Systems Research and Behavioral Science
Syst. Res. **25**, 447–465 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/sres.914

■ *Research Paper*

System Dynamics as Model-Based Theory Building

Markus Schwaninger[1]* and Stefan Grösser[1,2]

[1]*Institute of Management, University of St. Gallen, Switzerland*
[2]*Interfacultary Center for General Ecology, University of Berne, Switzerland*

This paper introduces model-based theory building as a feature of system dynamics (SD) with large potential. It presents a systemic approach to actualizing that potential, thereby opening up a new perspective on theory building in the social sciences. The question addressed is if and how SD enables the construction of high-quality theories. This contribution is based on field experiment type projects which have been focused on model-based theory building, specifically the construction of a middle-range theory—not a general one. The process of model building and validation is analysed from a theory-building perspective. The resulting theory is evaluated by means of a set of criteria for high-quality theories. As a conclusion, the insights thus gained are presented and condensed in a tentative set of heuristic principles for model-based theory building.

Keywords system dynamics; modelling and simulation; model-based theory building; case study

INTRODUCTION

This contribution aims to explore the issue of theory building with reference to system dynamics (SD) modelling. Our specific goal is to introduce the concept of model-based theory building and demonstrate its great potential on the basis of a detailed case study.[1] The question, which has moved us for years, is: 'Does system dynamics enable the building of high-quality theories, and how'? We are not comparing theories created with the help of SD versus those brought forth by means of other methodologies. Our approach is to gain insights into the nature of theory-building processes supported by SD and to evaluate an SD-enabled theory using general quality criteria for theories.

SD modelling is often colloquially referred to as 'theory building'. What is meant by theory building in this context has, however, hardly been subject to discussion. Even so, we have identified some work which has taken up the issue. Early on, Forrester conceptualized SD

*Correspondence to: Markus Schwaninger, Institute of Management, University of St. Gallen, Dufourstr. 40a, CH-9000 St. Gallen, Switzerland. E-mail: markus.schwaninger@unisg.ch
[1]The case study is one of a population of 36 case studies with similar characteristics. Here, we particularly refer to the selected case.

Submitted 2 July 2007
Accepted 2 February 2008

modelling and simulation in essence as a discipline of theorizing, involving experimental work, that is 'designed to prove or disprove the initial hypothesis' (1961, p. 450). Hanneman (1988) conceived of computer-assisted dynamic modelling as theory building. Schwaninger (2003) laid out the modelling endeavour as the building of local theories based on conceptual frameworks—the archetypes which emanate from the SD community (*cf*. Wolstenholme, 2003; Senge, 2006). Finally, Karlöf and Lövingsson (2006) even equated practitioners' problem solving with theory building. However, it is necessary to delve deeper into the concept of theory construction based on formal models. This is the gap we will address.

The paper is structured as follows. We will begin by conceptualizing theory building, modelling, the range and quality of theory. Thereupon, experiences with model-based theory building will be presented by means of a detailed case study. That case study refers to a modelling project as the locus of theory building. A discussion will ensue, and the paper closes with brief conclusions.

CONCEPTS

Theory Building

Our basic idea was that theory building, in principle, is more than an exercise in academic abstractions, but rather an activity fundamental to the survival of societies, organizations and even individuals (*cf*. Schwaninger and Hamann, 2005). We conceive of a theory as a structured, explanatory, abstract and coherent set of interconnected statements about a reality. More specifically, Davis *et al*. define *theory* as 'consisting of constructs linked together by propositions that have an underlying, coherent logic and related assumptions' (2007, p. 481). The term derives from the Greek 'theorein'—to look at, to ponder, to investigate. Concomitantly, the noun 'theoria' refers to notions such as observation, inquiry, reflection and research. Theorizing, in this sense, is to observe what is going on in the real world, to reflect about it or experiment with

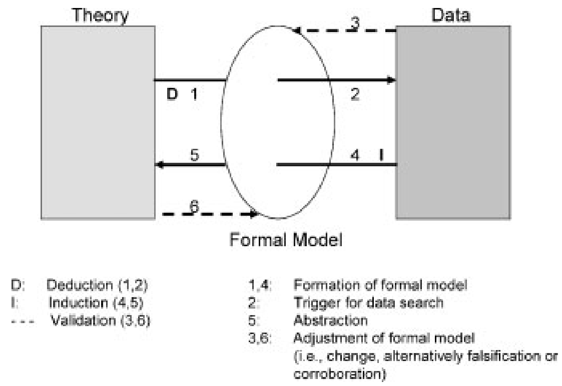

Figure 1. Induction and deduction in modelling and theory building

it and to draw systematic conclusions, which have practical implications.[2] Theory building as conceived of here, consists of generating and formalizing a theory in order to orientate action.

However, there are different kinds of theory building (see Figure 1). The differences become visible if we take the source of knowledge as the criterion of distinction. The common way of building theories is to rely on observation as the source of knowledge. One way is to gather the data by observing real-life events, studying cases, carrying out surveys, etc. The data are then analysed, usually explored statistically, interpreted and presented. Successive tests then lead to the formulation of the theory. This is the '*data-to-theory*' process which we call *induction*. The other approach is to build a theory in a way that certain assumptions are stated from which a theory is derived along a logical sequence of steps. Later on data may be collected to test the theory. This is the '*theory-to-data*' approach which we call *deduction*. As this description shows, both deduction and induction are components of the research process, which in principle complement each other.

When data are gathered, one already relies on a theoretical framework; and when starting a deductive venture, one has already observed

[2]Often a distinction is made between a reference to a system 'out there' and a subjective interpretation by a human (e.g. Checkland, 1981; Lane, 1999; Zagonel, 2004). The stance taken here is that the modelling of social systems as discussed in this paper always involves the subjective interpretation by an observer who, in this case, is a modeller.

Syst. Res. 25, 447–465 (2008)
DOI: 10.1002/sres

some real-life events which influence the choice of categories and the logic of the deduction.

As far as our chosen theory-building processes are concerned, they do not involve deduction or induction alone, but they utilize both in combination. Methodologically, the processes are focused on building a formalized, quantitative model and carrying out simulation with computer support. This is also referred to as computer-supported theory building. In the cases we will refer to, the SD methodology for modelling and simulation has been used as a conceptual and instrumental device supporting the construction of theories.

Why do we insist on model-based theory building? Would not a model in itself be a theory and *vice versa*? Indeed, a broad concept of 'model' would subsume theories, frameworks and the like. The concept of 'model' we are using here, however, is a more operational one. In the context of this contribution, we define 'model system' as a sufficiently accurate representation of a real system, in which the constituent variables and their functional relationships, including the underlying assumptions, are formalized and therewith made transparent.

From the stance of model-based theory building, we are considering the model, if it is generic and formalized, as a constituent part of the theory.[3] Therewith, theories which emanate from a process of theorizing that is based on explicit, formal modelling[4] have the potential to be stronger—in terms of both robustness and reach—than theories without this property, which are largely based on implicit, mental models.[5] In this connexion, Ashby's law of requisite variety must be quoted: 'Only variety can absorb variety'.[6] Translated to the context of this contribution, the strength of a theory hinges on its richness in relation to the complexity of the reality it deals with. In a similar vein, the Conant–Ashby theorem indicates that the effectiveness of an operator cannot be higher than the power of the model on which his or her operation is based.[7]

Model-supported theory building is, first of all, about the construction and validation of a model. The theory, then, emerges as the model is built and validated. That process is not a linear sequence from assumptions to hypotheses to empirical investigations, as postulated in descriptions of ideal-typical theory-building procedures (*cf*. Merton, 1968, p. 68). The process is iterative, and iterations can occur between any pair of steps of the procedure.

What is of interest, at this point, is the *How*, the *Why* and the *When* of theory emergence. The answers given here follow an evolutionary logic (*cf*. Popper, 1972; Dopfer and Potts, 2008). How does the theory emerge? It emerges via variation and selection; options are created, tried out and selected. Why do certain theories appear? They come up because they were, in principle, the stronger options which were selected. When do the theories emerge? They are brought about continually, and therefore they can emerge at any time.

