On Temporal Abstraction in Qualitative Physics

- A Preliminary Report -

4/9/93

Peter Struss
Technical University of Munich, Computer Science Dept.
Orleansstr. 34
W-8000 Munich 80
Germany
struss@informatik.tu-muenchen.de

Abstract

Abstraction is a fundamental element of qualitative reasoning about physical systems and crucial for coping with complexity of real problems. Temporal abstractions form an important class that comprises a variety of different transformations. We establish goals for a formalization of abstractions, discuss some examples of temporal abstraction, and show how they can be formalized in a theory of relational behavior models of physical systems.

1 Introduction

Transforming representations of problems into new ones is often crucial for finding solutions efficiently, sometimes for finding them at all. An important class of such transformations are abstractions. Abstraction is a conceptual generalization obtained by eliminating individual and arbitrary characteristics while maintaining those distinctions only that are essential in a particular context. It is frequently, and often unconsciously, applied in human reasoning and problem solving and promises to be good means for coping with complexity in automated reasoning and problem solving.

Whilst many AI researchers agree upon this, the attempts to formalize abstraction are quite diverse, sometimes reflecting different understanding of the nature of abstraction. They are also not comprehensive. In particular, we are lacking a theory of a quite distinctive and important type of abstraction: temporal abstraction.

In [Struss 91,92], we have presented a theory of multiple modeling and model transformations in the context of model-based and qualitative reasoning. It covers abstraction to some degree, and we have demonstrated the importance of such a rigorous theory for practical purposes: it helped to analyze the precise logical relations between alternative models used for fault localization in power transmission networks. However, the behavior models presented in [Struss 92] do not contain time. In the example, they only express the information which switches in the network were automatically opened (to isolate a short circuit), but not when exactly this happens or in which order. Since this may be a quite important aspect for determining the proper diagnoses, we better try to understand what is being lost by the use of the static models compared to a "gold standard" model that captures the temporal information. Thus, the practical task created the need for including temporal transformations in the theory. Other examples will be mentioned in this paper.

Beyond this, a theory of temporal abstraction contributes to a fundamental understanding of the existing qualitative reasoning systems, because they always include (implicitly or explicitly) some transformation of time, for instance, by turning time as a real number line into a mere ordering on qualitative states of a system.

The use of abstraction has been investigated and formalized in other areas, such as theorem proving and planning, and this work (in particular, [Giunchiglia-Walsh 89,92]) has been advocated as a theoretical basis in the area of qualitative reasoning ([Weld 92]). In order to clarify our notion of abstraction, we argue in the following section that this theory, being a useful formal theory of theory transformations, is too general to capture some of the essential characteristics of abstraction transformations. In particular, this is done by discussing the question whether abstraction may introduce new theorems and inconsistencies that are not present in the "ground theory".

Section 3 summarizes then our theory of multiple relational models that tries to preserve more of the content of abstraction and whose form is more adapted to specific representations in qualitative reasoning about physical systems. Different kinds of temporal abstraction are discussed in section 4, and finally, we demonstrate that the theory presented provides a foundation for formalizing at least some of them.

Although we are aware of the fact that results presented in this paper are still preliminary, we believe they may stimulate the discussion of this important area in the qualitative reasoning community.

2 Can an Abstraction Prove More Theorems and Provide False Proofs?

Based on work in theorem proving and planning, [Giunchiglia-Walsh 89,92] propose a comprehensive theory of abstraction. In their work, an abstraction is considered to be **any mapping**

$f: \quad \Sigma_1 \rightarrow \Sigma_2$

between two formal systems $\Sigma_i = (\Lambda_i, \Omega_i, \Delta_i)$, given by a triple $f = (f_\Lambda, f_\Omega, f_\Delta)$ of total functions

$f_\Lambda: \quad \Lambda_1 \rightarrow \Lambda_2$,

$f_\Omega: \quad \Omega_1 \rightarrow \Omega_2$,

$f_\Delta: \quad \Delta_1 \rightarrow \Delta_2$,

where Λ_i denotes the respective language, Ω_i the set of axioms, and Δ_i the deductive machinery.

They analyze, as special cases, mappings that are theorem increasing ("TI-abstractions"), i.e. for any well-formed formula ϕ holds

if $\vdash_{\Sigma_1} \phi$, then $\vdash_{\Sigma_2} f(\phi)$,

and theorem decreasing ("TD-abstractions"):

if $\vdash_{\Sigma_2} f(\phi)$, then $\vdash_{\Sigma_1} \phi$.

They have to cope with the "false proof problem": for a consistent Σ_1, a "TI-abstraction" may be inconsistent. Assume, for instance, that two different constants c_1, c_2 in Σ_1 which are positive real numbers are mapped onto the same constant in Σ_2, e.g.

$f(c_1) = f(c_2) = +$.

If f maps Σ_1's axiom $\neg(c1 = c2)$ onto

$f(\neg(c1 = c2)) = \neg(f(c_1) = f(c_2)) = \neg(+ = +)$,

(because f is assumed to be "negation preserving": $f(\neg\phi) = \neg f(\phi)$), Σ_2 is inconsistent.

At this point, we would like to pay some tribute to the terminology debate. In our opinion, the false proof problem, the existence of "theorem increasing abstractions", and even the fundamental definition of abstraction offered by [Giunchiglia-Walsh 92] contradicts a generally agreed meaning of the term. In their theory, abstraction can be an arbitrary mapping between formal systems (based on total functions for the elements), and indeed in [Giunchiglia-Walsh 92], the authors characterize their intuition of abstraction as being any mapping from one representation to another one, "throwing away details", but "preserving certain desirable properties". While these are necessary properties of an abstraction transformation, they are not sufficient for capturing its full meaning (and its practical importance!).

