Mathematical Modelling, Vol. 1, pp. 123–139
Printed in the U.S.A.
0270-0255/80/010123-17$02.00/0

TESTABILITY AND MEANING OF MATHEMATICAL MODELS IN SOCIAL SCIENCES

H.-J. ZIMMERMANN

Lehrstuhl für Unternehmensfoschung
Institut für Wirtschaftswissenschaften
RWTH Aachen, Templergraben 64, 5100 Aachen, W. Germany

Communicated by X. J. R. Avula

(*Received January 1980*)

Abstract—The terms "model," "theory," and "law" are being used with a variety of meanings in practice as well as in theory. Depending on the definition of the contents of these terms they have to satisfy different requirements. The author focuses his attention on mathematical models and their use in Operations Research and Management Science. After pointing out the difference between "theories," "laws," and "models," criteria for the quality of models are defined and problems of model building discussed. It is considered particularly important that factual models—as opposed to formal models—are designed in such a way that they are testable. A factual model which is not testable can neither be confirmed nor falsified and can therefore never claim to be factually true. Problems and ways of building and testing models in the social sciences are then illustrated by considering approaches to model human decision making in fuzzy environments.

1. THE MEANING OF MODELS

Models, theories and scientific laws

The terms "model," "theory," and "law" have been used with a variety of meanings, for a number of purposes, and in many different areas of our life. It is, therefore, necessary to define more accurately what we mean by models, theories and laws to describe their interrelationships and to indicate their use before we can specify the requirements they have to satisfy and the purposes for which they can be used. To facilitate our task we shall distinguish between definitions which are given and used in the scientific area and definitions and interpretations as they can be found in more application-oriented areas, which we will call "technologies" by contrast to "scientific disciplines." By technologies we mean areas such as "Operations Research," "Decision Analysis," and "Information Processing," even though these areas call themselves sometimes theories (i.e., decision theory) or science (i.e., computer science, management science, etc.). This is by no means a value statement. We only want to indicate that the main goals of these areas are different. While the main purpose of a scientific discipline is to generate knowledge and to come closer to truth without making any value statements, technologies normally try to generate tools for solving problems better and very often by either accepting or basing on given value schemes. Let us first turn to the area of scientific inquiry and consider the following quotation concerning the definition of the term "model": "A possible realization in which all valid sentences of a theory T are satisfied is called a model of T" [1].

Harré [2, p. 86] calls: "A model, *a*, of a thing, *A*, is in one of many possible ways a replica or an analogue of *A*," and a few years later, "In certain formal sciences such as logic and mathematics a model for, or of a theory is a set of sentences, which can be matched with the sentences in which the theory is expressed, according to some matching rule. . . . The other meaning of 'model' is that of some real or imagined thing or process, which behave similarly to some other thing or process, or in some other way than in its behavior is similar to it" [3, p. 173]. He sees two major purposes of models in science: (i) Logical: to enable certain inferences, which would not otherwise be possible to be made; and (ii) Epistomological: that is to express and enable us to extend our knowledge of the world. Models according to Harré are either used as a heuristic to simplify a phenomenon or to make it more readily handable and explanatorily where a model is a model of the real causal mechanism.

Leo Apostel [4, p. 4] provides us with a very good example for various definitions of models as tuples of a number of components when defining: "Let then $R(S, P, M, T)$ indicate the main variables of the modelling relationship. The subject S takes, in view of the purpose P, the entity M as a model of the prototype T." For the four components of his definition he gives a number of examples which are quite informative concerning the use of models in science and which can be summarized as follows:

Subjects (S) and purposes (P):

1. "For a certain domain of facts let no theory be known. If we replace our study of this domain by the study of another set of facts for which a theory is well known, and that has certain important characteristics in common with the field under investigation then we use a model to develop our knowledge from a zero (or near zero) starting point.
2. For a domain D of facts, we do have a full-fledged theory but one too difficult mathematically to yield solutions, given our present techniques. We then interpret the fundamental notions of the theory in a model, in such a way that simplifying assumptions can express this assignment.
3. If two theories are without contact with each other we can try to use the one as model for the other or introduce a common model interpreting both and thus relating both languages to each other.
4. If a theory is well confirmed but incomplete we can assign a model in the hope of achieving completeness through the study of this model.
5. Conversely, if new information is obtained about a domain, to assure ourselves that the new and more general theory still concerns our earlier domain, we construct the earlier domain as a model of the later theory and show that all models of this theory are related to the initial domain, constructed as model, in a specific way.
6. Even if we have a theory about a set of facts, this does not mean that we have explained those facts. Models can yield such explanations.
7. Let a theory be needed about an object that is too big or too small, too far away or too dangerous to be observed or experimented upon. Systems are then constructed that can be used as practical models, experiments which can be taken as sufficiently representative of the first system to yield the desired information.
8. Often we need to have a theory present to our mind as a whole for practical or theoretical purposes. A model realizes this globalization through either visualization or realization of a closed formal structure."

Thus, models can be used for theory formation, simplification, reduction, extension, adequation, explanation, concretization, globalization, action or experimentation.

Entity (M) and model type (T): M and T are both images or both perceptions or both drawings or both formalisms (calculi) or both languages or both physical systems or vice versa. M can also be a calculus and T a theory or language.

Apostel believes that all models which can be constructed by varying the contents of the four components form a systematic whole: "Models are used for system restructuration because of their relations with the system (partial discrepancy); because of their relationship among each other (partial inconsistency at least multiplicity); because of their relationship with themselves (local inconsistent or locally vague)."