A SD model is a theoretical statement. The theory is created as the model is accomplished. For that construction process, the source of knowledge is, in the first place, a reservoir of mental models, available in the modellers and their interlocutors, which have to be elicited. These mental models can be of an inductive or a deductive origin; usually they are both. The SD model itself emanates to a large extent from deductive reasoning, but the validation process partially relies on real-world data by which confidence is built into a model and its quality is enhanced. As the behavioural validation feeds back to model calibration and may even lead to changes in the model structure (Barlas, 1996), this component of the process is of an inductive nature.

[3]This expresses a view of theory building in a wider sense, and is in line with Forrester's claim that a 'general model' is 'a theory of a particular class of system' (1968, p. 607).
[4]A formal model can be of quantitative nature (e.g. numeric), but it can also be of a qualitative type (e.g. graphical).
[5]The benefits of formal theory building have been widely shown in the literature. See for example Grove and Meehl, 1996, Homer, 1996, Hannan *et al*., 2007, Lane, 2008.
[6]In the original: 'Only variety can destroy variety' (Ashby, 1956).
[7]In the original: 'Every good regulator of a system must be a model of that system' (Conant and Ashby, 1981). In this context it is noteworthy that, according to experimental research, the structures of mental models used by decision makers are strongly predictive of their performance (*cf*. Ritchie-Dunham, 2002, and literature related therein).

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

Modelling

We conceive of modelling as a process by which models are built, whereby we will focus on computational models. A model is a representation[8] or a construction of a reality. If perception is an activity, as opposed to a passive happening (Neisser, 1976), then a model is a subjective construction. More precisely, it is normally a conceptual construction of an issue under study. According to the constructivist position, modelling is constructing a (new) subjective reality. The modeller is an observer who, by the act of either observing or modelling, creates 'a new world' (*cf*. von Foerster, 1984).

In the following, we shall particularly refer to SD models. SD modelling is about constructing models as continuous feedback systems. These are formal, structural models. They incorporate hypotheses about the causal connections of parameters and variables as functional units, and the outcomes of their interactions. In the case of theory building, SD models must be fully transparent, that is they have to be available in the form of white boxes instead of black boxes, which would be counterproductive for this purpose. This way, a model is a strong device for supporting the process of theory building. It mostly evolves in the form of a discourse, in which different people are involved. In other words, what keeps the discourse going is usually not a completed model, but a model in its different stages, built by a group. Such group model building has already been the object of deeper studies (Vennix, 1996).

The formalization of mental models by SD—qualitative and quantitative[9]—fosters transparency. It also increases refutability, which is a prerequisite for submitting them to scientific debate. Theory building in the human and social sciences is not primarily meant to be an exercise of underpinning hypotheses, nor anything like 'proving' their truth. It is a process by which assumptions and theories, that is systems of hypotheses, are specified, and then are submitted to tests. These tests are essentially endeavours of falsification, as established in Popper's critical-rationalist theory of science (Popper, 2002). If falsified, then the theory is refuted. If, however, the attempt of falsification is not successful, then it can be temporarily maintained.

The basic value of a model or a simulation outcome is that they embody propositions which can be refuted. The point is not primarily if a proposition is true or false, but whether it provides an anchor around which arguments can be built.

SD models greatly enhance falsifiability—each interrelationship can be tested, both logically and empirically. In this sense, an SD model is a candidate for a theory. This consideration is applicable to properly constructed models, that is those models which make the underlying assumptions explicit, operationalize the variables and parameters and are submitted to adequate procedures of model validation (Forrester and Senge, 1980; Barlas, 1996; Sterman, 2000).

Range of Theory

Academics are, in the first place, interested in creating *general theories*.[10] General theory is any theory which attempts a highly generic, often overall explanation of a whole range of phenomena, for example of social systems. It often strives for a unification of several separate theories. An example from institutional economics would be a theory of social and economic change.

Middle-range theories are 'theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behaviour, social organization and social change' (Merton, 1968, p. 39). These theories of the middle range consolidate different hypotheses or findings (Merton, 1957, pp. 280, 328), and they are empirically grounded. Following up on the example of general theory, an instance of a

[8]Such representation can be descriptive or anticipative or prescriptive.
[9]We adhere to the view that also qualitative approaches to formalization do exist, for example graphical and verbal approaches, if the criteria for formalization are consistency, precision, inequivocality, etc.

[10]A similar concept used in the social and human sciences is the one of 'grand theory' (*cf*. Skinner, 1985).

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

middle-range theory would address connections of specific hypotheses from both the economic and the social domains. In our view, the consolidation provided by a theory of the middle range could also be one of different local theories, for example by expanding the realm in which a theory applies (see also Paich, 1985).

Local theories are those theories that apply to a particular context, explaining behaviours encountered in specific instances. An example here would apply the principles of economic change to a specific social system, for example an organization or society, by explaining how the current situation of that unit emerged. In this sense, theory building is not only the domain of theoreticians but also of practitioners.

Quality of Theory

Following up on these conceptual reflections, we shall refer to a set of criteria that helps to evaluate the quality of a theory. Usually, two or three criteria are quoted in this respect: validity, reliability and eventually objectivity. We have looked out for a more complete and operational set of criteria for theory evaluation. We found a sufficiently concrete list in Patterson's eight criteria for evaluating theory (Patterson, 1986). The original list has been completed with definitions by Holton and Lowe (2007). These are taken on here with small modifications and a slight extension of the set of criteria in attendance to the focus of model-based theory building (criteria one and eight have been added):

1. refutability: ability of a theory to be falsified (refuted) or supported;
2. importance: a quality or aspect of having great worth or significance; acceptance by competent professionals may be indicative of importance;
3. precision and clarity: a state of being clear; hypotheses can easily be developed from the theory;
4. parsimony and simplicity: uncomplicated; limitation of complexity and assumptions to essentials;
5. comprehensiveness: covering completely or broadly the substantive areas of interest;
6. operationality: specific enough to be testable and measurable;
7. validity: valid, accurate representation of the real system under study;
8. reliability: free of measurement errors;
9. fruitfulness: statements are made that are insightful, leading to the development of new knowledge;
10. practicality: provides a conceptual framework for practice.

This list is compatible with other suggested sets of criteria, for example with Bacharach's concept of organization theories, where falsifiability and utility—corresponding to the criteria 1, 9 and 10 above—are denoted as 'the two primary criteria upon which any theory may be evaluated' (Bacharach, 1989, p. 500). In the following discussion of the case study, we will, first, analyse the modelling process from a theory-building perspective, and second, evaluate the resulting model by means of the evaluation criteria for high-quality theories.

CASE STUDY: BUILDING A THEORY OF PRODUCT LAUNCH STRATEGIES

Context

The following case study analyses a modelling project in which a SD simulation model for a real-life corporate problem was developed. The resulting model will be evaluated from the perspective of theory building. It stems from a population of 36 case studies which were essentially theory-building endeavours. The selected case study demonstrates a typical model-building venture by which the features of a theory-building process can be fleshed out. We present a revelatory[11] exemplar, that is an instance, which illustrates something that is—in principle—encountered in the whole population of cases: the specific nature and features of a theory-building process, in particular, the combination of deduction and induction, under the

[11]Single case studies, as opposed to multiple case studies, are indicated if they are revelatory (Yin, 2003, p. 45f.).

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

condition of compliance with the criteria of high-quality theory and good modelling practice.

The project under study was realized in cooperation with a Swiss industrial company, which is one of the two world leaders in its field. The company faced the problem that at the time of a product launch it had no means of estimating the potential sales development. Furthermore, it lacked knowledge about the factors that significantly influence the development of product sales. The project was of high relevance because the short product life cycles require fast and resolute decisions. The purpose of the model was to optimize the company's product launches in order to maximize the monetary return of each product generation. It was aimed at understanding which factors influence the product launches; in particular, at identifying those that have the strongest impact on the financial outcome of a product launch. Moreover, the model should yield general policy insights to guide future product launches. For us, the external partners, the crucial objective of the project went beyond mere practical empiricism; the search for a theory of industrial product launches pervaded the whole endeavour.

