Modulated Modules: Designing Behaviors as Dynamical Systems

Herbert Jaeger
GMD, St. Augustin
herbert.jaeger@gmd.de
herbert@arti.vub.ac.be

June 1995

Abstract: A recent development in behavior-oriented robotics is to view behaviors in an agent as coupled dynamical subsystems. When one designs such subsystems a difficulty arises. The desired dynamics of the subsystem is likely to become qualitatively disrupted due to the superimposed dynamics of the behavior's interaction with other behaviors and the dynamics of sensor input. A mathematical technique is presented for making subsystems to some degree immune against such qualitative disruptions. By adding a suitable compensation term to the original ODE specification of the subsystem, the dynamics of its coupling variables lead to a benign *modulation* instead of a potential *disruption*. For the combined modulative effects of shift, amplitude variation, and velocity variation, the compensation term is given in an explicit, ready-to-use form. This is a step toward a modular design of behavior systems from interacting subsystems.

Zusammenfassung: Ein jüngerer Ansatz in der behavior-orientierten Robotik besteht darin, Behaviors in einem Agenten als gekoppelte dynamische Teilsysteme zu modellieren. Wenn man solche Teilsysteme im Zuge einer Roboterkonstruktion entwirft, entsteht die Schwierigkeit, daß die beabsichtigte Dynamik des Teilsystems durch dessen Interaktion mit anderen Teilsystemen und mit sensorischem Input qualitativ verändert wird. Die vorliegende Arbeit beschreibt, wie Behavior-Subsysteme bis zu einem gewissen Grad gegen solche qualitativen Veränderungen immun gemacht werden können. Durch die Hinzufügung eines geeigneten Kompensationsterms zu den ursprünglichen Systemgleichungen werden die Auswirkungen der Kopplungsdynamik auf *Modulationen* beschränkt, die den qualitativen Typ des Systemverhaltens erhalten. Für einige solcher Modulationen (Shift, Amplituden- und Geschwindigkeitsvariation) wird der Kompensationsterm in einer gebrauchsfertigen, expliziten Form angegeben. Dies stellt einen Schritt in Richtung eines modularen Entwurfs von Behavior-Systemen aus gekoppelten Teilsystemen dar.

1 Introduction

Behavior-oriented robotics, albeit successful in many respects, is still lacking a principled formal methodology. This seriously impedes progress in the field. Continuous dynamical systems theory has been proposed for filling the methodological gap in several target articles [6] [2]. The PDL language for programming robots, which has been developed at the VUB AI Lab and which is used in several European robotic labs, has been designed as a tool for realising behaviors as semi-continuous processes of real- valued quantities [7]. However, the target articles deal with the subject on a level that is too general for practical use, and current PDL programming practice typically clings to a traditional programming style that renders a system-theoretic analysis impossible. A practical yet principled methodology for designing behavior systems as dynamical systems is still missing.

Modeling agents as dynamical systems is quite a challenge (comprehensive discussion in [4]). One of the problems lies in a general property of dynamical systems: A system made up from nonlinearly coupled subsystems usually cannot be understood in terms of the behavior exhibited by the subsystems in isolation. This general phenomenon seriously obstructs a modular design of robots. If one tries to design some behavior systems (e.g., for forward movement or obstacle avoidance) first, and then to couple them together, one will typically find that they behave in a qualitatively different, unpremeditated fashion afterwards.

In spite of this general instability phenomenon in ad-hoc constructed dynamical systems, behaviors in animals often show a remarkable qualitative stability. For instance, gait patterns typically maintain their qualitative identity over a wide range of internal and external conditions. A walking pattern stays a walking pattern even though it might be sped up or slowed down, or changed in its step length, or be otherwise modulated quite quickly and strongly. Only when internal or external conditions drift beyond certain values, the walking pattern changes qualitatively, e.g. by turning into a running pattern.

This paper describes a simple formal mechanism for making dynamical subsystems to a certain degree "immune" against perturbations, as observed in natural behavior systems. Due to this mechanism, arbitrarily fast and reasonably strong variations in the coupling variable values do not result in a qualitative disruption of the subsystem's behavior; rather, they lead to a "benign", albeit possibly fast and strong *modulation* of its dynamics. I do not make any claims concerning the biological relevance of the mechanism. It is merely intended as a practically helpful contribution to a modular design of behavior-based robots.

The article is organized as follows. Section 2 provides the formal framework for modeling agents as continuous dynamical systems, and behaviors as subsystems therein. Section 3 presents an example of how a subsystem's dynamics gets disrupted through a seemingly harmless interaction dynamics. Section 4 specifies the "qualitative stabilization" mechanism in detail and generality. The basic idea is to add a compensation term to the original subsystem specification. This term measures the dynamics of the variables used for coupling the subsystem with the rest of the system, and generates a compensation for the potentially disruptive effects of these variables' dynamics. A general theorem is presented which describes how the compensated, coupled subsystem's dynamics can be understood in terms of the dynamics it shows when it is decoupled. In fact, this theorem boils down to an exercise in system transformations; it is not far from trivial. In section 5, a special case is treated which is relevant for many practical applications. A ready-to-use mathematical recipe is given for a non-disruptive modulation of dynamical systems with respect to shift, amplitude and velocity. Section 6 discusses some limitations of the mechanism and relates it to ongoing work in behavior-oriented robotics.

2 A framework for modeling behaviors as dynamical subsystems

We follow [6] and [2] in the basic ideas on how to model an agent as a continuous dynamical system. Namely, we describe it in terms of continuous *system variables* $x_1, \ldots, x_n$ and *coupling variables* $c_1, \ldots, c_m$.

If the model is used as a blueprint for design, the system variables must denote quantities describing the agent proper, like proprioceptive sensor readings, actuator control variables or motivational quantities. If the formal model is to be used as a tool for the analysis of an already existing agent, system variables can also refer to external descriptive dimensions, like position or direction of movement. We will be concerned with the first case only.

The coupling variables refer to quantities whose dynamics is determined from outside the agent, and whose function it is to dynamically couple the agent into its environment. The most important case of coupling variables are sensor readings. One might also wish to take into account others, like mechanical forces exerted on the agent body.

Given a suitable choice of system variables and coupling variables, the general scheme of a dynamical system model for an agent is shown in (1).

