VALIDATION IN SOFT OR, HARD OR AND SYSTEM DYNAMICS: A CRITICAL COMPARISON AND CONTRIBUTION TO THE DEBATE

Özge Pala, Jac A.M. Vennix, Jack P. C. Kleijnen
Tilburg University, The Netherlands
Tel: ++ 31 13 466 22 62
O.Pala@kub.nl
Nijmegen University, The Netherlands
Tel: ++ 31 24 361 62 91
J.Vennix@maw.kun.nl
Tilburg University, The Netherlands
Kleijnen@kub.nl

Abstract

This paper deals with the validation ideas of three fields, namely hard OR, soft OR, and system dynamics. Validation is an important, yet a very controversial, issue on model based fields. Even though everyone seems to agree that the validation of models must be assessed, the viewpoints on what validation is and how it should be established differ widely. The purpose of this paper is to (a) describe the similarities and differences (both within and between) the three fields (b) to provide explanations for the differences, in order to (c) create mutual awareness and understanding, and (d) to increase our understanding of the issues surrounding validation. We will first start with the validation opinions of the three fields and then describe and explain the differences in these opinions and their reasons.

Introduction

Model building and simulation are accepted ways of analyzing and helping to solve managerial problems. Interestingly enough, although model building and simulation are accepted ways of doing research in MS/OR, viewpoints on the validity of models differ widely. Compare for instance the following statements:

- "to validate any kind of model means to prove the model to be true" (Naylor and Finger, 1967).
- "the absolute worth of the model can be no greater than the worth of its objectives... Validity, as an abstract concept divorced from purpose, has no useful meaning". (Forrester (1961, p. 116)
- Validation means "…the process of establishing confidence in the soundness and usefulness of a model with respect to its purpose." (Forrester and Senge 1980).
- "Model validation is not now an issue of great moment in the development of SSM" (Checkland, 1995).

Some seem to be convinced that the validity of a model can be proven, others seem to hold the opinion that a model's validity can never be demonstrated. Some even consider model validity not to be of great importance. An intriguing question is what accounts for these differences in opinion. The purpose of this paper is to (a) describe the similarities and differences (both within and between) the three fields, namely hard OR, soft OR, and system dynamics, (b) to provide explanations for the differences, in order to (c) create mutual awareness and understanding, and (d) to increase our understanding of the issues surrounding validation.

The paper starts with a discussion of the validity opinions in the three fields. From this discussion it will become clear that there is a great deal of confusion with regard to the concept of validity. Next differences and similarities will be described and explanations for differences of opinions between the three fields will be offered.

Validation in hard or classical OR

Particularly within the field of hard OR there has been quite some debate on validation (cf. Ackoff, 1956; Naylor and Finger, 1967; Fishman and Kiviat, 1968; Van Horn, 1971; Shannon, 1975; Landry et al., 1983; Gass, 1983; Miser, 1993; Sargent, 1994; Balci,1994; Landry et a. 1996).[i] Several authors have provided historical accounts of the developments in the field with regard to validation (e.g. Dery et al., 1993; Landry et al, 1983). In this paper we will not take a historical perspective, but rather discuss a number of viewpoints with regard to validation which have recurred throughout the last decades. These relate to: validity as accuracy of representation, validity as confidence in the model, validity in relation to the model's purpose, validity as usefulness.

Validity as accuracy of representation

A central viewpoint is validity as correctness of representation of reality. In this viewpoint a model is seen as a partial representation of reality and validity as the degree of likeness or accuracy of representation of reality in the model.

The most notable representatives of this idea are Naylor and Finger, whose three stage model of verification still figure prominently in most OR textbooks (Winston 1991). The basic idea is: "to validate any kind of model means to prove the model to be true" (Naylor and Finger, 1967). In their view, any relationship or data which have not been subjected to empirical verification is meaningless. Over time the view on validation got influenced by other philosophical schools. But still the central theme is accuracy of representation, and although in recent years concepts of usefulness have entered the discussion, the idea of representation is still predominant.

When it comes to establishing the degree of representativeness most authors consider comparison of model predictions with data from the real system as the ultimate validation test (Ackoff, 1956, Naylor and Finger, 1967; Fishman and Kiviat, 1968; Gass,1983; Balci, 1994). In Naylor and Finger's multi-stage procedure the final decision concerning the validity of the model is based on its predictions, i.e. the ability to reproduce historical data.

Van Horn (1971) however explicitly points out that the objective is not to prove the model to be valid, but rather to '…validate a specific set of insights not necessarily the mechanism that generated the insights.' (p. 248).

