconditions for the filling process might solve the problem. This is less satisfactory because the limit analysis would no longer include reaching the capacity as a possible end of the process.

### 7.6 Automobile Transmission

Describing the geometry of gears in a standard transmission (or the fluid parts of an automatic transmission) is clearly beyond the scope of this theory. However, the mechanism for describing functional dependencies introduced here should provide a useful way to express the results of such deductions.

The first thing to note is that a transmission has several states. Call these states *neutral*, *first*, *second*, *third*, and *reverse*. These states are the only states for a transmission:

(taxonomy neutral(tr) first(tr) second(tr) third(tr) reverse(tr))

and if we identify the direction of rotation with a particular direction in the quantity space:

(implies neutral(tr) (not ( $\alpha_0$ \* speed(driven) speed(driver))))

(implies (or first(tr) second(tr) third(tr))  
( $\alpha_0$  speed(driven) speed(driver)))

(implies reverse(tr)  
( $\alpha_0$  speed(driven) -speed(driver)))

The notation for alternate situations introduced in Differential Qualitative analysis (see above)

allows the difference in rates for different states to be expressed:

V  $S_1, S_2, S_3$  C situations

(implies (and T[first(tr)  $S_1$ ]

T[second(tr)  $S_2$ ])

DQ[speed(driven),  $S_1, S_2$ ] = -1)

(implies (and T[second(tr)  $S_2$ ])

T[third(tr)  $S_3$ ])

DQ[speed(driven),  $S_2, S_3$ ] = -1)

## 8. Discussion

This paper has presented a new theory about common sense physical reasoning, Qualitative Process theory. To summarize:

- o Processes are the cause of changes in physical situations. A process is specified by the individuals it occurs between, the preconditions and quantity conditions that must be true the process to occur, the relations it imposes on those individuals and the influences it provides on their quantities.

- o An appropriate qualitative description of quantity for reasoning about processes is the Quantity Space. The relationship of the quantity with the other elements in the Quantity Space defines its value.

- o Processes provide a language for writing physical theories. In particular, the primitives are simple processes (which define the "nature of" objects and things like flows and state changes), the means of combination are sequentiality and shared parameters, and the means of abstraction are naming these combinations, including encapsulating a piece of the process history for a situation as a new process.

- o Several kinds of qualitative conclusions can be drawn using the constructs of QP theory, including reasoning about the effects of combined processes, the limits of processes, and alternative situations. It also provides a new perspective on causal reasoning, and should allow true qualitative "time domain" analysis.

- o Interesting phenomena in common sense reasoning appear to be described reasonably well by QP theory, including motion, materials, and oscillation.

### 8.1 Perspective

The present theory has evolved from several strands of work in Artificial Intelligence. The first strand is the work on envisioning, started by deKleer [deKleer, 1975](see also [deKleer, 1979][Forbus, 1981]). Envisioning is a particular style of reasoning about physical situations. Situations are modelled by collections of objects with qualitative states, and what happens in a situation is determined by running simulation rules on the initial qualitative states and analyzing the results. The weak nature of the information means that the result looks like a directed graph of qualitative states which corresponds to the set of all possible sequences of events that can occur from the initial qualitative state. This description

itself is enough to answer some simple questions, and more precise information can be used to determine what will actually happen if so desired.

While a powerful idea, the assumptions of envisioning as it has been developed thus far are too restrictive. The qualitative state representation of what is happening to an object is impoverished; the processes which they represent often involve several objects at once in an interdependent fashion. The use of qualitative simulation rules means that the only time information about events consists of local orderings, making new interactions between things happening in the situation ("collisions") hard to detect. Simulation rules are also a rather opaque way to encode knowledge about how things can happen in a situation. The rules themselves do not describe the mechanism by which the state transformation is accomplished (except implicitly), thus making it difficult (or impossible) to reason about changes in the assumptions which underly the rules. Qualitative Process theory should provide the basis for building much more flexible systems.

The second strand of work concerns the representation of quantity. Most AI schemes for qualitative reasoning about quantities violate what I call the *relevance principle* of qualitative reasoning - qualitative reasoning about something continuous requires some kind of quantization to form a discrete set of symbols; the distinctions made by the quantization must be relevant to the kind of reasoning being performed.<sup>1</sup> Almost all previous qualitative representations for quantity violate this principle. One exception is the notion of quantity introduced by deKleer as part of Incremental Qualitative analysis (discussed previously). For more general physical reasoning a richer theory of quantity is necessary. IQ analysis alone does not allow the limits of processes to be deduced. For instance, we could use it to deduce that the water in a kettle on a lit stove would heat up, but we couldn't deduce that it could boil. IQ analysis does not represent rates, so we could not deduce that if the fire on the stove were turned down the water would take longer to boil (Differential Qualitative analysis). The notion of quantity provided by QP theory should be useful for a broader range of inferences.