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \dots, x_n, c_1(t), \dots, c_m(t)) \\ &\dots \\ \dot{x}_n &= f_n(x_1, \dots, x_n, c_1(t), \dots, c_m(t)) \end{aligned} \tag{1}$$

In vector notation, (1) can be written more concisely as follows:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{c}(t)) \tag{2}$$

A concrete instantiation of this general scheme is given in (3):

$$\begin{aligned} \dot{x} &= y + 3c(t) \\ \dot{y} &= -x^2 - c(t) \end{aligned} \tag{3}$$

Some comments might be in place here for the benefit of readers not yet quite accustomed to dynamical systems:

- According to mathematical usage, the fact that *all* system variables and coupling variables appear on the right-hand side in each line in the general scheme (1) does not imply that all of them have to appear in every line of concrete instantiations of (1). (3) illustrates this fact.
- (3) is, in fact, an instantiation of (1): put $x_1 = x$, $x_2 = y$, $c_1(t) = c(t)$, $f_1(x_1, x_2, c_1(t)) = x_2 + 3c_1(t)$, and $f_2(x_1, x_2, c_1(t)) = -x_1^2 - c_1(t)$.
- The coupling variables are functions of time. They depend on time in a manner that is not specified within (1). Therefore, the dependency on time t is explicitly noted as $c_j(t)$.
- Of course, the system variables are also functions of time. However, in their case, the dependency on time is specified via (1). In order to arrive at explicit time functions $x_i(t)$, the system (1) of ordinary differential equations must be *solved.*

Coupling variables are formally similar to *input parameters* and *disturbance parameters*, as known from control theory, and to *control parameters*, as used in system theories in the natural sciences or in the mathematical theory of dynamical systems. All of them are free parameters in the rhs of system equations. However, the intuitive perspective on these parameters, and the appropriate mathematical methods for handling them, differ in all cases.

Input parameters, in a control theoretic setting, are used for an effective control of the system. By varying them dynamically, the system is steered towards a desired performance. The time scale on which the input parameters vary can be the same as the time scale of the system variable dynamics.

Disturbance parameters, again in a control theoretic perspective, are typically considered to reflect random "process noise". They are handled with stochastic methods. Non-random disturbances, which are also admitted into the framework of control theory, are assumed to be sufficiently small, such that they can be done with via local linearization (cf. [8], p. 283, 74). Furthermore, such disturbances do not appear in the form of free parameters; they are thus not relevant for the present discussion.

Finally, control parameters are typically required to vary on a time scale that is much slower than the system variable dynamics. In mathematical analysis, they are treated as constants. Bifurcation theory – one of the finest parts of modern mathematical system theories – is concerned with the comparison of system dynamics exhibited at different, but constant, values of control parameters.

Coupling variables are unlike all of these kinds of free parameters. First, they are not "used" to achieve a desired agent performance. They cannot be "controlled". Thus, the methods of control theory for dealing with input parameters are inapplicable. Second, they are not simply "noise". It would obviously be highly inappropriate to model sensor input or the influences of interacting behavior subsystems as noise. Finally, coupling variables typically vary on a time scale comparable to that of the system variables. Therefore, the analytic tools developed in bifurcation theory are inappropriate.

Unfortunately, no mathematical methods at all seem available to deal with coupling variables. This does not come as a surprise. One might argue as follows: "The system dynamics depends on the coupling variable dynamics, and the coupling variable dynamics is basically arbitrary – therefore, the system dynamics is basically arbitrary – yet non-random –, and thus cannot be the object of mathematical study". Control theory has taken one route out of this impasse, by considering input parameters that are not arbitrary but can be manipulated, and disturbance parameters that can be treated as random processes. Bifurcation theory has taken another available route by assuming that that control parameters are static, not dynamic.

The best what can be done for systems with arbitrary and dynamic coupling influences seems to be an analysis of extremely general properties. An example is provided by [3], where the occurence of chaotic states in such systems is defined.

The present paper indicates that it is in fact possible to arrive at a practically useful perspective on arbitrarily and dynamically coupled systems. The route that I wish to propose is to define and investigate systems that can cope with an arbitrary and fast coupling dynamics, in the sense of becoming modulated without getting qualitatively disrupted. This puts the present paper into perspective. In control theory and bifurcation theory, the restrictions necessary for enabling a mathematical investigation are applied on the side of the coupling variables' dynamics (controllable, or stochastic, or static, respectively). By contrast, I restrict the type of the very systems under investigation.

What I mean by "modulated", "qualitatively disrupted" etc. will become clear in the following sections. I touched these points here only in order to clarify the status of coupling variables.

The scheme (1) describes an entire agent. However, in this paper I focus on behaviors, which shall be modeled as subsystems of (1). A subsystem of (1) is specified by a subset of the system equations, say by the first k ones:

$$\begin{array}{rcl} \dot{x}_1 & = & f_1(x_1, \ldots, x_n, c_1(t), \ldots, c_m(t)) \\ & \cdots & \\ \dot{x}_k & = & f_k(x_1, \ldots, x_n, c_1(t), \ldots, c_m(t)) \end{array} \tag{4}$$

where $k \leq n$. Taken as a system in its own right, the subsystem (4) of (1) has system variables $x_1, \ldots, x_k$. What are its coupling variables? There are two kinds of them now. First, the subsystem (4) inherits the coupling variables $c_1(t), \ldots, c_m(t)$ of the supersystem (1). Second, the other supersystem variables $x_{k+1}, \ldots, x_n$, which appear in the rhs of (4), assume the role of additional coupling variables. They couple the subsystem (4) into the supersystem (1). From the perspective of (4), the dynamics of $x_{k+1}, \ldots, x_n$ is arbitrary and likely to lie on the same time scale as that of the subsystem variables $x_1, \dots, x_k$. Thus, from the subsystem's perspective, they behave essentially in the same way as the original coupling variables $c_1(t), \dots, c_m(t)$. Therefore, we are justified in rewriting (4) as follows:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \dots, x_k, x_{k+1}(t), \dots, x_n(t), c_1(t), \dots, c_m(t)) \\ &\cdots \\ \dot{x}_k &= f_k(x_1, \dots, x_k, x_{k+1}(t), \dots, x_n(t), c_1(t), \dots, c_m(t)) \end{aligned} \tag{5}$$

Up to a renaming of variables, (5) turns out to be equivalent to (1). For the purpose of mathematical investigation, we may therefore use the notationally simpler form (1) as a dynamical system scheme for a behavior, interpreting $c_1(t), \dots, c_m(t)$ as variables that couple the behavior subsystem into the agent supersystem, or as variables that couple it into the environment. Thus, in the sequel we will be concerned with dynamical systems like (1).

3 Qualitative disruption and qualitative stability: an example

In this section, the phenomenon of qualitative disruption, and a method for preventing it, will be illustrated with an example.