Validation as process and types of validation

Sargent (1982) and Landry et al. (1983) point out that model building and validation ought to be intertwined into a model validating process based on the various stages in the process of model building: the problem situation, the conceptual model, the formal model, and the solutions and recommendations. A problem situation consists of some aspects of the real world that are problematic. The conceptual model, which is "the coherent 'mental image' of the problem", indicates the objectives to be reached, and the relationship between the elements of interest. It is a result of the perceptions and value judgments of both model builders and decision-makers. The formal model is a translation of the conceptual model into mathematical symbols and/or computer codes. The purpose is to experiment with the problems and/or to obtain solutions (satisfactory or optimal) for formulating recommendations. The solution can be considered as the output of the modeling-validating process. It is obtained from the formal model using solution techniques and procedures. Then a decision is taken. A decision is a conclusion to which solution or recommendation will be implemented.

Each of these interrelated stages has a different type of validity associated with it. Below, we will explain the different validation stages, stated primarily by Landry et. al., but we will also discuss how Sargent sees the same stages. This, we believe, might also be helpful in seeing the differences within the field.

Conceptual Validation is concerned with testing the correctness or relevance of the theories and assumptions underlying the conceptual model of the problem for the intended purpose. According to Sargent, this should be done by using mathematical analysis and statistical methods. Even though Landry et. al. define the conceptual model in the same way, we think they have a different way of looking at it. According to them, the aim of this validation is to ask questions like "Are we looking at the problem situation from the appropriate perspective? To what extent are the constructs representative of the situation as perceived by the actor?". This is a way to make sure that the elements and relationships have been included as the users of the model see them and also to agree on the tools and techniques that are available for use in the coming phases.

Interestingly and understandably enough Landry et. al. point out that it is quite possible to come up with different models for the same situation depending on the backgrounds of the users and different epistemological approaches taken by the model developers and users. If one accepts such a point of view, then proposing a generally accepted validation procedure turns out to be very hard. A subjective, situation-dependent approach would be hard to accept for the core OR areas where things are done in a more defined, 'scientific' way.

Logical Validation is concerned with the capacity of the formal model to describe correctly and accurately the problem situation as described in the conceptual model. It should be checked whether any important variable or relationship is excluded from the model. This type of validation also includes verification, i.e. checking whether the computer model is programmed correctly. Actually Sargent calls this step 'computerized model verification'.

Experimental Validation refers to the quality and the efficiency of the solution mechanism. Hard OR gives many solution techniques to its users. Sometimes, however, the theoretically most appropriate solution technique may not be the best choice when it is considered in terms of the requirements of time, data, effort, and cost. Experimental validation deals with the efficiency of obtaining a solution, and the sensitivity (robustness) of the solution to the changes in the model's parameters. The general approach is to find the solution, devise an experimental design, and perform sensitivity analysis.

Operational Validation determines the quality and the applicability of solutions and recommendations with respect to the intended user and the problem situation. Actually, operational validity can be seen as assessing the usefulness of the model, because as a result information will be generated to help the users to accept or reject the results. Sargent mainly suggests comparison of the real and model generated data.

Data Validation evaluates the appropriateness, accuracy, sufficiency, completeness, unbiasedness, and availability of the data necessary for the problem solving process.[ii]

Validation and context / model purpose / problem type

One line of thought in the debate on validity of models is concerned with the purpose of the model. Landry et al. point out that the "…validity of a model is something to be considered within a context. Without a context, the concept of validity has no ground on which it can stand. (p. 219). Sargent (1994) and Balci (1994) provide similar views and emphasize validation with respect to purpose. Their definition of validation is "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model"**.** As we will see in the discussion on validation in system dynamics this is one of the core issues.

Based on the model by Landry et al., Oral and Kettani (1993) developed their tetrahedron model. It is actually an extended form of the Landry et. al.'s model, in the sense that it extends the model into three dimensions, the extra dimension being different types of OR problems. Each type of problem is put on one facet of the tetrahedron. Oral and Kettani argue that different types of problems necessitate focus on different phases of model building. Thus, the validation types which are required differ from one problem type to another. As Oral and Kettani point out: "This is rather a natural suggestion since the process of validation is dependent very much on the nature of the

OR problem being considered. One cannot expect to pursue exactly the same validation approach for all facets of the tetrahedron simply because the basic objectives of modeling vary from one facet to another in considerable ways." (Oral and Kettani, 1993, p. 232).

The authors distinguish four facets.