The final strand relevant to the theory is the Naive Physics enterprise initiated by Pat Hayes [Hayes, 1979]. The goal of Naive Physics is to develop a formalisation of our common sense physical knowledge. From the perspective of Naive Physics, Qualitative Process analysis corresponds to a cluster -

---

1. For an example of this principle applied to spatial reasoning, see [Forbus, 1981].

a collection of knowledge and inference procedures which is sensible to consider as a module. The introduction of explicit processes into the ontology of Naive Physics should prove quite useful. For instance, in the axioms for liquids [Hayes, 1979b] information about processes is encoded in a form very much like the qualitative state idea.<sup>1</sup> This makes it difficult to reason about what happens in situations where more than one process is occurring at once - Hayes' example is pouring water into a leaky tin can. In fact, difficulties encountered in trying to implement a program based on the axioms for liquids were a prime motivation for developing Qualitative Process theory.

## 8.2 Common Sense Physical Reasoning

Qualitative Process theory should be a useful tool in the development of Naive Physics:

- o reasoning about the results and limits of processes are obviously part of our common sense knowledge of the physical world.
- o Important phenomena such as motion and the effects of material compositions for objects can be modelled with it.
- o it provides a highly constrained account of physical causality (all changes are due to a finite vocabulary of processes) and a useful notion for representing causal connections ( $\alpha_0$ ).
- o it provides a highly constrained role for the use of experiential and default knowledge in physical reasoning - resolving influences and choosing or ruling out alternative endings to a particular episode.

It is interesting to speculate on what other representations for quantities might be useful in physical reasoning. "Real" numbers and IQ values can be thought of as opposite ends of a spectrum of representations for quantities, with the quantity space notion somewhere in the middle. Another candidate would be a representation of "order of magnitude" estimates, which would increase the comparability of various quantities but still not requiring exact information.

Qualitative Process theory should also contribute to the utility of the concept of histories Hayes

---

1. See for example axioms 52 through 62.

introduced to describe change. A history is a piece of space-time, temporally extended and spatially bounded. By contrast, situations in the situational calculus are spatially unbounded and temporally a point. The history representation trades the frame problem for two new problems: the intersection problem and the local dynamics problem. Histories can interact only when they overlap, thus making the problem of determining unexpected interactions the problem of intersecting the pieces of space-time which comprise the history. This however assumes that barring interactions, the history for an object can be generated locally. Qualitative Process theory should provide a useful language for writing the required "dynamics theories".

It should also be possible to test QP theory for psychological adequacy. If the strong Physical Causality conjecture holds, each person should have an identifiable process vocabulary. If the process vocabulary for an individual can be determined, then predictions about errors on specific problems can be made and checked. Devising such a two part experiment appears complex, however.

### 8.3 Reasoning about Engineered Systems

Many engineered devices are implemented as physical systems, and thus are subject to physical laws. A qualitative understanding of such systems involves our common sense physical knowledge. I have been applying Qualitative Process theory to reasoning about the physics of steam plants as part of the STEAMER project [Stevens, et. al., 1981].<sup>1</sup> It appears to have some important advantages:

- o because it is more powerful than IQ analysis, computing the behavior of a system from a description of its structure should be possible for more complex systems than before. In particular, a true qualitative "time domain" style of analysis should be possible.

- o the notions of quantity and functional dependence have been useful in thinking about more abstract functional descriptions (such as COMPARATOR and FEEDBACK-LOOP), because signals in a large class of engineered systems are continuous.

---

1. STEAMER is a joint project of NPRDC and BBN, to develop intelligent computer aided instruction techniques to train propulsion plant officers for the Navy.

o The qualitative nature of its descriptions appear similar to those used by students in understanding physical systems and often by experts in explaining them, making its conclusions appear useful for teaching.

