Substantial progress has been made in building and simulating qualitative models. However, the results of qualitative simulation may be difficult to understand and programs for explaining them have been few and have had a number of limitations. This paper describes a system, Expound, which provides causal natural language explanations of events in probably "faithful" abstractions of behaviors produced by qualitative simulation. Unlike previous explanation systems, Expound explains events rather than individual states, producing richer and more comprehensive explanations. It also does not require numeric simulation, handles behavior branching, and explains changes other than changes in active processes or in the model. The main steps in building explanations and basic data structures produced are discussed.

Behavior Abstraction

Several early works on qualitative reasoning saw explanation as one of the field's main goals in addition to simulation—that is, systems might state not just what happened but also why. E.g., (de Kleer & Brown 1984). Substantial progress has been made in simulation and model building (Kay 1996; Clancy 1997). However, explanation has proved more difficult and the number of explanation systems has been small. (Forbus 1984, 1990) and (Falkenhainer & Forbus 1990, 1992) explained behaviors in terms of active processes. (Gautier & Gruber 1993a, 1993b) causally explained transitions between different qualitative models. None of these efforts explained qualitative changes accompanied by changes in processes or the model. All used numeric simulation to eliminate branching and thus did not handle branching or purely qualitative behaviors. And none provided explanations that covered multiple qualitative states. (Clancy, Brajnik & Kay 1997) answered questions related to explanation, such as why alternative behaviors are not occurring.

Behavior Abstracting

The first need in explanation is for a comprehensible statement of what happened—a compact description of the system's behavior. Qualitative simulation can produce many branching behaviors, even with tools such as chatter elimination (Clancy & Kuipers 1997a, 1997b). It is often difficult to obtain a coherent overview of the behavior graph from this data. Automatic construction of a comprehensible description is essential. Our system requires abstraction criteria in the form of a selection of the variables of primary interest (typically state variables) and aspects of their behavior to focus on (typically either the qmag or the qdir). It then produces an abstract behavior graph by grouping into one abstract state all adjacent base (original) states that do not differ in these aspects. In (Mallory, Porter & Kuipers 1996; Mallory 1998) we prove that, with certain refinements, this provides a faithful abstraction in that it neither adds nor omits behaviors from the base behavior graph. As suggested above, the criteria for such abstractions are natural and simple to supply and could be supplied automatically. The user will typically build several different abstractions that highlight the behavior of individual state variables or pairs of variables in the model.

At this point we will introduce a running example—a model of the glucose-insulin system in the human body from (Ironi & Stefanelli 1994; Clancy & Kuipers 1994). An abstraction of the model is shown as an influence graph (Forbus 1984) in Figure 1 (the model simulated had twice as many variables). Simulation was performed using Qsim (Kuipers 1994). The behavior graph produced by Qsim of the system perturbed from equilibrium has 188 states and 63 behaviors and all but
Building Explanations

Once we understand what happened, we can proceed to explain why. Our objective is a causal explanation, which explains why a state (or selected aspects of it) occurs.

The process of simulation seems an obvious foundation for explanation but it usually fails for complex behaviors. For small models and simple behaviors, techniques such as propagation in Qsim succeed and provide a foundation for explanation. But propagation often fails, especially in larger models. Qsim then resorts to eliminating sometimes thousands of possible alternatives, leaving a handful of consistent ones. This says why things don’t happen, not why they do, and thus is not a source of concise or satisfying explanations.

Events support better explanations.

Indeed states themselves, even abstract states, are not the best foundation for causal explanation. Instead we focus on events, such as maxima, minima, critical points, etc. Events occur at time points (Clancy & Kuipers 1997a, 1997b) but their descriptions and explanations encompass the surrounding time intervals. They are larger scale, more abstract phenomena than states and also more interesting. Explanations of events explain not only the individual states involved but also the evolution of the behavior from one state to the next. They thus provide additional depth.

This additional depth is considerable. The simple example in Table 1 suggests why. It shows a bathtub departing from and returning to equilibrium. We will explain the behavior of the variable *amount* in terms of the variables *inflow* and *outflow*, whose behaviors we will take as given.