By now two things should have become obvious:

1. There is a very large variety of types of models, which can be classified according to a number of criteria. For our deliberations one classification seems to be particularly important: The interpretation of a model as a "*formal model*" and the interpretation as a "*factual, descriptive model.*" This corresponds to Rudolph Carnap's distinction between a logical and a descriptive interpretation of a calculus [5]. For him a *logically true* interpretation of a model exists, if whenever a sentence implies another in the calculus, in the interpretation whenever the first sentence is true, the second is equally true and whenever a sentence is refutable in the calculus, it is also false in the model. An interpretation is a factual interpretation if it is not a logical interpretation, which means that whether a model is true or false does not only depend on its logical consistency but also on the (empirical) relationship of the sentences (axioms of the model) to the properties of the factual system of which the model is supposed to be an image. The second interpretation of a model is the one that is quite common in the empirical sciences and it is the one, which we will primarily be referring to in the following.
2. There is certainly a relationship between a model and a theory. This relationship, however, is seen differently by different scientists and by different scientific disciplines. We will now try to specify this relationship because theories, to our mind, are the focal point of all scientific activities.

For Harré [3, p. 174] "A theory is often nothing but the description and exploitation of some model," or "Development of a theory on the other involves the superimposing of one model on another" [2, p. 99].

Wartofsky [6, p. 286] summarizes: "Theories are models which are taken either as imaginary constructions or as conjectures about the real nature of things." Suppes [7, p. 166] sees the distinction between theory and model in "That a theory is a linguistic entity consisting of the set of sentences and models are non-linguistic entities in which the theory is satisfied." Nagel [8] describes a theory as consisting of three distinguishable components:

1. An abstract calculus. That is the logical skeleton of the explanatory system;
2. A set of rules that in affect assign an empirical content to the abstract calculus by relating it to the concrete materials of observation and experiment; and
3. An interpretation or model for the abstract calculus, which supplies some flesh for the skeletal structure in terms of more or less familiar conceptual or visualisable materials.

White [9], eventually, simply points out that "There is a need to logically separate a model and a theory and that they play supporting roles in decision analysis, viz., some theory is needed so that aspects of models can be tested and that some model is needed so that the affects of changes can be examined. In particular validation of a model needs a theory." Thus, there seems to be a very intimate relationship between a model and a theory in scientific inquiry. Both, probably to varying degrees, base on hypotheses and these hypotheses can either be formal axioms or scientific laws. These scientific laws seem to me to fundamentally distinguish models and theories in scientific disciplines from the type of models (sometimes also called theories) in the more applied areas: "An experimental law, unlike a theoretical statement invariably possesses a determinate empirical content which in principal can always be controlled by observational evidence obtained by those procedures" [8, p. 83]. These laws as scientific laws assert invariance with respect to time and space. The tests to which such hypotheses have to be put before they can assert to be a law depend on the philosophical direction of the scientist. Karl Popper, as probably the most prominent representative of the "critical rationalism," believes that laws are only testable by the fact of their falsifiability. Popper holds further that a hypothesis is "corroborated" (rather than confirmed) to the degree of the severity of such tests. Such a corroborated hypothesis may be said to have stood up to the tests thus far without being eliminated. But the test does not confirm its truth. A good hypothesis in science, therefore, is one which lends itself to the severest test, that is, one which generates the widest range of falsifiable consequences [6,10].

Models in operations research and management science

The area of Operations Research will be considered as an example of a more application-oriented discipline, which I called technology, in which modelling plays a predominant role. Even though one might dispute, whether Operations Research is a science or a technology, I will follow Symonds, who, as the President of the Institute of Management Sciences, stated: "Operations Research represents the problem solving objective; management science is the development of general scientific knowledge" [11, p. 385].

What, now, is a model in Operations Research? Most authors using the term model take it for granted that the reader knows what a model is and what it means. Arrow, for instance, uses the term model as a specific part of a theory, when he says, "Thus the model of rational choice as built up from pair wise comparisons does not seem to suit well the case of rational behavior in the described game situation" [12, p. 21]. He presumably refers to the model of rational choice, because the theory he has in mind does not give a very adequate description of the phenomena with which it is concerned, but only provides a highly simplified schema. In the social and behavioral sciences as well as in the technologies it is very common that a certain theory is stated in rather broad and general terms while models, which are sometimes required to perform experiments in order to test the theory, have to be more specific than the theories themselves. "In the language of logicians it would be more appropriate to say that rather than constructing a model they are interested in constructing a quantitative theory to match the intuitive ideas of the original theory" [7, p. 169]. Rivett in his book *Principles of Model Building* [13] offers three different kinds of classifications of models; when enumerating the models which he suggests to put into the different classes, he no longer uses the term "model" but talks of "problems in this area" and "the theory of this area" as a not-to-well-defined entity of knowledge. Ackoff suggests as a model of decision making a six-phases-process which is supposed to be a good picture (model) of the real decision making process [14]. This is only one example of quite a number of very similar models of decision making.

From the point of view of purposes or functions of theories or models, respectively, there seems to be hardly any recognizable difference between these two terms. Mehlberg [15], for instance, suggests that there are four major functions of a theory: 1. The summarizing function, whereby a number of consequences can be deduced from a theory; 2. The predictive function; 3. The controlling and explanatory function; and 4. the information function relating to observable variables and their role in model validation. Without any problems we could accept the same functions as those of models.

If we consider the size of some of the models used in Operations Research, containing more than 10,000 variables and thousands of constraints we can easily see what does *not* distinguish a theory from a model: It is not the complexity, it is not the size, it is not the language and it is not even the purpose. In fact there seems to be only a gradual distinction between theory and model. While a theory normally denotes an entire area or type of problem, it is more comprehensive but less specific than a model (for instance: Decision Theory, Inventory Theory, Queueing Theory, etc.), a model most often refers to a specific context or situation and is meant to be a mapping of a problem, a system or a process. By contrast to a scientific theory, containing scientific laws as hypotheses, a model normally does not assert invariance with respect to time and space but requires modifications whenever the specific context, for which a model was constructed, changes.