Abstraction is a **process of generalization**. It steps **from individual objects to concepts** of these objects that **capture their essence** but **eliminate their individual and incidental properties**. More technically speaking, it creates equivalence classes of objects (or of existing concepts to build more abstract concepts). Of course, what is considered essential can depend on the context, task, or perspective.

The commonality of the behaviors of all individual resistors (w.r.t. some electrical phenomena, under certain contextual restrictions, etc.) is condensed to the concept of "the resistor". It does not contain more; it does not contain anything that does not hold for each single resistor (otherwise this would be no resistor). It might even not contain some aspects that are common to all resistors, but not essential under the given perspective (e.g. that they are material objects). How could this concept, this abstraction be used to derive any theorem about resistors that is not a theorem for each individual resistor?* How could it be inconsistent with anything that was not already inconsistent with the individual resistors? So, in our view, abstraction cannot be TI and must exclude the false proof problem. It may lose some theorems about individual objects, even though they may be true for all instances of a concept (if it treats the respective properties as irrelevant). Abstraction will have to be a subset of "TD-abstractions".

Let us emphasize that this standpoint is not an absolute criticism of [Giunchiglia-Walsh 89,92]. It is a very useful formal theory of theory transformations. But it is far too "abstract" to capture the essentials of abstraction that make it a powerful means for coping with complexity and still produce correct results. Preservation of truth is a property of abstraction we propose to maintain as a crucial property (as does [Weld-Addanki 92]), if one wants to construct a theory of abstraction (in our strict sense) within the framework of [Giunchiglia-Walsh 89,92]. We believe that this is possible. It requires a careful analysis of the types and interesting properties of the mapping f.

For instance, since abstraction by its very nature eliminates distinctions between some individuals (and also, perhaps, predicates), preservation of negation is probably not a general feature: things that are **not** the **same** can be mapped onto the **same** abstract concept.

It seems to us that the theory of granularity [Hobbs 87], in constructing equivalence classes of objects that cannot be distinguished by any predicate that is considered to be relevant, is closer to what we would like to consider as abstraction (we are not sure whether the reconstruction of this work in [Giunchiglia-Walsh 89,92] , which is TI, is a faithful one).

----

* There may be statements about the concept as a concept (e.g. that it is a concept useful for X) that are not true for the individuals represented, but they are "meta", i.e. not in the image of f.

In this paper, we do not attempt this specialization of [Giunchiglia-Walsh 89,92] . This would be more general than required by our goals in the area of qualitative physics. Furthermore, we believe that there is a different way better suited to describe the particular transformations that are commonly applied to models of physical systems. In the following section, we summarize the formalism for relational models we developed in [Struss 91,92] for structuring sets of multiple models and, in particular, for using this structure in model-based diagnosis.

3 Relational Models and their Transformations

Our approach is dual w.r.t. the one discussed in the previous section in the sense that, rather than viewing the problem from the perspective of **theories and proofs**, it analyzes different transformations applied to the set of **models** (in the logical sense)*

Different **representional spaces** for the behavior of a physical system, which may be an atomic constituent (component) or some aggregate, are given by different vectors of local variables v_i:

$\underline{v} = (v_1, v_2, ..., v_k).$,

and one or more domains of $\underline{v}$:

$DOM(\underline{v}) := DOM(v_1) \times DOM(v_2) \times ... \times DOM(v_k)$.

(In contrast to [Struss 91,92], we omit the indexing of variables, models etc. with the respective component, because it is not relevant to our discussion).

A behavior of the system is described by specifying the set of possible values of $\underline{v}$, i.e. by a relation $R \subseteq DOM(\underline{v})$. As a logical formula, the respective behavior model can be regarded as the statement that R contains (exactly) the values that can be observed in real situations:

Definition 3.1 (Strong Behavior Model, Complete Behavior Model)

A relation $R \subseteq DOM(\underline{v})$ specifies a strong behavior model by

$B(R) \Leftrightarrow$

$\forall \underline{v}_0 \in DOM(\underline{v}) \quad ((\exists s \in SIT\ Val(s, \underline{v}, \underline{v}_0)) \Leftrightarrow \underline{v}_0 \in R)$,

and a complete behavior model of C by

$M(R) \Leftrightarrow$

$\forall \underline{v}_0 \in DOM(\underline{v}) \quad ((\exists s \in SIT\ Val(s, \underline{v}, \underline{v}_0)) \Rightarrow \underline{v}_0 \in R)$.

(In [Struss 92], we focus on complete models, because they suffice for consistency-based diagnosis; in [Struss 92a], we introduce its obvious dual concept, sound models). Here, $Val(s, \underline{v}, \underline{v}_0)$ means that $\underline{v}$ has the value $\underline{v}_0$ in the situation s, and if $\underline{v}_0 = (v_{01}, v_{02}, ..., v_{0k})$, then

$Val(s, \underline{v}, \underline{v}_0) \Leftrightarrow \bigwedge_i Val(s, v_i, v_{0i})$

holds. Note that we do not postulate that the value be unique; from

$\exists s \in SIT\ (Val(s, \underline{v}, \underline{v}_0) \wedge Val(s, \underline{v}, \underline{v}_1))$

we cannot infer $\underline{v}_0 = \underline{v}_1$. This does not only allow us to handle multiple domains for $\underline{v}$. Even if $\underline{v}_0$ and $\underline{v}_1$ are from the same domain, they may be different. For instance, in the domain of intervals of real numbers

$Val(s, x, (1, 3)) \wedge Val(s, x, (2, 5))$

is perfectly consistent, if $Val(s, x, (a, b))$ means

$\exists r \in (a, b) \subseteq \mathbb{R}\ Val(s, x, r)$.