Consider the following two-dimensional system:

$$\begin{aligned} \dot{x}_1 &= -(x_2 - c_2(t)) \\ \dot{x}_2 &= x_1 - c_1(t) \end{aligned} \tag{6}$$

This is a non-autonomous system which instantiaties (1). (The terms *autonomous system* and *non-autonomous system* are used in this article in the sense of dynamical systems theory, i.e. an autonomous system is a *time-invariant*, or *stationary*, dynamical system. This has nothing to do with the notion of autonomous systems in the sense of Artificial Life). Let us fix the coupling variables $c_1(t), c_2(t)$ at some values that remain constant over time, i.e. let us consider the autonomous system

$$\begin{aligned} \dot{x}_1 &= -(x_2 - c_2) \\ \dot{x}_2 &= x_1 - c_1 \end{aligned} \tag{7}$$

with constants c_1, c_2. It can be solved analytically. As solution curves for (7), we get circles that are concentric around the point (c_1, c_2), which is the only fixed point (see fig. 1(a)). The system displays a pure rotation around this point with angular velocity 1, i.e. one full revolution takes 2π time units. From fig. 1(a) one sees that c_1 and c_2 can be interpreted as shift variables.

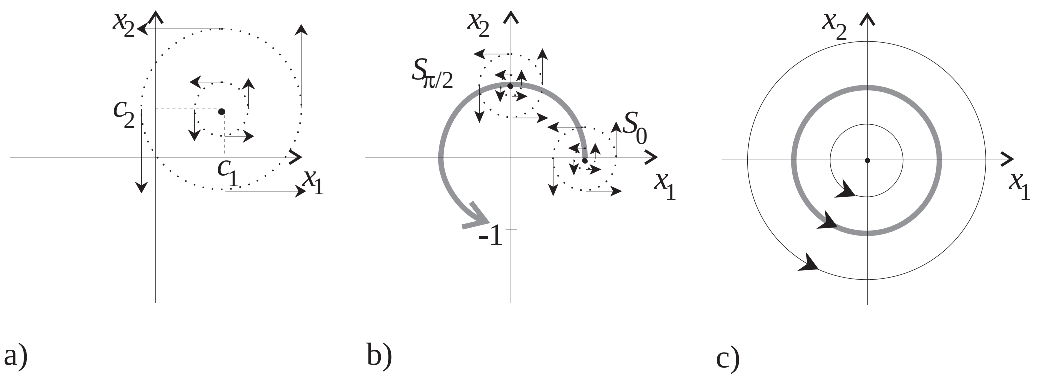

Figure 1: An example of qualitative disruption, part 1. For details compare text.

Now we let the shift variables c_1 and c_2 vary in time in a particular manner, namely, we let the point (c_1, c_2) perform a circular motion around the origin with radius 1 and angular velocity 1, starting in $(1, 0)$ (shaded arrow in fig. 1(b)). This shift motion can be expressed by

$$\begin{aligned} c_1(t) &= \cos t \\ c_2(t) &= \sin t, \end{aligned} \tag{8}$$

which yields a non-autonomous system

$$\begin{aligned} \dot{x}_1 &= -(x_2 - \sin t) \\ \dot{x}_2 &= x_1 - \cos t. \end{aligned} \tag{9}$$

It would be nice if we could understand the dynamics of (9) in terms of some kind of superposition of the dynamics of stationary, shifted systems (7) and the shift dynamics (8). Naively, we might reason as follows: “the pure rotational system is periodically shifted according to the shift operation, since the shift dynamics (8) means a periodic, cyclic displacement of the plane". In particular, we would expect the fixed point of the rotational system, which in $t = 0$ lies at $(c_1, c_2) = (1, 0)$, to follow in time the shift movement, i.e., to follow the shaded trajectory in fig. 1(b). Continuing this line of argumentation, we would end up with expecting the system (9) to exhibit trajectories like in fig. 1(c).

In fact, something quite different happens. (10) provides the true solutions of (9). One of them (with integration constants $C_1 = C_2 = 0$) is shown in fig. 2.

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = C_1 \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} + C_2 \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} + \begin{pmatrix} t \sin t \\ -t \cos t \end{pmatrix} \tag{10}$$

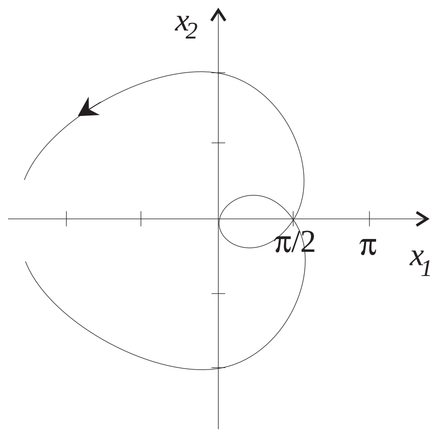

Figure 2: An example for qualitative disruption, part 2. For details compare text.

The trajectory shown in fig. 2 spirals to infinity with $t \to \pm\infty$. An examination of (10) reveals that *all* trajectories of (9) have this property. By contrast, trajectories in the autonomous systems (7) are cyclic and hence stay within a bounded region with $t \to \pm\infty$. Furthermore, the shift dynamics (8) periodically shifts a given point in the plane on a circle with radius 1, i.e., leads to a displacement that again stays in a bounded region with $t \to \pm\infty$. Thus, to our mild surprise, we find that the combined effects of two spatially bounded movements result in an unbounded movement. The autonomous system (7) has been *qualitatively disrupted* by shifting it periodically.

Although it is intuitively clear that the system (9) behaves qualitatively different from autonomous systems (7), I am not able to give a general definition of qualitative disruption. In the next section, however, a precise definition for a particular kind of *qualitative stability*, a complementary phenomenon, will be provided. Qualitatively stable, non-autonomous systems in the sense of that definition cannot, for instance, have unbounded trajectories where the associated autonomous systems dynamics and the coupling dynamics are both bounded.

What has gone wrong with our argumentation that led us to expecting the dynamics of fig. 1(c)? Intuitively, we interpreted the shift dynamics as a motion that is superimposed on the dynamics of the autonomous systems — as fig. 1(b) suggests. But in fact, the shift "movement" does not turn up at all where it is required, namely, in (9)! Intuitively, in fig1. (b) we see the point $(1,0)$ "move" on the unit circle according to the shift operation. This circular movement must be *added* to the rhs of (9) in order to make the system behave according to our intuitions. The shift dynamics (8) has the derivative

$$\begin{aligned} \dot{c}_1(t) &= -\sin t \\ \dot{c}_2(t) &= \cos t \end{aligned} \tag{11}$$

If we add this *compensation term* to the non-autonomous system (9), we get

$$\begin{aligned} \dot{x}_1 &= -(x_2 - \sin t) + -\sin t &= -x_2 \\ \dot{x}_2 &= x_1 - \cos t + \cos t &= x_1, \end{aligned} \tag{12}$$

which now yields trajectories as in fig. 1(c).

4 Compensating for coupling dynamics

In this section, the observations from the previous example are rigorously worked out. It is shown how and when a dynamical system with coupling variables can be "compensated" for the effects of the coupling dynamics, such that the resulting dynamics can be understood as a transparent superposition of the coupling dynamics and of the dynamics of a particular, "representative" autonomous system.

