# GETTING THE ENVISIONMENT RIGHT

**Benjamin Kuipers**
**Tufts University**
**Medford, Massachusetts 02155.**

## ABSTRACT

The central component of commonsense reasoning about causality is the *envisionment*: a description of the behavior of a physical system that is derived from its structural description by qualitative simulation. Two problems with creating the envisionment are the qualitative representation of quantity and the detection of previously-unsuspected points of qualitative change. The representation presented here has the expressive power of differential equations, and the qualitative envisionment strategy needed for commonsense knowledge. A detailed example shows how it is able to detect a previously unsuspected point at which the system is in stable equilibrium.

## THE ENVISIONMENT

Causal reasoning --- the ability to reason about how things work --- is central to expert performance at problem-solving and explanation in many different areas. The concept of an *envisionment*, developed by de Kleer [1,2], has provided a framework for most subsequent research on causal reasoning. A physical system is described by a *structural description* consisting of the context-independent behavioral repertoires of its individual components, and their connections in this context. The envisionment describes the potential behaviors of the system, and is produced by qualitative simulation of the structural description. It can be used directly to solve problems or answer questions, or can be further analyzed to produce a *functional* description to explain *why* the system works the way it does. However, there is significant disagreement on several key questions about the structure of the envisionment process:

(1) How should continuously variable quantities be described for qualitative simulation?

(2) How should the envisionment detect previously unsuspected points at which qualitatively significant changes take place?

De Kleer [1,2] does qualitative perturbation analysis by describing quantities in terms of the sign of the derivative (the IQ value) alone, but this is clearly too weak for other kinds of causal reasoning. Forbus [3] gets considerably greater power by representing each quantity in terms of the sign and magnitude of both its amount and its derivative. In practice, the power of his system depends only on the ordinal relations among quantities. Hayes [4] defines a modular *quantity space* in which inference about quantities takes place, but remains agnostic about its properties. None of the above systems use qualitative reasoning to discover previously unsuspected points where qualitatively significant changes take place, although de Kleer's "roller coaster" envisionment [1] is able to localize a change within a region before turning the problem over to a quantitative problem-solver.

In this paper, I present a simple but very general descriptive language for structural descriptions, and a qualitative simulation process for producing the envisionment. Within the causal structure description, a system is described as a collection of *constraints* holding among time-varying, real-valued *quantities*. A *value* is a description of the real number corresponding to a quantity at a given *time-point*. This description consists of the ordinal relations holding among the different values known to the envisionment, and the *IQ value* (the sign of the time derivative: +, 0, -) of the quantity at that time-point. A constraint consists of rules for propagating information describing the current value among the values of the related quantities. The mechanism is inspired by the scheme developed by Steele [6], modified to propagate ordinal and IQ value assertions rather than integers. The three types of constraints used in the example below are:

**Arithmetic**: $(X = Y + Z)$ The values of the quantities must have the indicated relationship within any time-point.

**Functional**: $(Y = M^{+}(X))$ Y is a strictly increasing (or decreasing $(M^{-})$) function of X. $M^{+}{}_{Z}$ indicates that zero corresponds to zero.

**Derivative**: $(Y = \frac{d}{dt} X)$ At any time-point, Y is the rate of change of X.

The envisionment consists of a finite set of time-points representing the qualitatively distinct states of the system, and values for each quantity at each time-point. Thus, the set of values that are part of the envisionment, and the ordinal relations that hold among them, can only increase as new information is propagated across constraints from quantity to quantity: the qualitative simulation is monotonic.

The qualitative simulation propagates information across the constraints to complete the description of the state of the system at the current time-point. After the propagation of information among values has settled down, the envisionment process examines the set of changing values in the current time-point to determine the next qualitatively distinct state. Determining the next state depends critically on the concept of *distinguished value*. Initially, zero is the only distinguished value, but if the IQ value of a quantity becomes zero, that value (a critical point)

This research was supported in part by NIH Grant LM 03603 from the National Library of Medicine.

becomes a new distinguished value. Two of the rules (the system currently has seven) for creating a new time-point to represent the next qualitatively distinct state of the system are:

Move To Limit: If the current value of a changing quantity is not distinguished, and there is a distinguished value in the direction of change, let the value of that quantity in the next time-point be the next distinguished value.

