DISCUSSION: A CRITICAL COMPARISON OF THE MODELS
H. SPADA (CHAIRMAN); G.H. FISCHER,
R. GRONER, W.F. KEMPF, J.M. SCANDURA,
H. SCHEIBLECHNER, P. SUPPES, D. VORBERG,
W.H. WULFECK II

SPADA:
Our intention in planning this round table discussion was to allow for a lively exchange of views about the models discussed at this symposium so far. By inviting outstanding proponents of quite different lines of research - let me just mention mathematical learning theory, deterministic theorizing, structural learning, logistic models of learning and thought and automaton models and register machines - we tried to create the basis for a fruitful and critical comparison of the models, in which different aspects are taken into consideration.

The discussion is going to start with general introductory comments by Prof. Suppes, Prof. Scheiblechner, Dr. Groner and Prof. Fischer. Afterwards we shall enter the open discussion, which in my opinion should be centered on three main areas: firstly, a critical comparison of the most relevant psychological and formal assumptions of the models; secondly, comments on possible ways and fields of application of the models to questions of science instruction - I think we should not forget this point at the last meeting of this symposium; and thirdly, a look at the future development of the models.

May I just give the word to Prof. Suppes for his first comment?

SUPPES:
I am going to make four comments and then deal with the items you mentioned. Before I begin, let me say how much I enjoyed being here and participating in the discussions. It has been a pleasure to learn all that I've learnt and to be able to criti-
cize all what I have criticized. Both are pleasures. So it has been a very constructive and enjoyable occasion for me.

I'm impressed with the discussion of the thesis of specific objectivity and yet not fully persuaded. One of the things, that I have talked to Prof. Fischer about, was to get a general formulation, in order to see how much the formulation can be separated from the specific Rasch-models. And I'm not clear at the end of this symposium, how much I should accept, to what degree I should accept, as a very strong methodological postulate, the principle of specific objectivity. It's a matter I think that I would like to discuss myself and to hear more about in detail. But one way of putting some of my doubts: it isn't quite clear to me at what level we want to draw the separation. So that, as far as I can see, if we think of the person as composed of parameters for different abilities we might insist on separation among those parameters. That the theory does not provide for that kind of separation, might be difficult.

But a more important remark is my second remark. It is not entirely clear to me, but seems like a serious problem. If the parameters that are introduced for individuals are not constants but random variables, then as I understand the formulation, the thesis of specific objectivity might well not be satisfied. That is the separability requirements, if the constants are what I would think they are in fact - namely, random variables, and have a proper distribution. I would raise the question, whether the thesis of specific objectivity can be satisfied when the individual parameters expressing individual differences are not constants but random variables. Let me say, by the way, that I'm very positive about many aspects of what I have heard, and I think that much of the discussion of the principle is something that I would support.

Historically, it's rather interesting in the way we might think of the principle of specific objectivity as a kind of new imperialism. In the past we have had a very strong intellectual imperialism of behavior learning theories. And it is characteristic of those classical learning theories not to admit within the theory anything of a highly specific nature about individual differences. For those coming from testtheory, which has quite a different thrust from classical experimental psychology and classical learning theory, one can see why the principle of specific objectivity is regarded as so important. But I remark that it is a principle that was not brought to the surface in classical learning theory or more broadly in many aspects of classical experimental psychology.

My third remark concerns the discussion of rules. It seems to me terribly important to be very careful in distinguishing knowing a rule from knowing when to use a rule. And this is a point that must be made in a theoretically explicit way, because there are deep learning results that contrast the distinction between knowing a rule and knowing when to use a rule. It is extremely easy to have a generation of problems, or ideas, of concepts, of solutions from a small number of finite rules, that clearly one knows in the ordinary sense of knowing the rule, but one cannot reduce to a rule the metatheoretic problem, when to apply the rule. That problem, it seems to me, needs very careful attention in this discussion of rules. Because one way of expressing a criticism of too much emphasis on explicit rules is precisely that one cannot reduce to a rule, the point of application of a rule. We must be careful in teaching not to overemphasize the importance of rules. If you consider the case of written language and teaching students how to use language of written form, we have a very good example in the sense that no one, at least no one to my knowledge, has proposed an explicit set of rules governing when to apply a rule. That point can be made formally. In the earlier discussion, I made that point about mathematical proofs. It is extremely easy to understand and to know the rules of a proof. It is in general not possible in fact to reduce to rules, the rule for when to use a rule in
The fourth remark is about the linear logistic test model (LLTM) and the generalization to the LLRA (linear logistic model with relaxed assumptions). I found impressive the applications of this model in the various kinds of situations that have been described here. And I enjoyed very much the detailed presentations of experimental results. I would have, however, the following general remark about this model. From the standpoint of the way I like to think about science, it too quickly reduces to a single equation. There is not, as far as I can see, a framework in which the model is discussed, an appropriate richness of possibilities for deriving different equations. I have in mind, for example, the classical and important way, in which differential equations were derived in classical physics or in psychology in classical S-R-Theory learning models, in which, for different reinforcement schedules, one can derive quite different equations for the prediction of behavior. Too often, it seems to me, the only thing that varies in the LLTM or its relaxed version is, for example, the number of the product: is the product a product of m terms or of n terms? There is not enough flexibility to adapt in a deep way to the structure of different problems if it reduces, essentially, to this characteristic equation - exponent of an expression divided by one minus exponent (or with product terms) - I am skeptical that a model of this form is rich enough for a wide variety of problems. I urge upon the advocates of it, to think about the ways in which it can be applied (to yield the same ideas, which seem to me very attractive) in a more general way.