- The prototype Facet: Well conceived and known types of problems that do not need too much further conceptualization. Therefore the conceptual model is less important in these type of problems. For example traveling salesman, vehicle scheduling etc.
- The theoretical Facet: This deals more with the formation of abstract concepts and generalizations, and less with a current and immediate managerial situation. Thus, there is no problem situation of concern. Basic research aspects of OR can be considered in this category.
- The descriptive facet: the emphasis is here more on '…understanding the system or organization in which managerial situations arise.' (p. 220) The authors place system dynamics on this facet.
- The pragmatic facet: this facet represents the situation where the formal model plays a secondary role. This situation comes closer to situations with which researchers in soft OR are confronted.

Validity as usefulness

Landry et. al. (1983) point out that during the early years of OR, the validation issue included ideas like usefulness, usability, representativeness, and cost considerations, whose relative importance varied. In the following years, model validity lost its usefulness aspect and came to mean only representativeness, meaning "…'the extent to which the model fits the real system' either in terms of structure and mechanism or in terms of output, depending on the context of the problem." (p. 207). OR turned out to be technique-oriented rather than problem-oriented. OR people grew further away from the problem owners. Meanwhile, towards the 1970s, the problem types changed into socio-economical ones. This led to an unsatisfactory result: many models were built, but so few were used (Tobin et. al. 1995). Gass (1981) argues that the reason for the failure was that the analyst could not make the users believe in the credibility of the results obtained.

In recent years the idea of usefulness enters again in the discussions on model validation. This is obvious in the tetrahedron model of Oral and Kettani. It is even more obvious in Landry et. al. (1996) who focus on model legitimization in operational research. These ideas come quite close to things like organizational platform for change.

Summary

From the viewpoints of authors in hard OR we may identify a number of shared ideas:

- Validation should be integrated into the process of model-building
- The meaning of validation depends on the problem situation and model purpose
- It checks consistency with the real system, primarily by checking the model's predictions against the behavior of the real system.

- Emphasizing the use of statistical methods.
- It somehow includes the idea of usefulness, although there seems to be an implicit distinction between scientific and useful models. It seems as if some authors assume that scientific models cannot be useful and vice versa.

Validation in System Dynamics

System dynamics is a theory on the relationship between structure and behavior of social systems. Its most important core assumptions are:

- Social systems are information feedback systems;
- (Feedback) structure drives behavior, but the human mind is ill-equipped to trace out the dynamic consequences of a complex feedback structure. As Forrester (1995) formulates it: "The human mind is not adapted to interpreting how social systems behave...A computer model can reliably determine the future dynamic consequences of how the assumptions within the model interact with one another... Generally, behavior is different from what people have assumed".
- Hence, mathematical models are required and simulation becomes necessary, because the complexity of these differential equation models impedes analytical solutions.

SD is a method for translating mental models of users into a computerized model.[iii] Since the purpose of system dynamics modeling is to help the design of an improved system (Forrester, 1961, 115), the appropriate setting of the purpose/objectives of the modeling process is important.

As was the case with hard OR, in the literature on validation in system dynamics there are a couple of important issues: validation and model purpose (usefulness), validation as process and validation as confidence.

Validation and purpose

Understandably view on the importance of model purpose also affects the validation perspectives of system dynamicists. Validation takes on the meaning of effectiveness in improving the real system that is modeled, rather than absolute ‘correctness’. A model should be a "good enough" representation of reality, depending on the goal of the model. This point of view coincides with some hard OR researchers (see above).

System dynamicists believe that it is meaningless to judge the validity of a model without a clear view of model purpose (Richardson and Pugh, 1981). The appropriate level of detail, i.e. system boundaries, system variables, and the assumed system interaction among the variables, is determined depending on the purpose and objectives that are set beforehand. Forrester (1961, p. 116) states the importance of objectives as follows "the absolute worth of the model can be no greater than the worth of its objectives... Validity, as an abstract concept divorced from purpose, has no useful meaning". Any criticism on the model should keep in mind the focus, boundaries and purpose for which the model was built.

Validity with respect to purpose requires that the internal structure represent those aspects of the problem that are relevant to the objectives (Barlas, 1996). Being causal-descriptive (white-box) models by their nature, producing the right behavior is not the only concern; the model should also explain how the behavior is generated, i.e. the dynamic behavior should be a consequence of the system structure. Therefore, "validity" means validity of the internal structure of the model, not only its output behaviorIn system dynamics this principle is known as "the right behavior for the right reason". This is where the purpose of the model becomes important.