Move From Point: If the current value of a changing quantity is distinguished, then let the next value be an undistinguished value in the given direction, closer to the starting point than any other distinguished value.

When the description of the system's current state is not sufficiently complete to determine the next state uniquely, the envisionment branches on the possible states of a particular IQ value or ordinal relation. If the qualitative simulation is unable to proceed, it may summarize the structural description to reduce its complexity at the cost of losing some information. The simulation terminates upon recognizing a contradiction, intractible branching, a cycle, or a quiescent system.

When all the qualitatively significant points are specified in advance, Forbus [3] shows how an envisionment process can determine the possible behaviors of a system. In order to demonstrate the power of this descriptive language and simulation process, the following example shows how, without external information, the simulation process deduces the existence of a previously-unsuspected distinguished value, and shows that a system moves to a stable equilibrium about that value.

Consider a simple physical system consisting of a closed container of gas (at temperature T) that receives heat from a source (at $T_s$) and radiates heat into the air ($T_a$). The problem is to deduce the existence of an equilibrium temperature ($T_e$) between the temperatures of the heat source and the air, and to show that the system moves to a stable equilibrium about that temperature. Tables 1, 2, and 3 show the different stages of the qualitative simulation as it creates the envisionment.[1] Table 1 shows how the envisionment of the double-flow system branches in order to derive missing IQ values, how a new distinguished point is discovered on one of the branches, and how a set of corresponding values is discovered when several quantities take on distinguished values simultaneously. Table 2 shows how the structural description is summarized when the first envisionment bogs down at an intractible branch. Table 3 shows how the summarized structural description, and the newly-discovered correspondence, allow the successor time-points on the remaining two branches to be determined uniquely so the envisionment can be completed. Diagnosis of a stable equilibrium takes place using the final envisionment structure, by showing that a perturbation from the final quiescent state places the system into one of the previously described states from which there is a restoring change.

1. The envisionment diagrams (Tables 1 and 3) are read from top to bottom, each line following from those above. Each cell corresponds to a single time-point. Time progresses from top to bottom, and alternate branches are side by side.

Stepping back to consider the general problem of representing commonsense knowledge of causality in physical systems, it is useful to highlight certain points.

(1) The structural description language has approximately the expressive power of differential equations, plus the ability to specify functional constraints as additional states of partial knowledge of a relationship.

(2) The description of quantities in terms of ordinal assertions and IQ values provides a qualitative representation capable of high resolution where the problem demands it, and very low resolution elsewhere, without requiring a premature commitment about which values should be distinguished.

(3) The accuracy and compactness of the envisionment depends on the set of distinguished values that indicate potential qualitative changes. The ability to create new distinguished values corresponding to critical points of the time-varying quantities is important to discovering previously unsuspected points of qualitative change, and avoids the need for premature commitments.

(4) Time is represented explicitly by the structure of the set of time-points, rather than implicitly in the dynamic behavior of the simulator, so the value of each quantity at each time-point is tied into the network of ordinal assertions.

(5) Each inference is irrevocable, so the state of knowledge becomes monotonically better specified as the simulation runs.

At the time this is written, the propagation, envisionment, and summarization components have been completely implemented but the perturbation analysis of the stable equilibrium is done by hand. This paper is a summary of [5], which provides a complete specification for the representation and qualitative simulation.

## ACKNOWLEDGEMENTS AND REFERENCES

Christopher Eliot is responsible for the implementation of this system, and has made substantial contributions to its design. Ken Church, Ken Forbus, Ramesh Patil, and Peter Szolovits have also provided helpful comments.