As a preparation, we briefly leave our topic of non-autonomous dynamical systems with time-dependent coupling variables and take a closer look at autonomous systems with constant control parameters. We start by recapitulating the well-known notion of *qualitative equivalence* (or *topological equivalence*; for a more detailed treatment cf. [1]).

Definition 1. Two autonomous dynamical systems that have the same phase space $\mathcal{D}$ are *qualitatively equivalent* if there exists a continuous bijection $\varphi : \mathcal{D} \to \mathcal{D}$ which map the phase portrait of one onto that of another in such a way as to preserve the orientation of the trajectories.

Qualitative equivalence is an equivalence relation on dynamical systems. Therefore, all the members in a family of dynamical systems are pairwise qualitatively equivalent if they are all qualitatively equivalent to any one particular, *representative* member in the family. It should be noted that the bijection φ required in definition 1 is in general not uniquely determined.

A typical phenomenon in dynamical systems with control parameters, i.e. systems of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{c}) \tag{13}$$

is that if the control parameters are selected from within a certain range $\mathcal{C}$, the resulting systems are qualitatively equivalent (a phenomenon called *structural stability*). Only if control parameters surpass certain critical values, the system undergoes a qualitative change called a *bifurcation*. For the remainder of the article, we assume that $\mathcal{C}$ is a range of control parameter values that yields qualitatively equivalent systems for a scheme like (13).

We collect some further notations and standard assumptions. Let the system (13) be denoted by the shorthand notation $S_{\mathbf{c}}$, and let $(S_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}}$ denote the family of qualitatively equivalent systems that we get by allowing $\mathbf{c}$ to be taken from $\mathcal{C}$. Let S_{rep} be a representant of that family. By convention, the control parameters of S_{rep} are denoted by $\mathbf{c}_0$, and the system variables of S_{rep} by $\tilde{\mathbf{x}}$. Let $\varphi_{rep\to\mathbf{c}} =: \varphi_{\mathbf{c}} : \mathcal{D} \to \mathcal{D}$ be a bijection that maps S_{rep} on

$S_{\mathbf{c}}$ as required in definition 1. Additionally we require that $\mathcal{C}$ is an open set and that the parametrised family of mappings $(\varphi_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}}$ is continuously differentiable. More precisely, the latter means that the function

$$\begin{aligned} \lambda\mathbf{c}\lambda\mathbf{x}\varphi_{\mathbf{c}}(\mathbf{x}) \;\; : \;\; & \mathcal{C}\times\mathcal{D}\rightarrow\mathcal{D} \\ & (\mathbf{c},\mathbf{x})\mapsto\varphi_{\mathbf{c}}(\mathbf{x}) \end{aligned}$$

is continuously differentiable. (The lambda notation for functions is not common in differential geometry, but should be familiar to readers with an AI background. I use it in this article because it clearly shows which parameters in a functional expression are considered as function arguments, and which others as constants - this will be particularly handy further below in (17), where time occurs both as an argument and as a constant within one functional expression). We sum up what we have got so far in a definition:

Definition 2. A triple $((S_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}}, S_{rep}, (\varphi_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}})$ that satisfies the conditions in the above remarks is a *$\mathcal{C}$-family.*

As an example, we reconsider the scheme of systems (7). We have two control parameters c_1, c_2, which means that $\mathcal{C}\subseteq\Re\times\Re$, where $\Re$ denotes the reals. The domain of the systems is the plane, i.e., $\mathcal{D}=\Re\times\Re$. The shorthand notation for the system (7) is $S_{(c_1,c_2)}$. In fact, we have $\mathcal{C}=\Re\times\Re$, since *all* members of the family $(S_{(c_1,c_2)})_{(c_1,c_2)\in\Re\times\Re}$ are qualitatively equivalent. Taking the unshifted system $S_{(0,0)}$ as our representant S_{rep}, the bijections $\varphi_{(c_1,c_2)}$ can be taken (among other possibilities) as the shifts of the plane, i.e.

$$\varphi_{(c_1,c_2)}(x_1,x_2) \;\; = \;\; (x_1+c_1, x_2+c_2). \tag{14}$$

After these preparations, we return to our main concern, non-autonomous systems with time-dependent coupling variables. We consider a $\mathcal{C}$-family $((S_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}}, S_{rep}, (\varphi_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}})$. Now, we allow the variables $\mathbf{c}$ to vary in time. We require this variation to be differentiable and to stay within $\mathcal{C}$, but otherwise we impose no restrictions on it. I.e., we consider a differentiable time function, which we denote with the same symbol $\mathbf{c}$ that we used for the control parameters,

$$\begin{aligned} \mathbf{c} \ : \ & \Re \to \mathcal{C} \\ & t \mapsto \mathbf{c}(t). \end{aligned}$$

Interpreting $\mathbf{c}(t)$ as coupling variables gives us again (2), which is repeated here for convenience:

$$\dot{\mathbf{x}} \quad = \quad \mathbf{f}(\mathbf{x}, \mathbf{c}(t)) \tag{15}$$

According to our assumptions, for every $t \in \Re$ it holds that $\mathbf{c}(t) \in \mathcal{C}$. Thus, if we take snapshots of (15) at various times t_i, the resulting autonomous systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{c}(t_i))$ are all qualitatively equivalent. We have learnt from our example in the previous section that the non-autonoumous system (15) need not display the same qualitative type of dynamics. However, as we will presently show, adding a suitable compensation term to (15) will essentially achieve this goal. More precisely, this will give us a method to project trajectories in the non-autonomous system on trajectories in the representant system S_{rep}.

In order to understand this compensation term, let us take a look at the schematic diagram of an exemplary one-dimensional system in fig. 3. In the upper half, the figure shows a trajectory $\tilde{x}(t)$ in the representant system S_{rep}. The vector field $\dot{\tilde{x}}$ is drawn into the upper diagram repeatedly at time intervals of length 1 (of course, they exist at every point on the time line, since we have continuous time).

The lower half of the diagram shows the non-autonomous system as we would like it to get with the help of the compensation term. The vector fields at times t_i belong to the autonomous systems $S_{\mathbf{c}(t_i)}$ (cf. the shaded slice in the figure). The mappings $\varphi_{\mathbf{c}(t)} : \mathcal{D} \to \mathcal{D}$ transform the phase space at time t according to definition 1, which in the simple case of fig. 3 boils down to the requirement that the unique fixed point in S_{rep} is mapped on the unique fixed point of $S_{\mathbf{c}(t)}$.