(In this case, we may want to infer a relation weaker than equality, namely that the two values have a non-empty intersection).

New behavior models can be obtained from an existing one in two ways:

- by transforming the relation $R \subseteq DOM(\underline{v})$ to some relation R' in the same representation. Of course, if this transformation is not the identity, the property of a strong behavior model is lost, while a complete model property may survive,
- by transforming the representation space:

  $\tau: DOM(\underline{v}) \rightarrow DOM'(\underline{v}')$

  and, thus, obtaining a relation $R' = \tau(R) \subseteq DOM'(\underline{v}')$ from $R \subseteq DOM(\underline{v})$. In contrast to [Giunchiglia-Walsh 89,92] , here a set of (logical) models is transformed under a mapping of the spaces of interpretations.

The former allows us to directly express a variety of common modifications of descriptions of system behaviors, such as linear approximation (replacing R by the graph of a piecewise linear function) and introducing tolerances (by expanding R). The latter provides the basis for exploiting other, equally frequently applied, mappings between different representations, for instance switching from Cartesian to polar coordinates. In [Struss 92], we illustrated such mappings also by structural aggregation (by dropping internal variables) and mapping domains into equivalence classes (e.g. real numbers to intervals between landmarks) using a real world example (These transformations are closely related to those in the independent work on hierarchical diagnosis reported in [Mozetic 92] which is more in the spirit of [Giunchiglia-Walsh 89,92]).

Again, the question is raised what preconditions ensure the preservation of behavior model properties. A quite obvious and basic conditions turns out to suffice, namely that the Val-predicate is preserved in both directions:

Definition 3.2 (Representational Transformation)

A mapping

$\tau: DOM(\underline{v}) \rightarrow DOM'(\underline{v}')$

is a representational transformation, iff

$Val(s, \underline{v}, \underline{v}_0) \wedge \underline{v}_0 \in DOM(\underline{v}) \Rightarrow Val(s, \underline{v}', \tau(\underline{v}_0))$

and

$Val(s, \underline{v}', \underline{v}'_0) \wedge \underline{v}_0 \in DOM'(\underline{v}')$

$\Rightarrow \exists \underline{v}_0 \in \tau^{-1}(\underline{v}'_0)\ Val(s, \underline{v}, \underline{v}_0)$.

----

* Because the term "model" is used in different ways in logic and qualitative physics, what has sometimes lead to confusion, we will try to use "behavior model" or "component model" for descriptions (theories) of the behaviors of a physical system.

Intuitively speaking, the second condition states that the target domain does not introduce values that are not grounded in the original one (Actually, this definition becomes slightly more complicated for domains that allow for multiple values, such as intervals, see [Struss 93]).

Theorem 3.1
If τ: $DOM(\underline{v}) \rightarrow DOM'(\underline{v}')$
is a **representational transformation**, then the **image of a strong model is a strong model**:
$B(R) \vdash B(\tau(R))$.

(For the preservation of complete behavior models the second condition of Def. 3.2 is sufficient). Obviously, we are getting closer to our concept of an abstraction transformation.

Definition 3.3 (Abstraction)
An abstraction is a **representational transformation** that is
- **surjective**
(i.e. each element of $DOM'(\underline{v}')$ is the image of an element of $DOM(\underline{v})$) and is
- **not injective**
(i.e. some elements of $DOM'(\underline{v}')$ are the image of more than one element of $DOM(\underline{v})$).

Although we feel these properties still to be too weak to capture all aspects (more like a necessary condition rather than a definition), the definition expresses that
- behavior model abstraction is induced by a **transformation of representations** (in accordance with the intentions of [Giunchiglia-Walsh 89,92]), as opposed to "surgery" applied to a particular behavior model in one representation (e.g. by approximations),
- this transformation preserves and grounds valid values,
- it eliminates certain distinctions between tuples of values,
- it preserves the truth of the strong behavior model property.

(As a side remark, we mention that Def. 3.2 makes the concept insensitive to what the mapping does to portions of the domain that do not occur in real situations. This seems to be a nice feature, because it makes the abstraction "pragmatic", or grounded in reality. This can be exploited by reasoning based on working hypotheses as illustrated for diagnosis in [Struss 92]).

So far, in previous presentations of our theory and in this paper, we used static views on physical systems to illustrate the concepts. It is now time to ask whether modeling dynamic systems and formalizing temporal abstraction can be handled within this framework.

4 Types of Temporal Abstraction

Unfortunately, there seem to be quite different methods to be subsumed under the heading of temporal abstraction. Describing behaviors as sequences (as done in qualitative simulation), viewing a slow process as providing constant conditions from the perspective of a faster one ([Kuipers 87], [Iwasaki 92]), identifying a behavior as cyclic ([Weld 86]), and characterizing a signal as changing at least once in an interval ([Hamscher 91]) appear to be quite different operations, but still share that some aspect of time has been "abstracted away".