Focusing on the states alone we can say the following:

<table>
<thead>
<tr>
<th>State</th>
<th>S-1 [t0 t1]</th>
<th>S-2 [t1 t2]</th>
<th>S-3 [t2 t3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount</td>
<td>std</td>
<td>inc</td>
<td>std</td>
</tr>
<tr>
<td>Inflow</td>
<td>std</td>
<td>inc</td>
<td>std</td>
</tr>
<tr>
<td>Outflow</td>
<td>std</td>
<td>inc</td>
<td>std</td>
</tr>
</tbody>
</table>

Table 1: A behavior fragment for the amount in a bathtub and its inflow and outflow. The qdirs of three variables are shown in three adjacent states.

- During S-1, inflow equals outflow so amount is steady.
- During S-2, inflow exceeds outflow so amount is increasing.
- During S-3, inflow equals outflow so amount is steady.

This contains the first explanation and also explains how inflow comes to exceed outflow and how they return to equality. It thus explains the cause of amount’s changes. The richer explanation is enabled by focusing on events rather than states. An explanation for an event not only explains the variable values for the time point and the surrounding intervals but also provides a focus uniting these explanations in an overall picture of the transition from one interval to the next.

In addition, what is interesting at the beginning of time interval state S-2 (inflow is increasing faster than outflow) differs from what is interesting at the end (the reverse is true). Focusing on the events reveals this; focusing on the states does not.

Events are thus the focus of Expound’s causal explanations. Which events should be explained is enough a matter of the user’s interest that Expound leaves that selection to the user, informed by abstract behavior graph. Our direction is suggested by Figure 4, which shows Expound’s output explaining maximum of insulin shown in Figure 3.

From behavior graphs to events

Unfortunately, the path from abstracted behaviors to events is more difficult than it might seem. Here we will
only outline the problems; the details are in (Mallory 1998, chapter 6).

**Passages.** Identifying an event requires identifying its path through the behavior graph, which we call a passage. A passage's canonical form is a pair of time intervals separated by a time point, at which the event occurs. The passage may be a pair of abstract time interval states separated by an abstract time point state, but behavior abstraction and branching can make identifying passages more difficult. Table 1 has two passages—from state S-1 to state S-2 through time point t1, and from state S-2 to state S-3 through time point t2; the base time point states have been abstracted into the abstract states S1 and S3. When the behavior branches, analysis becomes more complex. Not all the base states in an abstract time interval state may lead to each of the following time point states, so the set of base states in a time interval may be distinct from the abstract state; a pair of time interval states joined by multiple time point states should sometimes be represented as one passage; and a pair of time interval states joined by one or more time point states and also by a sequence of states should sometimes be represented as two or more passages. Expound builds a graph of passages from the abstract behavior graph to handle these and other cases and represent information about passages.

**Edge qvalues.** Identifying an event also requires the qualitative values (qvalues) of variables, particularly their qdirs, to be determined as precisely as possible before and after the event. Behavior abstraction, including chatter elimination, often produces compound gmags and qdirs for variables in abstract states. But, for example, if we only know that a variable was increasing or steady before an event and steady or decreasing afterward, we do not know whether the variable reached a maximum, departed from steady, arrived at steady, or experienced no event at all. However, there is often some sub-interval of time adjacent to a time point during which the qualitative value of the variable must be more narrowly defined. In Table 1, the derivative of the amount has qdir (inc std dec) during the interval (t1 t2), but it must be increasing immediately after t1 and decreasing immediately before t2 if it is to be 0 at t1 and t2 but positive in between. This and other methods may be applied to narrow the qualitative values of variables during the “edges” of intervals around an event. We call these values edge qvalues. We refer to the edge interval before the time point as the prologue.
Figure 3: An abstract behavior graph showing the qualitative values of insulin.

During the passage from state A-30 through state A-31 to state A-32, the amount of insulin reaches a maximum and begins decreasing.

Before the event, in state A-30, production of insulin exceeds elimination of insulin, so the amount of insulin is increasing. However, production is decreasing and elimination is increasing. Eventually production falls below elimination and the amount of insulin begins decreasing.

At this point, the amount of insulin reaching a maximum causes elimination to reach a maximum and to begin decreasing. However, production is decreasing more rapidly than elimination is decreasing and the amount of insulin continues to decrease.