Intuitively, what we want to achieve is that trajectories $\tilde{x}(t)$ in the representant system are mapped on trajectories $x(t)$ in the non-autonomous system via $\varphi_{\mathbf{c}(t)}$. I.e., we would like the following to be true:

$$x(t) \quad = \quad \varphi_{\mathbf{c}(t)}(\tilde{x}(t)) \tag{16}$$

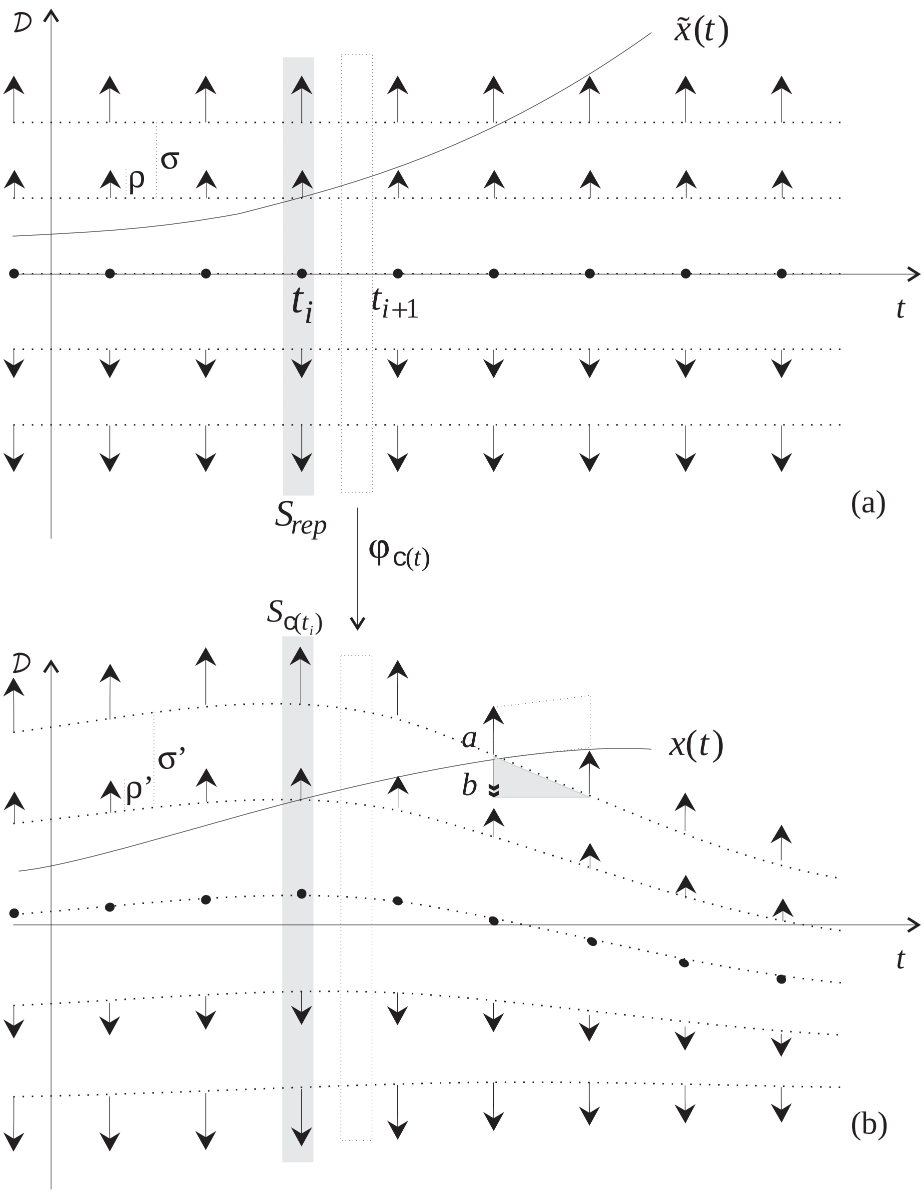

Figure 3: How the compensation term works. For details compare text.

It is clear what we have to do in order to arrive at this result: we have to give the trajectories in the non-autonomous system an additional bend that makes them follow the temporal bend of the phase space transformations $\varphi_{\mathbf{c}(t)}(\mathcal{D})$. This is achieved by additively superimposing on the vectors a of $S_{\mathbf{c}(t)}$ a vector b that makes up for the phase space bending. The vector b is our compensation term. It depends both on time and on the point $x \in \mathcal{D}$ in the non-autonomous system. Formally, it is the temporal derivative of $\varphi_{\mathbf{c}(t)}$ at the place $\varphi^{-1}_{\mathbf{c}(t)}(x)$:

$$b(x,t) \;\; := \;\; \left(\frac{d}{dt} \left(\lambda\tau \, \varphi_{\mathbf{c}(\tau)} \left(\varphi^{-1}_{\mathbf{c}(t)}(x) \right) \right) \right)(t) \tag{17}$$

By adding this compensation term to the rhs of (a one-dimensional version of) (15), we arrive at our desired compensated system, for whose trajectories (16) holds:

$$\dot{x} \;\; = \;\; f(x, c(t)) + b(x,t) \tag{18}$$

By convention, the compensated system (18) is denoted by $S_{\mathbf{c}(t)}$. Note that if the time function $\mathbf{c}(t)$ is a constant, i.e. if $\mathbf{c}(t) \equiv \mathbf{c}$, then $b(x,t)$ vanishes and we get $S_{\mathbf{c}(t)} = S_{\mathbf{c}}$.

We have tacitly assumed in these considerations that the bijections $\varphi_{\mathbf{c}(t)}$ are *time-preserving*, i.e., that they do map trajectories on trajectories not only in an orientation-preserving, but also in a time-preserving way. Intuitively, this means that if the phase space is stretched through $\varphi_{\mathbf{c}(t)}$ at some point $\tilde{x} \in \mathcal{D}$ by some factor s, then the derivative $\dot{\tilde{x}}$ must be augmented proportionally. In fig. 3, this means that at all corresponding points $\tilde{x}$ and $\varphi_{\mathbf{c}(t)}(\tilde{x})$ the ratio σ'/σ must be equal to the ratio ρ'/ρ. If we express these ratios in terms of suitable derivatives, we find that $\varphi_{\mathbf{c}(t)}$ is time-preserving if the following holds for all $\tilde{x}$:

$$\frac{\partial}{\partial \tilde{x}} \varphi_{\mathbf{c}(t)}(\tilde{x}) \;\; = \;\; \frac{f(\varphi_{\mathbf{c}(t)}(\tilde{x}), \mathbf{c}(t))}{f(\tilde{x}, \mathbf{c}_0)} \tag{19}$$

This is a very special case, and we shall get rid of this restriction later. However, for the time being we stick with it, since the present concern is understanding the basic nature of the compensation term, which can be seen most clearly in the time-preserving case.

We collect our argumentation in a theorem, assuming the most simple case that both $\mathcal{D}$ and $\mathcal{C}$ are one-dimensional.