The project started in May 2007 and lasted for six months. The interaction with the company hinged on two gatekeepers who were the main interlocutors of the external modellers, that is the authors with the support of three students, for whom the modelling was part of their training. The modelling work was carried out by the external partners in interaction with the internal gatekeepers, who on their side involved additional company members. For the sake of brevity and a clear distinction, we will use the terms *core modelling team* and *researchers* for the external group, while the internal people involved actively in the modelling process will be called *client team*.

Modelling Process

The modelling process followed standard procedures and good modelling practice as laid out, for instance, by Richardson and Pugh (1981), Wolstenholme (1990) and Sterman (2000). The core modelling team interacted with the client team in several group model-building sessions. Before the first, the issue statement as well as the purpose statement of the model had been crafted in a dialogue between the internal gatekeepers and the external partners, via internet. These preliminary statements were discussed, changed and constituted in the first joint session. The group effort on that occasion also led to a first conceptualization of the dynamic hypotheses, which will be expounded shortly.

In addition, reference modes for two variables were established—monthly product sales rate and revenues, to guide the modelling endeavour. Figure 2 already shows the time series for sales rate based on the empirical values from internal company data. In the following sessions, the qualitative dynamic hypotheses were formalized as quantitative, equation-based dynamic hypotheses, which were subject to discussion and enrichment. The clients were 'walked through' the model in detail in order to ensure both the model's face validity and the acceptance by our corporate partners—essentially a deductive exercise. These were 'theoretical' sessions for practical probation. The reflections offered on these occasions led to improvements of the early model versions and to the identification of additional data requirements. Several subsequent modelling sessions helped to parameterize and test the model, as well as perform scenario and policy analyses. These interactions with the client team occurred via personal and online collaborations as well as offline correspondence.

During the modelling endeavour, process facilitation and communication with the client was given priority over the precision of the model, particularly in using parameters which were reasonable to the problem owners. The modellers focused on keeping the model as parsimonious as possible and, at the same time, to avoid black box modelling, which easily could have occurred given the complexity of the market situation and the company's size, structure and product portfolio. Figure 3 shows the main feedback structure of the model that emerged within about 4 months. The diagram also highlights that the core of the model is a Bass diffusion structure (*cf*. Sterman, 2000).

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

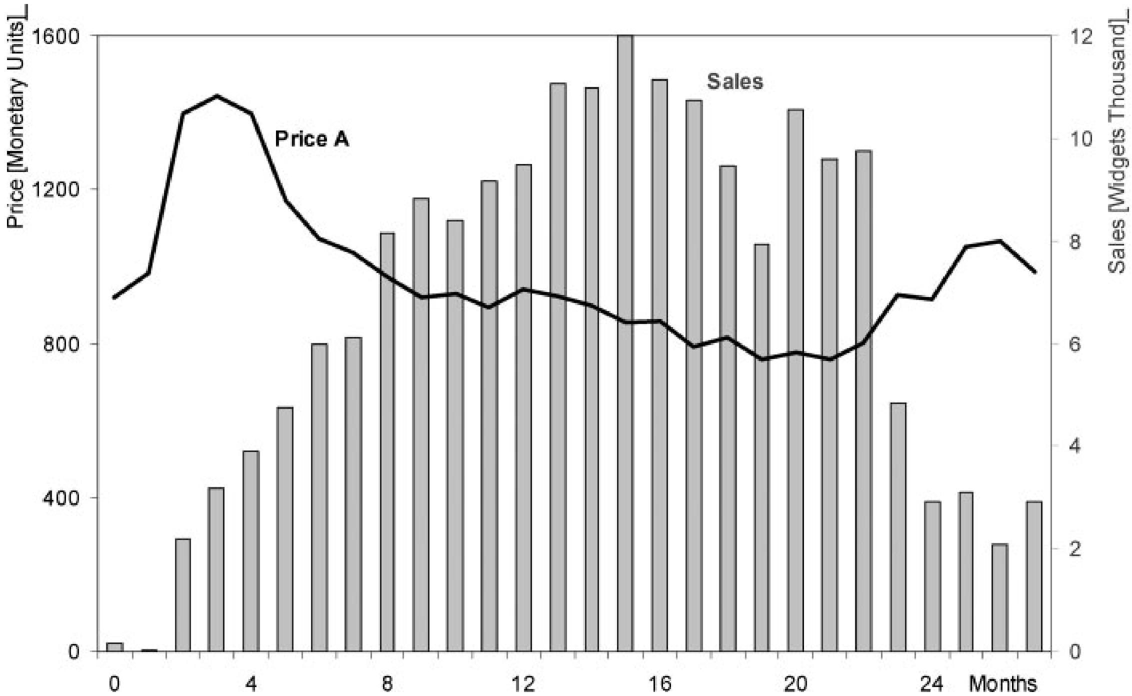

Figure 2. Data about product sales, used as a reference mode for the product launch model, in relation to the exogenous price of the product

The purpose of the initial modelling was to develop an explanation of the historical sales of a typical product of the company, as graphed in Figure 2, and not to investigate the future. We started out with initial hypotheses, which were 'improved', maintained or eliminated in an evolutionary process, in which some of the hypotheses 'survived'. One of those dynamic hypotheses was that growth in the number of customers, and thus growth in the monthly sales rate, was the result of improvements in the perceived attractiveness of the company's product relative to the products of competitors, rather than reputation effects:

$$AR_t = \max\{ARA_t + AWM_t, PC_t\} \qquad (1)$$

where AR is the adoption rate, ARA the adoption from relative attractiveness, AWM the adoption from word of mouth and t is the timing of the launch.

Attractiveness was defined as a function of quality of service, product features, effect of marketing activities and product assembly time, all of which were funded or influenced by the company's revenues. The influences of all of these attributes on the product attractiveness were weighed according to the judgment of product experts of the company. The product assembly time was positively influenced by cumulative learning effects; Figure 4 provides details about the attributes of the company's product attractiveness.

A second dynamic hypothesis was that only a positive influence of word of mouth on the adoption rate existed; we thereby abstracted from possible negative effects of word of mouth:

$$AWM_t = \frac{\bar{c} \times \bar{i} \times (TC_t - PC_t) \times PC_t}{TC_t} \qquad (2)$$

where AWM is the adoption from word of mouth, c the contact rate, i the adoption fraction, TC total customers and PC are potential customers.

However, negative effects were included in the model in that long product assembly times reduce the product attractiveness and lead to customer losses (*cf*. Figure 4). The last dynamic

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

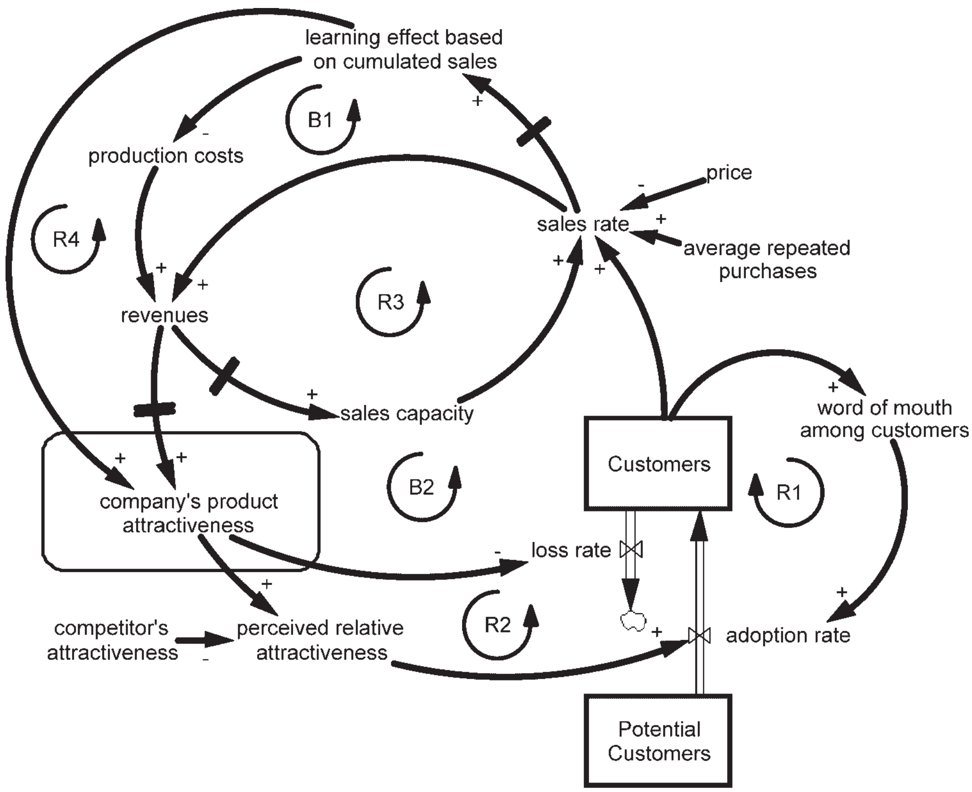

Figure 3. Main feedback structure of the product launch model near completion. The rounded rectangle summarizes several product attributes that are detailed in Figure 4