Those were the remarks I was thinking about during the day and I think that I will stop here and take up afterwards the themes you mentioned.

SCHEIBLECHNER:
There were three broad classes of models being discussed at our meeting: the Markov-chain-models, the logistic type of model and the deterministic structural type of model. Each of the types of models has its specific advantages and its specific problems, and I would like to mention the specific advantages and problems which I see.

I think for the Markov type of model, all that I would call an incidental parameter causes major difficulties. For the logistic type of model, I would agree that very often we use a very implausible learning theory. And for the deterministic structural type of model, I see a problematic relationship to observations, or to put it otherwise: I would even say that all you can do with the structural deterministic type of model, you are able to do also with the logistic type of model. Therefore, I will not speak about this type of model any longer.

The advantages of the approaches are: 1) for the Markov type of model you have very often a psychologically convincing learning theory, and the feature of discrete changes is one which especially appeals to me; 2) for the logistic type of model I see the great advantage of the ease of handling incidental parameters, e.g. individual parameters.

What I would regard as most interesting are the trials realizing a bridge between the three approaches. And I think one trial was Mr. Kempf, who proposed the DTM, which I consider being a bridge between Markovian Types of models and logistic types of models. And I also suggested at least some approximation to this: the recursive formulation of learning assumptions within the framework of logistic models.
GRONER:
I would like to comment further on some points raised by the speakers before me and to emphasize the special advantages of the Markovian type of model.

In the area of learning and thinking, their principal advantage has been in describing dynamic and discrete processes in time. Already in some early applications to learning processes, it was shown that in a variety of situations, they allowed for a very compact and parsimonious description of entire empirical frequency distributions, powerful in handling the effects of stimulus variation and reinforcement. When it was shown relatively soon that these models were equally successful in the more complex area of thinking, more effort was spent for the exact definition of the underlying psychological process. Logically sufficient mechanism under the label of "strategies" were postulated, which were capable of generating the solutions of the thought problems. A further advancement involves the formalization of the strategy concept into an algebraic automaton model, where use can be made of the isomorphisms between probabilistic automata and Markovian processes, thus combining formal precision with empirical testability.

As already mentioned in our paper (GRONER & SPADA, this volume), individual differences could also be handled by Markovian models through individual parameters. However, such a parameterization seems to me as equally inappropriate as the introduction of "incidental" individual parameters in the refined Rasch methodology: there is substantial psychological evidence that, in most cases, interindividual differences do not behave as simple fixed constants. Therefore the notion of an underlying "ability" doesn't appeal to me, even if put into quotation marks and estimated by conditional maximum likelihood methods. Either one goes the hard way and tries to explain the full range of individual behavior by linear and nonlinear concepts (e.g., attitude-treatment interactions, individual production systems in the sense of NEWELL & SIMON, etc.), or one looks for an alternative which doesn't run across with present day theorizing.

In our application to rule learning, the assumption that the interindividual variation can be explained by the same mechanisms as the intrindividual variation, i.e., a gradual rule acquisition process, seems like an intuitively appealing idea to me. If such an assumption (maintained, e.g., by PIAGET) is valid, then the vector of starting probabilities is a sufficient statistic for the interindividual differences.

FISCHER:
The speakers before me, I think, have given a rather general and at the same time fairly deep discussion of the types of models and the problems we are faced with. Therefore, I would only like to pose some rather specific questions, for instance with respect to the lectures of the second day, which I hope can be answered in the discussion.

It was my feeling that the controls of the linear logistic models in the applications mentioned by the speakers have not been severe enough. There is the demand of comparing results from different subsets of items that had, for instance, been raised by Mr. Scheiblechner on a previous occasion. It's my feeling that in the applications discussed here, such severe controls of the model have not been carried out with sufficient rigor. Therefore, I would ask the speakers of the second day to briefly state once more the controls or tests of the model structure they actually applied in the empirical analysis.

As to the papers of this morning - there I would like to except Mr. Reulecke - I have the feeling that there is a big discrepancy between the practical classroom situations and the resulting types of observations on one hand, and the speculations on a hierarchy of rules or cognitive operations in the deterministic theories on the other. I do not see how this gap can be filled. So I would like to ask the speakers of this morning to describe once more - just in a few sentences - precisely the experimental observations as well as the statistical or other
empirical analyses and how these are mapped into the theory.

I would also like to address one question to Prof. Scheiblechner. I refer to the recursive part of the learning model in your second paper; I do not quite understand what happens if the change of bias parameter, being a continuous variable, becomes very small. Then, I think, the direction of change becomes more and more undetermined. What becomes of the model in this case?

These are some concrete problems I would like to have discussed in the following.

SPADA:
I think we should now start the open discussion. Of course you all are invited to enter. There have been some questions to which I think we have to revert to. One of them, concerning the fit of the models, I would like to answer myself later on. But first there is an urgent remark.

SCANDURA:¹
I would like to respond to some of the questions that have been raised. First, I would like to comment briefly on specific objectivity. The major value of specific objectivity, as I see it, and I say this on the basis of only limited familiarity with the concept, is not so much as a basis for model building but as a basis for separating empirical effects due to individual differences and those due to test items. Although one can introduce additional structure into the basic Rasch formulation (which follows from specific objectivity), there is little by way of a deep psychological justification for such structure. (This is essentially the same point made by Prof. Suppes who argued that all Rasch-type models are simple variants on the same

¹ Editor's note: Prof. Scandura revised the first draft of this part of the discussion, taken as tape record, very carefully. He added some additional comments, too, which of course couldn't be answered in the following contributions.