Theorem 1. Let $((S_c)_{c\in\mathcal{C}}, S_{rep}, (\varphi_c)_{c\in\mathcal{C}})$ be a $\mathcal{C}$-family, $c : \Re \to \mathcal{C}$ a differentiable time function, $\dot{x} = f(x, c(t))$ a one-dimensional system analogous to (15), $b(x,t)$ a compensation term as specified in (17). Let the bijections $\varphi_c(t)$ be time-preserving as specified in (19). Then, the solutions $x(t)$ of the compensated system

$$\dot{x} = f(x, c(t)) + b(x,t) \tag{20}$$

are related to the solutions $\tilde{x}(t)$ according to

$$x(t) = \varphi_{c(t)}(\tilde{x}(t)). \tag{21}$$

Proof. All we have to do is to check that (21) in fact solves (20). To this end, we first observe that $\varphi_{c(t)}(\tilde{x}(t))$ can be formally rewritten as a function $\varphi(c(t), \tilde{x}(t))$ with two arguments $c(t)$ and $\tilde{x}(t)$, which allows us to compute its temporal derivative as follows:

$$\frac{d}{dt}\varphi_{c(t)}(\tilde{x}(t)) = \frac{\partial\varphi}{\partial c}\frac{dc}{dt} + \frac{\partial\varphi}{\partial\tilde{x}}\frac{d\tilde{x}}{dt} \tag{22}$$

On the other hand, if we insert the rhs of (21) into the rhs of (20), we get

$$\begin{aligned}
& f(\varphi_{c(t)}(\tilde{x}(t)), c(t)) + b(\varphi_{c(t)}(\tilde{x}(t)), t) = \\
[(17):] \quad =& \; f(\varphi_{c(t)}(\tilde{x}(t)), c(t)) + \left(\frac{d}{dt}\lambda\tau\,\varphi_{c(\tau)}\left(\varphi^{-1}_{c(t)}(\varphi_{c(t)}(\tilde{x}(t)))\right)\right)(t) \\
=& \; f(\varphi_{c(t)}(\tilde{x}(t)), c(t)) + \left(\frac{d}{dt}(\lambda\tau\,\varphi_{c(\tau)}(\tilde{x}(t))\right)(t) \\
=& \; f(\varphi_{c(t)}(\tilde{x}(t)), c(t)) + \frac{\partial\varphi}{\partial c}\frac{dc}{dt} \\
=& \; \frac{f(\varphi_{c(t)}(\tilde{x}(t)), c(t))}{f(\tilde{x}(t), c_0}f(\tilde{x}(t), c_0) + \frac{\partial\varphi}{\partial c}\frac{dc}{dt} \\
[(19):] \quad =& \; \frac{\partial\varphi}{\partial\tilde{x}}\frac{d\tilde{x}}{dt} + \frac{\partial\varphi}{\partial c}\frac{dc}{dt}
\end{aligned} \tag{23}$$

This is equal to the rhs of (22), which completes the proof. $\square$

Theorem 1 treats the simplest of all cases. However, it fully captures the basic idea, and more general cases essentially only add technicalities. The theorem (and its generalizations) also formalizes what I mean, in a special sense, by qualitative stability of non-autonomous systems: namely, that the system's vector fields at different times are all qualitatively equivalent (which holds for the uncompensated system, too), *plus* that trajectories in the non-autonomous system can be derived from trajectories in an autonomous representant system by applying the same transformations that lead from the representant system to the non-autonomous systems.

I shall now give, without spelling out the proof, the theorem in its full glory, i.e. for n-dimensional phase space $\mathcal{D}$, $m-$dimensional coupling variable space $\mathcal{C}$, and for non-time-preserving $\varphi_{\mathbf{c}}$.

Theorem 2. Let Let $((S_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}}, S_{rep}, (\varphi_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}})$ be a $\mathcal{C}$-family, $\mathbf{c}: \Re \to \mathcal{C}$ a differentiable time function, and $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{c}(t))$ a system like in (15). Let

$$\mathbf{b}(\mathbf{x}, t) \;\; := \;\; \mathrm{D}\left(\lambda\tau\, \varphi_{\mathbf{c}(\tau)}\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(t) \tag{24}$$

be the compensation term analogous to (17), and

$$\dot{\mathbf{x}} \;\; = \;\; \mathbf{f}(\mathbf{x}, \mathbf{c}(t)) + \mathbf{b}(\mathbf{x}, t) \tag{25}$$

the compensated system $S_{\mathbf{c}(t)}$. Let $\mathbf{x}(t)$ be a solution of $S_{\mathbf{c}(t)}$ which passes through $\mathbf{x}_0$ in t_0. Let $\tilde{\mathbf{x}}(t)$ be the solution of S_{rep} which passes through $\varphi^{-1}_{\mathbf{c}(t_0)}(\mathbf{x}_0)$ in t_0. Then, a monotonously increasing *time rescaling function* $T: \Re \to \Re$ exists, such that

$$\mathbf{x}(t) \;\; = \;\; \varphi_{\mathbf{c}(t)}\, \tilde{\mathbf{x}}(T(t)). \tag{26}$$

The time rescaling function satisfies the following condition:

$$T(t) = t_0 + \int_{t_0}^{t} \frac{\mathrm{D}\left(\lambda\xi\, \varphi^{-1}_{\mathbf{c}(\tau)}(\xi)\right)(\mathbf{x}(\tau)) \cdot f(\mathbf{x}(\tau), \mathbf{c}(\tau))}{f\left(\varphi^{-1}_{\mathbf{c}(\tau)}\mathbf{x}(\tau), \mathbf{c}_0\right)}\, d\tau \tag{27}$$

.

Alternatively, with a different time rescaling function T', time can be rescaled in the argument of $\mathbf{x}$, and we get the following complementary version of (26) and (27):

$$\mathbf{x}(T'(t)) \quad = \quad \varphi_{\mathbf{c}(t)}\,\tilde{\mathbf{x}}(t). \tag{28}$$

$$T'(t) = t_0 + \int_{t_0}^{t} \frac{\mathrm{D}\left(\lambda\xi\,\varphi_{\mathbf{c}(\tau)}(\xi)\right)(\tilde{\mathbf{x}}(\tau)) \cdot f(\tilde{\mathbf{x}}(\tau), \mathbf{c}_0)}{f\left(\varphi_{\mathbf{c}(\tau)}\,\tilde{\mathbf{x}}(\tau), \mathbf{c}(t)\right)}\,d\tau \tag{29}$$

.

Like its simplicistic companion, theorem 1, this theorem states that trajectories in the compensated system essentially are of the same qualitative type as trajectories in the representant system. The former can be derived from the latter by applying the same transformation $\varphi_{\mathbf{c}(t)}$ that also, loosely speaking, transforms the representant system into the non-autonomous system.