hypothesis was that the sales volume depended on the sales capacity, which on its part is a function of the company's revenues (Formula 3). Other determinants of sales volume were price and average repeated purchases:

$$SV_t = \min\{FTP_t + RP_t, SC_t\} \quad (3)$$

where SV is the sales volume, FTP the value of first time purchase, RP the value of repeat purchases and SC is the sales capacity.

All of these three dynamic hypotheses stood the empirical tests, as will be delineated in the next section.

Model Validation

As suggested by SD experts (e.g. Forrester and Senge, 1980; Barlas, 1996; Sterman, 2000), the validation of the model was an integral part of every modelling activity. During the iterative model development, we constantly questioned our assumptions by means of several direct and indirect structure tests. Each attempt to falsify the provisional theory resulted in a better understanding of the model structure, and the relationship between structure and behaviour. Furthermore, these trials also helped the core modelling team to refine and calibrate the model, whereby confidence in the resulting simulations and structural analyses was enhanced. In the following, we concentrate on three validation issues.

The first, and perhaps most important, aspect of validation was that the model's structure and its behaviour emerged directly in interaction with the group of managers who would finally use the model to support their decisions. The shared understanding of how a product launch occurred, and what effects on the company as

Syst. Res. 25, 447–465 (2008)
DOI: 10.1002/sres

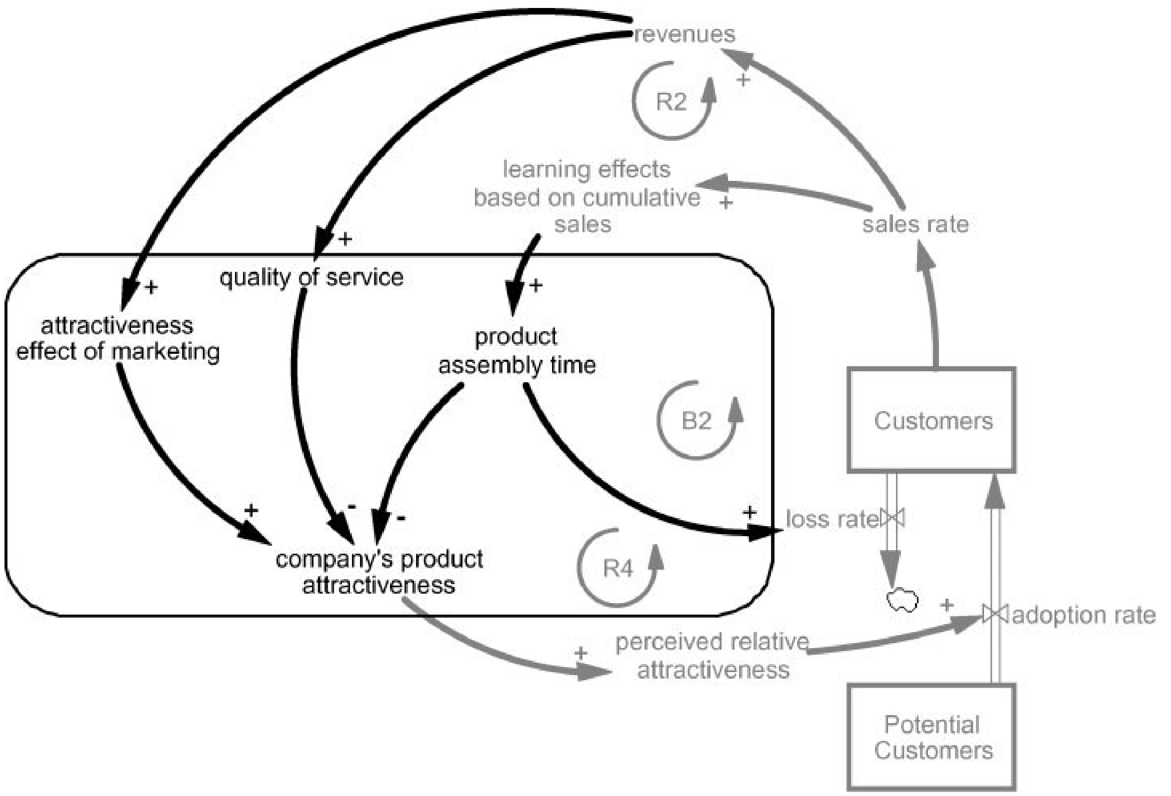

Figure 4. Detailed representation of the attributes of the company's product attractiveness (for the complete feedback loops see Figure 3)

well as potential customers it would have, emerged from discussions with initial sketches being drawn by hand on whiteboard or on a projected computer screen. This procedure was important to establish the credibility of the simulation modelling and nurture a feeling of ownership in the model among the group (*cf.* Richardson and Pugh, 1981, p. 355; Vennix, 1996). The inclusion of the client team in the modelling phase is a procedural measure for improving the validity of the model.

Second, the inductive validation of propositions or hypotheses about causal relationships and assumptions about parameter values utilized the standard direct structure tests and drew on a variety of data sources. Following Randers, 'model testing should draw upon all sources of available knowledge' (1980, p. 295). We used numerical data from the company's records. However, some parameters and functional relationships were not directly available in the existing data records. We had to rely on additional sources: (1) specific observations of company processes by members of the client team, (2) expert interviews and (3) surveys among experts who were non-core project participants. It is noteworthy that the requirements for the additional data emerged during the model creation process. We sampled the data as required in order to validate new assumptions or causal relationships. The gatekeepers supported the modellers in this process and gathered independent estimates from knowledgeable managers. These data were analysed for consistency and reliability before they were utilized for the model formulation. In this way virtually all data needs were met and the researchers were able to create a simulation model on solid empirical grounds that would be accepted by the client.

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

Third, sensitivity analyses and other indirect structure tests (also called structure-oriented behaviour tests: Barlas, 1996) enhanced the confidence of managers in the simulation model. The analyses supported the validation of model assumptions for which it was difficult to obtain numerical empirical data by demonstrating the model's behaviour insensitivity to changes in parameters. Moreover, the analyses showed that the derived policy insights were largely independent within plausible ranges of parameter values. One of these insights was that the mode of the product launch strategy of the competitors—either parallel or sequential—severely impacts on the company's sales development over the life cycle. The reason is that an early mover advantage can hardly be compensated over the relatively short life cycle of 24 months.