(mostly linear) theme.)

In effect, one would almost seem to have a formalism looking for an appropriate theory. The results reported by Hilke, Kempf and Scandura (this volume) suggest that the deterministic structural learning theory might provide a useful basis for developing analogous Rasch models. (In my opinion, however, not necessarily shared by my colleagues, the proposed stochastic theory reflects given states only, in what is, in fact, a potentially dynamic situation. Dr. Kempf's important DTM would formally allow for external aspects of the dynamism involved but not the internal changes involved in such things as deriving needed solution rules. The complexities involved in actually constructing and testing such analogues also would be considerable.)

Following up on this point, I would like to pose a counter-comment to Prof. Scheiblechner. Suppose one can account for behavior deterministically - under idealized conditions. This is something which we have been able to do in many of our experiments. It also seems possible to specify deviations from ideal conditions in a way which might allow predictions in new situations. (Our limited data (Voorhies & Scandura, 1976) on this point, while not conclusive, appears promising.) My question to Prof. Scheiblechner, then, is: Why would one want to replace a deterministic account of very specific behavior, with an account in terms of probabilities, which inherently gives less information?¹

Second, recall Prof. Suppes' comment concerning the difference between knowing a rule and knowing when to use it. I think this distinction is, of course, a very critical one. The distinction is an essential part of any complete characterization of any rule. In particular, and as a minimum, one must specify both the operation to be carried out and the domain of applicability of that operation. Neither alone is sufficient.

In psychology, it is often easier to identify the operation of
a rule than the domain of applicability. This is particularly true of higher order rules, more informally called "heuristics". Consider the higher order rule shown in Figure No. 1

![Flow diagram for the higher order finite differences rule which acts on restricted rules and generates rules of the form n → an + d](image)

Although the operation (flow diagram) is specified in this case, the domain strictly speaking is not. In the SCANDURA (1974) study, the experimental inputs were triples of instances such as

- 1 → 3
- 4 → 12
- 5 → 15

The higher order rule operates on such triples and generates as outputs rules of the form "n → an + d". But, this higher order rule will not work with all such triples. For example, consider

- 1 → 1
- 4 → 16
- 5 → 25

In order for the rule to succeed universally the domain would have to be limited to triples of the form

\[
1 \rightarrow a + d \\
m \rightarrow am + d \\
m + 1 \rightarrow am + a + d.
\]

Whereas it would be possible to devise such a rule in this case and in others even more complex (e.g., SCANDURA, DURIN, & WULF-ECK, 1974), I do not believe, for the most part, that human knowledge is that algorithmic. In general, I believe that human knowledge, especially of the higher order variety, consists of rules with only more or less accurate domains.

This does not mean, however, that one cannot have a theory of how people use and acquire "imperfect" as well as perfect knowledge. Indeed, psychologically speaking, I don't think that there is any difference between the two. In the structural learning theory, at least, it makes no difference whether the rules of competence, or of knowledge, are perfect or imperfect. Their use and acquisition is governed by the same principles.

Now, Prof. Suppes, I believe, is referring to quite a different kind of imperfection. This imperfection is mathematical in nature and follows directly from a well-known theorem by the logician CHURCH. Namely, there exist classes of tasks (problem domains) for which no algorithmic solution exists.

In short, there is an important difference between mathemati-
cal constraints and psychological relevance. No serious psychologist ever thought that anyone knew everything there is to know. It is, in fact, the introduction of imperfect as well as perfect higher and lower order rules which makes it possible to "account" for problem domains of often unclear scope. The non-algorithmic and often unpredictable nature of such knowledge is what gives structural accounts a "lifelike" and creative character. On the other hand, it is important to know that some task domains are sufficiently complex, that no strictly algorithmic account is possible. In such cases (as with most realistic task domains), one must rely in evaluating alternative rule accounts on empirical comparisons of relative generative power.

Finally, Prof. Fischer raised the question of how experiments are mapped into the structural learning theory. I am afraid that any really convincing answer to this question could take the time remaining this afternoon. Perhaps the best way to answer here is to say that explicit attention has been given to making the theory operational (i.e., testable) and to refer the interested person to the rather considerable literature which has developed over the past decade. Some of the most directly relevant empirical studies are briefly summarized in the concluding section of my chapter.

SPADA:
There has been a question, referring to the speakers of the second day, especially concerning the problem of model controls. I would like to answer it myself. For one type of item I have been working with - the direction of rotating wheels - there have been quite a few very sharp model controls. Firstly, I compared the direct estimations of the item parameters (estimated by the model of Rasch) with the estimations resulting if the hypotheses about the task structures and the structural parameters of the linear logistic model are taken into account. And as I discussed in my paper yesterday, this test showed

that there were some statistical deviations. Two other very interesting tests were the comparison of the estimations of the structural parameters of the linear logistic model in different samples of subjects, and even more interesting of course, the comparison of the estimations in different samples of items. The correspondence there between the parameters was really good for four (respectively five) of the five (respectively six) operations. It wasn't good for operation 4. The same was true for the learning parameters. Beside this, I also analyzed data from different experimental procedures; for example, I worked with different item sequences, which I think is a very sharp test of the stability of the learning parameters. And in these controls - as I showed in my second paper yesterday - it turned out, that the same operations which corresponded quite well in the other comparisons corresponded well here too. For operation 4, deviating estimations resulted again.