In the following I will concentrate on models rather than on theories. Realizing that there is quite a variety of types of models, I do not think that it is important and necessary for our purposes to distinguish models by their language (mathematics or logic is considered to be a modelling language), by area, by problem type, by size, etc. One classification, however, seems to be important: the distinction of models by their character. Scientific theories were already divided into formal theories and factual theories. For models, particularly in the area of the technology in which values and preferences enter our considerations, we will have to distinguish among the following:

1. *Formal Models:* These are models, which are purely axiomatic systems from which we can derive if-then-statements and the hypothesis of which are purely fictitious. These models can only be checked for consistency but they can neither be verified nor falsified by empirical arguments.
2. *Factual Models:* These models include in their basic hypotheses falsifiable assumptions about the object system, i.e., conclusions drawn from these models have a bearing on reality and they, or their basic hypothesis, have to be verified or can be falsified by empirical evidence.
3. *Prescriptive Models:* These are models which postulate rules according to which processes have to be performed or people have to behave. This type of model will not be found in science but it is a common type of model in practice.

The distinction between these three different kinds of models is particularly important when using models: All three kinds of models can look exactly the same but the "value" of their outputs is quite different. It is therefore rather dangerous not to realize which type of model is being used because we might take a formal model to be a factual model or a prescriptive model to be a factual model and that can have quite severe consequences for the resulting decisions.

As an example, let us look at the above mentioned Ackoff model of decision making. Is that a formal, a factual, or a prescriptive model? If it is a formal model, we can not derive from it any conclusion for real decision making. If it is a factual model then it would have to be verified or falsified before we can take it as a description of real decision making. The assertment, however, that decision making proceeds in phases was already empirically falsified in 1966 [16]. Still a number of authors stick to this type of model. Do they want to interpret their model as a prescriptive model? This would only be justified, if they could show that, for instance, decision making can be performed more efficiently when done in phases. This, however, has never been shown empirically. Therefore, we can only conclude that authors suggesting a multiphase scheme as a model for decision making take their suggestion as a formal model and do not want to make any statement about reality, or that they are using a falsified, i.e., invalid and false, factual model.

2. BUILDING, TESTING AND USING MODELS

The model-building process

Talking, thinking or writing about model building can mean two different things: Either one focusses on the product, i.e., the model, and investigates which properties it should have; or one concentrates on the model building process with the aim to find efficient procedures to build models. By far most publications do the former. Only very few authors probe the model building process at all. Even if they do, their conclusions are not too helpful. Figure 1, which is taken from Rivett [13] is a model of the model building process.

Looking at the figure it becomes very quickly obvious that the suggested model building process assumes a certain type of model, which we would call "factual model." For formal models or prescriptive models other model building activities would have to be performed.

We would assume that a model-building process is not much different from a problem

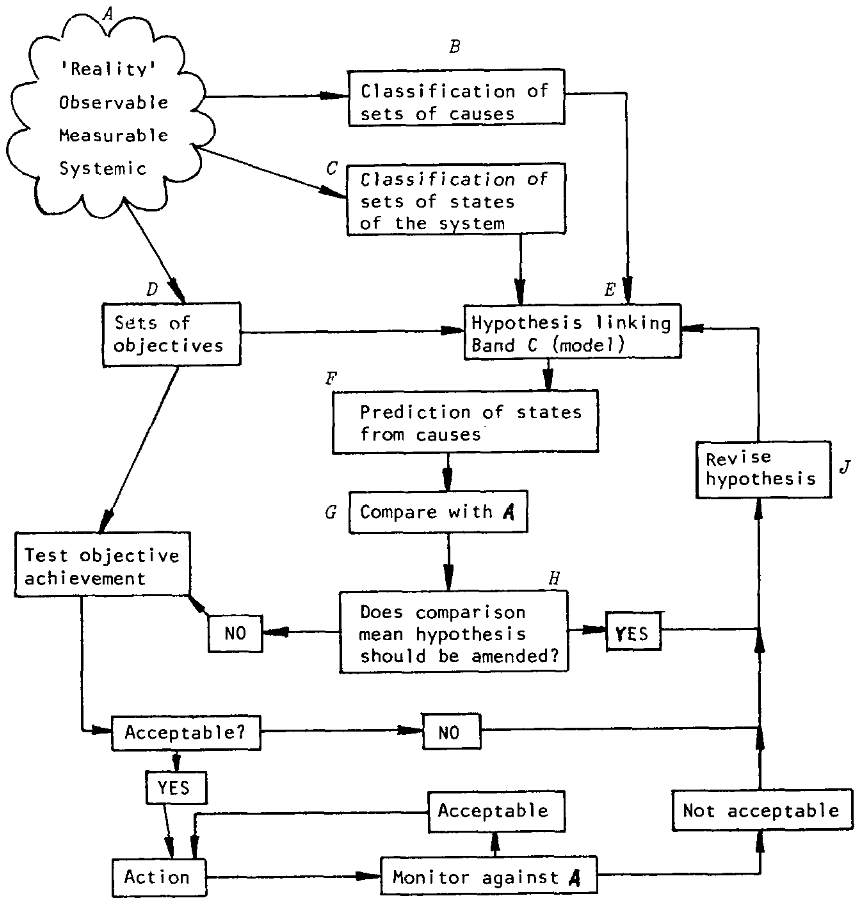

Fig. 1. The model-building process.

solving process or a decision making process and would therefore have to be primarily considered as an information processing process. It has already been mentioned that so far many authors still advocate the "phase-concept" as a model of decision making. Even though it has been shown empirically that decision making processes are very complex, creative, multioperational, multitemporal, and multipersonal information processing processes. This, however, can at best be a starting point: The not yet existing theory of model building would certainly not (only) try to find out empirically how models are being built but it should and would be a prescriptive theory developing or indicating ways to construct models properly and efficiently. First developments in this direction may be found in the area of experimental design and of EDP-program development.

Quality of models

The quality of a model depends on the properties of the model and the functions which the model is designed for. In general, models will have to have at least the following three major properties: formal consistency, usefulness, and efficiency.

By logical consistency I mean that all operations and transformation have been performed properly and that all conclusions follow from the hypothesis. This consistency has to be demanded of all types of models, whether they are formal, factural, or prescriptive.

By usefulness I mean that the model has to be helpful for the function it has been designed for, and by efficiency I mean that the model as the tool to achieve an end, has to fulfill the desired function at a minimum of effort, time, and cost.