Figure 4: Output from Expound explaining a maximum in the amount of insulin. This explanation was produced with reference to the constraints in Figure 1.

of the passage, the edge interval after the time point as the epilogue, and the time point itself as the vertex.

Events. With passages and edge values computed, the events in each passage are identified. Table 1 contains six events— at t1, all three variables depart from steady, and at t2 they all arrive at steady again. Figure 3 illustrates maxima, minima, arrival at steady (quiescence), and crossing a landmark (norm). The only other event involving a change in qdir is a critical point. With no change in qdir, a variable may cross a landmark or otherwise shift its magnitude or it may simply experience a null event (no change).

From events to reasons

To derive explanations, we need to look at them more closely. Here we discuss the definition and construction of structures for representing the reasons for events, which underlie the explanations in natural language.

Influences have a direction of causation. Causal explanation relates behavior to the model. When a constraint has a direction of causation, the behavior of the dependent variable is caused by the behaviors of the remaining, independent variables and we use this interpretation to construct causal explanations.

In the QPT perspective, all influences have causal directions (Forbus 1984, 1990) (Falkenhainer & Forbus 1990, 1992). In Qsim terms, unless a variable is exogenous, there is a single constraint on the variable, and it is the collection of all the influences in which the variable is the dependent variable. These influences will either be all indirect (differential) or all direct (monotonic functional). In all the other constraints in which the variable participates, it is an independent variable.

This is in accord with Iwasaki and Simon’s demonstration that it is usually possible to deduce the causal direction of a constraint without using any knowledge from the domain (Iwasaki & Simon 1986, 1994). As a practical matter, we have found little difficulty in identifying the causal directions of constraints (Rickel & Porter 1997) and have supplied that information to Expound.

Thus to explain an event, we relate the behavior of the event variable to the behavior of the independent variables in the variable’s constraint.

Causal classification of variables. The first step in this process is to assign each of the independent variables one of several possible causal classifications. The causal classification states the independent variable’s causal role in the event. Specifically, an independent variable may be classified as—

- **Causing** the behavior of the dependent variable, by itself or in concert with other independent variables;
- **Opposing** the behavior of the dependent variable, which is caused by other independent variables;
- **Steady** and thus not contributing to the behavior of the dependent variable (unless it and all other independent variables are also steady); or
- having some other less interesting causal relationship to the dependent variable in case of chatter and certain other circumstances.

The causal classification of a variable may vary from prologue to epilogue.

Take for example the maxima of insulin and its rate of elimination at state A-31 in Figure 3, explained in Figure 4. The influences in Figure 1 show that the rates of production and elimination of insulin determine the
amount of insulin and that the rate of elimination is in turn a function of the amount of insulin.

For monotonic functional influences, the qdirs of the independent variables explain the qdir of the dependent variable. The dependent variable is the signed sum of the independent variables, or a monotonic function of the independent variables, so increasing independent variables with a positive sign and decreasing independent variables with a negative sign tend to cause the dependent variable to increase, and similarly for other signs and qdirs. Independent variables with qdirs that tend to cause the observed qdir of the dependent variable are classified as Causing and those that tend to cause the opposite qdir are classified as Opposing. Independent variables that are steady do not affect the qdir of the dependent variable unless all the independent variables are steady, causing the dependent variable to be steady also. In the insulin example, the rate of elimination is a function of the amount, so the qdir of the amount always causes the qdir of the rate of elimination.

For differential influences (I+ and I−) the situation is less intuitive. A variable subject to differential influences is constrained by its derivative. We will call such a variable an amount, as it is in our examples. An amount is increasing, steady, or decreasing as its derivative is positive, zero, or negative, respectively. Thus a change in qdir of the amount corresponds to a change in sign of the derivative, and the cause of the latter is the cause of the former. But in any event in the qdir of an amount with simple (atomic) qdirs, the amount is steady at the vertex and non-steady before and/or after. Thus the derivative is zero at the vertex and non-zero before and/or after and the explanation may focus on why the derivative becomes zero or diverges from zero or both. But the latter is adequately explained if we explain the qdir of the derivative. Thus a qdir event for an amount is explained by the qdir of its derivative.