Coming back to one comment of Prof. Suppes, who emphasized that knowing a rule isn't the same as knowing when to apply a rule, I think this problem is not crucial for the tasks analyzed by me so far, e.g., with regard to the task area 'direction of rotating wheels' I distinguished between five operations about rules and one operation - a branching decision in the sense of Prof. Scandura - stating which other operations are relevant for the solution of a special problem. This branching decision could also be formulated as a rule, which says when to apply which other rules. The content of this rule is that, with closed gears, it is necessary to analyze all transformations of the movement from wheel to wheel to decide whether a movement is possible at all. More complicated operations concerning the proper use and sequence of other operations didn't occur in the task area studied by me.

My last statement refers to an interesting result of an empirical analysis of two of the discussed models (the Rasch model and the deterministic structural learning theory).
Let us have a look at the data of just two of the balance problems I mentioned yesterday (of course I compared all the items). The frequency of 144 indicates that 144 students solved item B 22 correctly and didn’t solve B 24 correctly. 161 students solved item B 24 correctly but not item B 22. The data of 949 students were analyzed. In accordance with the deterministic theory and the hypotheses about the task structure, which says that the structures of these two items are equal, and without learning assumptions, both frequencies should be zero. There shouldn’t be any person solving one of the items without solving the other one, because both items are psychologically – on the basis of the operations – identical. According to an all-or-none learning process, there should be a zero in the cell, where, in fact, a frequency of 144 was found because there shouldn’t be any students solving item 22 without solving item 24 – item 22 was presented before 24.

Corresponding to the assumptions of the linear logistic model (respectively the Rasch model) and the hypotheses about the task structures, the two frequencies should be different from zero and approximately equal if no learning is supposed. Assuming some incremental learning process, it seems plausible that item 24 was solved correctly more often than item 22.

And it is interesting and almost a little bit implausible on psychological grounds – I would agree with Prof. Scheiblechner on this point – that the ratio of the two frequencies is just the amount it should be on the basis of a probabilistic formulation along the lines of the Rasch model. As you know, the ratio of \( n_{ij} \) divided by \( n_{ij} \) equals the ratio of the corresponding Rasch-item parameters if the model is valid. In fact, the parameters which were estimated by another algorithm (see Fischer, 1974, 230-239) – starting from data organized another way and not from these frequencies – yielded a quotient of 1.10 and the empirical finding here was 1.12. (161 divided by 144). This is, of course, only the comparison of two items, but the results were quite similar for all the items I compared.

KEMPFF:
I would like to add something to this and to what Prof. Scheiblechner has said before.

Let me start with Prof. Scheiblechner's comment, stating that what can be done with the deterministic structural learning theory can also be done with the LLTM. I fully agree with this, and I think that this afternoon's presentation (cf. HILKE, KEMPFF & SCANDURA, this volume) has shown that quite clearly. But the converse is also true: what can be done with the LLTM, in principle, can also be done with the deterministic structural learning theory – given the chance to establish so called "idealized conditions".

Now, as you mentioned, Dr. Spada, in your data there is a clear deviation from this. But as far as I know about the conditions of data collection in your study, this is not surprising at all.

However, I think establishing "idealized conditions" is an essential problem for testing the deterministic theory, because one needs a theory about the possible influencing conditions in order to make statements about whether they meet the requirements. For instance, the "idealized conditions" might include "memory free" conditions. So, if you want to test the deterministic structural learning theory, you have to have a theory which tells you the conditions under which you can perform an
experiment that is not influenced by memory. Thus you need a
toory of structural learning. I would like to ask you, Prof.
Scandura, whether you do have such a theory.

FISCHER:
I would like to answer at least partially the questions raised
by Prof. Suppes insofar as they refer to my paper. I fully
agree that it would be desirable to have a very general theory
of the consequences resulting from the postulate of specific
objectivity. RASCH (1961, 1968) published several fundamental
results on this topic, but apparently it is difficult to give a
completely general definition of specific objectivity and to
derive results on very weak assumptions. It is hardly possible
to go into detail here.

Insofar as I apply the concept of specific objectivity in my
paper to the context of measuring change, it boils down to the
following assumptions: the measurement of (i.e., the statistical
results related to) change should be essentially independent
of the distribution of the individual abilities in the sample
of subjects and of the distribution of difficulty parameters
in the sample of items. Starting from these two assumptions,
it is not necessary to specify the Rasch model as an addi-
tional assumption, because the Rasch model follows as a nec-
essary consequence from these two assumptions. In order to
carry out this derivation, you need some technical assumptions,
related, of course, to the number of response categories, the
dimensionality of the parameter spaces and the probabilistic
nature of the observations. But I don't think that these as-
sumptions are really of much concern for the empirical scien-
tist.

That what is arbitrary to a certain extent in the LLTM and
the LLRA is the linear structure within the parentheses. You
could have a nonlinear structure for instance; you could have
interaction terms, which can be introduced easily; but statisti-
cally speaking, it will be hard to estimate interaction ef-
fects, because then you need a great amount of statistical in-
formation. Linear functions may be considered as a first order
approximation to whatever more complicated mathematical forms
within the argument of the logistic functions.