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

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

[3] K. D. Forbus. 1981. Qualitative reasoning about physical processes. Proceedings of the Seventh International Joint Conference on Artificial Intelligence. Vancouver, B.C.

[4] P. J. Hayes. 1978. Naive physics I: Ontology for liquids. Department of Computer Science, University of Essex, unpublished manuscript.

[5] B. Kuipers. 1982. Representing the Structure of Causal Relationships. Medford, MA: Tufts University Working Papers in Cognitive Science, No. 18.

[6] G. L. Steele, Jr. 1980. The definition and implementation of a computer programming language based on constraints. MIT Artificial Intelligence Laboratory TR-595.

Table 1. Double heat-flow.

* The structural description of the heat-flow system is shown in Table 2(a).
* In time-point (1), starting with the condition that $T_a < T < T_s$, ordinal assertions propagate through the network, but fail to provide information about `net flow`.
* In order to allow the derivative constraint to derive IQ values, the envisionment is split into cases according to the sign of `net flow`. In the branches, with `net flow` specified, IQ values propagate through the network to complete the description.
* Time-point (1E) is quiescent, with all IQ values `steady`, so new distinguished values are created, and the correspondence between quantities taking on distinguished values is recorded.

(net flow: 0) <=> (inflow: flow*)
<=> (outflow: flow*)
<=> ($\Delta T_a$: $\Delta T_a$*)
<=> ($\Delta T_s$: $\Delta T_s$*)
<=> (T: $T_e$)

* Time-points (1G) and (1L) each contain *six* changing values. However, not enough is known to show that they arrive at their limits simultaneously, making the required case split intractibly large, so the envisionment halts.

constant($T_a$)
constant($T_s$)

(1)

$T_a < T < T_s$

| | |
|---|---|
| $\Delta T_a > 0$ | $\Delta T_s > 0$ |
| outflow > 0 | inflow > 0 |

net flow = unknown

Case Split: relation(net flow, 0)

| (1G) | (1L) | (1E) |
|---|---|---|
| net flow > 0 | net flow < 0 | net flow = 0 |
| inflow > outflow > 0 | 0 < inflow < outflow | inflow = outflow > 0 |
| $T_a < T < T_s$ | $T_a < T < T_s$ | $T_a < T < T_s$ |
| $\Delta T_a$, $\Delta T_s$ > 0 | $\Delta T_a$, $\Delta T_s$ > 0 | $\Delta T_a$, $\Delta T_s$ > 0 |
| increasing(T) | decreasing(T) | steady(T) |
| increasing($\Delta T_a$) | decreasing($\Delta T_a$) | steady($\Delta T_a$) |
| increasing(outflow) | increasing(outflow) | steady(outflow) |
| decreasing($\Delta T_s$) | increasing($\Delta T_s$) | steady($\Delta T_s$) |
| decreasing(inflow) | increasing(inflow) | steady(inflow) |
| decreasing(net flow) | increasing(net flow) | steady(net flow) |

Table 1. Envisioning the double heat-flow system.