We mentioned two important aspects with respect to validation, namely a sound structure and purpose of the modeling study. In the light of this, the question "is the model valid?" can be decomposed into two relevant questions (Richardson and Pugh, 1981):

- "Is the model suitable for its purpose?" and
- "Is the model consistent with the slice of reality it tries to capture?"

Consistency with reality means that there must be a correspondence between the parameters and the structure of the model and those of the real system. That is, the mechanisms that generate the dynamic characteristics in the model must be the same as those in the real system. Only in this way the model can generate the nature of dynamic characteristics that are of interest for the problem at hand. So the confidence depends on how acceptable the model is as a representation of the organizational and decision-making details of the actual system. And it is confirmed by the correspondence of the total model behavior to that of the actual system (Forrester, 1961).

Richardson and Pugh also see validity from the utility and effectiveness points of view: can the model and the results be used? So in a sense they define the degree of confidence as the degree to which the model helps to generate insights, enhances understanding, and influences its audience. They state as follows "if somewhere along the way in a modeling study the feeling grows that new insights have been obtained, the model can hardly be judged invalid". This, according to them, makes subjectivity also an unavoidable aspect of validation, because it is the individual persons who will decide whether the model built is a useful one or not.

Validity as process and confidence

Some authors have taken validation to imply a certain degree of confidence in a model (cf. Shannon, 1975), which is built up during the model construction process. Most experienced system dynamicists would acknowledge that validation is not a one-time event. It is a process supported by a number of tests at the structural and behavioral characteristics. In addition validation is seen as gradual 'confidence building' Therefore, an accepted general definition of validation would be: "the process of establishing confidence in the soundness and usefulness of a model with respect to its purpose" (Forrester and Senge 1980, Barlas 1996).

Problems with validity as confidence:

- Gass (1981, 1983) argues that model confidence is not an attribute of a model, but of a particular user: it is an information-based opinion or judgment for a given decision environment. The level of confidence varies from user to user, due to differences in application requirements and subjective judgmental preferences. If we tie validity of the model to the confidence generated in the users, then a model can have different validity levels. Moreover, let's think of a model without a designated user who will judge the model. Since the priori confidence level is zero, and confidence is in the eyes of the user, does that mean that the validity level of this model will always remain at zero? One can argue that the model developers can judge whether to put confidence in the model. But then that would be judging the "truth" of the model with the same beliefs that have built the model, which is not an independent way of judging validity.
- What is the level of confidence that is required such that a model is judged to be "good enough"? How is confidence built?
- If confidence rather than validation is the issue, isn't it postponing the question of validity?

Validation Tests

According to system dynamicists model validation cannot be totally formal and quantitative, because it signifies usefulness with respect to a purpose. Therefore, it is important to test the validity, or usefulness, of the purpose and objectives, which are subjective entities. So objective, quantitative, and formal tools go together with subjective, qualitative, and informal tools. The subjective, qualitative ones are those that are dispersed throughout the modeling process, whereas the quantitative tools are used at the end of model construction, immediately before policy design simulations (Barlas, 1996).

The validation tests are grouped into two classes, namely structural and behavioral. Structural validity tests check whether the structure of the model matches the structure of the system being modeled (Shreckengost, 1985). Form of each equation, selection of system boundaries, variables and relationships between the variables must be tested (Forrester, 1961). All important factors of the real system must be represented in the model, and all elements of the model must have counterparts in the real system.

Barlas (1989, 1996) partitions the structural validity into two groups: direct structure tests and structurally-oriented behavior tests. Direct structure tests are strong tests since they evaluate the model structure directly. These tests do not use simulation. Basically, they take each relationship (mathematical equation or any form of logical relationship) individually and compare it with the available knowledge (empirical or theoretical) about the real system (Barlas, 1996). Structure-oriented behavior tests assess the validity of the structure indirectly, by applying certain behavior tests on model-generated behavior patterns. They do involve simulation, and can be applied to the whole model as well as to its isolated sub-models.

Once the structure of the model is believed to be adequate, one can pass on to behavior validity tests. These tests check if the model is capable of producing an acceptable output behavior. Output validation involves demonstrating that the model is able to reproduce the dynamic time patterns that have been observed in the behavior of the real system (Barlas, 1990). So, the crucial point is the emphasis on pattern prediction rather than point prediction. Barlas (1989) mentions five components present in complex behavior patterns: trends, periods of oscillations, phases of oscillations, average values, and amplitudes (variations). He also develops a multi step behavior validation procedure (Barlas 1989, 1996).