I don't know whether these answers satisfied you; at least I
have tried my best.

SCANDURA:
One of the things that has impressed me about this symposium is
the relative depth of the questions and responses. Many times
one gets only superficialities.

In specific response to the comments and presentations by Dr.
Spada and Dr. Kempf, perhaps I should quote one sentence from
what I said this morning; I think that it is directly relevant:
"One might be tempted to propose that a stochastic theory pro-
vides a better account of certain data than a deterministic the-
ory, if the stochastic theory does a better job of predicting
average behavior, which it is designed to do, than a determin-
istic theory predicts the non-idealized behavior of individuals
in specific situations, which it is not designed to do." This
quotation, of course, is simply a way of elaborating on what
Dr. Kempf has said.

With regard to Dr. Kempf's question, I would like to answer
in two parts. To give you a better feeling for what is meant
by "idealized conditions", let me give you a simple example. In
one of our experiments, we taught children rules for converting,
say, from measure A to B and from B to C. We also taught some
of the children a higher order composition rule (restricted to
a suitable domain). Then, we tested them on A-C problems. The-
oretically, according to the control mechanism outlined in my
paper, one would predict that every subject who learned all
three rules would succeed on the A-C problem under "idealized
conditions". The question is: What are the "idealized condi-
tions"?

In a word, idealized conditions serve as boundary conditions
which must be met if the theory is to apply. Thus, in the structural learning theory, the control mechanism is the sole determinant of what happens in a given situation only under certain very specific conditions. (1) One must know not only what relevant rules have been learned, but which are immediately available to the human processor. What does this mean operationally? It means that training must proceed to a certainty or as close to a certainty as possible. Just as one would try hard to create a perfect vacuum in comparing the rates of fall of an iron ball and a feather (see Section 1 in my chapter), one would try hard to insure that assumed rules are indeed learned and available to the subject at the time of testing. One would not want to set a learning criterion (e.g., three consecutive successes) independently of what circumstances require. The critical thing is "to get as much of the air out" as possible. (2) Notice also that having learned a rule at some previous time, even to a very high criterion, and having it immediately available in the processor (in working memory), is quite another thing. Empirical conditions should be set up to come as close as possible to guaranteeing availability. This can be accomplished, for example, by having cues for the assumed rules immediately and at all times in front of the subjects. No secrets!

Another constraint has to do with processing capacity. Here I refer to the widely accepted belief that people can only process a small number of cognitive elements at the same time. In the experiments we have run under "memory free" conditions, all assumed rules are assumed to be available. There are no limitations on processing capacity.

Therefore, to obtain information about the control mechanism, unencumbered by processing capacity, one must also be sure that the available rules do not take up too much space in the processor. Otherwise, the processor may get overwhelmed in the process of shifting control and so on.

In order to reduce processing load, one may try to insure some degree of overlearning (with respect to the assumed rules).

Overlearning has the effect of chunking rules and other cognitive elements into a smaller number of units. One might also minimize the role of processing capacity, for example, by providing subjects with paper and pencil and all the time they need.

In summary, insuring idealized conditions amounts to insuring the availability of assumed rules and eliminating the role of processing capacity (and other constraints on cognition). In experiments run under such conditions, we have been able to get close to one hundred percent correct prediction. If the conditions are changed, then the results will deviate from perfect prediction just to the extent that the conditions deviate from the required idealized conditions.

The principle is exactly the same as would be the case in comparing the rates of fall of an iron ball and a feather. Their rates of fall may be expected to deviate from being equal just to the extent that one has failed to create a perfect vacuum.

There are, of course, many empirical situations where processing capacity (and other cognitive constraints) may play an important role. In such situations, one has two choices. One can admit failure - an inability to account perfectly for the behavior, and turn unashamedly to stochastic theorizing. Or, one can attempt to add more structure to the deterministic partial theory so that it accounts perfectly for behavior in a greater variety of (but never all) empirical situations. This is exactly what introducing a fixed capacity processor does to the structural learning theory.

SUPPES:
I just have two remarks in response to what Prof. Fischer has said. I certainly agree that the derivation of the linear logistic model and its generalizations is not arbitrary. I think what Prof. Scheiblechner was saying is important. I am seeking for ways to put it. I quite agree, that the learning models
have neglected individual differences. The important point is that in stochastic process models the linear logistic model would be regarded only as averaging; it gives a mean probability at the time the item was given. One would anticipate there were effects. And in the items that everything is not independent. The only case in which this would correspond to the probability of the sample path would be when independence of items exists. And, of course, in actual behavior the sequential effects are substantial. It is precisely the sequential effects - as a way of putting this - that we want to predict under different conditions from a model. When we have the problem of accounting for a variety of sequential effects, it is exactly this, exactly the problem of predicting the sequential effects, that gets us away from some model of the logistic form. We can write the model in this form, but we have to have a derivation of equations for the different sequential effects, that we believe are taken into account by the model. The only case, where that is not true, is in the case when everything is independent. Then what holds in the mean holds also sequentially because of the strong independence.

But in almost all behavior, I would think such independence is actually violated. It is quite true we can attempt to construct tests that satisfy such strong independence, but in actual solving of problems sequential effects, I think, are very strong.

FISCHER:
Only one question, if I may. Do you mean complete stochastic independence or local stochastic independence?

SUPPES:
What do you mean by local?