Functions of models

We have already considered functions of models in the area of scientific activities. Here models are primarily used to develop, improve, and test theories. One should, however, point to a distinction which is made in the philosophy of science, between the discovery phase and the confirmation phase of a theory. While in the discovery phase almost all methods and models which could possibly initiate a quasitheory or pseudotheory are permitted; models and methods have to meet rather high scientific standards if they are to be used in the confirmation phase of a theory. In the area of technology and practise the situation is not much different. While the initiation of a model design might heavily rest on heuristic methods and on intuition, the later stages of model development might require similar methods as are used in the confirmation phase of a scientific theory.

As mentioned above, scientists see as the major functions of a theory the summarizing function, the predictive function, the controlling and explanatory function and the information function. All of these functions could also be considered to be proper functions of models in the not scientific area. To be able to be somewhat more specific I will turn again to the area of Operations Research and decision making.

In order to prevent misunderstandings I shall exclude from my considerations "algorithms," which have the very well defined objective of finding specific solutions (i.e., optimal solutions, satisfactory solutions, feasible solutions) in existing models. This means that they do not "model" in any way, even though they are part of information processing and they cause information processing.

By contrast to the major goal in science, to generate true and additional knowledge, the objectives in the nonscientific area, i.e., the practice of management, engineering, teaching, etc., depend very much on the specific area. In management, for instance, models will primarily be used to facilitate decision making, problem solving, communication, and control. Three kinds of models will be necessary for these purposes: factual models, formal models, and prescriptive models. Because of the limited information storage and processing capacity of people, their need for models will be higher the more complex the problems, decisions, or systems under consideration.

In *decision making* and *problem solving* factual models will be needed to describe, to explain and to predict phenomena and consequences. For "conditional predictions" formal models will also be useful in order to obtain "if-then-statements," for instance, in the framework of simulations. Formal models will also be useful and necessary for the area of communication within the decision making process and for relaying the resolutions or conclusions of the decision or problem solving process to the "actors." One should assume that prescriptive models are the most common in decision making. This, however, is only true if one calls all "decision models," i.e., models which contain an objective function, by which an optimal solution can be determined, prescriptive models. To my mind this is not quite appropriate because these kinds of models only prepare suggestions for possible decisions and the normative or prescriptive character is acquired only after the "solution" has been declared a decision by the authorized decision maker. A much more important feature of these models seems to me, that they have to describe or define properly the conditions which limit the action space (such as capacities, financial resources, legal restrictions, etc.).

The area in which prescriptive models are primarily used is, to my mind, the area of control and delegation. A plan, a budget, a projection, or any other kind of instruction and any kind of model, which is used in such a function, is a genuinely prescriptive model.

I can now restate the notion of the quality of a model more precisely; I already mentioned that consistency is one of the necessary conditions for quality. Usefulness of a model will have to be defined for each of the three different types of models differently: 1. While a factual model can be called useful, if it is "factually true" (by contrast to logically true), i.e., if it maps the object system with an appropriate precision (which can only be tested empirically), the model also has to generate knowledge, i.e., the user of a model should gain knowledge which he would not have gained without using the model or which he did not have available before using the model. 2. Formal models can neither by verified nor falsified empirically. Such a model will be considered useful, if activities, such as teaching, explaining, communicating, become more efficient by using the model than without it. 3. Prescriptive models can also not be verified or falsified. They are the more useful, the more effectively they help to enforce the desired behavior, to control predefined performance measures, and to define ranges within which decision makers have freedom to decide.

The modelling language

I do not want to discuss the problem of the lack of communication between scientists and practitioners because of language difficulties. This is certainly a serious problem but it is not a part of model building, but belongs in the area of diffusion of theories and concepts.

What I want to discuss is the appropriateness of the language used for a specific model or a specific theory. The language is, roughly speaking, the link between the thinking of human beings, the object system, which is to be modelled, and the model itself. Statements in a model are formulated by means of linguistic expressions. That means that the model language semantically and syntactically has to be well suited to grasp those aspects of reality which are included in the model; the perspective of a scientific theory is already contained in its linguistic terms.

Let us look at the appearance of scientific theories and of practical problems which we want to model: "Most scientific systems do not have a clear distinction between logical and descriptive constants. In fact, a science seems to be written in a mixture of everyday language describing experiments, mathematical equations describing calculations, and semi-formulized deduction if the science is sufficiently advanced to contain a theoretical part" [4, p. 29].

"The problems of management exist only in the minds of managers and in the minds of their advisers. There is, therefore, a completely subjective personal basis to our science. It is not the case that the company has an inventory problem in the sense that if you visit it you can see it there in concrete form. . . . Even within our own minds as we approach a practical problem, we will be unsure how to show which form of model is likely to be optimal . . . Since the problems as seen in each of our minds will be different and mutual comparison will be impossible, it follows that we have a significant, even dangerous, freedom in the way in which we select the models which will represent decision making situations" [13, p. 129].

On the other hand a mathematical model is defined to be a collection of statements about a set of variables from which the truth or falsity of other statements can be deduced [9].

The utter importance of the modelling language is recognized by Apostel, when he says: "The relationship between formal languages and domains in which they have models must in the empirical sciences necessarily be guided by two considerations that are by no means as important in the formal sciences:

(a) The relationship between the language and the domain must be closer because they are in a sense produced through and for each other;
(b) extensions of formalisms and models must necessarily be considered because everything introduced is introduced to make progress in the description of the objects studied. Therefore we should say that the formalization of the concept of approximate constructive necessary satisfaction is the main task of semantic study of models in the empirical sciences" [4, p. 26].

Because we request that a modelling language is unequivocal and nonredundant on one hand and, at the same time, catches semantically in its terms all that is important and relevant for the model we seem to have the following problem. Human thinking and feeling, in which ideas, pictures, images, and value systems are formed, first of all has certainly more concepts or comprehensions than our daily language has words. If one considers, in addition, that for a number of notions we use several words (synonyms) then it becomes quite obvious that the power (in a set theoretic sense) of our thinking and feeling is much higher than the power of a living language. If in turn we compare the power of a living language with the logical language, then we will find that logic is even poorer. Therefore it seems to be impossible to guarantee a one to one mapping of problems and systems in our imagination and a model using a mathematical or logical language. One might object, that logical symbols can arbitrarily be filled with semantic contents and that by doing so the logical language becomes much richer. I will show that it is very often extremely difficult to appropriately assign semantic contents to logical symbols.