The importance of the model building process: Group Model-Building

Related to the discussion of usefulness and usability of models in system dynamics is the fact that almost from the start system dynamicists have been concerned with the participation of the client in the process of model construction. Over the decades this has given rise to an approach known as Group Model-Building. This can be linked with the implementation problem and the paper by Landry et al (1996) on legitimization.

Validation in Soft OR

Soft OR is a generic term to denote a number of methodologies aiming at resolving messy problems. Among the most well known are: SSM and SODA.

It is hard to talk about validation in soft OR, because validation is not an issue for the soft OR people. They do not believe that models can be checked on the basis of their representativeness of the real world (Checkland, 1995). A model cannot be a description of the world. In his paper on model validation in soft systems practice Checkland (1995) points out that the meaning of the word model in hard OR has the connotation of ‘model of part of the real world’. In SSM models are not considered to be models of part of the world but rather “…they are only *relevant to debate about* the real world and are used in a cyclic learning process. (p. 47). Checkland perceives models as entities that are relevant to debate about the world; entities that serve to clarify ideas, illuminate topics, and open up discussion points. Checkland calls SSM an explicit inquiring and learning system. So models in SSM are 'epistemological devices'. He says it clearly: "Models are not would-be descriptions of the world, and hence they cannot be tested by checking how well they represent the world, since this is something they do not purport to do".

According to Checkland, the validity question in SSM is whether a model is a good device for learning or not. There are two criteria: (i) the model should be relevant to the topic; (ii) a model should be competently built. The question of relevance is something to be answered during the modeling study. This means that, if the users learn during the study, then the model is judged to be relevant to the problem at hand. The question of competence is answered by asking the question 'is the model defensible?'. The measures of performance of the model must be linked to the world view it is expressing. If links between the world view of the users and the model can be demonstrated and defended, then the model can be regarded as good enough.

While Checkland states the above two questions of relevance and competence, he does not give any guidelines to any particular tests to judge the validity. Actually, he himself admits that model validation is not an issue of great moment in the development of SSM. If the model looks plausible to the model users and if the users get the feeling that they are learning something, then the model is assumed to be valid. The validity check of plausibility is basically done by face validity (even though Checkland does not use such a name). However, if we understand it right, the people who build the models, do these checks. Such a check is not an independent way of doing validation.

Cognitive maps of SODA are tools for reflective thinking and problem solving. In some cases they are developed by aggregating the maps from individuals; in other cases by building a map directly with a group, and still in other cases by inference from documentary evidence that relates to an organization or to an industry. Eden (1992) says that in some cases the individuals are involved in validating their own maps, and in others the link between data collection and map is managed solely by the researcher. Eden mentions that the process of establishing the status of a map involves exploring two things: (i) the link between the model used and its internal coherence; (ii) its adequacy in relation to purpose:

First, the link between the model and its internal coherence should be checked. SODA approach was developed to reflect the personal construct theory. Therefore, cognitive maps can be validated by checking the degree to which they conform to underlying ideas of construct theory. Moreover, the models should be amenable to transparent analysis. For more information on the analysis of the cognitive maps, interested readers can refer to Eden et. al. 1992.

The model should be a basis for cognitive negotiation with the group (Eden, 1992). Therefore, if the maps serve to move a group towards negotiation and commitment to action, then the maps can said to be valid. This is also true for SSM.

Sometimes, the cognitive maps of SODA are mixed with the causal-loop diagrams of SD. For SD, the aim is to form causal maps formed by feedback structures that dictate dynamics. However, the existence of feedback loops in a cognitive map of SODA can have two meanings. Firstly, the existence of a loop may be a coding accident that needs correction; secondly loops imply the possible existence of dynamic consideration within cognition. So as feedback loops SD models stem from the theory of SD, this is not the case for SODA. The existence of a feedback loop can be a possible error in coding. Thus, checking the existence of causal loops and their reasons can be regarded as a part of validation.

It is quite hard to place soft OR somewhere in this framework. Soft OR people do not take validation for serious. Their main focus is on helping the decision-makers clarify their perception about the problem situation and on increasing communication amongst the decision-makers. In this sense, we may talk about the usefulness aspect of the models, rather than the validity. Checkland (1995) puts this in words: "Model validation is not now an issue of great moment in the development of SSM". We think that this nature of soft OR methodologies has certain drawbacks.

Differences between the three fields

In this section we will deal with the differences in the problem type and modeling aims of the three fields. It is argued that these differences have implications for differences in their validation views.

Problem type

The most conspicuous difference between the three fields is the type of problem they attempt to help solve.