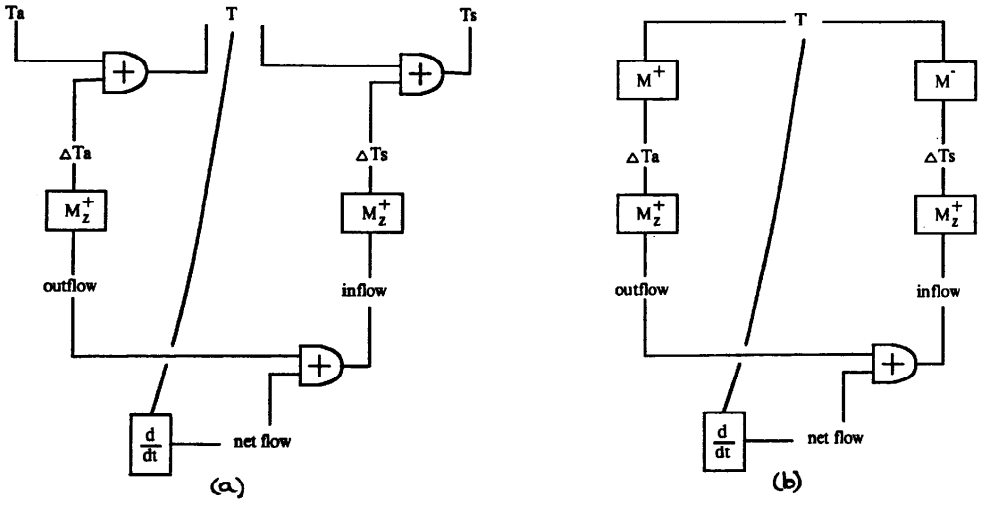
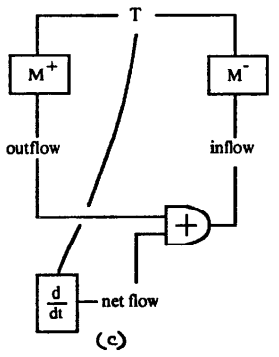



Table 2(a,b): The arithmetic and functional parts of the causal structure description are simplified in three steps, applying the following simplification rules.

(a-b) x + y = z & constant(y) => z = $M^+$(x)
(a-b) x + y = z & constant(z) => y = $M^-$(x)
(b-c) y = $M^+$($M^+$(x)) => y = $M^+$(x)
(b-c) y = $M^-$($M^+$(x)) => y = $M^-$(x)
(c-d) y = $M^-$(x) - $M^+$(x) => y = $M^-$(x)



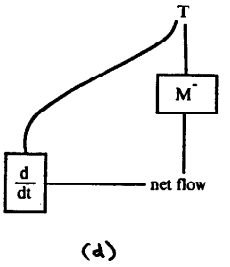
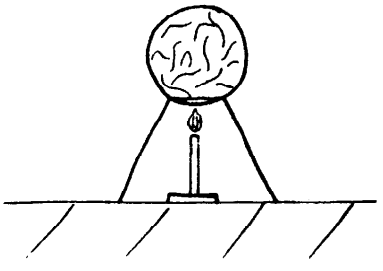




Table 2(c,d): The final step in summarizing the heat-flow description. The resulting structural description is less informative than the original, but equally valid.

Table 3. Summarized heat-flow.

* The summarized structural description is shown in Table 2(d).
* In time-point (1), ordinal assertions propagate as before, and the need for IQ values prompts a case split.
* Time-point (1E) is quiescent as before.
* The previously-determined correspondence makes it possible to infer the relation between T and $T_e$ in time-points (1G) and (1L).
* Since time-points (1G) and (1L) each contain only two changing quantities and their limits are known to correspond, their subsequent states, (2G) and (2L), are easily and unambiguously determined by the Move To Limit rule.
* Since the three branches of the split have identical end states, they are joined to create state (2).



| |
|---|
| ($T$: $T_e$) <=> (net flow: 0) |
| (1)<br>$T_a$ < $T_e$ < $T_s$<br>$T_a$ < T < $T_s$<br>net flow = unknown |
| Case Split: relation(net flow, 0) |

| (1G) | (1L) | (1E) |
|---|---|---|
| net flow > 0<br>$T_a$ < T < $T_e$<br>increasing(T)<br>decreasing(net flow) | net flow < 0<br>$T_e$ < T < $T_s$<br>decreasing(T)<br>increasing(net flow) | net flow = 0<br>T = $T_e$<br>steady(T)<br>steady(net flow) |
| (2G)<br>T = $T_e$<br>net flow = 0<br>steady(T)<br>steady(net flow) | (2L)<br>T = $T_e$<br>net flow = 0<br>steady(T)<br>steady(net flow) | |

| |
|---|
| Case Join: identical outcomes on all branches |
| (2)<br>net flow = 0<br>T = $T_e$<br>steady(T)<br>steady(net flow) |

Table 3. Envisionment of the summarized double heat-flow description.