The usefulness of the mathematical language for modelling purposes is undisputed. However, there are limits of the usefulness and of the possibility of using the classical mathematical language, based on the dichotomous character of set theory, to models in particular systems and phenomena in the social sciences: "There is no idea or proposition in the field, which can not be put into mathematical language, although the utility of doing so can very well be doubted" [17]. Schwartz [18] brings up another argument against the unreflected use of mathematics, if he states: "An argument, which is only convincing if it is precise loses all its force if the assumptions on which it is based are slightly changed, while an argument, which is convincing but imprecise may well be stable under small perturbations of its underlying axioms." I will return to this argument later in this paper.

Data quality and measurement

The saying "garbage in—garbage out" is well known and speaks for itself. The following quotation from Josiah Stamp [19, p. 236] points in the same direction: "Governments are very keen in amassing statistics. They collect them, add them, raise them to the n^{th} power, take the cube root and make wonderful diagrams. But what you must never forget is that every one of these figures comes in the first instance from the village watchman who just puts down what he damn pleases."

It must, however, be born in mind that the effort put into deriving and obtaining numerical values or relations must be geared to the value of the model and that when data is scarce it may still be useful to draw conclusions from not fully satisfactory input data. In this case a tentative look at the dependence of the solution from the quality of the input data may be very advisable.

The quality of the input data is closely related to the question of operational definitions for the relevant variables. The processes of defining variables and their operational indicators and measurement are intertwined. To quote White [9, p. 102] "We take 'measurement' to be a special aspect of a 'definition'. One might take the view that measurement is the actual procedure for assigning the real numbers which constitute the measure. However, as pointed out in a previous section, this is the quantification process

Table 1. Hierarchy of scale levels

Type of scale	Permissible transformation: Verbal	Permissible transformation: Formal	Invariance	Example
Nominal scale	One-to-one function	$x_i \neq x_j => x_i = x'_j$	Uniqueness of values	License plates
Ordinal scale	Monotonic increas. f.	$x_i \leqq x_j => x'_i \leqq x'_j$	Rankorder of values	Marks
Interval scale	Affine function	$x' = a \cdot x + b$	Ratio of differences	Temperature (C°, F°)
Ratio scale	Similarity function	$x' = a \cdot x$	Ratio of values	Length (cm, inch)
Absolute	Identity	$x' = x$	Values	Frequency

and in itself does not constitute a measure unless it is a homomorphism. The homomorphism then defines the measure." Very often when modelling in the area of social sciences one will find that relations, data or values are stated in very vague ways. Goals, for instance, may be stated as "trying to achieve satisfactory profits," data such as "the South of the country is much poorer than the North" and relations such as "his investment strategies were much more risky than those of his competitors." Very often these variables are measured subjectively and point scales are used to transform the "measurements" into numerical values. Even though it is necessary to include variables in the model which are considered important but very hard to operationalize and to measure, the quality of the input data might have very limiting effects on their transformation which can be permitted in the model. Rather than neglecting these kinds of data one should consciously determine which scale quality these data have and then make sure that only admissible transformations are being used when processing these data in the model. Table 1 sketches the hierarchy of scale levels including the permissible transformations for each of the levels.

Testability of models

I already mentioned that part of the usefulness of models is their formal and factual truth. Formal truth can in most instances and to a certain degree be guaranteed by applying appropriate logical and mathematical methods. This formal truth might even be provable (it has to be born in mind however, that some models, for instance large EDP-programs, can never be guaranteed to be absolutely faultless). Factual truth is not established that easily. Factual truth means that the model contains only factually true relations and hypotheses, that it contains all relevant relations and variables and that there is no factual inconsistency of hypothesis. One prerequisite for testing a model is that it's hypothesis, parameters, and relations are testable, a requirement which is known as "principle of ontological relativism."

White distinguishes three different sorts of testing [9, p. 131]: 1. Testing while the model is being build—*a priori* testing; 2. Testing on the basis of the outputs of the analysis; and 3. Testing by implementation.

In the case of *a priori* testing the model is synthesized from its components which might already contain either validated variables and relationships or testable hypotheses which can be tested on the basis of existing data or by setting up experiments before proceeding further in the modelling process. Two observations seem to be appropriate.

A. There is a difference, whether testing is being done for scientific purposes or for practical purposes; and if it is done for scientific purposes one will have to distinguish the initiating phase from the confirmation phase. The following statement by White holds true for scientific testing in the confirmation phase but not necessarily for scientific testing in the initiation phase or for testing models which do not assert invariance in the scientific sense. "In testing relationships there are certain barriers, which we can not transgress, viz.:

(i) we can never prove general infinite deterministic empirical relations;
(ii) we can only disprove general infinite deterministic empirical relations;
(iii) we can never prove nor disprove probabilistic relations.

As a result of this we recognize the Principle of Epistemological Relativism which denies the existence of apodictic truth in empirical investigations and the Principle Perspective Relativism which states that all empirical theories are relative to input hypotheses. These hypotheses themselves are not absolute. We can continue to regress in an attempt to explain hypotheses at one level by others at a higher level. But we always remain in a relative position although we may gain more confidence by doing so."

B. The testability of the components of a model—in the scientific and in the practical context—depends largely on the operational definition of the hypotheses. In this sense, observation and formal analysis prior to model building can very often improve the testability of hypothesis. Let us illustrate this with the following example. In decision analysis one normally distinguishes between decision making under certainty, decision making under risk and decision making under uncertainty; one assumes that in decision making under risk the decision maker is able to store and process probability distribution functions. Here probabilities ought to be interpreted as Koopman-type probabilities, i.e., probabilities as expressions of belief and not in the frequentistic sense. This hypothesis is hardly testable because a situation of decision making under risk is not homogeneous with respect to the available information at all. An improvement in the testability of hypotheses could be achieved if one would distinguish, for instance, among:

a. Decision making when quantitative probabilities are known (interval-scale);
b. Decisions when interval probabilities are known (hyper-ordinal scale);
c. Decisions when qualitative probabilities are known (ordinal scale);
d. Decisions when partially ordered nominal probabilities are known (preordinal scale);
e. Decisions when nominal probabilities are known (states are known but not truth ratable);
f. Decisions when only some of the nominal probabilities are known.