FISCHER:
I use the concept of "local stochastic independence" as it is defined, e.g., in the book by LORD & NOVICK (1968): if all the dependencies can be expressed as change of underlying parameters, then local stochastic independenc holds. This is the case in the linear logistic models, since the learning or effects of instruction are described by certain changes of the parameters. Besides, stochastic independence is assumed (experimental independence, to use one more concept from LORD & NOVICK).

SUPPES:
Yes. That is the sense I mean when I say it is a stochastic process. Knowing the parameters does not imply independence of the responses. That is exactly the characteristic. What I am claiming is that this is precisely a feature of much actual behavior.

It is a scientific hypothesis that it is true in data. I would be willing to wage a rather substantial bet that you cannot remove all dependencies by the parameters.

I think that the parameters that characterize the state of the organism are also affected by the inputs. So it is a characteristic feature, for instance, of a probabilistic automaton that knowing everything that is knowable about the construction of the machine would not permit you to make the best sequential predictions. So if any of the behavior of a complex nature has the characteristics of a probabilistic automaton, my thesis would be established.¹

¹ Editor's note: The assumption of local stochastic independence doesn't rule out inputs changing the parameters of the model. What is ruled out are changes in the reaction probabilities of an individual caused directly by his proceeding reactions (cf. the two papers of SCHEIBLECHER, this volume).
SCHIBLECHNER:
Two meanings of dependence and independence were raised in this discussion. And I agree fully with Prof. Fischer's point; but the model of Mr. Kempf, the dynamic model, was just an attempt - I don't want to say whether it was a successful attempt or not - but it was an attempt to make a third meaning of dependence or independence built into this type of model. In Mr. Kempf's model, the next observations will be dependent not only on the parameter which was present before, but also on the realization of the random variable, which occurred before. That was just what Mr. Kempf attempted to put into his models.

KEMPF:
Of course I did, but still I would not regard the dynamic test model (DTM) as a "learning theory" by itself.

The basic DTM I presented yesterday is just the Rasch model extended by some simple form of dynamic processes, dependencies of the probability distributions of an individual's responses on the number of his prior positive responses. The basic DTM does not make any reference to the internal structure of the items. Unlike the LLTM, it does not attempt to explain the item difficulties in terms of the difficulties of underlying cognitive operations involved in the solution process.

Nonetheless, such further specification of the DTM is possible, and a first attempt in this direction was published in KEMPF (1974a, b): a structural dynamic model which asymptotically approaches the LLTM. This mutual allowance to the LLTM yielded a parametric structure, however, which is somewhat odd in view of the constraints between item and transfer parameters in the DTM.

A more successful approach in this direction will be published in a forthcoming paper (cf. KEMPF 1976b) and has a one-dimensional special case of the Rasch model for m>2 categories of answers as its asymptotic form: the "additive logistic test model" (ALTM, cf. KEMPF & NIEHUSEN 1976). When compared with the LLTM, however, both the ALTM and its dynamic extension, the ADTM ("additive dynamic test model"), are more restricted with respect to further generalizations (e.g., the measurement of experimental effects on the difficulties of the underlying operations can be performed independently of the selection of items only if a multivariate analysis of the data is admissible).

Though such a specification of the DTM is possible, the type of transfer or learning assumptions included in the model is still rather simple. Just recently I have proved fairly general conditions for the existence of sufficient statistics in "dynamic" test models, which make clear that there are quite stringent restrictions with regard to the type of dependencies among responses that can be described within models that allow for conditional inference (cf. KEMPF 1976a).

Summarizing this, I want to say that "dynamic" test models may be useful in educational contexts, e.g., for evaluation purposes, but they do not constitute a "learning theory" by itself; nor does the LLTM.

SUPPES:
I think we perhaps pushed that as far as we can push it usefully in this discussion.

My second remark was about rules. I do want to try to make very clearly the point. I agree very much about instructing and teaching the rules. I disagree very strongly with the view that one is to teach successfully in a complete way second order rules. It seems to me a very important aspect of teaching, to make clear to the student, that there will be no mechanization and no explicit rules for the choice of what rule to apply when. In most subject matters it's only in trivial cases that such mechanization is going to be the case. Only in trivial cases can we give explicit the set of rules saying when to apply what. That, it seems to me, is absolutely fundamental. And I am quite prepared to argue about it.
FISCHER:
I fully agree with Prof. Suppes' last remarks, but I would like
to make an additional remark. In the applications of the linear
logistic model that have been reported here, not the knowledge
of the operations but the actual availability of the operations
in the complex tasks was measured by the "basic" or "elementary"
parameters $n_j$ in the model. I thought I should stress that.

SUPPES:
But may I comment? Let's take the paper contributed by Mr. Spa-
da - I read this paper again last night. Now those are exactly
the cases, in which you do have an algorithm. And I think this
is also true to some extent of your limited set of construc-
tions. It would not be true in general.

I suppose one of the problems that I might in fact want to
pose for the linear logistic model or its generalizations is
how to analyse data in which knowing the necessary steps is
not the case: the classic example would be mathematical proofs,
where every student does something different. My argument is:
that this is what most learning is about; that kind of learning,
ot algorithmic learning.

FISCHER:
I fully agree.

SUPPES:
I mean we are all in the same camp with our studies. I have al-
so done extensive studies of algorithms.

SPADA:
I fully agree that non-algorithmic learning is not the field of
the linear logistic model. But without a priori hypotheses about
the steps of problem-solving processes, a lot of different
classes of models could not be applied. They cannot be applied
to cases where every student works with another strategy on the
task and where you have no chance, experimentally, to go into
details.

