

## A FRAMEWORK FOR A QUALITATIVE PHYSICS

John Seely Brown  
Johan de Kleer  
Xerox Parc

Change is a ubiquitous characteristic of the physical world. But what is it? What causes it? How can it be described? Thousands of years of investigation have produced a rich and diverse physics which provides many answers. Important concepts and distinctions underlying change in physical systems are state, cause, law, equilibrium, oscillation, momentum, quasistatic approximation, contact force, feedback, etc. Notice that these terms are qualitative and can be intuitively understood. Admittedly they are commonly quantitatively defined. The behavior of a physical system can be described by the exact values of its variables (forces, velocities, positions, pressures, etc.) at each time instant. Such a description, although complete, fails to provide much insight into how the system functions. The insightful concepts and distinctions are usually qualitative, but they are embedded within the much more complex framework established by continuous real-valued variables and differential equations. Our long-term goal is to develop an alternate physics in which these same concepts are derived from a far simpler, but nevertheless formal, qualitative basis.

The motivations for developing a qualitative physics stem from outstanding problems in psychology, education, artificial intelligence, and physics. We want to identify the core knowledge that underlies physical intuition. Humans appear to use a qualitative causal calculus in reasoning about the behavior of their physical environment. Judging from the kinds of explanations humans give, this calculus is quite different from the classical physics taught in classrooms. This raises questions as to what this (naive) physics is like, and how it helps one to reason about the physical world.

In classical physics the crucial distinctions for characterizing physical change are defined within a nonmechanistic framework and thus they are difficult to ground in the common-sense knowledge derived from interaction with the world. Qualitative physics provides an alternate and simpler way of arriving at the same conceptions and distinctions and thus provides a simpler pedagogical basis for educating students about physical mechanisms.

Artificial intelligence and (especially) its subfield of expert systems are producing very sophisticated computer programs capable of solving tasks that require extensive human expertise. A commonly recognized failing of such systems is their extremely narrow range of expertise and their inability to recognize when a problem posed to them is outside this range of expertise. In other words, they have no common-sense. In fact, expert systems usually cannot solve simpler versions of the problems they are designed to solve. The missing common-sense can be supplied, in part, by qualitative reasoning.

A qualitative causal physics provides an alternate way of describing physical phenomena. As compared to modern physics, this qualitative physics

is only at its formative stages and does not have new explanatory value. However, the qualitative physics does suggest some promises of novelty, particularly in its explicit treatment of causality - something modern physics provides no formalism for treating.

Our proposal is to reduce the quantitative precision of the behavioral descriptions but retain the crucial distinctions. Instead of continuous real-valued variables, each variable is described qualitatively - taking on only a small number of values, usually +, -, or 0. Our central modeling tool is the qualitative differential equation, called a confluence. For example, the qualitative behavior of a valve is expressed by the confluence  $\partial P + \partial A - \partial Q = 0$  where Q is the flow through the valve, P is the pressure across the valve, A is the area available for flow, and  $\partial Q$ ,  $\partial A$ , and  $\partial P$  represent changes in Q, A, and P. The confluence represents multiple competing tendencies: the change in area positively influences flow rate and negatively influences pressure, the change in pressure positively influences flow rate, etc. The same variable can appear in many confluences and thus can be influenced in many different ways. In an overall system each confluence must be satisfied individually. Thus if the area is increasing but the flow remains constant, the pressure must decrease no matter what the other influences on the pressure are. A single confluence often cannot characterize the behavior of a component over its entire operating range. In such cases the range must be divided into subregions, each characterized by a different component state in which different confluences apply. For example, the behavior of the valve when it is completely open is quite different from that when it is completely closed. These two concepts, of confluence and of state, form the basis for a qualitative physics, a physics that maintains most of the important distinctions of the usual physics but is far simpler.

#### ENVISION

One of our central tenets of our methodology for exploring the ideas and techniques is to construct computer systems based on them and to compare their results with our expectations. The program ENVISION has been run successfully on hundreds of examples of various types of devices (electronic, translational, hydraulic, acoustic, etc.). Although we view constructing working programs as an important methodological strategy for doing research, the existence of a working implementation contributes little to the conceptual coherence of the theory - it is an existence proof at best.

Although the algorithms ENVISION uses are not the primary focus of this talk, a brief description of its inputs, outputs, and success criteria clarifies the stage for the following conceptual presentations. ENVISION's basic task is to derive function from structure for an arbitrary device. It relies on a single library of generic components and uses the same model library to analyze each device. The input to ENVISION is a description of a particular situation in terms of (1) a set of components and their allowable paths of interaction (i.e., the device's topology), (2) the input forces applied to the device (if any), and (3) a set of boundary conditions which constrain the device's behavior. ENVISION produces a description of the behavior of the system in terms of its allowable states, the values of the system's variables, and the direction these variables are changing. Most

importantly, it produces complete causal and logical analyses for that behavior. Both of these analyses provide explanations of how the system behaves with the causal analysis also identifying all possible feedback paths.