In most cases, hard OR problems are well-defined. For example, if the client is a supermarket owner, then he comes with a problem of, say, customer dissatisfaction with respect to the waiting times. The client knows what the problem is and has a possible solution (of e.g. increasing the number of servers) in mind. But how to do it, how much it would cost etc. are not known. Thus a modeling study is done to find the optimum number of servers to meet a certain amount of demand and its cost. This is a situation where 'what to do' is known but 'how to do it' is not known. In soft OR, however, 'what to do' is not known either (Checkland 1981, Eden 1982). According to Eden, clients do not want to use soft OR methodologies when the problem is well-structured and when the social relationships surrounding the problem are not important. He gives an example: "the client felt an unease about the way in which his business (recruitment and employment agency) was developing, and that he was having problems deciding how to improve the situation". SD is somewhere in-between hard OR and soft OR. It requires a problem definition from the beginning (reference mode of behavior). However, 'what to do' is not known as in hard OR. SD aims at understanding the causes of the problem and thereafter come up with 'what to do'.

As the aspects of the problem the model tries to capture change, the model characteristics also change. If one compares the problem domains of each field, one realizes that hard OR tries to capture tangible parts of a system. These are those that are easier to visualize and quantify. "Reality" becomes easier to define and represent. Moreover, using animations as a validation tool becomes possible. For example, looking at typical queuing models of discrete-event simulation, it is possible to represent a queue and servers. It is also easy to model a production line, and find the optimum number of products to be produced in order to meet a certain level of demand. However, if the implementation of this production line means that half of the employees will be fired and this introduces some political and personal concerns, then formulating a mathematical model becomes harder and in certain cases impossible. Such intangible, 'soft' parts, which cannot be easily quantified, are the elements Soft OR concentrates on.

As hard OR and soft OR concentrate on tangible and intangible parts respectively, SD takes a position in-between. SD has its own way of quantifying the intangible elements, named as soft variables. Quantification becomes possible because the exact values of the variables and the parameters do not have a vital importance as in hard OR. This decreases the importance of parameter estimation in SD model validation. This cannot be understood by hard OR people.

Purpose of modeling

As a consequence of dealing with different types of problems different purposes play a role. In SD, for instance, the aim is to have a general understanding of the source of any problematic behavior which may have been observed or feared. SD models are built in order to understand policies which produce the fundamental behavior modes of the system, that is, "the forces that determine the basic tendencies towards growth, fluctuation and decline" (Forrester, 1959, p.35). The aim is to come up with robust policies. Policy is a line of argument rationalizing the course of action. Barlas and Diker (1996) have developed an interactive simulation game based on a SD model to analyze a range of problems concerning a university administration system (growing student-faculty ratio, poor teaching quality, low research productivity, etc.) and to come up with certain policies for overcoming these problems. A policy question in this case can be whether concentration on graduate study rather than undergraduate has a positive effects on the amount and/or quality of research.

Hard OR people, on the other hand, are interested in decisions rather than policies. Referring to the above example, how many students to accept into graduate and undergraduate programs, or how much money to allocate to each department for this month would be questions for hard OR.

Aiming at policies or decisions has certain implications. The first one is the focus on pattern versus point prediction. Since the aim of SD modeling is finding robust policies and designing improved systems, the interest is in the behavior of the system over time: will a certain change improve or degrade the system behavior? Checking behavior means focusing on pattern. Hard OR, on the other hand, focuses on point-prediction, due to the interest in exact values. A second implication is the focus on parameter versus structure. SD's belief that the causes for problems can be found within the system structure makes the structural validity crucial. When emphasizing structure, SD deals less with parameter estimation, which is of paramount importance for hard OR. The difference is again the focus on long-term (policies) versus short-term (decisions) concerns: even though short-term values may be sensitive to parameter changes, long-term behavior of a system is insensitive to parameter adjustments. Large socio-economic feedback models, which are the concern of SD methodology, are pattern insensitive to a majority of parameter values. This is called the negative feedback compensation. Therefore, SD people head for robust structural changes.

Coming to soft OR, we can say that the aim is threefold: trying to understand the problem, creating agreement in a management team and last but not least, creating a platform for the resulting strategic decision. Soft OR methodologies do not start with a clear problem definition, whereas hard OR and SD do. Through employing their methodology they hope to find out what the problem is. Therefore, the primary aim of soft OR is problem identification. The only validation they seem to apply is face validity. Actually, this is not a big deal either, because in all the methodologies the models are built together with the clients. So actually, it is the client who builds the model and if something "wrong" is seen, it is corrected during the model building process.