It is obvious that the information storage and processing requirements which a human being would need in order to decide "rationally" are quite different in the above cases and that the permissible operations in the model will also be different depending on the type of probability which can be assumed to exist.

If the testing is done on the basis of the outputs of the analysis the decision maker might already be able to indicate that the output of the analysis is not satisfactory, probably because important relations or variables have been omitted. If the decision maker or expert rates the output of the model satisfactory, it gains the status of face-validity, sometimes in practice the most we can hope for.

Ideally a model should now be tested by implementation, i.e., by comparing actual with predicted results. This, however, in many instances is impossible for several reasons:

a. *Changes of environments:* Factors such as sales, price levels, etc. might have changed while the model was built and implemented and therefore the observed results after implementation of the model can no longer be compared with the predicted results.
b. *Changes in performance:* If, for instance, the model is tested after implementation by running the old procedure parallel to the model and if the old procedure included human activities, the performance of these activities might be improved by the persons because they know that the "new" model is being compared with their performance, which would probably drop again, if and when the operation of the new procedures would be terminated.
c. *Risk and uncertainty:* It is obvious that if procedures have been designed to optimally

decide in situations of risk or uncertainty, the "real" result can not meaningfully be compared with the probabilistic prediction.

d. *Optimality:* If only one solution is actually implemented there is, of course, no way to compare this with other alternatives. In many cases the optimal solution with which the model solution could be compared is not known at all because it is not computable or because optimality was defined subjectively in a way which is not reproducable objectively.

3. AXIOMATIC JUSTIFICATION VS. EMPIRICAL TESTINGS

Consistency and degree of confirmation

It has already been pointed out that all kinds of theories and models can be and ought to be tested for consistency. In formal analysis it might even be possible to prove consistency which does not mean that models and theories, the consistency for which has not been proven yet, are not formally correct. For "factual" or "substantial" theories and models empirical testing of basic hypotheses, relations, and resulting outputs is absolutely necessary in order to achieve a certain degree of confirmation of the theory or the model. This fact is often neglected when working with theories and models. If, for instance, the hypothesis of "rationality" in decision making models is "justified" by defining rationality by more basic axioms such as transitivity, reflexibility, existence of an ordering, etc., which seem quite plausible and natural, then the model or the theory might become more testable but certainly not better confirmed. To confirm the model would require empirically testing either the main hypothesis or the presumably more operational basic axioms. This, of course, still does not determine uniquely the methods which can be used for testing hypotheses. These methods will depend on the area in which the model is being used (physics, engineering, management) and the purposes for which the model has been built. Thus, in scientific inquiry probabilistic tests might not be acceptable because scientific laws assert deterministic invariance. These methods, however, might be the only available ones for testing models in areas such as management, sociology, and political decision making.

Cost and benefits of testing

Let us first consider scientific research. In the phase of generating theories we can certainly talk of different benefits being derived from different kinds of testing. In the confirmation phase of the theory the testing of hypotheses is no longer desirable but an absolute necessity. Since the social sciences are considered to be still in the pre-scientific state of an empirical science, the major problem is still to formulate quasi-theories in a testable way which in many instances will mean that they have to use mathematical models extensively. In building these models it should always be borne in mind that their hypotheses will have to be operational and testable, if the standing of scientific theories and scientific laws is ever intended.

Cost of empirical research can be looked at from two different points of view. Either one compares the total effort which has been put into empirical research in this area with the corresponding effort in natural sciences and engineering, or one compares typical (specific) research efforts in theoretical research with that of empirical research. The first type of comparison will certainly lead to the conclusion that research efforts are by magnitudes larger in the natural sciences and in engineering than they are in social sciences. The second type of comparison will certainly show that the typical research effort is much higher, particularly in the social sciences, for empirical research than for formal research. This might partly explain the rather small scope of empirical research in the social sciences compared to the formal theoretical research, which has been going on for the last 50 years. I will illustrate this point later in more detail.

In the area of practical decision making the effort for empirical research and testing of decision models should be determined by the predicted benefits of such research and testing. Here, however, we have to face a very serious problem. Benefits of empirical research and testing can be quite substantial but it is very often either extremely difficult or impossible to prove in advance the benefits of such activities. Very often the benefits are hard to quantify because they might be derived from a higher transparency of the system under consideration, an improvement of interpersonal relationships, more flexibility, etc.

Compared to scientific research and testing the efforts of empirical testing in practice will be rather low. The reason for this is primarily that the scientific demand for invariance or invariant validity of theories and laws does not exist in practice: Models have to be valid under certain circumstances and for certain times (i.e., they are context- and time-dependent) and they will have to be modified whenever the context or the time changes. In addition, the accuracy and dependability of the conclusions derived from such models, and with it the necessary effort, can be determined autonomously by the decision maker. Since the practitioner by contrast to the scientist, does not necessarily strive for new knowledge in the scientific sense he can rely more heavily on hypotheses which are already sufficiently confirmed.

Summarizing, we can state that empirical research and empirical testing is desirable to varying degrees in science as well as in practice. In the confirmation phase of scientific inquiry, however, it is absolutely necessary for all theories and laws which assert "factual truth" and not only "formal or logical truth." Some of the above observation shall now be illustrated by the following example.