The success criteria for ENVISION are also important. Our physics is qualitative and hence sometimes underdetermines the behavior of a system. In these cases ENVISION produces a set of behaviors (we call these interpretations). At a minimum, for a prediction to be correct, one of the interpretations must correspond to the actual behavior of the real device. A stronger criterion follows from observing that a structural description, abstracted qualitatively, of a particular device implicitly characterizes a wide class of different physically realizable devices with the same device topology. The stronger criterion requires that for the predictions produced by envisioning to be correct then (1) the behavior of each device in the class is described by one of the interpretations, and (2) every interpretation describes the behavior of some device of the class.

We present a new unifying framework: the qualitative differential equation. This new research builds on the ideas of qualitative value, component models, and envisioning, developed in our earlier investigations. These concepts and our methodology are discussed in more detail in our earlier papers. Some of the important concepts developed more recently within our qualitative framework are:

Quasistatic approximation: Most modeling, whether quantitative or qualitative, makes the approximation that behavior at some small time scale is unimportant. In modern thermodynamics this concept is central to the definition of equilibrium. Until now qualitative physics has treated this modeling issue in both an ad hoc and tacit manner. In our formulation quasistatic assumptions play a theoretically motivated and explicit role.

Causality: The behavior of a device is viewed as arising from the interactions of a set of processors, one for each component of the "device." The information-passing interactions of the individual components are the cause-effect interactions between the device's components. Within this framework causal accounts are defined (as interactions that obey certain meta-constraints) and their limitations explored.

Mythical causality and mythical time: Any set of component models make some assumptions about device behavior (i.e., quasistatic assumptions) and hence cannot, in principle, yield causal accounts for the changes that must occur between equilibrium states of a system. In order to handle this problem we have defined new notions of causality and time (i.e., mythical causality and mythical time) cast in terms of information-passing "negotiations" between processors of neighboring components.

Generalized machines: Many physical situations can be viewed as some kind of generalized machine, whose behavior can be described in terms of variable values. These variables include force, velocity, pressure, flow, current, and voltage.

Proof as explanation: Physical laws, viewed as constraints, are acausal. We discuss how a logical proof of the solution of a set of constraints is a kind of acausal explanation of behavior.

Qualitative calculus: Qualitative physics is based on a qualitative calculus, the qualitative analog to the calculus of Newton and Leibniz. We define qualitative versions of value, continuity, differential, and integral.

Episodes: Episodes are used to quantize time into periods within which the device's behavior is significantly different.

Digital physics: Each component of a physical system can be viewed as a simple information processor. The overall behavior of the device is produced by causal interactions between physically adjacent components. Physical laws can then be viewed as emergent properties of the universal "programs" executed by the processors. A new kind of physical law might thus be expressible as constraints on these programs, processors or the information-flow among them.

Structure to Function: We want to be able to infer the behavior of a physical device from a description of its physical structure. The device consists of physically disjoint parts connected together. Each component has a type, whose generic model (i.e., laws governing its behavior) is available in the model library. The structure of a device is described in terms of its components and interconnections. The task is to determine the behavior of a device given its structure and access to the generic model as specified in the model library.

No-function-in-structure: The goal is to draw inferences about the behavior of the composite device solely from laws governing the behaviors of its parts. This view raises a difficult question: where do the laws and the descriptions of the device being studied come from? Unless we place some conditions on the laws and the descriptions, the inferences that can be made may be (implicitly) pre-encoded in the structural description or the model library.

The no-function-in-structure principle is central: the laws of the parts of the device may not presume the functioning of the whole. Take as a simple example a light switch. The model of a switch that states, "if the switch is off, no current flows; and if the switch is on, current flows," violates the no-function-in-structure principle. Although this model correctly describes the behavior of the switches in our offices, it is false as there are many closed switches through which current does not necessarily flow (such as, two switches in series). Current flows in the switch only if it is closed and there is a potential for current flow. One of the reasons why it is surprisingly difficult to create a "context-free" description of a component is that whenever one thinks of how a component behaves, one must, almost by definition, think of it in some type of supporting context. Thus, the properties of how the component functions in that particular supporting context are apt to influence subtly how one models it.

Locality: The principle of locality demands that the laws for a part cannot specifically refer to any other part. A part can only act on or be acted on by its immediate neighbors and its immediate neighbors must be identifiable a priori in the structure. To an extent, locality follows from no-function-in-structure. If a law for part of type A referred to a specific other part of type B it would be making a presupposition that every

device which contained a part of type A also contained a part of type B. The locality principle also plays a crucial role in our definition of causality.