SUPPES:
Why not? What do you mean with details? Students give mathemati-
cal proofs. You have an enormous amount of details. The prob-
lem is to understand how to deal with them. It is not a problem
of lacking details. - It is a problem of understanding. Because
for each student you can have a protocol, consisting of a se-
quence of steps. The difficulty is that different students do
different things.

SCANDURA:
It's interesting how people with different perspectives look at
the same questions.

I think the main point concerns the possibility of identifying
higher order rules explicitly. As far as I can see, there
has been no attempt to date, at least not explicitly, to iden-
tify what I would call higher order rules in the context of the
LLTM. I think that it might be done, but so far it hasn't been
done. The question about whether one should attempt to teach
more than lower order rules is one about which Prof. Suppes and
I, and I guess Prof. Fischer, seem to disagree. I believe that
higher order rules can and should be taught for several reasons:
(1) It is possible to identify such higher level rules.
(2) It is possible to teach higher order rules successfully so
that students may learn and apply them in quite new situ-
ations.
(3) As one teaches more and more higher order rules, the ten-
dency is for the student to be able to transfer to new situ-
ations for which no higher order rules were ever taught.
That is, students seem to get better at being able to in-
vend their own higher order rules as a result of having
seen a variety of examples. This tendency was apparent,
for example, in Mr. Wulfeck's experiment.
SUPPES:
I think that it is appropriate to teach rules as heuristic strategies. I agree that you can teach heuristic strategies. But one of the characteristics of a heuristic strategy is, if it is indeed heuristic (this is not a matter of faith, but can be proved) - that the heuristic rules are incomplete. It is a very important thing for students to understand that heuristic rules are not complete. In the case of languages it seems to me even more fundamental. We are not going to tell the student by higher order rules, when to use an adverb to modify an adjective; we are going to give them examples, we are going to give them some general overview, but we are not going to be able to give them higher order rules that have even the range, say, of higher order rules or heuristics in the case of mathematical proofs. And it's terribly important, it seems to me, that the student be left with the very clear impression that it is not possible to reduce the writing of German or the writing of English in a proper literary style or scientific style, to a set of explicit rules.

SCANDURA:
You can formulate that as a higher order rule.

SUPPES:
What?

SCANDURA:
What you just said!

WULFECK:
It is the case that a very large proportion of the material taught in schools (from U.S. grades 1 through 12) and colleges, and practically all of the science and mathematics, could be accounted for within a Structural-Learning-type approach. Most of the material commonly taught in the schools is not of the "highly creative" variety, but is "algorithmic", and the structural learning theory has great potential for research and instructional design, with respect to structured and algorithmic subject matters.

SPADA:
It would be very interesting to have comments by Prof. Dörner or Prof. Lüer. They studied and developed information-processing theories and computer simulations with reference to tasks of the type just mentioned- tasks which could only be handled by the statistical models discussed in this conference with great difficulties.

SCHIEBLECHNER:
I want to answer a question Prof. Fischer asked me: How do you find the direction of change for the bias parameter? The bias parameters and discrimination parameters were computed for each repetition of the experiment separately. Therefore they are known numerical values which are very seldom equal; if they are equal, no prediction is made. And I just use the estimates - the numerically known estimates at time t-1 and t - to see the direction of change. This is the random variable (Gsi), which of course has measurement error. These are errors in my regression model - recursive regression model - so to speak.

FISCHER:
But in the long run, the changes in the bias parameter I think must become smaller and smaller. And therefore the model eventually breaks down.
SCHEIBLECHNER:
Then the subject has no more problem with bias correction, because it was the content of the model, to estimate the change in this bias and adapt this bias. If he has succeeded with this, then he can achieve further success with his discrimination. If the bias is close to zero, there only remains an improvement in discrimination.

FISCHER:
Yes, but your parameter (n) begins to show random fluctuation, depending on the direction of change, which itself behaves more and more randomly. You must stop the model before reaching an asymptotic level of performance. Is this correct?

SCHEIBLECHNER:
I could not say exactly in which situation the random error of the estimate is largest. I could not say this exactly, but this is one point for all logistic models which could also be discussed. Should we consider the matrix of coefficients Q either as known constants or as random variables, or estimates of random variables? I am of the opinion, that it be more appropriate for all applications to consider the coefficients $q_{ij}$ as estimates of parameters which actually are unknown; e.g., if Mr. Spada says: rule 1 is present here in this example or it is not present, then you can - for example - graphically plot the example in different manners and it will be psychologically different: the difficulty will be different to perceive where rule 1 has to be applied here. Therefore, our model matrix, composed of zeroes and ones, actually is a rough estimate of the degree to which the rule is present. And if this is true, then the model matrix itself is a matrix of incidental parameters, and then the estimation theory, as elaborated up to now, breaks down again.

FISCHER:
I agree in principle, but there is one important difference between the general situation you outlined now for the LLTM on one hand and your application on the other, because the true changes during the learning process in your application almost surely become smaller and smaller, which means that your coefficients $c_t$ and $w_t$ behave more and more randomly. On the contrary, the coefficients of the matrix Q in the LLTM may also contain some random noise, but that random noise hopefully is small and remains constant. Your application is a case where Q in the course of time becomes more and more indeterminate.

SCHEIBLECHNER:
It also can be reverse, but you are right: there is a systematic trend in the amount of error.