4. MODELLING DECISIONS IN FUZZY ENVIRONMENTS

The problem

The importance of choosing an appropriate modelling language has already been mentioned. The problem considered in this context is essentially due to two well known facts:

1. In natural languages the meaning of words are very often vague. The reason for this can also be that the meaning of a word is well defined but that when using the words as labels for sets the boundaries within which objects belong to the set or do not become fuzzy or vague. The same vagueness can of course be found in words which are used to express relations, so called connectives.
2. Natural languages cannot be substituted by formal languages in the area of empirical research. Many scientists have already pointed to the fact that formal languages are rather simple and poor and have to be considered as very specific languages useful for specific purposes. Mathematics and logic as languages have, for instance, proved to be extremely useful for modelling purposes in the natural sciences and in engineering but not quite as adequate for the social sciences.

The "Theory of Fuzzy Sets" suggested by L. A. Zadeh for the first time in 1965 [20] seems, as a formal language, better suited to map the vagueness of natural languages than classical mathematics which bases on two valued logic. For this purpose, a vague notion of a natural language can be interpreted as a subjective category and can be mapped by a fuzzy set, which is defined as follows:

Definition—If $X = \{x\}$ is a collection of objects denoted generically by x then a fuzzy set A in X is a set of ordered pairs

$$A = \{[x, \mu_A(x)] \mid x \in X\}.$$

$\mu_A(x)$ is called the membership function, grade of membership, degree of compatibility or degree of truth of x in A which maps X to the membership space M. [When M contains only the two points 0 and 1, A is nonfuzzy and $\mu_A(x)$ is identical to the characteristic function of a nonfuzzy set.] The range of the membership function is a subset of the nonnegative real numbers whose supremum is finite. For "normalized fuzzy sets" the membership function is limited to values between 0 and 1. A nonempty fuzzy set A can always be normalized by dividing $\mu_A(x)$ by $\mathrm{Sup}_x \mu_A(x)$.

In 1970 fuzzy sets were used to model human decisions [21]. An indication to interprete this model as a factual model is the frequent reference to real decision-making environments, such as "much of the decision making in the real world takes place in an environment in which the goals, the constraints and the consequences of possible actions are not known precisely." As in day-to-day language a decision was defined as linking goals and constraints by "and" (i.e., the decision should satisfy the goal *and* the first constraint *and* the second constraint etc.). The meaning of "and" is defined in classical logic as a connective such that a sentence, which consists of other sentences linked by "and" is only true, if and only if, all subsentences are true. This "and" is denoted symbolically by $\wedge$ which again is the symbol for the intersection of two sets in classical set theory.

In the theory of fuzzy sets the membership function of the fuzzy set "intersection" of other fuzzy sets is defined by applying the minimum operator to the membership functions of all intersecting fuzzy sets. Consequently a decision in fuzzy environments was defined as the intersection of all fuzzy sets involved (goals and constraints) which, if taken as a factual model of human decision making, means that human beings when making decisions aggregate different goals, which can be interpreted as subjective categories, by applying the minimum operator. This hypothesis is a contradiction to most of the models in multicriteria decision making in which generally the aggregation by weighted addition is assumed. Neither for one way nor for the other empirical confirmation was available. To test, or at least to justify, the use of any of those connectives when modelling human decisions, two approaches were used:

1. The reduction of the hypothesis (of using the minimum operator or other operators) to either empirically tested or at least, axioms with face-validity.
2. Direct empirical research.

The results of the first approach shall be sketched very briefly because they are of interest in connection with the results of the empirical research; they are characteristic for the approaches most frequently used in utility theory and formal decision theory.

Axiomatic approaches vs empirical investigation

Deductive derivation of meaning of connectives. Starting from ten plausible axioms Hamacher [22] derived four basic nonredundant axioms for the connective "and" (conjunction) interpreted as the intersection operator, which also seemed to be sensible from the point of view of rational decision making:

Axiom 1—The connective C (conjunction) is associative, i.e., $A \wedge (B \wedge C) = (A \wedge B) \wedge C$, (necessary to extend C from 2 to n statements).

Axiom 2—C is continuous (secures stability of C for small changes of μ_A and μ_B).

Axiom 3—C is injective in each argument $[(A \wedge B) \neq (A \wedge C)$ if $B \neq C]$.

Axiom 4—$\bigwedge_{x \in (0,1]} D(x,x) = x \Longleftrightarrow x = 1$.

A, B, and C denote fuzzy sets and x is the degree of membership of an element with respect to the appropriate fuzzy set.

Hamacher proved that if C is an algebraic function, a rational function and a polynom in x and y then only the following mathematical description of the operator is possible:

$$C(x,y) = \frac{x \cdot y}{\gamma + (1 - \gamma)(x + y - xy)}, \gamma \geq 0.$$

γ is an arbitrary constant. For $\gamma = 1$ the above formula reduces to

$$C(x,y) = x \cdot y,$$

which obviously contradicts the assertion that the minimum operator is used to model the "and."

Empirical testing [34]

1. *Experimental design and pretest:* It had already been established that formal education had an important influence on the meaning that persons attach to the word "and." It was also known that the degree of membership had the quality of an absolute scale and that human beings are not reliable and valid measuring devices on this level. By choosing an appropriate experimental design, selecting or developing appropriate measuring methods, and by choosing appropriate "test examples" it had to be insured that:
— Biases due to former education were eliminated by selecting an appropriate random sample of test persons;
— The intersection could really be interpreted as the aggregation of subjective categories by selecting appropriate test examples and by insuring that the fuzzy sets (subjective categories) to be intersected were of equal weight and independent from each other;
— The required scale level was at least approximately achieved by using especially designed indirect measuring methods.

2. *Testing the hypothesis:* Because of obvious reasons the minimum operator and the product operator were tested as hypothetical connectives corresponding to "and." It turned out that at the required significance level both operators had to be rejected even though the minimum operator showed a better fit than the product operator. In discussions with colleagues, however, another surprising observation was made: No one was able to come up with practical examples of decision making in which the "and"-connective was really used in the sense of the logical "and." It was felt that in linking goals in the sense of an "and" certain compensatory effects had always to be assumed. The logical "and" however, does not allow any compensation between the degrees of membership of elements with respect to different fuzzy sets.