If abstraction preserves some properties while treating others as unessential, we may start by asking what the properties of time are that might be kept, weakened, or thrown away by abstraction. We can describe a particular temporal ontology by a triple
$(T, (A, d), <)$
of a universe of time instances T, a metric
d: $T \times T \rightarrow \mathbb{R}_0^+$,
an algebraic structure $A = (+, -, \ldots)$ containing, at least, addition and subtraction:
$+,-: T \times T \rightarrow T$,
and an ordering relation
$< \in T \times T$.
Often, we consider the ground representation of time to be given by $T = \mathbb{R}$ with the usual arithmetic, metric and order. We combine A and d, because, usually, they are tighly related. e.g. by defining
$d_{\mathbb{R}}(a, b) := |a - b|$
on $T = \mathbb{R}$.

Also d and < are related, e.g. by
$0 < a \wedge 0 < b \wedge d(0, a) < d(0, b) \Rightarrow a < b$.

It should be possible to characterize different types of temporal abstractions by the changes they apply to one or more of the three elements.

Changing the set of time instances T:

Enforced by principled limitations of our technical equipment, one frequently applied treatment of time is ignorance of all time instances except for some set of measure zero, more explicitly: sampling at a finite set of time points $S \subset \mathbb{R}$ while preserving the metric and order:
$(\mathbb{R}, (A_{\mathbb{R}}, d_{\mathbb{R}}), <_{\mathbb{R}}) \rightarrow (S, (A_{\mathbb{R}}, d_{\mathbb{R}}), <_{\mathbb{R}})$.

(Note, however, that S is not necessarily closed under $A_{\mathbb{R}}$, e.g. if it is not an equidistant sample of time points). Another nice example is given in [Hobbs 85], where continuous time is mapped onto a set of time instances determined by the end of events in the blocks world.

Dropping metric properties:

The transformation of continuous functions into sequences of qualitative states is quite fundamental for qualitative simulation and envisioning. This step saves something of the ordering, but eliminates metric information, and can be described by a mapping
$(\mathbb{R}, (A_{\mathbb{R}}, d_{\mathbb{R}}), <_{\mathbb{R}}) \rightarrow (\mathbf{Z'}, <_{Z})$

of temporal representations, where $Z' \subset Z$ is a (contiguous) subset of the integers. Thus, for instance, oscillations with different frequencies are collapsed into a sequence (..., 0, +, 0, -, 0, +, ...) of qualitative values (signs).

Switching metrics:

Combining changes that take place with speed (and, hence, temporal granularity) of different orders of magnitude has been investigated and used e.g. in [Weld 86], [Kuipers 87], [Iwasaki 92], where the faster process is seen as instantaneous from the viewpoint of the slower one, which in turn appears to provide constant conditions for the fast process. As one way for formalizing this technique we propose to provide an explicit link between different metrics:

$d_{T_1} \rightarrow d_{T_2}$.

Changing everything:

In qualitative reasoning, often intervals are used instead of real numbers representing time points. In this case, algebraic operations, metric, and ordering have to be defined accordingly, but are expected to preserve the structure on $\mathbb{R}$ to some degree. For instance, if two time points $p_1, p_2 \in \mathbb{R}$ are mapped onto intervals $i_1, i_2 \in I(\mathbb{R})$, then

$p_1 < p_2 \Rightarrow i_1 \leq i_2$

should hold.

Dropping (almost) everything:

Sometimes, neither metric nor ordering information matters, and time is abstracted away entirely. For instance, our power network diagnosis system DPNet ([Struss 92]) only exploits the information which switches were opened (to isolate a short circuit), but not when exactly this happens or in which order. XDE for trouble-shooting of digital circuits ([Hamscher 91]) contains a number of temporal abstraction operators, such as counting the number of changes in a signal in a given interval (Because of the Boolean domain, the ordering is implicitly determined in this case). This reflects, among other things, that an observer has easy access to this abstract type of information only.

The question we will be discussing in the following section is whether and how we can formalize these kinds of temporal abstractions in our theory of abstraction based on relational behavior models as outlined in section 3.

5 Behavioral Abstraction of Relational Models

5.1 A Naive(?) Approach to Temporal Abstraction

As a matter of fact, the properties and their transformations we considered in the previous section are not very specific for time, but could be related at least to any variable that is considered to have the real numbers as the ultimate ground representation. So, one might be tempted to simply introduce time as another variable in the vector $\underline{v}$:

$\underline{v} = (v_1, v_2, ..., v_k, t)$

and describe the behavior over time by a relation

$$R \subseteq DOM(\underline{v}) := DOM(v_1) \times DOM(v_2) \times ... \times DOM(v_k) \times T .$$

However, we have to be careful with this representation, because it already involves a step of abstraction: consider two functions

$x = f_1(t) = t^3$ and $x = f_2(t) = -t^3$

describing the possible behavior of a system (Fig. 5.1).

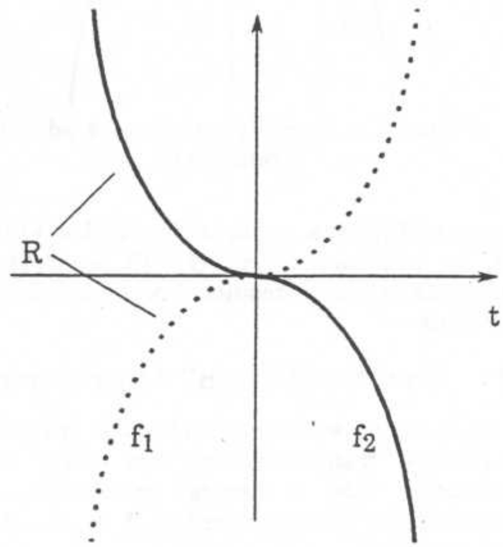

Figure 5.1 A relation R specified by f_1 and f_2

Choosing

$\underline{v} = (x, t)$, $DOM(\underline{v}) = \mathbb{R} \times \mathbb{R}$

as the representation leaves us with a relation

$$R = \bigcup_{t \in \mathbb{R}} \{t^3, -t^3\} \times \{t\} .$$

But R represents, besides $\{f_1, f_2\}$, also the set of continuous functions $\{f_3, f_4\}$ with