Things are complicated by the fact that the transformations $\varphi_{\mathbf{c}(t)}$ need not be time-preserving. This makes a time rescaling necessary. The formal derivation of this rescaling can be done in the same vein as the arguments that led to (19). The fact that $\mathbf{x}(t)$ occurs in the rhs of (27) is owed to the circumstance that the transformations $\varphi_{\mathbf{c}(t)}$ can induce different speedups or slowdowns at different places in $\mathcal{D}$. If the temporal speedup or slowdown induced by these transformations is homogenous over $\mathcal{D}$, (27) is much simplified. In particular, it becomes independent from $\mathbf{x}(t)$ on its rhs, which renders (27) useful for an effective computation of $\mathbf{x}(t)$. We will exploit this in the next section. If, finally, $\varphi_{\mathbf{c}}$ is time-preserving, we have $T = T' = \mathrm{id}$, which is the situation of theorem 1.

The fact that we have to deal with an n-dimensional phase space $\mathcal{D}$ and an $m-$dimensional coupling variable space $\mathcal{C}$ is reflected in the occurence of Jacobians D in (24), (27), and (29). Applying the chain rule, and spelled out in matrix form, the compensation term reads like follows:

$$\mathrm{D}\left(\lambda\tau\,\varphi_{\mathbf{c}(\tau)}\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(t) =$$

$$
\begin{aligned}
&= \mathrm{D}\left(\lambda\chi\,\varphi_\chi\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(\mathbf{c}(t))\cdot \mathrm{D}\mathbf{c}(t) \\
&= \begin{pmatrix} \left(\frac{\partial}{\partial c_1}\lambda\chi\,\varphi_{1\chi}\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(\mathbf{c}(t)) & \cdots & \left(\frac{\partial}{\partial c_m}\lambda\chi\,\varphi_{1\chi}\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(\mathbf{c}(t)) \\ \cdots & \cdots & \cdots \\ \left(\frac{\partial}{\partial c_1}\lambda\chi\,\varphi_{n\chi}\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(\mathbf{c}(t)) & \cdots & \left(\frac{\partial}{\partial c_m}\lambda\chi\,\varphi_{n\chi}\left(\varphi^{-1}_{\mathbf{c}(t)}(\mathbf{x})\right)\right)(\mathbf{c}(t)) \end{pmatrix} \cdot \\
&\quad \cdot \begin{pmatrix} \left(\frac{\partial}{\partial \tau}\lambda\tau\, c_1(\tau)\right)(t) \\ \cdots \\ \left(\frac{\partial}{\partial \tau}\lambda\tau\, c_m(\tau)\right)(t) \end{pmatrix}
\end{aligned} \tag{30}
$$

The functions $\varphi_{i\chi}$ that appear in (30) are the component functions of φ_χ.

Let us briefly see how theorem 2 applies to the example from section 2. We take the unshifted system $S_{(0,0)}$ as representant S_{rep}, and the bijections $\varphi_{(c_1,c_2)}$ as in (14). They are obviously time-preserving, and therefore, $T = \mathrm{id}$. Evaluation of (26) leads to the following result:

$$
\begin{aligned}
\mathbf{x}(t) &= \varphi_{(c_1,c_2)(t)}\,\tilde{\mathbf{x}}(t) \\
&= (\tilde{x}_1(t) + c_1(t), \tilde{x}_2(t) + c_2(t)) \\
&= (\tilde{x}_1(t) + \cos t, \tilde{x}_2(t) + \sin t)
\end{aligned}
$$

In particular, we see from this equation that the fixed point solution in the origin of S_{rep} is mapped on the unit circle solution of the non-autonomous system, just as we intuitively wished it to happen in fig. 1(c).

A final remark. The compensation term can stabilize the qualitative type of a behavior subsystem only as long as the variations of the coupling variables stay within $\mathcal{C}$. If these limits are transgressed, the compensated system can, of course, be driven into bifurcations (which remain to be formally defined for the kind of non-autonomous systems considered here!).

5 Modulating velocity, amplitude, and shift

In this section, we treat a special case of theorem 2, where the compensation term and the time rescaling function are given in an explicit form that can be directly used for numerical algorithms. This special case concerns the effects of modulating a dynamical system with respect to the combined effects of shift, amplitude, and velocity. It is shown how a behavior subsystem within an robot system can be dynamically modulated with respect to these three modes of variation by varying coupling variables at arbitrary time scales, without qualitatively disrupting the subsystem. These are three important characteristics of a behavior, and typically all of them vary dynamically in animal behaviors. Thus, the results reported in this section have considerable practical value.

We start from an n-dimensional, autonomous system:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \ldots, x_n) \\ &\cdots \\ \dot{x}_n &= f_n(x_1, \ldots, x_n) \end{aligned} \tag{31}$$

We call this system S_{rep}; it will soon become clear that this system assumes the role of a representant system in the sense of the previous section. We wish to transform the vector field specified by (31) in three steps. They are illustrated in fig. 4 with two exemplary vectors:

- First, the original system S_{rep} (fig. 4(a)) is to be shifted by amounts $s_i \in \Re$ in the dimensions x_i, where $i = 1, \ldots, n$ (fig. 4(b)).
- Then, it is to be expanded homogenously by a factor of $a > 0$, which yields an expansion of the trajectorie's amplitudes. In fig. 4(c), this factor takes a value of 2.
- Finally, its vectors are multiplied by a scalar $v > 0$, which amounts to changing the system's velocity by a factor of v. In fig. 4(c), this factor takes a value of 1/3.

These transformations, executed in this order which leads to the total transformation seen in fig.4(d), are realized in (31) as follows:

$$\begin{aligned} \dot{x}_1 &= vaf_1(x_1/a - s_1, \ldots, x_n/a - s_n) \\ &\cdots \\ \dot{x}_n &= vaf_n(x_1/a - s_1, \ldots, x_n/a - s_n) \end{aligned} \tag{32}$$

In our terminology, the system (32) is the system $S_{(v,a,s_1,\ldots,s_n)}$, or $S_{\mathbf{c}}$ for short, where $\mathbf{c} = (v, a, s_1, \ldots, s_n)$. We find that the representant system S_{rep}

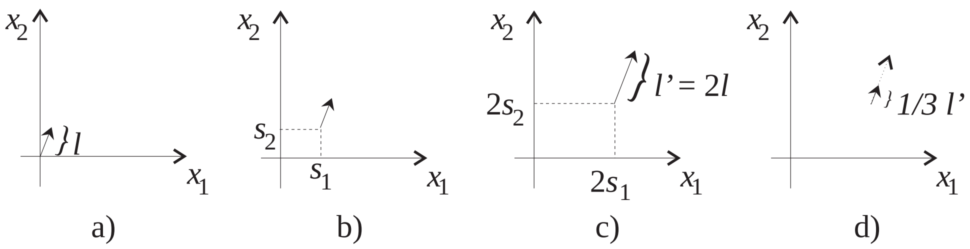

Figure 4: The effects of shift, amplitude variation, and velocity variation on the vector field of S_{rep}

from (31) is the same as $S_{(1,1,0,\ldots,0)}$, which can also be expressed by stating that $\mathbf{c}_0 = (1, 1, 0, \ldots, 0)$.