Applications other than teaching are imaginable. If extension theories were provided to interface the basic QP theory descriptions with quantitative descriptions of what is actually happening in a system several new possibilities arise. Controlling systems should ultimately be possible, using the condition mechanism to express desired and undesired operational characteristics. More immediately feasible would be an interpretation module, which would gather data from instruments and build theories about what the underlying processes that generate those data are. Such a module could be used as part of a diagnosis program or as a "hypothesizer" that could serve as a devil's advocate during the operation of a complex system. For example, the incident at the Three Mile Island reactor probably wouldn't have happened if the operators had thought of the alternate explanation for the overpressure in the reactor vessel - that instead of being too high, the level of cooling water was too low, thus causing a boiling that raised the pressure.<sup>1</sup>

#### 8.4 Economic modelling and Support Systems

Many non-physical systems are often modelled with continuous parameters and processes, notably economic theories. A theory of physical reasoning might provide useful leverage in understanding such systems in several ways. First, physical limitations often constrain such systems (storage capacities, transportation capacities, time required for processes such as crop growth or manufacture, etc.). Secondly, economic systems are often described by analogy with physical systems (Samuelson, for instance, cites the aphorism "the central bank can pull on a string (to curb booms), but it can't push on a string (to reverse deep slumps)"[Samuelson, 1973]). Finally, non-physical processes themselves might be usefully described using a theory like the present one.

Several caveats are in order. First, unlike physical systems, there is no real agreement on what are valid process descriptions in domains like economics. Secondly, changes in circumstances may dictate

---

1. [Pew et.al.] hypothesizes premature commitment by operators to a particular theory about the state of the plant as a common source of human errors in plant operation.

changing process vocabularies (certain stock transactions may be deemed illegal, for instance). This means that the set of possible influences is essentially unbounded. These application areas are therefore much harder than physical reasoning.

### 8.5 Acknowledgements

Gerald Sussman, Dave McAllester, Johan deKleer, Al Stevens, Dedre Gentner, Bruce Roberts, and Patrick Winston have all influenced the development of this theory in various ways.



## 9. Bibliography

Allen, J. "A General Model of Action and Time" University of Rochester Computer Science Department TR-97, November, 1981

deKleer, Johan "Qualitative and Quantitative Knowledge in Classical Mechanics" TR-352, MIT AI Lab, Cambridge, Massachusetts, 1975

deKleer, Johan "Causal and Teleological Reasoning in Circuit Recognition" TR-529, MIT AI Lab, Cambridge, Massachusetts, September 1979

deKleer, Johan and Brown, John Seely "Assumptions and Ambiguities in Mechanistic Mental Models" to appear in Mental Models, D. Gentner and A. Stevens, editors.

diSessa, A. "Phenomenology and the Evolution of Intuition" to appear in Mental Models, D. Gentner and A. Stevens, editors.

Forbus, K. "A Study of Qualitative and Geometric Knowledge in Reasoning about Motion" TR-615, MIT AI Lab, Cambridge, Massachusetts, February, 1981

Forbus, K. "A Conlan Primer" BBN Technical Report No. 4491, prepared for Navy Personnel Research and Development Center, 1980.

Forbus, K. "Using Qualitative Simulation to Generate Explanations" BBN Technical Report No. 4490, prepared for Navy Personnel Research and Development Center, 1980.

Forbus, K. "Qualitative Reasoning about Space and Motion" to appear in Mental Models, D. Gentner and A. Stevens, editors.

Forbus, K. "Qualitative Reasoning about Physical Processes" Proceedings of IJCAI-7, 1981

Hayes, Patrick J. "The Naive Physics Manifesto" in Expert Systems in the Micro-Electronic Age, edited by D. Michie, Edinburgh University press, May 1979

Hayes, Patrick J. "Naive Physics I - Ontology for Liquids" Memo, Centre pour les etudes Semantiques et Cognitives, Geneva, 1979

McCloskey, M. "Naive Theories of Motion" to appear in Mental Models, D. Gentner and A. Stevens, editors.

McDermott, D. "A Temporal Logic for Reasoning about Processes and Plans" Yale Computer Science Dept. TR 196, 1981

Moore, R. "Reasoning from Incomplete Knowledge in a Procedural Deduction System" MIT AI-TR-347,

Cambridge, Massachusetts, December 1975

Moore, R. "Reasoning about Knowledge and Action" MIT PhD thesis, February, 1979.

Pew, R., Miller, D. and Feeher, C. "Evaluation of Proposed Control Room Improvements Through Analysis of Critical Operator Decisions" Electric Power Research Institute Report NP-1982, August, 1982

Reiter, R. "A logic for Default Reasoning" Artificial Intelligence, Vol. 13, 1980, pp 81-132.

Samuelson, P. "Economics", McGraw-Hill, N.Y., 1973

Stansfield, J. "Conclusions from the Commodity Expert Project" MIT AI Lab Memo 601, November, 1980

Stevens, A. et. al. "Steamer: Advanced Computer Aided Instruction in Propulsion Engineering" BBN Technical Report No. 4702, July 1981

Williams, M., Hollan, J. and Stevens, A. "Human Reasoning about a Simple Physical System". to appear in Mental Models, D. Gentner and A. Stevens, editors.

END

DATE  
FILMED

4-82

DTIC