Class-Wide Assumptions: Those assumptions that are idiosyncratic to a particular device must be distinguished from those that are generic to the entire class of devices. For example, the explanation of the pressure regulator's behavior ignored turbulence at the valve seat, Brownian motion of the fluid molecules, and the compressibility of the fluid; these are however all reasonable assumptions to make for a wide class of hydraulic devices. We call such assumptions class-wide assumptions and they form a kind of universal resolution for the "microscope" being used to study the physical device.

Given this definition for class-wide assumptions, the no-function-in-structure principle can be stated more clearly: the laws for the components of a device of a particular class may not make any other assumptions about the behavior of the particular device that are not made about the class in general.

Although as originally phrased, no-function-in-structure is unachievable, its essential idea is preserved through the use of class-wide assumptions. An presupposition behind no-function-in-structure is that it is possible to describe the laws and the parts of a particular device without making any assumptions about the behavior of interest of the device. There is no neutral, objective, assumption-free way of determining the structure of the device and the laws of its components. The no-function-in-structure demands an infinite regress: a complete set of engineering drawings, a geometrical description, and the positions of each of its molecules, all make some unwarranted assumptions for some behavior that is potentially of interest. Thus, we admit that assumptions in general cannot be avoided in the identification of the parts and their laws, which is why class-wide assumptions are crucial.

Class-wide assumptions play two important roles in our qualitative physics. First, they play a definitional role. Formalizing the idealization (i.e., qualitative physics) demands that we be explicit about which assumptions we are making. Second, and as important, they are important for building expert systems. In constructing an expert system to design, operate or troubleshoot complex devices it is critical to clearly state what assumptions are being used in modeling the given device. Thus, when the unexpected situation or causality occurs, these assumptions can be examined to determine whether the "knowledge base" can be relied on.

The most common kind of class-wide assumption is that behavior of short enough duration can be ignored. Under this assumption the "settling" behavior by which the device reaches equilibrium after a disturbance need not be modeled. As "short enough" is a relative term, this assumption can be made at many levels. This assumption plays a major role in studying the heating and cooling of gases. In classical physics it is called the quasistatic approximation. For example, the lumped circuit formulation of electronics makes the quasistatic assumption that the dimensions of the physical circuit are small compared to the wavelength associated with the highest frequency of interest. Other examples of class-wide assumptions are

that the mean free path of the fluid particles is small compared to the distances over which the pressure changes appreciably and that the rate of change of the fields is not too large.

A sophisticated reasoning strategy concerns when and how to change the class-wide assumptions when reasoning about a particular device. Such concerns are critical for troubleshooting where faults can force devices into fundamentally new modes of operation. However, even a simple analysis can sometimes require departures from the usual set of class-wide assumptions. For example, it is sometimes important to remove a class-wide assumption for some localized part of the device, such as two wires running close together which should be modeled as a transmission line.

The Importance of the Principles: Violating the no-function-in-structure principle has no direct consequences on the representation and inference schemes presented. Although the form of the structure and the laws are chosen to minimize blatant violations of the no-function-in-structure principle, it is possible to represent and draw inferences from arbitrary laws--in fact it is too easy.

Without this principle our proposed naive physics would be nothing but a proposal for an architecture for building hand-crafted (and thus ad-hoc) theories for specific devices and situations. It would provide no systematic way of ensuring that a particular class of laws did or did not already have built into them the answers to the questions that the laws were intended to answer. That is not to say that the hand-crafted theories are uninteresting--quite the reverse, and the architecture proposed in this talk may well be appropriate for this task. This is especially true for constructing an account of the knowledge of any one individual about the given physical situation. We are doing something quite different; we want to develop a physics--not a psychological account--which is capable of supporting inferences about the world.

Another purpose for the principles is to draw a distinction between the "work": our proposed naive physics does and the "work" that must be done (outside of our naive physics) to identify the parts and laws. Only after making such a distinction is evaluation possible. Without making the distinction, a reader could always ask, in response to some complexity in an example, "Why didn't they model it differently?"; or in response to some clever inference in an example "They built this into their models." As the principles define what can and what cannot be assumed within the models, the criticisms implied by these two questions are invalid. Of course, the principles themselves are open to challenge.

#### Our Basic Strategic Move

The essence of doing physics is modeling a physical situation, solving the resulting equations and then interpreting the results in physical terms. Modeling a physical situation requires a description of its physical structure. Although there does not exist a general methodology for describing the structure of all physical situations, system dynamics fortunately, provides a methodology for describing a large and interesting collection of physical systems. Thus we initially focus our attention on this class of situations and on how behavior arises from structure. This

move combined with our use of causality as an ontological principle results in a very mechanistic world view. Every physical situation is regarded as some type of physical device or machine made up of individual components, each component contributing to the behavior of the overall device.

**INVITED ADDRESS: ASPECTS OF COGNITIVE REPRESENTATION**

Fred Dretske, University of Wisconsin