Finally, behavior reproduction tests compared the reference modes that were based on historical real-world company data with the output of the simulation model. Forrester and Senge suggest a battery of tests in this category. However, most of them are not applicable to models that try to replicate one instance of life cycle dynamics, as in our case (Forrester and Senge, 1980: 219). Hence, we applied the behavior reproduction test in its basic version and compared the reference mode pattern with the pattern of the model output. The model reproduced the product life cycle pattern with high accuracy. Figure 5 shows the historical and the simulated time series for the product revenues (based on n=26 data points, R^2=0.9967 for the corrected historical time series; see below). To assess the model's behavior validity, we utilized the mean square error (MSE) and Theil's inequality statistics (Sterman, 1984). MSE provides a measure of the total error and Theil's statistics specify how it breaks down into three components: bias (U_M), unequal variation (U_S) and unequal covariation (U_C). An inspection of the historical data series revealed an interesting deviation from the normal life cycle behavior for $t \in [20, 22]$. Discussions with company experts could not clarify the reasons for the exceptional behavior. Hence, we treat the data points as outliers and substitute them with the average value of the long-term trend. Given the corrected historical data series of revenues, the MSE is 0.35%, the inequality statistics are: U_M=0.01, U_S=0.01, and U_C=0.98. The major part of the error is concentrated in the unequal covariation component, while U_M and U_S are small. This signifies that the point-by-point values of the simulated and the historical data do not match even though the model captures the dominant trend and the average values in the historical data. Such a situation might indicate that the majority of the error is unsystematic with respect to the purpose of the model, and it should therefore not be rejected for failing to match the noise component of the data (Sterman, 1984: 56). The residuals of the historic and simulated time series show no significant trend and strengthen the assessment that the structure of the model captures the fundamental dynamics of the issue under study.

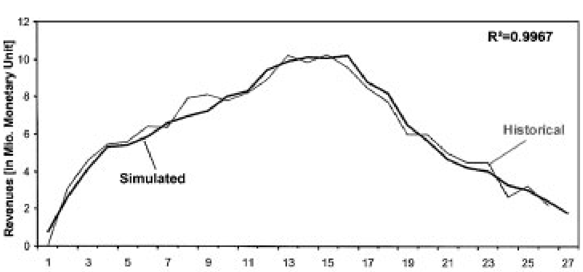

Figure 5. Comparison of historical and simulated time-series for the product revenues. The explained variance is close to 100% ($R^2 = 0.9967$).

The final version of the model is relatively small, containing about 50 equations. However, it not only provided a close fit to the available product-level data, but it also did so for the right reasons: a set of parameter values and causal relationships that appeared robust and accurate to the client. Even though the model was small and condensed, it triggered important insights. That was what had been required at the outset.

The model provides a theory which explains the evolution of the sales of a typical industrial product launch as a function of a set of drivers, that is parameters, which are the explicit expression of certain assumptions. Alternative theories had been considered along the way, but were successively substituted by better ones. The selection filter which orientated the choices made en route was the process by which the internal validity was ascertained. Aspects of external validity and generalizability will be treated in the following section.

Discussion and Evaluation of the Case

Our aim for this section is to discuss the model-building process and its results in terms of the theoretical and methodological considerations laid out in the first part of the paper.

We chose a single case design, selecting a revelatory case from a sample of 36 cases. Our main argument is that a SD model can be a crucial factor in a theory-building process and that the model itself becomes an integral component of the theory created in that process. We do not propose that this would apply to SD models only. Davis *et al*. (2007) have shown that other methodologies as well possess large potential for the construction of model-based theories. We have supported our argument only by means of the evidence collected in the application of SD.

The purpose of our work was to understand product launches of the case company, not of a specific product. Beyond that, we tried to identify the class of systems in terms of situations to which the model could be applied (*cf*. Bell and Senge, 1980, p. 66; Lane and Smart, 1996, p. 94), and to model that particular class. We are confident that we have captured a highly generic system structure that can be used not only to understand the case described here, but virtually any industrial product launch. To test this assertion, we used a second set of reference data representing a product launch in a geographically distant country. The calibration process required only the adaptation of parameter values of the Bass diffusion structure (contact rate, adoption fraction, initial potential customers) and some product attributes such as repeat purchase rate, initial purchase rate and price of product. This is plausible since the two countries (one within Europe, the other within North America) certainly differ in their information dissemination characteristics. This second application follows what Yin calls 'replication logic', which allows cumulating knowledge across cases and building support for the theory under test (Yin, 2003, p. 37). Therefore, the results of the second application have increased our confidence that the created theory is applicable to a whole class of cases—product launches in industrial firms. In this sense, it is more than a local theory. As it consolidates several hypotheses and applies to a large range of cases, it is closer to a middle-range theory.

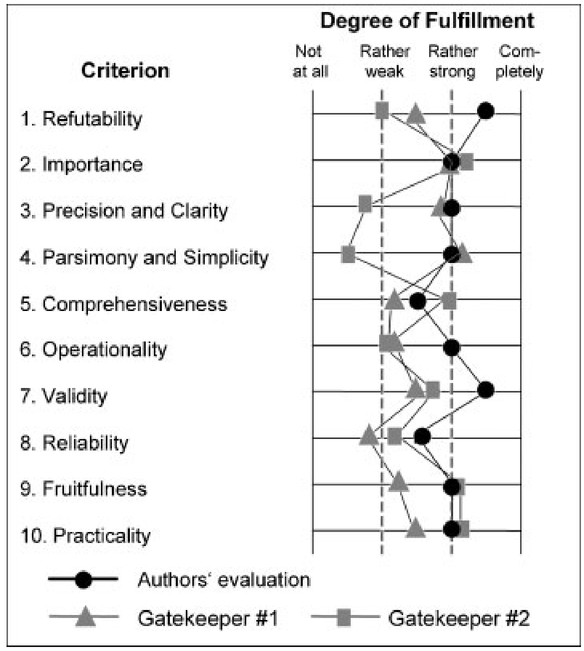

Figure 6. Summary of three views of the product launch theory

In the following, the quality of the attained theory will be evaluated utilizing the 10 evaluation criteria of the quality framework developed in the first chapter: a summary can be found in Figure 6:

- *Refutability*. A sine qua non of a quantitative SD approach is the formulation of an explicit set of mathematical equations. These stand for testable propositions, enabling researchers to check how well their assumptions match available data about the overall system behaviour. In the simulation model, the underlying structural and behavioural assumptions are formalized and therefore testable. In other words, in principle they can be falsified.

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

- *Importance*. This attribute of a theory can be assigned only by the target group for which the theory has been created. The primary targets were the project managers of the Swiss production company. They continually alleged, during the modelling sessions, that the theory under construction was very helpful from their perspective; hence, the mere acceptance of the theory by the client was given. After the formal end of the cooperation, the project was even heaved to the top management level, being presented during the strategy meeting towards the end of the year. In a next iteration of the project, it will—most likely, as the gatekeepers assured us—be supported by top management executives. This is another hint on the importance of our theory-building work.
- *Precision and clarity*. Throughout the project, the researchers paid close attention to the consistent and precise use of the terms and concepts which appear in the emerging theory. Moreover, the concepts used were operationalized, explained and documented in a glossary to increase the theory's clarity. In the end, the theory furnished a precise and consistent conceptual apparatus. Furthermore, the relationships between the theory's concepts are defined explicitly and unambiguously. In this way, the development of testable hypotheses is greatly facilitated.
- *Parsimony and simplicity*. This criterion, in essence, advocates the avoidance of multifariousness as opposed to the capability for absorbing complexity, which has to be fostered as Ashby's law postulates. Theories therefore should be as simple as possible, but not simpler than that. In principle, SD is a methodology for capturing and managing dynamic complexity, not the complications of combinatorial complexity (Sterman, 2000). One of the dangers of SD modelling, however, is that it may lead one to succumb to pressures towards the creation of overly large models that try to contain the combinatorial richness of the real system but not the dynamic complexity. The guideline that one should model a problem or issue, not a system with all its details (*cf*. Richardson and Pugh, 1981; Sterman, 2000), aims precisely at avoiding that trap. In the case under study, the modellers abided by this principle, and they also had to defend it against severe initial resistance from the client team. The team members called for incorporating more and more details in the model, which conformed to their day-to-day experience and to an event-oriented perception. Finally however, the parsimonious approach proved to be appropriate and useful. The parsimony of the theory is expressed in the limited size of the model. Its simplicity lies in the small set of components it consists of, which is fully transparent and focused on the essentials.
- *Comprehensiveness*. This criterion establishes whether the theory is sufficiently broad to cover the substantive issues of interest. Regarding the theory developed in our case, we are confident that the level of comprehensiveness is reasonable, because the boundary of the simulation model was permanently subject to tests whether the model scope was adequate or not. At the beginning of the project, the scope was rather narrow. During the course of the project, more details were added until the model contained structural elements the deletion of which barely influenced the behavioural outcome. At that stage, the modellers began to continuously apply the boundary adequacy test (Barlas, 1996) in order to examine the appropriateness of the level of detail of the model as well as its scope. For instance, the sales capacity had been modelled in more detail in an earlier version of the model until it became clear that the dynamics of hiring–firing and training of the sales force was not important to the situation-in-focus. We discovered that it was far more important to model the physical capacity of the product assembly, since this was one of the limiting factors for product customization and delivery: capacity severely influenced the perceived relative attractiveness of the company's products. In conclusion, through several adaptations of the model boundary, which were based on empirical data and discussions with the client team, we were able to create a rich theory. It captured the essential dynamics

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

of the system under study, which indicates a certain comprehensiveness in the model.