$f_3(t) = |t^3|$ and $f_4(t) = -|t^3|$,

which are obtained by "pasting together" the pieces of f_1 and f_2 at $t = 0$ (see Fig. 5.2). If we drop the property of continuity, R covers even an infinite set of functions. This means, this kind of representation, though capturing the possible values at each time point, does not preserve the actual changes over time. We discuss this issue at more length than the "naive" representation might deserve, because this kind of abstraction, which may lose the global behavior over time, is actually a cause for the generation of spurious behaviors of many qualitative physics systems. Technically, it is due to merely local transition analysis which does not reflect the full "history" of a behavior when determining the possible transitions from the most recent state into the next one (see

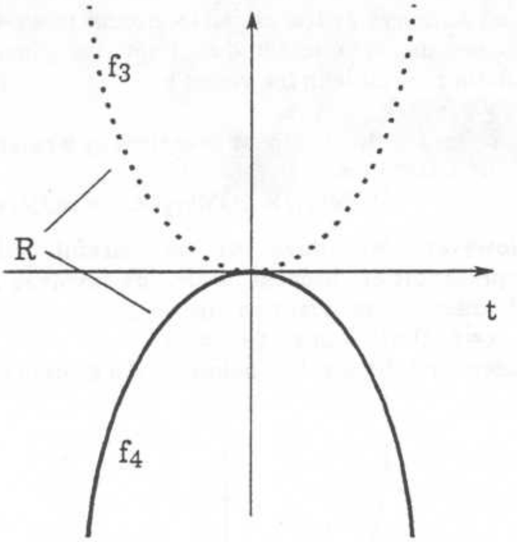

Figure 5.2 f_3 and f_4 are also covered by the relation R

[Kuipers 86]). Thus, we have identified an abstraction that is intrinsic to many QR systems, which sometimes leads to results more general than we appreciate.

5.2 The "Oscillation" Abstraction

To do it right, we now consider the problem of an abstraction transformation that maps different oscillations into a general qualitative concept represented by the sequence (..., 0, +, 0, -, 0, +, ...), or, as a relation,

$$\begin{aligned} R_{Q_3} = &\{(0, t) \mid t = 2n, n \in \mathbb{N}_0\} \\ &\cup \{(+, t) \mid t = 4n+1, n \in \mathbb{N}_0\} \\ &\cup \{(-, t) \mid t = 4n-1, n \in \mathbb{N}_0\} \\ &\subseteq \{0, +, -\} \times \mathbb{N}_0 = Q_3 \times \mathbb{N}_0 = DOM'((x, t)). \end{aligned}$$

Let us assume that two possible behaviors are described in the ground representation

$$DOM((x, t)) = [-1, 1] \times \mathbb{R}_0^+$$

by

$$R_R = \{(\sin t, t) \mid t \in \mathbb{R}_0^+\} \cup \{(\sin 2t, t) \mid t \in \mathbb{R}_0^+\}$$

(Fig. 5.3).

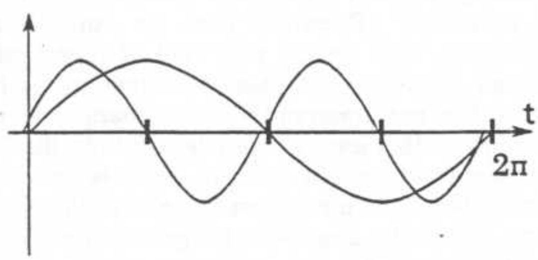

Figure 5.3 (sin t, t) and (sin 2t, t)

We notice that it is impossible to find a mapping

$$\tau_t: \mathbb{R}_0^+ \to \mathbb{N}_0$$

of the temporal spaces such that, in conjunction with the qualitative domain abstraction

$$\tau_q: [-1, 1] \to Q_3,$$

it forms a representational transformation

$$\tau = (\tau_q, \tau_t): [-1, 1] \times \mathbb{R}_0^- \to Q_3 \times \mathbb{N}_0.$$

For instance, the tuple $(0, 2\pi) \in DOM((x, t))$ is shared by both oscillations. But for sin t, it has to be mapped onto (0, 2), whereas sin 2t reaches 0 for the fourth time, and, hence, $(0, 2\pi)$ corresponds to (0, 8). In other words, the temporal transformation is not a unique one, but may be specific for each single behavior. We see that temporal abstraction is often **behavioral abstraction**, and it might seem that we have to drop our claim that abstraction is to be considered as a general mapping between representational spaces rather than tied to particular behaviors.

However, for the relational models, there are no restrictions imposed on the choice of the domains for the variables. They need not be sets of real numbers, integers, or qualitative values, as before. They can also be **sets of functions over time**. So, if T is some temporal universe, and $DOM_v(v_i)$ denotes the set of possible values v_i can take in principle at each time instance, let

$$F(T, DOM_v(v_i)) = \{f \mid f: T \to DOM_v(v_i)\}$$

be the set of functions in $DOM_v(v_i)$ over time and

$$F(T, DOM_v(\underline{v})) = F(T, DOM_v(v_1) \times DOM_v(v_2) \times \ldots \times DOM_v(v_k)).$$

If s_t is the "time slice" corresponding to $t \in T$ in situation $s \in SIT$, then

$$\forall f \in F(T, DOM_v(\underline{v})) \quad (Val(s, \underline{v}, f) \Leftrightarrow \forall t \in T\ Val(s_t, \underline{v}, f(t)))\,.$$

Then we can describe behaviors over time in the representational space

$$DOM(\underline{v}) = F(T, DOM_v(\underline{v}))\,,$$

and an abstraction will be a transformation of the function spaces. (Note that we assume the same temporal representation for each variable). This allows us to correctly handle some kinds of temporal abstraction, such as the "oscillation abstraction".