Furthermore, the derivative will be constrained by a functional constraint—the signed sum of the independent variables of the differential influences—so the derivative’s qdir is explained by the qdirs of its influencing variables, the inflows and outflows to the amount. These influencing variables will have the same causal classifications with respect to the derivative as independent variables in other functional constraints. We can thus use the same set of causal classifications for qdir events for variables subject to both differential and functional constraints. The wording of the explanations themselves will, of course, differ but the underlying causal classifications are the same.

We will apply this to the insulin example. As suggested by Figure 4, production is decreasing during the prologue and epilogue; this by itself could cause the event, so production is classified as Causing in both the prologue and epilogue. Elimination is also increasing during the prologue; this by itself could cause the amount to become steady, so elimination is classified as Causing in the prologue. However, elimination is functionally related to the amount of insulin (see Figure 1), so when the amount experiences a maximum and then decreases during the epilogue, so does elimination. This tends to counteract the decrease of the amount—if the decrease in elimination were the only change after the event in the variables affecting the amount, the amount would not decrease after the event. Elimination is thus classified as Opposing during the epilogue.

Reasons and the reason graph. As in Figure 4, we may wish to explain several related events together. In Figure 4 there are only two such events—maxima in the amount and rate of elimination of insulin. In other models there may be chains or trees of related events, as from one end of a causal chain to the other. For example, in a model of the relationship between the turgor pressure in a plant leaf and the opening of the stomates in the leaf, there is a chain of influences from the former to the latter. The explanation of how the former influences the latter needs to reflect this chain. (The model contains 12 variables and 19 influences in its abstracted form. It would require a half page to print here, so we omit it.) Expound’s explanation of how decreasing turgor pressure leads to decreasing stomatal opening is shown in Figure 6.

To produce coordinated explanations of such related events, Expound builds the reasons for an event as a graph. A reason is an edge from the event for some variable to the event for one of its influencing variables, labeled by the causal classifications of the influencing variable. The reason graph is the collection of all such edges for a particular passage. The reason graph for the two events explained in Figure 4 is shown in Figure 5.

Every variable in a model has an event in a passage, even if only a null event. The reason graph may have about as many reasons as the model has influences. However, it usually has considerably fewer because it is unnecessary to construct reasons for null events and for changes in qmag only (no qdir event) (they are explained by the qdirs of the event variables, which have no event to explain). The reason graph can be appropriately ordered, pruned, and traversed to obtain the best order for explaining several related events. For example, while the influence diagram behind Figure 6 is relatively linear, it does have several branches and cycles. The reason graph and the causal classifications of the reasons enable Expound to traverse it properly.

Explanations and evaluation

Natural language explanations are produced from the reason graph, state and edge qvalues, and other sources. See (Mallory 1998, chapter 8). Expound uses highly conditionalized templates to cover the large number of different forms of language that seemed clearest in various circumstances. Although not the focus of this work, the results have proved adequate for the examples used. Some additional inference is needed to determine—
Initially, in state A-0, the system is in equilibrium. During the passage from state A-0 to state A-1, the following chain of events occurs.

First, the pressure of water in the leaf symplast stops being steady and begins decreasing.
This causes synthesis of ABA in the mesophyll ... to begin increasing.
This causes the amount of ABA in the mesophyll ... to begin increasing.
This causes transport of ABA from the mesophyll to the guard cells ... to begin increasing.
This causes the amount of ABA in the guard cells ... to begin increasing.
This causes active transport of K+ from the guard cells to the accessory cells ... to begin increasing.
This causes the amount of K+ in the guard cells ... to begin decreasing.
This causes the opening of the stoma ... to begin decreasing.

They also show that Expound's results are available to those with modest familiarity with Qsim.

Conclusions

Building faithful abstractions of qualitative behaviors and focusing on events rather than states enables the construction of rich, high-level explanations of qualitative behaviors. Expound provides this while handling behavior branching and without requiring numeric information or limiting its explanations to changes in active processes or in the model. The basic information underlying the natural language explanations is a graph of reasons which relates behavior to the model and records the causal classifications of variable behaviors.

References


Citation Abbreviations