- *Operationality*. Operationality signifies that the theory is specific enough to be both testable and measureable. First of all, we made specific efforts to abide by the imperative that no variables should be used that do not have a counterpart in the real world. For example an initially used variable 'costs ratio' was discarded, because it was not sufficiently observable in the real world and it was actually not a variable used by the decision makers. Instead, we replaced it with the total unit costs. Since the product launch theory is one of a socio-technical system, it incorporates not only clear-cut factors but also soft variables that are difficult to measure (*cf*. Stouffer *et al*., 1950). For the development of the theory, we captured soft variables by using Likert scales (*cf*. Cowles, 2001). For example a questionnaire was distributed to knowledgeable experts in the client company. Their task was quantifying the impacts which different product attributes have on product attractiveness, etc. The created glossary testifies that each variable of the product launch theory is highly operational, in that it can be measured and tested with real-world data.
- *Validity*. Validity implies that the model depicts the real system under study with a high level of accuracy. In SD, sophisticated research about model validation has been conducted (*cf*. Forrester and Senge, 1980; Barlas, 1996; Sterman, 2000). The extensive validation work undertaken for our theory of product launch included direct and indirect structure tests. In this case, we conducted an iterative process of model construction, in which the assumptions were validated or falsified in many simulation experiments based on comparisons between the model output and available data, especially the reference modes about monthly sales rate (*cf*. Figure 2) and revenues. For instance, the cost/sales structure of the initial model contained only direct sales and cost of sales for the tangible products. The validation against the reference modes indicated that a second source of revenues existed, which increased in volume over time. In discussions with the client team, we identified product-related services as this second source. The dynamic change of the revenues was tied to increasing marginal revenues that were traced back to learning effects. The inclusion of this source of sales improved the fit of the model output and the reference modes. To summarize, based on numerical, written and verbal data, dynamic hypotheses were inductively created, implemented in the simulation model and then deductively tested and compared to the reference modes. Uncountable trial and error experiments were necessary to arrive at a model structure that appeared to be robust and that was able to reproduce the behaviour of interest both accurately and for the right reasons. Homer (1996) refers to such iterative processes of induction and deduction as *scientific modelling*.[12] The external validity of the model was examined as outlined at the beginning of this section. The results speak in favour of some degree of generalizability in our findings.
- *Reliability*. In the context of theory evaluation, reliability can be understood as the extent to which the consequences of a test of the theory remain constant, if repeated under identical conditions. The product launch model is a set of deterministic difference equations. It follows that reliability is warranted since re-tests of the theory will yield identical outcomes. However, reliability can also relate to the process of theory building and its replication. A major provision for increasing the reliability of case studies is to thoroughly document the process of conducting research (Yin, 2003, pp. 38 and 105), in our case, the development of the simulation model. We documented our simulation model in such a way that an external observer is able to trace the model development from the conceptualization phase until the final version.
- *Fruitfulness*. The question here is whether the theory-building process has led to important

[12]As we have observed, this iteration principle is applicable to the components of the whole process, for example issue and purpose formulation, structural validation, etc.

Syst. Res. 25, 447–465 (2008)
DOI: 10.1002/sres

insights and ultimately to the creation of new knowledge. We are looking at the fruitfulness of a theory in terms of its heuristic power, as defined earlier in this paper. In order to increase the heuristic power of the theory, we tried to broaden the base from which the interpretation and conceptualization work was done: the independent researchers and the members of the client team held several group model-building sessions in which the conceptualization and formulation of the model was conducted. Throughout the modelling process, and especially towards the official end of the project, the model was used extensively to explore what-if situations in order to draw inferences and explore management implications. These were discussed in the group setting. Several new insights emerged. One of them concerned the role of product assembly time. This had been assumed to have an insignificant effect on company sales. The model gave a different answer. Product assembly time was the most important factor, the impact of which even surpassed product attractiveness. An examination on the basis of company data corroborated this counterintuitive result of our simulations. A second insight, as previously mentioned, concerned the severe impact of competitors' mode of product launch on the development of company sales. These insights were new and important knowledge for the client. Finally and perhaps most important, our model integrated several partial theoretical components (from production theory and marketing) concerning product launches, which had existed to date, in a more comprehensive view of the subject under study (see also above: *comprehensiveness*).

- *Practicality*. Practicality is given, if a theory provides a conceptual framework for practice. The project team in our case consisted of researchers and responsible managers of the client company. Thus, a prerequisite for a theory under development to be of practical relevance was established. Furthermore, the members of the client company were included in each modelling activity. By these means, it was ensured that the model incorporated concepts that both explain the dynamic behaviour of interest and can to a large degree be influenced or manipulated by the decision makers. The design of the process of theory formation was critical. The process generated a theory that provided decision makers with a practically relevant framework of the essential control and policy variables.

To summarize this account, we have plotted our evaluations per criterion on a four-point measurement scale (Figure 6). The scale reaches from a degree of non-fulfilment to one of complete fulfilment, which corresponds to the ideal state. The intermediate values are relative to the ideal state. After filling in our evaluation form, we asked the two gatekeepers independently to give their assessments. The respective profiles of the three parties are represented in the chart: gatekeeper one (triangular marks), gatekeeper two (square marks), authors (round marks).

If we analyse the grading, we find that uniformity of evaluation is given in only one instance, the *importance* of the model, which is rated as high ('rather strong') by all three parties involved. Four other criteria are rated nearly equally by the authors and one of the gatekeepers—*precision / clarity*, *parsimony/simplicity*, *fruitfulness*, *practicality*. The ratings of all four stand for 'rather strong', as well. With regard to the criteria of *operationality*, *validity* and *reliability*, the rankings of all three parties differ from one another slightly. However, the relative levels of the three parties' rankings compared to one another remain fairly constant across the three criteria. We should not interpret these curves excessively, however. The questionnaires were filled in by the two gatekeepers individually, aided only by a written explanation of the 10 criteria. On that basis errors of measurement cannot be excluded.

It is not possible to fulfil all of the 10 criteria to the maximal possible degree since some of them exhibit tradeoffs; for instance, *parsimony and simplicity* and *comprehensiveness*'. While the first favours theories with as few assumptions as possible, the second tries to widen the scope of the theory to include different subject areas and more detail. In the case under study, a good balance between the degree of achievement of each criterion for itself and in combination has in our view been attained.

The elaborated theory is more general than it might appear at first glance. It refers to more than was expected at the beginning. At the outset, the purpose of the project was to build a model that would support short-term decision making about product launches. The outcome was a model which provided decision makers with more than that: a theory about the interrelationships between the factors essential to a class of situations and their longer term implications.

The modelling procedure was one of scientific modelling, which 'is distinguished from other approaches largely by the quality of evaluation and revision performed and by an insistence upon empirical evidence to support hypotheses and formulations' (Homer, 1996, p. 1). The final model falls into the category of 'canonical situation models', being fully formulated and calibrated. It is based on a case study reduced to its essentials, so as to enable the causal explanation of the dynamic behaviours generated by the underlying structure (Lane, 1998, p. 937). We may assume that it is applicable to a whole class of systems, but we do not suggest that it proposes universal laws (*cf*. Lane, 2000). The mature model can be considered a theory,[13] not the model in its pre-mature stages. It is an integrative theory, which builds on components, for example the Bass diffusion model and the learning curve.