As a first step, we show that the "traditional" domain abstraction

$$\alpha: \quad DOM_v(\underline{v}) \to DOM'_v(\underline{v})$$

induces a behavioral abstraction.

We can map a function

$$f: T \to DOM_v(\underline{v})$$

onto its composition with α:

$$\tau_\alpha: DOM(\underline{v}) = F(T, DOM_v(\underline{v})) \to DOM'(\underline{v}) = F(T, DOM'_v(\underline{v}))\,,$$

where

$$\tau_\alpha(f) := \alpha \circ f: T \to DOM'_v(\underline{v})$$

and

$$\alpha \circ f(t) := \alpha(f(t))\,.$$

τ_α collapses functions that differ only in values that are mapped to the same image, into one function on the abstract domain. More generally:

Lemma 5.1

If $\alpha: DOM_v(\underline{v}) \rightarrow DOM'_v(\underline{v}')$
is an abstraction, then

$$\tau_\alpha: DOM(\underline{v}) = F(T, DOM_v(\underline{v})) \rightarrow DOM'(\underline{v}') = F(T, DOM'_v(\underline{v}'))$$

is an abstraction.

For instance, all functions $k*\sin t$, $k \in (0, 1]$, are mapped to the same function which has the value + over $(0, \pi)$, 0 at π, - over $(\pi, 2\pi)$ etc. (see first mapping in Fig. 5.4). So far, no property of time has been lost.

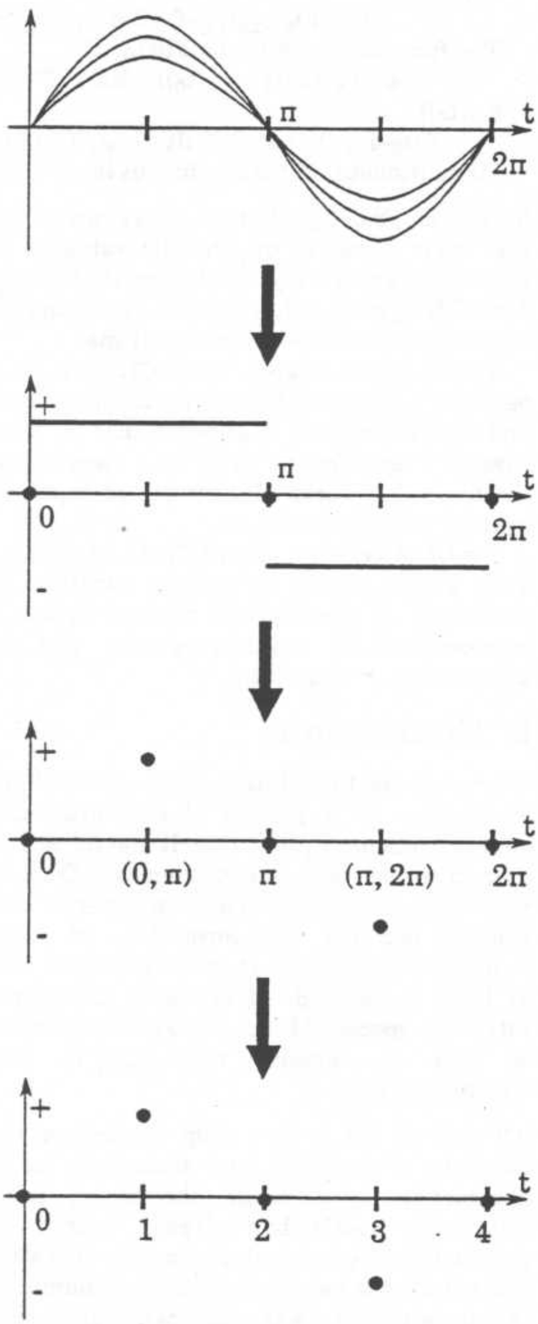

Figure 5.4 "The steps in constructing the oscillation abstraction"

But temporal abstraction may now be applied by collapsing maximal time intervals in which no changes in values occur into single time instances t' of a new temporal space, T'_f, which is still specific for each function f. Formally, we define

$$T'_f = \{ i \in I(T) \mid \exists \underline{v}'_0 \in DOM'_v(\underline{v}')\ f(i) = \{\underline{v}'_0\} \wedge (i' \supseteq i \Rightarrow f(i') \neq \{\underline{v}'_0\}) \},$$

where I(T) is the set of intervals of T. Then the function

$$f': T'_f \rightarrow DOM'_v(\underline{v}')$$

is well-defined by

$$f'(t') := f(t) \text{ for some } t \in t'.$$

This is the second mapping illustrated in Fig. 5.4. The spaces T'_f still contain the metric information of T. But under certain reasonable assumptions (that exclude pathological cases such as sin 1/t), there exists a subset $\mathbf{Z}'_f \subseteq \mathbf{Z}$ of the integers and a bijective mapping (the third one in Fig. 5.4)

$$\tau_Z: \mathbf{Z}'_f \rightarrow T'_f$$

for each f. Defining

$$\tau'(f) = f' \circ \tau_Z: \mathbf{Z}' \rightarrow DOM'_v(\underline{v}')$$

as a (partial) function on $\mathbf{Z}' = \cup \mathbf{Z}'_f \subseteq \mathbf{Z}$, we can obtain the desired representational transformation

$$\tau': F([-1, 1], \mathbb{R}_0^-) \rightarrow F(Q_3, \mathbb{N}_0)$$

that maintains the ordering, eliminates metric properties, and performs the "oscillation abstraction". It is also the basis for other temporal abstractions, e.g. used in XDE ([Hamscher 91]). For instance, the predicate that counts the number of changes, maps f to $|\mathbf{Z}'_f|$ -1.