An obvious way to define $\varphi_{(v,a,s_1,\ldots,s_n)} =: \varphi_{\mathbf{c}} : \Re^n \to \Re^n$ is to put

$$\varphi_{\mathbf{c}}(x_1, \ldots, x_n) = (a(x_1 + s_1), \ldots, a(x_n + s_n)) \tag{33}$$

Finally, if we take $\mathcal{C} = \Re^+ \times \Re^+ \times \Re \times \cdots \times \Re$, it is easy to see that $((S_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}}, S_{rep}, (\varphi_{\mathbf{c}})_{\mathbf{c}\in\mathcal{C}})$ makes a $\mathcal{C}$-family according to definition 2.

Definition 3. A $\mathcal{C}$-family of the kind just described is called a *vas-family.*

The compensation term for non-autonomous systems formed from a *vas*-family can be stated in explicit form. Let $\mathbf{c}(t) = (v(t), a(t), s_1(t), \ldots, s_n(t))$ be a time function. Then, a straightforward calculation shows that the compensation term (30) takes the following form:

$$\begin{pmatrix} \dot{a}(t)\frac{x_1}{a(t)} + a(t)\dot{s}_1(t) \\ \cdots \\ \dot{a}(t)\frac{x_n}{a(t)} + a(t)\dot{s}_n(t) \end{pmatrix} \tag{34}$$

The time rescaling function (27) depends only on the temporal development of v. We get the following specialization of (27):

$$T(t) = t_0 + \int_{t_0}^{t} v(\tau)d\tau \tag{35}$$

The compensation term (34) contains the temporal derivatives of a and s_i. They must be effectively computed when a compensated *vas*-system is used as a behavior module in a robot. A quick and dirty method for doing this would be to let the subsystem lag behind in its response to the coupling variable dynamics a bit, which allows one to use the history of the immediate past for calculating the derivatives from the already observed history of these variables. A rather more satisfying, but also more difficult approach is to estimate the derivatives by a suitable filter. It seems to me that biological behavior subsystems often combine the filtering and the lagging strategy.

It should be noted that v can be allowed to vary over *all* reals, although the bijections $\varphi_{\mathbf{c}}$ reverse the orientation of trajectories if v becomes negative, and even lead to a complete standstill if v is zero. However, all the results reported are still valid for any differentiable $v : \Re \rightarrow \Re$. This is useful since it allows to put the system into "reverse gear" while the trajectories can still be understood in terms of (26) (or (35), to be more specific).

6 Conclusion

In this paper I have started to explore methods that help to design behaviors as dynamical systems. I have developed an approach to making behaviors, to some extent, immune against qualitative disruption due to coupling dynamics. The methods described enable the designer to construct behavior subsystems as "modulated modules". The important special case of shift, amplitude and velocity modulation is worked out in a ready-to-use form.

The work reported here is only a beginning. Besides *vas*-families, other standard modulations appear to be easily accessible to an explicit treatment, e.g. rotations or amplitude variations that are different in different phase space dimensions. The observation that v can be allowed to take negative values cries for a closer investigation, since it might open the way for useful generalizations of qualitative equivalence. It would also be very interesting to investigate behavior systems that afford of several different qualitative dynamical modes with bifurcations in between, like e.g. gait patterns of walking systems. One might wish to stabilize each of the modes with separate compensations. And last but not least, of course, one should compare the compensation mechanism described here with biological mechanisms.

It seems to me that there exist at least two different, equally important functions of modulation mechanisms like the one described in this paper.

First, they are useful for the *control* of a behavior subsystem by others. For instance, a motivation subsystem might increase running speed by increasing the value of the velocity coupling variable v of the running subsystem. Second, they can serve the *adaptation* of a behavior to external circumstances. For instance, a walking behavior might adjust its stance length according to variations in its amplitude factor a, which are effected through sensor input corresponding to the ease of walking on different types of ground. The difference between these two functions is mainly one of perspective. If modulation is viewed as a function of control, the focus lies on the active demands generated by external modules. If it is considered in the service of adaptation, the focus is on the modulated behavior itself.

This may all be quite nice as far as it goes. One must not forget, however, that the techniques reported in this paper cannot help with the design task proper: which behaviors an agent should have, and how they should interact with each other. Only after this basic design task has been accomplished, compensation techniques can be applied to stabilize the qualitative type of behaviors. But then, they should.

Acknowledgements. The basic ideas for the work reported in this paper were developed during a stay at the VUB AI Lab at the Vrije Universiteit Brussel on a grant from the CONSTRUCT project, which is funded by DWTC through the programme IUAP nr. 20. The ideas were worked out in detail within a working contract from GMD at St. Augustin. I am deeply indebted for support and encouragement given by Luc Steels and Thomas Christaller.

References

[1] D.K. Arrowsmith and C.M. Place. *Dynamical Systems: Differential equations, maps and chaotic behavior.* Chapman and Hall, 1992.

[2] R. Beer. A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72(1/2):173–216, 1995.

[3] M. Casdagli. A dynamical systems approach to modeling input-output systems. In M. Casdagli and S. Eubank, editors, *Nonliner Modeling and*

Forecasting, volume XII of *SFI Studies in the Sciences of Complexity*, pages 265–281. Addison-Wesley, 1992.

[4] H. Jaeger. Dynamische Systeme in der KI und ihren Nachbarwissenschaften. Arbeitspapiere der GMD 925, GMD, St. Augustin, 1995.

[5] T. Smithers. What the dynamics of adaptive behavior and cognition might look like in an agent-environment interaction system. In *Proceedings of the Workshop 'On the Role of Dynamics and Represenation in Adaptive Behavior and Cognition'*, pages 135–153. Universidad del Pais Vasco, San Sebastian, 1994.

[6] T. Smithers. What the dynamics of adaptive behavior and cognition might look like in agent-environment interaction systems. In R. Pfeifer, editor, *Practice and Future of Autonomous Agents. Workshop Notes of the NATO ASI on Autonomous Agents, Ascona.* Computer Science Dpt., University of Zurich, 1995.

[7] L. Steels. Mathematical analysis of behavior systems. In P. Gaussier and J.-D. Nicoud, editors, *From Perception to Action*, pages 88–95. IEEE Computer Society Press, Los Alamitos, CA, 1994.

[8] R.F. Stengel. *Stochastic Optimal Control.* John Wiley and Sons, 1986.