SPADA:
I think one more point should be discussed, namely, the application of the models to questions of science instruction. It was, in fact, touched upon by the question concerning the range of applicability of the linear logistic model. I would like to raise the question of the applicability of Markovian models. I think that the range of applicability of Markovian models is much more restricted and that these models, of course, do not account for tasks of the type Prof. Suppes mentioned, either. Maybe we can center on this question for a while.

VORBERG:
You think that the applicability of Markov models is more restricted than that of the LLTM. Could you go into detail about that?
SPADA:
Let's center on science instruction--on the problems which are interesting here. If you analyze the applications of Markovian models in psychology, I think you will see that it is not just per chance that they have only been applied in connection with very simple problems of concept identification and similar tasks. I think it is an inevitable consequence of this formalization--where you rule out quite a lot of factors (for example, individual differences; but also many others, of course) that application is centered on very special phenomena.

VORBERG:
I completely agree with you with regard to the applicability of the Markov models which have been proposed so far. It is clear that the Markov models, which have been developed in the experimental lab, cannot be applied directly to science instruction, and also that the results obtained there with these models are not immediately generalized to the area of science instruction.

However, I do not see any argument why the general class of Markov models which lent themselves easily to process interpretation should not be applicable to the problems of science instruction. Of course, in order to develop a model, it is first necessary to analyze the problem situation in terms of the processes supposed to be going on. Once this is done, the process assumptions can be incorporated in a formalized model; I do not see any reason why this could not be done in the framework of a Markov model.

It seems to me that the strategy, which has turned out to be so fruitful in the experimental lab, merits at least some consideration in the problem area we are discussing here. The finding of discrete performance levels and of stochastic transitions between different degrees of knowledge, which are reflected in the structure of Markov learning models with few states, might also characterize the learning processes occurring in science instruction. It seems strange to me that the very same learning assumptions, which have been shown to be wrong in most verbal learning tasks, namely, incremental learning occurring "automatically" on every trial, should be used as the basis of formal models of science instruction, whereas the recent results obtained from the analysis of learning processes in terms of Markov models are ignored. Unless we believe that every applied field of psychology requires its unique underlying theory, it might be worthwhile to critically examine the lessons to be learned from Markov learning models.

SUPPES:
I don't agree with Mr. Spada's remarks. It seems to me that individual differences are no real problem, because I am perfectly prepared to formulate now a Markov model with individual differences. It seems to me overemphasized regarding science instruction. The problems are not the individual differences. The problem which will apply about equally to the linear logistic models and to Markov models is to have the appropriate analysis, or appropriate protocols, of what the students are doing. The difficulty is much deeper than the level of Markov models. The difficulty, if you are observing students solving a laboratory task, is how to reduce the protocol, the very elaborate protocol in its complete detail to a sequence. For example, if you want to apply a Markov model, we would want to reduce the protocol to a sequence of used rules.

The problem for any of the models is to have a reasonable reduction of the data, so that we can theorize. I think the examples that you, Mr. Spada, have given are nice examples, because of their simplicity. They are not nice in the sense that they are very far removed from more complex behavior. And this difficulty, it seems to me, is not at all an inherent difficulty of Markov models or of linear logistic models for this purpose. Because, if we have the protocols, we have many ways--for example, probabilistic automata or richer devices--to analyze the protocols. The problem is to get the protocols in a shape
which is reasonable.

SPADA:
I think that, on one hand, we overestimate the complexity and the difficulty of science instruction contents a bit, for example, for grades five to nine. The study of Dr. Häußler on functions shows that for really relevant content for education in physics in these grades, there are ways to handle the problems.

But I would agree fully on one point, and I want to stress that point. Progress has to be made on the experimental side. With relatively easy tasks, which can be solved after only a few cognitive operations, the question of how to collect data is not too difficult. But if problems are analysed, for which a lot of operations have to be carried out by the student, just to present each item and to classify the answers as right or wrong would not allow receiving appropriate and informative data. Maybe we should try in these cases to make the sequences of operations shorter by means of a good experimental design; not by presenting easy items, but by for example formulating specific questions about some of the steps of the problem solving process. Or it could be useful to give the student some sequences of correct steps of the solving process and to let him think about the next steps.

SCANDURA:
I am personally persuaded by much of what has been said as to the potential applicability of various of the stochastic models that have been proposed. There is little doubt that one can apply Markov models and Rasch-type models in some practical situations (e.g. in learning vocabulary). The robustness of the models in such situations suggests that each approach has a kernel of truth. The critical question for science education, I think, is which strategy will be the most fruitful one for uncovering the "maximum truth". Is it better to elaborate and extend stochastic theories, or is it better to take a deterministic approach such as the one I have proposed? On this point, I think there is a basic philosophical difference of opinion among many people here. Along with some of the participants, and an apparently growing number of others, I believe that deterministic theorizing provides the most promising first step. Many others are skeptical and remain committed to stochastic theorizing. In the final analysis, we will have to judge each approach in terms of what it accomplishes - both in terms of understanding and improving science education.

SPADA:
It seems to me to be a nice concluding remark that we have to look for the - how was it called? - "maximum truth theory".

Maybe we should stop here if there are no more very deep comments. I want to express my thanks to all of you and I hope you will come back sometime to the IPN and that you will now have a good return journey. Last but not least, I want to thank all the members of the IPN who assisted creatively in making this symposium possible.