5.3 Multiple Temporal Granularity and Continuity

In order to formalize the interaction of processes with different speed and temporal granularity, we propose to provide metrics that relate temporal universes of adjacent levels of temporal granularity. So, let $(T_i, d_i, <_i)$, $i \in \{s, f\}$, be the temporal representations for a slow and a fast process (one may keep in mind the example from [Iwasaki 92]: a fast oscillation of a block on a surface which is very slowly wearing out due to the sliding block). What we would like to express is that at some time point, $t_0 \in T_s$, the fast process is active (at least, "for a while") before the slow one exhibits any change. The basic idea for formalizing this in our framework is to include t_0 and subsequent time points of T_s in T_f and to let the structure of T_f impose the desired order:

$$\forall t_f \in T_f\ \forall t_s \in T_s\ (t_0 < t_s \wedge d_f(t_0, t_f) < \varepsilon) \Rightarrow t_f < t_s,$$

i.e. regradless of how close the "next" t_s is under d_s, at the fine-grained level for some duration ε, time steps are squeezed in between. Hence, what is influenced by the slow process appears to be "frozen" (constant, that is) for the fast one, and the effects of the fast process are received by the slow one as instantaneous changes (i.e. happening without a time tick in T_s) (Fig. 5.5).

If a behavioral description at the coarse level is constructed as a restriction of the fine-grained

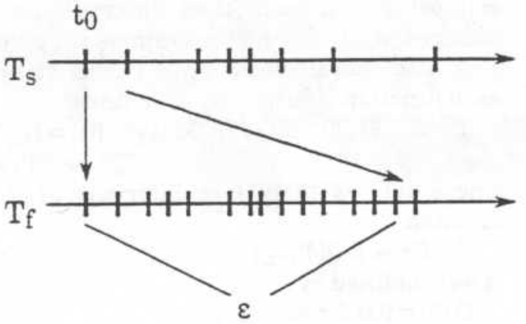

Figure 5.5 Relating T_s and T_f

description to $T_s \subset T_f$, then it is an abstraction according to the following straightforward lemma:

Lemma 5.2

Let $T' \subseteq T$ and

τ: $F(T, DOM_v(\underline{v})) \rightarrow F(T', DOM_v(\underline{v}))$

defined by

$\forall t' \in T'$ $\tau(f)(t') := f(t')$.

Then τ is an abstraction (provided $DOM_v(\underline{v})$ contains more than one element).

In our scenario, a model of the slow process omits a detailed description of the fast changes and represents a whole class of fast processes creating the same effects.

There are a number of remarks to be made at this point, indicating the need for further investigation.

First, we can avoid the metric and directly use the order to achieve an extreme version of the effect:

$\forall t_f \in T_f$ $\forall t_s \in T_s$ $(t_0 < t_f \wedge t_0 < t_s) \Rightarrow t_f < t_s$,

i.e. the fast process runs "to completion" (that means, in many interesting cases, to an equilibrium).

Second, we may be unable to specify ε, i.e. the duration of the fast process, directly. Rather, it may depend on the effects of this process or the slow one (e.g. crossing a certain threshold).

Often, this is related to a third issue: tolerance ranges. Up to a certain threshold, a slow change may be considered insignificant, but beyond this, it has to be taken into account. If we pretend we can interpolate at the coarse level between the time points of T_s (e.g. by treating a value constant), we obtain a simplification (or approximation) of a model, as opposed to an abstraction. It seems that this is what happens in [Iwasaki 92] with the sliding block. This can be mended if we explicitly represent the tolerances in the model (see [Struss 93]). Intuitively speaking, we replace a function $f(t) = \text{const}$ by a tolerance strip $f(t) \in (\text{const}-\varepsilon, \text{const}+\varepsilon)$.

Finally, changes over time with varying granularity raise the problem of continuity. In fact, this problem occurs with discrete domains, anyway, and many qualitative reasoning systems handle it by sets of transition rules (or in an algorithmic way) by allowing only changes to "adjacent values". Our fomalism enables us to give a general definition of continuous functions for arbitrary temporal universes and domains:

Definition 5.1 (Continuous Functions)

Let

$(T, d_T, <_T)$ and $(DOM_v(\underline{v}), d_{DOM}, <_{DOM})$

be metric spaces with a total order. Furthermore, for $t_0 \in T$, $f \in F(T, DOM_v(\underline{v}))$, and $rel \in \{<, >\}$, let

$\delta_{rel}(t_0) :=$

$\inf\{ d(t_0, t) \mid t \in T \wedge t \text{ rel } t_0 \}$,

$\varepsilon_{rel}(t_0) :=$

$\inf\{ d(f(t_0), \underline{v}_0) \mid \underline{v}_0 \in DOM_v(\underline{v}) \wedge \underline{v}_0 \text{ rel } f(t_0) \}$.