CONCLUSIONS

The case study presented here in detail can be considered a large successful modelling venture. It generated concrete benefits for real organizations. The outcome of the modelling process is a theory that is a structured, explanatory, abstract and coherent set of statements about a partial reality. In the modelling venture drawn upon here, SD was the methodology used. The case illustrates an evolutionary concept of theory development. The emphasis is on theory building, not on static theorizing. Here, theorizing becomes an interaction between modellers and model—a dialogue through which the theory is created and enhanced.

The research question posed at the outset can be answered affirmatively: the methodology of SD does enable the creation of high-quality theories. A condition is that the process abides by the rules of good modelling practice, model validation in particular (*cf*. Sterman, 2000).

In the following, we summarize our insights and formulate a set of recommendations. These do not constitute a theory, but heuristic principles for model-based theory building. We condense them into eight points which—in the light of the experiences from an 11-year period—appear to be very important:[14]

1. *Issue orientation*: theory building with SD should not aim primarily at 'grand theories', which try to capture all aspects of a system.[15] It is more effective to strive for a holistic model of a complex issue of interest and to do it with a clear purpose. The need to solve a problem or to understand a system better can be the motivation for the model-building venture. Such issue orientation[16] has a much higher potential for achieving a substantive, operational, local theory with heuristic power, than a broad orientation which explains 'everything and nothing'. However, there is also an often viable intermediate way—the development of special theories applicable to classes of systems or limited conceptual ranges (*cf*. Merton, 1968, p. 51). SD appears to be very appropriate for achieving this kind of middle-range theory.
2. *Formalization*: build theories on the basis of formal models. These enable appropriate testing of emerging theories. The use of formal models also increases the likelihood that progress is made in the theory-building process, because systematic testing enhances improvements via selection. In contrast,

[13]With the attribute *mature* we denote a model that is extensively validated and which largely fulfils the criteria for the assessment of theories as presented above.

[14]We assume most of these points could also be transferred to other methodologies of model building, for example agent-based modelling.

[15]No theory or model is 'complete for *all* purposes' (Putnam, 1981, p. 147).

[16]Often also called 'problem orientation' (see, e.g. Sterman, 2002).

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

refraining from formalization entails the danger of logical flaws and lack of precision. SD can achieve more than building pragmatic models; it could and should be used increasingly for theory building with the help of formal models.
3. *Generalization*: generalization is a relevant principle of theory building even if one does not aim at a general theory. Also, middle-range theories as well as local theories should strive for insights which are generic, that is which can be transferred to other situations of the same type. This way, external validity is enhanced, which also means that the theory attains higher value. As a rule, be as general as possible and only as specific as necessary. This implies an appeal to the system dynamicist to avoid excessive detailing, despite the ease of carrying it out.
4. *Validation*: The quality, and with it the heuristic power of a model, is a function of its validity. Validity is the degree to which a model represents what it is supposed to represent. In other words, it is the degree to which the operational definitions of the variables and functions of the model accurately reflect the variables and functions they are designed to represent.[17] Be meticulous about model quality. Strive for improving the validity of your model continually in order to enhance the correspondence between the evolving theory and the reality it is supposed to explain (or 'create'[18]). A thoroughly developed set of theory tests to support theory building is already available as part of the SD methodology.
5. *Explanation*: for theory building, the reproduction of a behaviour under study is not enough. Beware of mechanical validation. Make sure that the model does not only reproduce but also enables the explanation of the behaviour under study. A model by itself is not a theory yet; 'what is required is a model along with a plausible account of why the model produces the behaviour that it does' (Lane, 2008).
6. *Falsification*: the touchstone of a theory is falsification. The status of a theory can be claimed only if serious attempts to falsify it have been undertaken and foundered. To be able to meet that criterion, a theory must be formulated in a way that it is falsifiable. From that view, which represents the critical-rationalist stance, justifying hypotheses is not the proper way to build a theory; the path leads over to attempts at falsification. Formal models are probably the most powerful devices for systematic trials to falsify a theory; moreover, from a theory-building perspective that is their primary role. If system dynamicists adhere to falsification trials, thereby avoiding the confirmation trap, they are more likely to fully use the potential of SD for theory building.
7. *Process design*: the design of the modelling and theory-building process is crucial for the quality of the theory which can result. The trial of falsification should be iterative, triggering the selection of hypotheses and theories. As has been emphasized in SD research for many years, group model building is a key to process design. Internal and external agents have to cooperate. Agents with different disciplinary backgrounds should synergize. An aspect introduced here is that, thanks to the SD methodology, even young modellers with limited experience can be leveraged as full participants in the process, if they are properly coached.
8. *Concept of learning*: theory building is inherently an operation with the characteristics of a learning process.[19] It consists of a sequence in which better theories successively replace their weaker predecessors. Learning is a progression of states with an ever higher potential for effective action (Kim, 1993). Model and theory building can benefit enormously from an arrangement of actor-centred learning, by which the actor is an agent searching to mobilize his or her potentials in

[17]A validity of 100% is, in principle, unachievable for models of social systems (*cf*. Sterman, 2000).
[18]Models do not only represent what is, but they can also play a significant role in the construction of a new reality (*cf*. the constructivist position, e.g. von Foerster, 1984).
[19]Also, modelling has been conceptualized as a learning process ('modelling-as-learning') by Lane (1994).

order to learn. In addition, an SD-supported theory-building process probably has the potential for becoming one of the most effective vehicles for organizational learning (*cf*. Senge, 2006).

Among these principles, the one which is perhaps most often debated is formalization. There appears to be a widespread prejudice against the building of formalized models. It is considered too costly, cumbersome, complicated, etc. To take one indicative example: according to a study among German companies, the interviewees consider IT-supported simulation methods as the number-one success factor in strategy development. However, only less than 5% of the responding firms use simulation in the formation and evaluation of strategies, and only 20% in strategic analysis.[20]

The cases we have referred to here show that the bias against formal modelling is mistaken. In the light of the Conant–Ashby theorem, high-quality modelling is an imperative. Full-fledged modelling is possible in a disciplined way, even in small ventures, generating high value for the firms that engage in it. We maintain that building local or middle-range theories is not a task limited to researchers in universities and laboratories. It should also be an integral component of managerial practice.

The construction of theories is ultimately a natural activity that any organism needs in order to be viable. It enables both understanding of the world and living in it. It provides prospect and enables robust decisions. It is the basis for intelligent action and reaction. The best managers are theoreticians in this sense; they rely on model-based theory building for the benefit of their organizations. This leads to a paradoxical conclusion: *managers can be superior practitioners if they become better theoreticians.*

If it is compared to theory-building processes without underlying structural models, SD modelling is shown to enable more disciplined and consistent argumentation. It offers a language which, despite its high precision, impresses by its broad applicability and richness of expression. This idiom is more precise than spoken language, but not quite as rich. SD in its quantitative forms is a type of mathematical modelling with great accuracy. In its qualitative forms ('qualitative SD') it is less precise than many mathematical approaches, but it can greatly facilitate the identification of the essential variables of a complex system. Altogether, SD can be qualified as more powerful in addressing complex phenomena than the analytical methods of mathematics.[21] A rigorous comparison between them could well be the object of further research.

[20]The interviewees from 112 companies were staff members of Corporate Development und Strategy units, as well as board members and their assistants (Behnan *et al*., 2002).

EDITORS' NOTE

The ideas in this paper are explored further in the following Discussant's Comments piece.

ACKNOWLEDGEMENTS

The authors extend their thanks to Professor Dr. Andreas Größler for his helpful comments on this paper, presented during the Third European System Dynamics Workshop at the University of St. Gallen, in March 2007, and two anonymous referees for their valuable suggestions.

REFERENCES

Ashby WR. 1956. *An Introduction to Cybernetics*. Chapman and Hall: London.
Bacharach SB. 1989. Organizational theories: some criteria for evaluation. *Academy of Management Review* **14**(4): 496–515.
Barlas Y. 1996. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review* **12**(3): 183–210.
Beer S. 1966. *Decision and Control*. Wiley: Chichester.