Implications for validation

We have tried to show that these three fields deal with problems that have different characteristics. The differences between these fields are also shown in the table. We find it logical that different types of problems require different models, and different models should be validated in different ways! If SD did not differ from hard OR with respect to the aspects that hard OR criticizes SD, it would not be able to tackle the more complex problems (that they have originally intended to do) on which sufficient data are difficult to obtain and about which laws are difficult to formulate. One important implication of all three characteristics concerns the qualitative versus quantitative nature of validation. As the problem gets more ill-defined, deals with more intangible parts of a system, and gets larger in scope, the ambiguities surrounding the system structure increases. In hard OR, there is no uncertainty about the system structure. Thus, validation is mainly based on quantitative methods. SD faces more uncertainty about the structure. It is even more in the case for soft OR. Thus, qualitative validation increases as one goes from hard OR to SD and then to soft OR. For soft OR, quantitative nature of validation loses its meaning.

Table: Differences amongst hard OR, soft OR, and system dynamics

	Hard OR	Soft OR	System Dynamics
Problem Type	• Clearly defined problems: suitable for the application of prescriptive or descriptive techniques	• Problem is not clear at the outset, neither is the particular objective to be reached. • Systems that involve	• Problems that are dynamic and persistent in nature. • Complex, nonlinear, dynamic, multi loop

	• Those parts of the enterprise that have tangible forms	intangible parts of the enterprise, especially those involving human activity	feedback systems • Deals with both tangible (hard) and intangible (soft) parts of a system
Aim of Mo eling	• Problem *solving* • Experiment with the model and transfer results to the real world. Apply certain techniques and if possible optimize. • Reach decisions • Point prediction	• Problem *identification* • Clarify what the problems and weaknesses are in given circumstances. • Enhance users' understanding of the problem situation. • Does not predict anything.	• Identification of structure/behavior pairs • General understanding of the system structure • Policy re-design • Pattern prediction: *General dynamic tendencies* is of concern
Questions asked	• What system will meet *a given need?* • How to achieve (a *desired solution*)?	• What is the need? • What is the problem?	• Why does the given system cause problems and how can those be solved? • How can system performance be improved?
Model	• Models are seen as representations of the real world • Quantitative	• Models are used to debate about a problematic situation • Qualitative	• Models are seen as representations of the real world and are relevant for und standing the system structure • Qualitative and Quantitative

References

Ackoff, R. L. 1956. The Development of Operations Research as a Science. Operations Research 4 (3): 265-295.

Balci, O. 1989. How to Assess the Acceptability and Credibility of Simulation Results. Proceedings of 1989 Winter Simulation Conference, IEEE, Piscataway, NJ: 62-71.

Balci, O. 1994. Validation, Verification, and Testing Techniques throughout the Life Cycle of a Simulation Study. Annals of Operations Research 53: 121-173.

Banks, J., J. S. Carson. 1984. Discrete Event Simulation. Prentice Hall, Englewood Cliffs.

Barlas, Y. 1989. Multiple Tests for Validation of System Dynamics Type of Simulation Models. European Journal of Operational Research 42: 59-87.

Barlas, Y. 1996. "Formal Aspects of Model Validity and Validation in System Dynamics". System Dynamics Review 12 (3):183-210.

Barlas, Y. and V. Diker. 1996. Decision Support for Strategic University Management: A Dynamic Interactive Game. Proceedings of the 14th International System Dynamics Conference, Cambridge, MA. vol.1: 47-50.

Checkland, P. 1981. Systems Thinking, Systems Practice. Chichester, Wiley and Sons.

Checkland, P. 1995. Model Validation in Soft Systems Practice. Systems Research 12(1): 47-54.

Dery, R., M. Landry, C. Banville. 1993. Revisiting the issue of model validation in OR: An Epistemological view. European Journal of Operational Research 66: 168-183.

Eden, C. 1982. Problem construction and the influence of O.R. Interfaces: 50-60.

Eden, C. 1992. On the Nature of Cognitive Maps. Journal of Management Studies 29 (3): 261-265.

Eden, C., F. Ackermann, S. Cropper. 1992. The Analysis of Cause Maps. Journal of Management Studies 29 (3): 309-323.

Forrester, J. W. 1959. Advertising: A Major Breakthrough for System Dynamics. In Collected Papers of Jay W. Forrester. Cambridge, USA: Wright-Allen Press.

Forrester, J. W. 1961. Industrial Dynamics. Cambridge, Mass.: MIT Press.

Forrester, J. W. 1995. Counterintuitive Behavior of Social Systems. MIT System Dynamics Group Working Paper D-4468 (Original paper in Technical Review. January 1971. Alumni Association of MIT).