The function f is continuous in t_0, iff

$\forall (\varepsilon'_<, \varepsilon'_>) \supset (\varepsilon_<(t_0), \varepsilon_>(t_0))$ $\exists (\delta'_<, \delta'_>) \supset (\delta_<(t_0), \delta_>(t_0))$

$f((t_0-\delta'_<, t_0+\delta'_>)) \subseteq (f(t_0)-\varepsilon'_<, f(t_0)+\varepsilon'_>)$.

f is continuous, if it is continuous in all $t_0 \in T$.

Intuitively, the $\delta_{rel}(t_0)$ and $\varepsilon_{rel}(t_0)$ express "how close" the "next" (smaller or greater) value is in T and $DOM_v(\underline{v})$, respectively. Obviously, for the special case $T = DOM_v(\underline{v}) = \mathbb{R}$ with the usual metric and order, this definition coincides with the usual one:

$\forall \varepsilon > 0$ $\exists \delta > 0$ $f((t_0-\delta, t_0+\delta)) \subseteq (f(t_0)-\varepsilon, f(t_0)+\varepsilon)$,

because $\delta_{rel}(t_0) = \varepsilon_{rel}(t_0) = 0$ in $\mathbb{R}$. For discrete time and discrete values, it allows either no change or a change to an adjacent value in a step, as usual. But the definition works also for mixed representations, even for odd cases, such as

$T = DOM_v(\underline{v}) = (-\infty, -1) \cup [0, 0] \cup (1, \infty)$ with $d_{\mathbb{R}}$, $<_{\mathbb{R}}$.

This allows us to investigate whether and how continuity of functions is affected under different representational transformations and temporal abstractions, in particular.

6 Conclusions

We have tried to maintain some essential features of abstraction as a process of conceptualization and generalization, arguing that it should not introduce new theorems and inconsistencies. Our theory of relational descriptions of system behavior allows us to express common transformations of models and representations easily. It can also be used to formalize at least some kinds of temporal abstraction if we introduce spaces of functions over different universes of time as domains that can be subject to transformations.

All this is but a first step towards a systematic analysis of problems and techniques of temporal abstraction. Also, while the theory of multiple relational models has already helped to tackle practical problems involving models of **static** devices, this remains to be explored for the dynamic case. This is true especially when the behavior description of components is not a priori given as sets of functions over time, but in terms of (ordinary or qualitative) differential equations, as is the case in most qualitative physics systems.

Temporal abstraction is too important as a means for complexity reduction, and we cannot afford to neglect it or treat it too generally in the ongoing discussion about multiple modeling and automatic model generation.

Acknowledgements

Thanks to Oskar Dressler, Hartmut Freitag, Gerhard Friedrich, Roy Leitch, Wolfgang Nejdl, partners in the ARTIST project, and the reviewers for comments on this paper and related ones. This work was supported in part by BMFT (ITW 8506 E4, ITW 9001 A9), CEC (P5143, 3249), and CISE.

References

[Giunchiglia-Walsh 89]
Giunchiglia, F., and Walsh, T., *Abstract Theorem Proving*, IJCAI-89

[Giunchiglia-Walsh 92]
Giunchiglia, F., and Walsh, T., *A Theory of Abstraction*, In: Artificial Intelligence, 57(2-3), 1992

[Hamscher 91]
Hamscher, W., *Modeling Digital Circuits for Troubleshooting*, in: Artificial Intelligence, 51(1991)

[Hobbs 85]
Hobbs, J.R. , *Granularity*, IJCAI-85

[Iwasaki 92]
Iwasaki , Y., *Reasoning with Multiple Abstraction Models*. In: Faltings, B. and Struss, P. (eds.), Recent Advances in Qualitative Physics, 1992

[Kuipers 86]
Kuipers, B., *Qualitative Simulation* , Artificial Intelligence 29 (3), 1986

[Kuipers 87]
Kuipers, B., *Abstraction by Time-Scale in Qualitative Simulation*. AAAI-87

[Mozetic 92]
Mozetic, I., *Hierarchical Model-based Diagnosis*. In: Hamscher, W., de Kleer, J., and Console, L. (eds.), Readings in Model-based Diagnosis, San Mateo, 1992

[Struss 91]
Struss, P. *A Theory of Model Simplification and Abstraction*. In: Working Papers of the 5th International Workshop on Qualitative Reasoning about Physical Systems, QR-91, Austin, Tx, 1991

[Struss 92]
Struss, P. *What's in SD? Towards a Theory of Modeling for Diagnosis*. In: Hamscher, W., de Kleer, J., and Console, L. (eds.), Readings in Model-based Diagnosis, San Mateo, 1992.

[Struss 92a]
Struss, P. *A Theory of Abstraction and Simplification for Relational Models*. In: Workshop on Approximation and Abstraction, AAAI-92

[Struss 93]
Struss, P. *Multiple Models of Physical Systems - Modeling Intermittent Faults, Inaccuracy, and Tests in Diagnosis*. To appear in: Annals of Mathematics and Artificial Intelligence, 1993.

[Weld 86]
Weld, D.S., *The Use of Aggregation in Qualitative Simulation*. In: Artificial Intelligence 30(1), 1986

[Weld 92]
Weld, D.S., *Contribution to a Panel Discussion on Multiple Modeling*, QR-92, Edinburgh (Unpublished)

[Weld-Addanki 92]
Weld, D.S. and Addanki, S., *Task-Driven Model Abstraction*. In: Faltings, B. and Struss, P. (eds.), Recent Advances in Qualitative Physics, 1992