[21]Ultimately, the systems approach emerged in the face of growing rates of change and complexity in social and socio-technical systems (*cf*. Beer, 1966; Rapoport, 1986; Schwaninger, 2006). It provided conceptual breakthroughs such as the ideas of feedback, mutual causality and self-reinforcing processes (Richardson, 1999; Hammond 2003, p. 15) which have proved highly effective in dealing with complex phenomena. The formal mathematical modelling approach, which is much used, for example in economics, is actually appropriate for static, homogeneous, equilibrating worlds (Miller and Page, 2007, p. 20).

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

Behnan M, Gilbert D, Luber T. 2002. *Strategic Issues in German Industries: Ergebnisse der Tracking-Studie zur Evaluierung von Strategieprozessen in deutschen Top-Unternehmen*. The Galileo Consulting Group KG: Ingelheim am Rhein.

Bell JA, Senge PM. 1980. Methods for enhancing refutability in system dynamics modeling. In *System Dynamics. TIMS Studies in the Management Sciences*, Vol. 14, Legasto AA, Forrester JW, Lyneis JD (eds). North Holland: Amsterdam; 61–73.

Checkland P. 1981. *Systems Thinking, Systems Practice*. Wiley: Chichester.

Conant RC, Ashby WR. 1981. Every good regulator of a system must be a model of that system. In *Mechanisms of Intelligence. Ashby's Writings on Cybernetics*, Conant R (ed.). Intersystems Publications: Seaside, CA; 205–214.

Cowles M. 2001. *Statistics in Psychology: An Historical Perspective*. Erlbaum: Mahwah, NJ.

Davis JP, Eisenhardt KM, Bingham CB. 2007. Developing theory through simulation methods. *Academy of Management Review* **32**(2): 480–499.

Dopfer K, Potts J. 2008. *The General Theory of Economic Evolution*. Routledge: London.

Forrester JW. 1961. *Industrial Dynamics*. MIT Press: Cambridge, MA.

Forrester JW. 1968. Industrial dynamics—a response to Ansoff and Slevin. *Management Science* **14**(9): 601–618.

Forrester JW, Senge PM. 1980. Tests for building confidence in system dynamics models. In *System Dynamics*, Legasto AA Jr, Forrester JW, Lyneis JM (eds). North-Holland: Amsterdam; 209–228.

Grove WM, Meehl PE. 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. *Psychology, Public Policy, and Law* **2**(2): 293–323.

Hammond D. 2003. *The Science of Synthesis: Exploring the Social Implications of General Systems Theory*. University Press of Colorado: Boulder, CR.

Hannan MT, Pólos L, Carroll GR. 2007. *Logics of Organization Theory. Audiences, Codes, and Ecologies*. Princeton University Press: Princeton and Oxford.

Hanneman RA. 1988. *Computer-assisted Theory Building. Modeling Dynamic Social Systems*. Sage Publications: Newsbury Park, CA.

Holton EF, Lowe JS. 2007. Toward a general research process for using Dubin's theory building model. *Human Resource Development Review* **6**(3): 297–320.

Homer J. 1996. Why we iterate: scientific modeling in theory and practice. *System Dynamics Review* **12**(1): 1–19.

Karlöf B, Lövingsson FH. 2006. *ReOrganisation*. Springer: Berlin.

Kim D. 1993. The link between individual and organizational learning. *Sloan Management Review* **35**(1): 37–50.

Lane DC. 1994. Modeling as learning: a consultancy methodology for enhancing learning in management teams. In *Modeling for Learning Organizations*, Morecroft JDW, Sterman JD (eds). Productivity Press: Portland, OR; 85–117.

Lane DC. 1998. Can we have confidence in generic structures? *Journal of the Operational Research Society* **49**: 936–947.

Lane DC. 1999. Social theory and system dynamics practice. *European Journal of Operations Research* **113**: 501–522.

Lane DC. 2000. Should system dynamics be described as a 'hard' or 'deterministic' systems approach? *Systems Research and Behavioural Science* **17**: 3–22.

Lane DC. 2008. Formal theory building for the avalanche game: explaining counter-intuitive behaviour of a complex system using geometrical and human behavioural/physiological effects. *Systems Research and Behavioural Science* **25**(4): 521–542.

Lane DC, Smart C. 1996. Reinterpreting 'generic structure': evolution, application and limitations of a concept. *System Dynamics Review* **12**(2): 87–120.

Merton RK. 1968. *Social Theory and Social Structure* (enlarged edn). Free Press: New York.

Merton RK. 1957. *Social Theory and Social Structure*. Free Press: New York.

Miller JH, Page SE. 2007. *Complex Adaptive Systems. An Introduction to Computational Models of Social Life*. Princeton University Press: Princeton and Oxford.

Neisser U. 1976. *Cognition and Reality. Principles and Implications of Cognitive Psychology*. Freeman: San Francisco, CA.

Paich M. 1985. Generic structures. *System Dynamics Review* **1**(1): 126–132.

Patterson CH. 1986. *Theories of Counseling and Psychotherapy* (4th edn). Harper and Row: New York.

Popper KR. 1972. *Objective Knowledge*. Clarendon Press: London.

Popper KR. 2002. *The Logic of Scientific Discovery*. Routledge: London.

Putnam H. 1981. *Reason, Truth and History*. Cambridge University Press: Cambridge, MA.

Randers J (ed.). 1980. Guidelines for model conceptualization. In *Elements of the System Dynamics Method*. Productivity Press: Cambridge, MA; 117–138.

Rapoport A. 1986. *General System Theory. Essential Concepts and Applications*. Abacus Press: Turnbridge Wells, Kent/Cambridge, MA.

Richardson GP, Pugh AL. 1981. *Introduction to System Dynamics Modeling With DYNAMO*. Productivity Press: Cambridge, MA.

Richardson GP. 1999. *Feedback Thought in Social Science and Systems Theory*. Pegasus Communications: Waltham, MA.

Ritchie-Dunham JL. 2002. Balanced scorecards, mental models, and organizational performance: a simulation experiment. *PhD Dissertation*, University of Texas at Austin.

Schwaninger M. 2003. Modeling with archetypes: an effective approach to dealing with complexity. In

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres

Computer Aided Systems Theory—EUROCAST 2003, Moreno-Díaz R, Pichler F (eds). Springer: Berlin; 127–138.

Schwaninger M. 2006. The evolution of organizational cybernetics. *Scientiae Mathematicae Japonicae* **64**(2): 405–420.

Schwaninger M, Hamann T. 2005. Theory-building with system dynamics. In *Computer Aided Systems Theory—EUROCAST 2005*, Moreno-Díaz R, Pichler F, Quesada-Arencibia A (eds). Springer: Berlin; 56–62.

Senge PM. 2006. *The Fifth Discipline. The Art and Practice of the Learning Organization* (new edn). Doubleday: New York.

Skinner Q (ed.). 1985. *The Return of Grand Theory in the Human Sciences*. Cambridge University Press: Cambridge, MA.

Sterman JD. 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* **10**(2): 51–66.

Sterman JD. 2000. *Business Dynamics. Systems Thinking and Modeling for a Complex World*. Irwin McGraw-Hill: Boston, MA.

Sterman JD. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review* **18**(4): 501–531.

Stouffer SA, Guttman L, Suchman EA, Lazarsfeld PF, Star SA, Clausen JA. 1950. *Measurement and Prediction*. Wiley: New York.

Vennix JAM. 1996. *Group Model Building: Facilitating Team Learning Using System Dynamics*. Wiley: Chichester.

von Foerster H. 1984. *Observing Systems* (2nd edn). Intersystems Publications: Seaside, CA.

Wolstenholme EF. 1990. *System Enquiry—A System Dynamics Approach*. Wiley: Chichester.

Wolstenholme EF. 2003. Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review* **19**(1): 7–26.

Yin RK. 2003. *Case Study Research: Design and Methods* (3rd edn). Sage: Beverly Hills, CA.

Zagonel AA. 2004. Reflecting on group model building used to support welfare re-form in New York State. *Doctoral Dissertation*, SUNY-New York State University, Albany, NY.

Syst. Res. **25**, 447–465 (2008)
DOI: 10.1002/sres