Forrester, J. W., and P. M. Senge. 1980. Tests for Building Confidence in System dynamics Models. In System Dynamics. Ed. A. A. Legasto, J. W. Forrester and J. M. Lyneis. Amsterdam: North-Holland.

Gass, S. I. 1977. Evaluation of Complex Models. Computers & Operations Research 4: 27-35.

Gass, S. I. 1981. Concepts of Model Confidence. Computers & Operations Research 8 (4): 341-346.

Gass, S. I. 1983. Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. Operations Research 31: 603-631.

Kleijnen, J. P. C. 1995. Verification and Validation of Simulation Models. European Journal of Operational Research 82 (1), April: 145-162.

Kleijnen, J. P. C. 1998. Validation of Simulation, With and Without Real Data. CentER Discussion Paper. No. 9822. Tilburg.

Landry, M., J. L. Malouin, M. Oral. 1983. Model Validation in Operations Research. European Journal of Operational Research 14 (3), November: 207-220.
Landry, M., C. Banville, M. Oral. 1996. Model legitimisation in operational research. European Journal of Operational Research 92, 443-457.
Naylor. T. H., J. M. Finger. 1967. Verification of Computer Simulation Models. Management Science 14 (2), B-92-101.
Oral, M., and O. Kettani. 1993. The Facets of the Modeling and Validation Process in Operations Research. European Journal of Operations Research 66: 216-234.
Richardson, G. P., and A. L. Pugh III. 1981. Introduction to System Dynamics Modeling With DYNAMO. Portland, OR: Productivity Press.
Rosenhead, J. (ed.). 1989. Rational analysis for a problematic world: problem structuring methods for complexity, uncertainty and conflict. Chichester, Wiley and Sons.
Sargent, R. G. 1994. Verification and Validation of Simulation Models. Proceedings of 1994 Winter Simulation Conference, IEEE, Piscataway, NJ, 77-87.
Shannon, R. E. 1975. System Simulation: the art and science. New Jersey: Prentice Hall.
Shreckengost, R. C. 1985. Dynamic Simulation Models: How Valid Are They? In Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity. Ed. B. A. Rouse, N. J. Kozel, L. G. Richards. U.S. Government Printing Office.
Tobin N. R., K. Rapley, W. Teather. 1995. The Changing Role of OR. In: Understanding the Process of Operational Research: collected readings. Ed. P. Keys. John Wiley & Sons, New York. (First published in Journal of Operations Research Society 1980)
Whitner, R. B., O. Balci. 1989. Guidelines for Selecting and Using Simulation Model Verification Techniques. Proceedings of 1989 Winter Simulation Conference, IEEE, Piscataway, NJ: 559-568.
Winston, W. L. 1991. Operations Research: Applications and Algorithms. PWS-KENT Publishing Company, Boston.

---

[i] The discussion on validity of models has given rise to a host of concepts, which are used in conjunction with validity: e.g. "significance", "confidence", "suitability", "consistency", "soundness", "usefulness", "effectiveness", "plausibility", "reliability", "credibility", "usable" or "acceptable". The word seems to be moulded so that it can get into any shape. We propose the following working definitions. Verification: Checking the internal consistency of the computer program. It is the process of making sure that the computer model is free of programming errors ("bugs"). This includes checking the dimensions of the parameters. Validation: Checking whether the model (formal or conceptual) is an accurate, good enough representation of reality; checking the correspondence with the real system. Usefulness: is the model suitable to the purpose it is built for; is it capable of being put to use, whether it serves for a beneficial end or object? Usability: is the model economically feasible to solve, use, and is it accessible to those who wish to use it?

[ii] Many validation techniques have been devised by OR specialists. The usage of these tests depend on the type of the model and the problem situation. These range from face validity or expert opinion, to parameter validity or sensitivity analysis, to extreme condition test, and various statistical tests. More thorough explanations of validation tests and usage of statistical measures can be found in Whitner and Balci (1989), Balci (1989, 1994), Kleijnen (1995, 1998), Shannon (1975), Landry et. al. (1983), Banks and Carson (1984), Gass (1977, 1983).

[iii] System dynamics has two levels of validity: the validity of a model and the validity of its worldview: social systems are feedback systems and structure drives behavior. The same holds for soft OR, but probably not for hard OR, since it has no theory. It seems that for system dynamics there are at least two questions which need to be answered with regard to

validation. The first question relates to the validity of a single model for a particular problem situation. The second relates to the validity claim with regard to considering social systems to be feedback systems. See also Landry et al. 211.