3. *Modifying the hypotheses:* On the basis of the observations described above, the hypotheses as well as the test examples had to be modified in the following way.
   a. Fuzzy sets had to be selected for intersection, for which a certain compensation could be assumed when intersected in order to find the resulting fuzzy set.
   b. In addition to the minimum operator and the product operator the following operators were used as additional hypotheses: Maximum Operator, Algebraic Sum, Arithmetic Mean, Geometric Mean.
   c. A new generalized operator was developed as additional hypothesis and on the basis of the following considerations. The logical "and" does not allow any compensation; the logical (inclusive) "or" however, implies full compensation. We,

therefore, assumed that an operator implying a certain degree of compensation would have to lie between the two extremes: "logical and" and "logical or." Characterizing the "logical and" by the product operator and the "logical or" by the generalized algebraic sum the following generalized operator was derived.

$$\mu^{\gamma} = \left[\left(\prod_{i=1}^{m} \mu_i \right)^{(1-\gamma)} \left(1 - \prod_{i=1}^{m} (1 - \mu_i)^{\gamma} \right) \right], \quad \begin{array}{c} 0 \leq \gamma \leq 1 \\ 0 \leq \mu \leq 1 \end{array}.$$

In this operator μ_i are the normalized degrees of membership of the fuzzy sets and γ is a parameter indicating the degree of compensation. This parameter can easily be determined on the basis of a few observations when solving the above equation for γ. Further testing showed that the minimum operator performed as badly as the maximum operator; arithmetic and geometric mean performed better. The best results were obtained for the above mentioned generalized operator for $\gamma = 0.615$.

Results, interpretations, and conclusions

How can it be explained that the word "and," when used by human beings in connecting goals or constraints neither means "plus" in the mathematical sense nor "and" as defined in logic and characterized by the minimum or the product operator? The following interpretation seems to be plausible. Human beings connect goals and constraints context-dependently in many different ways, i.e., they use numerous "latent" operators. In our language however, which is much poorer than our thinking, only two words are available to indicate the linkage of subjective categories: "and" and "or." The meaning of these two words has been defined formally by logicians in a very definite way. If a human being wants to articulate the type of operator he has used in aggregating these categories, he can only choose one of these two words and he probably chooses the word, which according to his feeling comes closest to the operator he has really used (i.e., he chooses the word which expresses as well as possible the degree of compensation that is implied in his real operator). By contrast to the "logical and" we called the generalized operator "compensatory and," a connective which could never be used in verbal communication because it had not been defined before.

The above mentioned empirical tests were very time-consuming and much more expensive than the axiomatic treatment of the same problem. They are, however, only a first step in the direction of generating a formal (mathematical) modelling language, which is appropriate to model decisions or problems, which originate or exist "only in the minds of managers" and cannot be observed directly by looking at real things. The existence of such a language is, however, a prerequisite for models and theories, which assert to be "factual models" of decision making and problem solving, which are falsifiable and confirmable [23,24].

REFERENCES

1. A. Tarski, A general method in proofs of undecidability. In *Undecidable Theories*, (Edited by A. Tarski, A. Mostowski, R. M. Robinson). North Holland Publ. Co., Amsterdam (1953), p. 11.
2. R. Harré, *An Introduction to the Logic of the Sciences*. London, Melbourne, Toronto (1967).
3. R. Harré, *The Philosophies of Science*. Oxford University Press, London, Oxford, New York (1972).
4. L. Apostel, Formal study of models. In *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, (Edited by H. Freudenthal). D. Reidel Publ. Co., Dordrecht (1961).
5. R. Carnap, *Introduction to Semantics*. Harvard University Press, Cambridge, MA (1946).
6. M. W. Wartofsky, *Conceptual Foundations of Scientific Thought*. Macmillan, New York, London (1968).

7. P. Suppes, Meaning and uses of models. In *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, (Edited by H. Freudenthal). D. Reidel Publ. Co., Dordrecht (1961).
8. E. Nagel, *The Structure of Science*. Routledge & Kegan Paul, London (1969).
9. D. J. White, *Decision Methodology*. Wiley, London, New York (1975).
10. K. Popper, *The Logic of Scientific Discovery*. Hutchinson, London (1959).
11. G. H. Symonds, *Executive Readings in Management Science*, (Edited by M. K. Starr). Macmillan, New York, London (1965).
12. K. J. Arrow, *Social Choice and Individual Values*. Wiley, New York (1951).
13. P. Rivett, *Principles of Model Building*. Wiley, London, New York, Toronto (1972).
14. R. Ackoff, *Scientific Method: Optimising Applied Research Decisions*. Wiley, New York (1962).
15. H. Mehlberg, Theoretical and empirical aspects of science. In *Logic, Methodology and Philosophy of Science*, (Edited by E. Nagel, P. Suppes, A. Tarski). Stanford Univ. Press, Stanford, CA (1962).
16. E. Witte, Phasen-Theorem und Organisation komplexer Entscheidungen. *ZfbF* 625, 1968.
17. H. W. Brand, *The Fecundity of Mathematical Methods*. Dordrecht (1961).
18. J. Schwartz, The permicious influence of mathematics in science. In *Logic, Methodology and Philosophy of Science*, (Edited by E. Nagel, P. Suppes, A. Tarski). Stanford Univ. Press, Stanford (1962).
19. J. Stamp, Quoted by D. J. White, *Decision Methodology*. Wiley, London, New York (1975).
20. L. A. Zadeh, Fuzzy sets. *Info. and Cont.* **8,** 338–353 (1965).
21. R. Bellman and L. A. Zadeh, Decision making in fuzzy environments. *Mgt. Sci.* **17B,** 141–164 (1970).
22. H. Hamacher, *Über Logische Aggregationen nicht-binär explizierter Entscheidungskriterien*. R. G. Fischer, Frankfurt/M. (1978).
23. U. Thole, H.-J. Zimmermann and P. Zysno, On the suitability of minimum and product operators for the intersection of fuzzy sets. *Fuzzy Sets and Systems* **2,** 167–180 (1979).
24. H.-J. Zimmermann and P. Zysno, Latent connectives in human decision making. Working Paper No. 79/02, RWTH Aachen. *Fuzzy Sets and Systems* (1980) (Accepted for publication).